This page purposely blank.
Acknowledgements

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

Katherine Bruce, Honors Programs
Julie Teague, Marvin Jones, and Steffaney Cohen, Office of Institutional Research and Assessment
This page purposely blank.
# Table of Contents

**Executive Summary** ........................................................................................................ 1  
**Background and Scope** .................................................................................................. 3  
**Methodology** .................................................................................................................. 5  
  - **Assessment Tools** ....................................................................................................... 6  
  - **Sample Selection** ........................................................................................................ 6  
  - **Scoring** ..................................................................................................................... 7  
**Results for Basic Studies Courses** .................................................................................. 11  
  - **A Note about the Presentation of Findings** ................................................................. 11  
  - **Description of Sample** .............................................................................................. 11  
  - **Critical Thinking** ....................................................................................................... 13  
  - **Diversity** .................................................................................................................. 17  
  - **Information Literacy** ................................................................................................. 20  
  - **Thoughtful Expression (Written)** ............................................................................. 24  
**Results for Upper-Division Courses** ............................................................................. 29  
  - **Pilot 1 Information Literacy and Thoughtful Expression (Written)** ............................ 29  
  - **Pilot 2 Inquiry** .......................................................................................................... 35  
  - **Pilot 3 Thoughtful Expression (Oral)** ........................................................................ 38  
**Process Results** ............................................................................................................. 43  
  - **Reliability of Scores** ................................................................................................. 43  
  - **Process Feedback** ..................................................................................................... 48  
**Discussion, Limitations and Recommendations** ............................................................. 57  
  - **Critical Thinking** .................................................................................................... 57  
  - **Diversity** ................................................................................................................ 59  
  - **Information Literacy** ................................................................................................ 60  
  - **Thoughtful Expression (Written)** ............................................................................ 60  
  - **Pilot 2 Inquiry** ........................................................................................................ 61  
  - **Pilot 3 Thoughtful Expression (Oral)** ....................................................................... 62  
  - **Relative Strengths and Weaknesses across Rubrics** ................................................. 63  
  - **Methodology and Process** ....................................................................................... 64  
  - **Limitations** ............................................................................................................. 65  
  - **Recommendations** ................................................................................................. 66  
**Follow up on Previous Recommendations** ................................................................. 67  
**References and Resources** ............................................................................................. 69  
**Appendix A Rubrics Used** ............................................................................................. 71  
**Appendix B Dimension Means and Standard Deviations** ............................................... 79  
**Appendix C Correlations Between Rubric Dimensions** ................................................ 81  
**Appendix D Detailed Scorer Feedback** .......................................................................... 85
LIST OF FIGURES

Figure 1 Distribution of Scores for Critical Thinking, Applicable Scores Only........................................14
Figure 2 Distribution of Scores for Diversity, Applicable Scores Only......................................................18
Figure 3 Distribution of Scores for Information Literacy, Applicable Scores Only ..................................21
Figure 4 Distribution of Scores for Written Communication, Applicable Scores Only.........................25
Figure 5 Distribution of Scores for Pilot 1 Information Literacy, Applicable Scores Only.....................30
Figure 6 Distribution of Scores for Pilot 1 Written Communication, Applicable Scores Only..............33
Figure 7 Distribution of Scores for Pilot 2 Inquiry, Applicable Scores Only........................................36
Figure 8 Distribution of Scores for Pilot 3 Oral Communication, Applicable Scores Only.....................39

LIST OF TABLES

Table 1 Interrater Reliability for Basic Studies..........................................................................................46
Table 2 Interrater Reliability for All Pilot Studies......................................................................................47
Table 3 Basic Studies, Pilot 1 Scorer Feedback on Process........................................................................49
Table 4 Critical Thinking Percent of Sample Scored at Least 2 and at Least 3..................................58
Table 5 Diversity Percent of Sample Scored at Least 2 and at Least 3..................................................59
Table 6 Information Literacy Percent of Sample Scored at Least 2 and at Least 3............................60
Table 7 Thoughtful Expression (Written) Percent of Sample Scored at Least 2 and at Least 3........61
Table 8 Inquiry Percent of Sample Scored at Least 3.............................................................................62
Table 9 Oral Communication Percent of Sample Scored at Least 3.....................................................62
EXECUTIVE SUMMARY

This report provides the results of the General Education Assessment efforts for Fall 2010 and Spring 2011. The UNCW Learning Goals were assessed within Basic Studies courses and courses in the majors using AAC&U VALUE Rubrics and one locally created rubric, using the process recommended by the General Education Assessment Committee’s March 2009 recommendations. The Basic Studies sample consisted of 231 student work products from the following courses: HST 102, HST 103, HST 201, ENG 201, and WMS 210. The samples from the majors consisted of 116 student work products from NSG 327, PLS 401, and PSY 410.

BASIC STUDIES FINDINGS

CRITICAL THINKING
Student work products scored highest on two dimensions: CT1 Explanation of Issues and CT2.1 Selecting and Using Evidence. The CT 2.2 Critically Examining Evidence dimension was seen as not applicable to three of the five assignments scored, and scores were lowest on this dimension for those work products scored. The total number of hours completed by students was positively correlated with all dimensions except Critically Examining Evidence.

DIVERSITY
The score distributions on the two dimensions of this rubric were similar, although the scores for DV2 Contextual Importance and Implications tended to be slightly lower than those for DV1 Knowledge of Diverse Perspectives. The total number of hours completed and number of transfer hours were both positively correlated with DV2, indicating that skill in this dimension increases with number of hours completed (transfer students had on average more credit hours completed).

INFORMATION LITERACY
The scores for IL5 Access and Use Information Ethically were generally higher than for the other four dimensions, and lowest for IL3 Evaluate information and sources critically. There was a statistically significant difference between the performance of transfer students and freshman-start students on the Evaluate Information and Sources dimension.

THOUGHTFUL EXPRESSION (WRITTEN)
WC1 Context of and Purpose for Writing, WC4 Sources and Evidence, and WC5 Control of Syntax and Mechanics were the three higher-scoring dimensions. WC2 Content Development and WC3 Genre and Disciplinary Conventions showed slightly lower scores, with fewer student work products scoring threes and fours for these two dimensions. There were no statistically significant differences between the score distributions of transfer vs. freshman-start students, honors vs. non-honors students, or male vs. female students, and there was no statistically significant correlation between any dimension and total hours. However, UNCW hours was statistically significantly negatively correlated with three dimensions.
FINDINGS FROM PILOTS IN THE MAJORS

INFORMATION LITERACY
Scores were highest on IL2 Access Needed Information and IL4 Use information to accomplish a purpose, and lowest on IL3 Evaluate Information and Sources. There were no statistically significant differences between the score distributions of transfer vs. freshman-start students, honors vs. non-honors students, or male vs. female students. Hours completed was statistically significantly positively correlated with IL4 Use information to accomplish a purpose.

WRITTEN COMMUNICATION
Scores were lowest on WC2 Content Development. There were no statistically significant differences between means, medians, or the score distributions of transfer vs. freshman-start students, honors vs. non-honors students, or male vs. female students. Hours completed was statistically significantly positively correlated with WC1 Context of and Purpose for Writing, WC3 Genre and Disciplinary Conventions, and WC4 Sources and Evidence.

INQUIRY
The percent of performances scoring 3 or above was highest for IN1 Topic Selection and IN2 Existing Knowledge, Research, and/or Views. Demographic and preparedness information was not available for this sample.

ORAL COMMUNICATION
The percent of performances scoring 3 or above was highest for OC5 Central Message and lowest for OC3 Delivery. Scores on all dimensions were statistically significantly correlated with the number of UNCW Hours completed, but not with total hours completed. Scores on three dimensions were statistically significantly negatively correlated with three dimensions (OC1, OC4, and OC5). In addition, transfer student scores were statistically significantly lower that freshman starters for the same dimensions.

RECOMMENDATIONS
The following recommendations were adopted by the Learning Assessment Council (see page 66 for further information.

• Continue efforts to improve interrater reliability
• Present targeted content workshops where we invite on-campus experts in various disciplines to share some of the specific ways they are introducing the UNCW Learning Goals into their courses
• Introduce Learning Outcomes and guidelines for Explorations Beyond the Classroom and Thematic Transdisciplinary Clusters which will include a synthesis of information and aspects related to critical thinking. This will increase exposure to these learning goals.
• Develop a central information site to share summaries of reports, best practices, and other teaching and assessment resources.
• Reemphasize the desirability of including the relevant learning outcomes on syllabi of courses approved for the new University Studies curricula (or at least providing links to these goals on the syllabi).
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 2010 – 2011. In Fall 2010 and Spring 2011, the following learning goals were assessed: Critical Thinking, Diversity, Information Literacy, Inquiry, and Thoughtful Expression (both Oral and Written). This report outlines the methodology of and findings from five separate studies, and provides useful information on the abilities of UNCW students as measured during their basic studies as well upper-division courses, as seen through course-embedded assignments. This report also provides follow up information on the progress made on recommendations made last year.
METHODOLOGY

For the purposes of this report, general education assessment activities in academic year 2010 – 2011 are divided into two areas: assessment of student learning within the current Basic Studies courses, and assessment of student learning in upper-division courses.

Within the Basic Studies courses, the following questions were examined:

- What are the overall abilities of students taking basic studies courses with regard to the UNCW Learning Goals of Critical Thinking, Diversity, Information Literacy, 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 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?

Within the upper division courses, the following questions were examined:

- What are the abilities of students taking upper-level courses with regard to the UNCW Learning Goals of Information Literacy, Inquiry, Thoughtful Expression (Written), and Thoughtful Expression (Oral)?
- What are the relative strengths and weaknesses within the subskills of those goals?
- Are there any differences in performance based on 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. A 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 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 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 all UNCW Learning Goals except Diversity, the Association of American Colleges and Universities (AAC&U) Valid Assessment of Learning in Undergraduate Education (VALUE) rubrics (Rhodes, 2010) were used. 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.

- For Critical Thinking, the VALUE Critical Thinking rubric, with one modification, was used;
- for Information Literacy, the VALUE Information Literacy rubric was used;
- for Thoughtful Expression (Written), the VALUE Written Communication rubric was used;
- for Thoughtful Expression (Oral), the VALUE Oral Communication rubric was used; and
- for Inquiry, the VALUE Inquiry and Analysis rubric was used.

A locally created rubric was piloted for assessing Diversity. Appendix A contains the versions of each of the rubrics that were used in the study.

SAMPLE SELECTION

The sampling method used lays the foundation for the generalizability of the results. No one part of the basic 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. A Fall 2008 survey helped determine which basic studies courses are most appropriate for assessing each of these goals. Program assessment reports, which include student learning outcomes aligned to the UNCW Learning Goals, are used to determine appropriate courses in the majors. For assessment within basic studies, courses were selected that not only met the learning goals, but were 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 were 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 were 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 were collected, and random samples of the work products were selected (sometimes consisting of all papers).

Prior to the start of the semester, the General Education Assessment Director met with course instructors to familiarize them with the relevant VALUE rubrics. Instructors were asked to review their course content and assignments, and to select one assignment that they felt fit the dimensions of the rubric(s) being used.

Each student enrolled in the selected course sections filled out a Student Work Product Cover Sheet, which acknowledged the use of their work for the purpose of general education assessment. These cover sheets were removed before scoring. The name and student ID information on the cover sheets was matched with student demographic information in university records for the purpose of analysis based on demographic and preparedness variables.

Sample selection for the three pilot studies in upper division courses was similar. In each case, we sampled courses that were part of the majors and that addressed the learning goals we wished to examine—Information Literacy, Thoughtful Expression (Written Communication), Inquiry, and Thoughtful Expression (Oral Communication). From each class, either all student work was used, or student work was randomly sampled.

**SCORING**

**SCORER RECRUITMENT AND SELECTION**

Scorers were recruited from UNCW faculty. A recruitment email was sent to chairs in the humanities on November 29, 2010, with a request that it be forwarded to all department faculty, for scoring of work collected in the fall for assessing Critical Thinking and Diversity. On March 23, 2011, a recruitment email for scoring of work collected in the spring for assessing Information Literacy and Written Communication (including the Pilot 1 papers) was sent to all chairs (in all colleges and schools) with a request that it be forwarded to all department faculty. The desire was to include reviewers from a broad spectrum of departments. The intent was to give faculty who do not teach in departments that offer basic studies courses the opportunity to see the work being done by students in the general education courses, and to give faculty who teach upper-level courses such as capstone courses within departments that do offer general education courses the opportunity to see the learning students experience as they begin their programs, as well as faculty who do teach basic studies courses. It was also important to have at least one faculty member from each of the departments from which student work products were being reviewed. Scorers were selected from those expressing an interest to make up a broad-
based panel consisting of full-time and part-time faculty. For the second pilot study, the scoring team was made up of the department assessment committee. For the third pilot study, the scoring team was made up of two course faculty, the Director of Assessment, and the General Education Research Assistant.

**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. But 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.

Volunteer scorers initially attended a two-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.

Additional assumptions were necessary for the Information Literacy rubric.

- Consistent use of citations and references means that citations are used when appropriate and provide enough information for the reader to find the source.
- “Common knowledge” refers to the common knowledge we would expect the intended audience to have.
- If there is evidence of plagiarism or other unlawful or unethical use of material, work will be scored with a zero for [rubric] Dimension 5 and not scored on the other dimensions.

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.

The work on common assignment-specific assumptions or guidelines was continued on the day of scoring. Scorers were assigned to groups of two or three. 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. 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.
RESULTS FOR BASIC STUDIES COURSES

A NOTE ABOUT THE PRESENTATION OF FINDINGS

Data from rubrics is ordinal, not interval-level, data. That is, although a level 2 is considered higher, or larger, than a level 1, it is not proper to assume that a student that scores at a level 2 is twice as knowledgeable as a student who scored at a level 1; nor can we assume that, whatever the difference is between these two categories, that it is exactly the same as the difference between levels 2 and 3. Therefore the findings in the body of this report are expressed as percentages of work products scored at each level. Medians and other percentiles are also given. Means and standard deviations are often provided by researchers for ordinal data. This information is given in Appendix B, as it may be helpful to some readers as a starting point to suggest further investigation using statistical methods appropriate to ordinal data.

DESCRIPTION OF SAMPLE

DESCRIPTION OF COURSES

A total of 231 student work products were collected from 12 course sections of the following 5 courses:

- HST 102 Western Civilization II (2 sections taught by 2 instructors)
- HST 103 Introduction to Global History (2 sections taught by 2 instructors)
- HST 201 American History I (2 sections taught by 2 instructors; 1 online section)
- ENG 201 College Writing and Reading II (5 sections taught by 4 instructors)
- WMS 210 Introduction to Women’s Studies (1 section taught by 1 instructor)

To the extent possible, the breakdown of sections taught by tenure-line faculty, lecturers, and part-time faculty were selected to be representative of those breakdowns for the course as a whole during the semester sampled. For HST 102 sections in Fall 2010, the majority (66.7%) of the total sections offered were taught by tenure-line faculty; one-third were taught by temporary instructors. The two sections sampled in Fall 2010 were both taught by tenure-line faculty. For History 103 and 201, all sections were also taught by tenure-track faculty, so all of our History course samples for the general education assessment process were taken from sections taught by tenure-line faculty. Of the Spring 2011 ENG 201 course sections, 39.4% were taught by part-time instructors, 30.3% were taught by teaching assistants, 24.2% were taught by lecturers, and 6.1% were taught by tenure-line faculty. Our sample was drawn from 20% TA-led sections, 40% part-time instructor-led sections, 20% professor-led sections, and 20% lecturer-taught sections. While these ratios are not precisely like the overall breakdown of the English 201 sections, this sample does allow sampling from courses led by each sort of teacher type. For WMS 210 in Spring 2011, there were two sections offered, both led by temporary instructors. We sampled from one of these sections.
**Sample by Rubric**

Work products were randomly selected from the above-described sample for scoring. The total number of work products in the final sample scored using each rubric was:

- Critical Thinking: 114 work products scored by 10 scorers
- Diversity: 113 work products scored by 10 scorers
- Information Literacy: 100 work products scored by 11 scorers
- Written Communication: 80 work products scored by 12 scorers

The total number of scores produced was larger than the total number of work products because 113 work products were scored using both the Critical Thinking and Diversity rubrics, and 63 were scored using both the Information Literacy and Written Communication rubrics. No work products were scored using more than two rubrics.

**Description of Students**

The 231 work products were produced by 231 unique students. Samples were randomly selected from course sections based on the average amount of time required to score the student work products and the number of scorers. The demographic breakdown of the participating students, compared in parenthesis to the overall undergraduate enrollment for AY 2010-2011 was:

- 57.6% (59.8%) female;
- 15.2% (10.9%) transfer students;
- 4.3% (4.0%) honors students;
- 5.2% (4.3%) African American;
- 0.9% (0.6%) American Indian;
- 0.9% (1.9%) Asian;
- 4.3% (4.6%) Hispanic;
- 1.3% (1.8%) of multiple race or ethnicity;
- 0.0% (1.4%) non-resident alien;
- 0.4% (0.1%) of Hawaiian or Pacific Island ethnicity;
- 84.4% (83.2%) white;
- 2.6% (2.3%) listed unknown or other ethnicity (UNCW OIRA, 2010). The only group that was not representative of all UNCW students is the percent of transfer students.

For those students with SAT score information (193), the mean Total SAT score was (compared in parenthesis to the overall undergraduate enrollment for AY 2010-2011) 1139 (1160), the mean SAT Math was 579 (590), and the mean SAT Verbal was 560 (570), which is slightly below the 50% percentile for Fall 2010 freshmen (UNCW OIRA, 2011). For those who took the ACT college placement test (60), the mean composite score was 24, which is the same as the 50% percentile for Fall 2010 freshmen (UNCW OIRA, 2011).

The mean total number of credit hours students had completed prior to the semester during which their work was collected was 41.5. It is important to note that there were a number of outliers with over 120 total hours (maximum was 155). The median number of hours was 44. This includes both UNCW hours and transfer hours. The median UNCW hours was 28 (mean 27.7), and the median transfer hours was 9 (mean 16.8). Broken down into groups, 33.8% had completed between 0 and 30 hours, 47.2% had completed between 31 and 60 hours, 12.5% had completed between 61 and 90 hours, and 6.5% had completed 91 or more hours.
CRITICAL THINKING

At the basic studies level, 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, 2009). The VALUE Critical Thinking rubric contains five dimensions that are aligned with the UNCW description of Critical Thinking. Based on feedback from scorers in Spring 2010, dimension 2 of the rubric was divided into two parts.

SUMMARY OF SCORES BY DIMENSION

Ten faculty scorers scored 114 work products from three courses from the Fall 2010 semester, HST 102, HST 103, and HST 201. Thirty-six work products (31.6%) were scored by multiple scorers. Scorers determined that CT2.2 was not applicable for two assignments (48 work products), and CT3 and CT4 were not applicable for one assignment (16 work products). Figure 1 provides the score distributions for each dimension for work products that were scored on that dimension (i.e., work products scored as NA are not included).
CRITICAL THINKING RESULTS BY DIMENSION FOR APPLICABLE SCORES ONLY

RESULTS BY DIMENSION
CT1 Explanation of Issues
This dimension was scored for all of the assignments. Scores on this dimension were the highest of all dimensions of Critical Thinking (along with CT2.1 Selecting and Using Evidence). Only three out of one hundred work products provided no explanation of the issue (scores of 0). Almost one in four work products stated the issue or problem with no clarification (scores of 1). Over one third of work products stated the issue or problem, but left some points ambiguous (scores of 2). Slightly more than one in three work products stated, described, and clarified the issue or problem (scores of 3 and 4).
CT2.1 Evidence; Selecting and Using
This dimension was scored for all assignments. Scores on this dimension were the highest of all dimensions of Critical Thinking (along with CT1 Explanation of Issues). Fewer than two out of one hundred work products provided no evidence (scores of 0). Slightly more than one quarter of the work products provided evidence that was taken from sources without evaluation or interpretation (scores of 1). Over one third of students provided evidence with some interpretation, but not enough to develop a coherent analysis or synthesis (scores of 2). Almost four in ten work products provided evidence that was interpreted, analyzed (score of 3) or synthesized (score of 4).

CT2.2 Evidence; Critically Examining
This dimension was deemed applicable for three out of the five assignments, resulting in 66 work products being scored for this dimension. Almost one in twenty work products provided no evidence (scores of 0). Over half of the work products scored provided evidence from experts that was taken as fact, without question (scores of 1). Just over one quarter of the work products demonstrated that viewpoints of experts were taken as mostly fact, with little questioning (scores of 2). Almost one in seven work products indicated that the viewpoints of experts were subject to questioning (scores of 3 and 4).

CT 3 Influence of Context and Assumptions
This dimension was viewed as applicable and was scored for four of the five assignments, resulting in 98 work products being scored for this dimension. Scores on this dimension were in the middle range of scores. Just over one out of ten work products demonstrated no awareness of assumptions (scores of 0). Three out of ten work products showed an emerging awareness of assumption and some identification of context (scores of 1). One quarter of work products questioned some assumptions (but overlooked others) and identified some relevant context (scores of 2). One third of the work products identified the student’s own and others’ assumptions as well as several relevant contexts (scores of 3 and 4).

CT4 Student’s Position
Like CT3, this dimension was scored for four of the five assignments (98 work products). Scores on this dimension were some of the lower scores, along with CT5 Conclusions and Related Outcomes. Four out of one hundred work products contained no statement of student position (scores of 0). Over one third of the work products provided a simplistic or obvious position (scores of 1). Just over than one quarter of work products provided a specific position that acknowledged different sides of an issue (scores of 2). Slightly more than one third of products not only acknowledged difference sides of an issue, but incorporated those positions and took into account the complexities of the issue (scores of 3 and 4).
CT5 Conclusions and Related Outcomes
This dimension was scored for all five assignments. Scores on this dimension were the lowest of the dimension scores. Just under half of the work products provided no conclusions (scores of 0) or provided oversimplified outcomes and conclusions that were inconsistently tied to some of the information discussed (scores of 1). More than one third of work products provided conclusions that were logically tied to information (because information was chosen to fit the conclusion) and identified some related outcomes clearly (scores of 2). Approximately one fourth of the work products provided conclusions logically tied to a range of information, including opposing viewpoints and identified related outcomes clearly (scores of 3 and 4).

Correlation Between Dimensions
All dimension scores were correlated with each other at the .01 level of significance. The magnitudes of correlations range from .443 to .836, with the highest correlation between CT4 Student’s Positions and CT2.1 Selecting and Using Evidence. This finding seems appropriate as a student’s position is developed partially through the evidence selected. See Appendix table C1 for a complete presentation of correlation coefficients. The large and statistically significant correlations between the scores on each dimension of the rubric may suggest a lack of independent scoring on the part of the scorers; however; they may simply represent the interdependence among all aspects of critical thinking.

Demographic and Preparedness Findings
There were no statistically significant difference between the means, medians, and the score distributions of transfer vs. freshman-start students, honors vs. non-honors students, or males vs. females. The samples of students with race/ethnicities other than white were too small to compare the groups.

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 (60.5% of the sample), 31 – 60 credit hours (28.1% of the sample), 61 – 90 (7.9% of the sample), and over 90 credit hours (3.5% of the sample). 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 CT2.1, CT3, CT4, and CT5, with all means higher for sophomores than freshman, but only CT1 and CT2.1 showed means increasing across all four categories. However, the sample contained only 9 juniors and 4 seniors. Looking at Spearman rho correlation coefficients, the number of total hours completed was positively correlated with five of the six dimensions (CT1 .202*, CT2.1 .303**, CT3 .262**, CT4 278**, and CT5 .273**); UNCW was positively correlated with CT1 (.188*), CT2.1 (.200*), CT3 (.219*), and CT4 (.202*); and transfer hours was positively correlated with CT2.1 (.266**), CT3 (.203*), and CT5 (.261**).
SAT-Verbal was positively correlated with CT1 (.271**) and CT2.2 (.347**). There were no significant correlations with GPA, ACT, or SAT-Math.

**Comparison Between Courses and Assignments**

All assignments came from 100- and 200-level history courses. One course was an on-line course. There were statistically significant differences in the means, median, and distributions of only two of the six dimensions, CT1 and CT2.2, between the classroom-based courses and the online course, with the classroom-based scores higher. One set of artifacts were from an in-class essay exam and the rest were term papers. There were statistically significant differences in the means and distributions of CT1, CT2.1, CT3, and CT4 between the term papers and the in-class essay (the in-class exam essay was not scored on CT 2.2), with the term papers scores higher.

**Diversity**

At the basic studies level, the UNCW Diversity Learning Goal is for students to describe the importance and implications of human diversity. For the purposes of this Learning Goal, “Diversity constitutes the knowledge, skills, and attitudes necessary to examine the importance and implications of cultural and ethnic human differences. Diversity examines the significance of historical, political, social, racial, ethnic, and cultural realities through critical thinking to understand and explain their implications in human endeavors” (UNCW Learning Goals, 2009).

**Summary of Scores by Dimension**

Ten faculty scorers scored 113 work products from three courses from the Fall 2010 semester, HST 102, HST 103, and HST 201. Thirty-five work products (31.0%) were scored by multiple scorers. Scorers determined that DV2 was not applicable for two papers. Figure 2 provides the score distributions for each dimension for work products that were scored on the Diversity dimensions.
**RESULTS BY DIMENSION**

**DV1 Knowledge of Diverse Perspectives**

This dimension was scored for all five assignments. Scores on this dimension were slightly higher than for DV2, Contextual Importance and Implications. Almost one in ten work products failed to identify facets of the perspective(s) being studied (scores of 0). Over one quarter of the work products began to identify facets of the perspective(s) being studied although the connection was vague (scores of 1). Just over a third of the work products identified some facets of the perspectives(s) being studied (score of 2). Three in ten work products discussed many significant facets or examined the complexity of the facets of the perspectives being studied (scores of 3 and 4).

**DV2 Contextual Importance and Implications**

This dimension was scored for all five assignments, though scorers indicated that this dimension was not applicable for the work products of two students. One in seven work products showed no
awareness of the importance and/or implications of human diversity within the disciplinary context (scores of 0). Slightly more than one in four work products showed a vague awareness of the importance and/or implications of human diversity within the disciplinary context (scores of 1). Just over one quarter of the work products identified some factors that illustrate the importance and implications of human diversity (scores of 2). About three out of ten work products either discussed factors that establish the importance and implications of human diversity with the disciplinary context (score of 3) or thoroughly examined and established the importance and implications of human diversity (score of 4).

**Correlation Between Dimensions**
The two Diversity dimension scores were correlated with one another at the .01 level, with a correlation coefficient of .859. See Appendix table C1 for a complete presentation of the correlation coefficients for the Diversity dimension scores.

**Demographic and Preparedness Findings**
There were no statistically significant differences between the means, medians, and the score distributions of honors vs. non-honors students, transfer vs. freshman-start students, or males vs. females. There were significant differences between means, medians, and score distributions of transfer students vs. freshman-start students for both DV1 and DV2. The samples of students with race/ethnicities other than white were too small to compare the groups.

To compare scores based on number of credit hours completed, two methods were used. First, students were grouped into four categories, having completed 0 – 30 credit hours (61.1% of the sample), those 31 – 60 credit hours (27.4% of the sample), 61 – 90 (8.0% of the sample), and over 90 credit hours (3.5% of the sample). 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. Looking at Spearman rho correlation coefficients, the number of total hours completed was positively correlated with DIV2 (.247**), and transfer hours was positively correlated with DIV2 (.230*).

There were no significant correlations with GPA, ACT, SAT-Verbal, or SAT-Math.

**Comparison Between Courses and Assignments**
All assignments came from 100- and 200-level history courses. One course was an on-line course. There were no statistically significant differences in the means, median, and distributions of DV1 and DV2 between the classroom based courses and the online course. One set of artifacts were from an in-class essay exam and the rest were term papers. There were statistically significant differences in the means and distributions of DV1 and DV2 between the term papers and the in-class essay, with the term papers scores higher.
INFORMATION LITERACY

At the basic studies level, the UNCW Information Literacy Learning Goal is for students to locate and effectively apply information using academic and technological skills (UNCW Learning Goals, 2009). For the purposes of this Learning Goal, “Information Literacy is the ability ‘to recognize when information is needed and to locate, evaluate and use effectively the needed information.’” (American Library Association, 1989). Demonstrating information literacy involves determining the extent of 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. The VALUE Information Literacy rubric contains five dimensions that are aligned with the UNCW description of Information Literacy.

SUMMARY OF SCORES BY DIMENSION
Eleven faculty scorers scored 100 work products from one course from the Spring 2011 semester, ENG 201. Sixteen work products (16.0%) were scored by multiple scorers. All five assignments were deemed as applicable for scoring using the Information Literacy rubric. There was one blank score for one work product, on dimension IL 2. Figure 3 provides the score distributions for each dimension for work products that were scored on the Information Literacy dimensions.
RESULTS BY DIMENSION

IL1 Extent of Information Needed

Only one out of 100 work products were scored a zero. One in ten student work products indicated difficulty defining the scope of the research question or thesis (score of 1). One in four work products incompletely defined the scope of the research question or thesis but did determine some key concepts (score of 2). Over three in ten of the work products defined the scope of the research question or thesis completely and determined key concepts (score of 3). Two in ten work products effectively defined the scope of the research question and effectively determined key concepts (score of 4).
IL2 Access Needed Information
Just fewer than one in ten work products either showed no evidence of accessing the needed information, or accessed information randomly or retrieved information that lacked relevance and quality (scores of 0 or 1). Slightly more than four out of ten student work products accessed information using simple search strategies and retrieved information from some relevant, though limited and similar sources (scores of 2). Over four out of ten of the work products accessed information using a variety of search strategies and from relevant sources, while also demonstrating the ability to refine the search strategy (scores of 3). Just fewer than one in ten work products accessed information using effective, well-designed search strategies and from the most appropriate information sources (scores of 4).

IL3 Evaluate Information and Sources
Scores for this dimension of Information Literacy tended to be slightly lower than those for the other dimensions. One out of 100 work products showed no evidence of evaluating information and it sources critically (scores of 0). One in ten work products indicated that information was taken from sources without any interpretation or evaluation of the materials and that the viewpoints of the authors were not subject to questioning (scores of 1). Just under half of the work products indicated that information was taken from source(s) with some interpretation/evaluation, but did not include a coherent analysis of the material and viewpoints of the authors were taken mostly as fact (scores of 2). Slightly more than four out of ten work products demonstrated that information was taken from source(s) with enough interpretation or evaluation to develop a coherent analysis or comprehensive analysis of the material, with viewpoints of authors questioned (scores of 3 or 4).

IL4 Use Information Effectively
As with the preceding Information Literacy dimensions, one out of 100 assignments received a score of zero for this dimension, indicating that the work products did not use information effectively to accomplish the specified purpose. One in ten work products communicated information from sources, but that information was fragmented and/or used inappropriately, with the intended purpose not being achieved (scores of 1). Four out of ten papers communicated and organized information from sources, but that information was not fully synthesized (scores of 2), and the same proportion of papers communicated, organized, and synthesized information from sources, achieving the intended purpose (score of 3) or did so with clarity and depth (score of 4).

IL5 Access and Use Information Ethically
This dimension deals with ethical use of information and tended toward higher scores than the other four Information Literacy dimensions. Three out of 100 work products received a score of zero. For any score for this dimension greater than zero, 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. Fewer than ten out of a hundred work products received scores of 1, indicating the work 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). One third of the work products demonstrated two types of information-use strategies (scores of 2). Four out of ten work products showed evidence of three types of information-use strategies (scores of 3). One fifth of the work products showed evidence of all types of information-use strategies (scores of 4).

**Correlation Between Dimensions**

All dimensions scores were correlated with each other at the .01 level. See Appendix table C2 for a complete presentation of correlation coefficients.

**Demographic and Preparedness Findings**

There were no significant differences between honors vs. non-honors students, or males vs. females. There was a statistically significant difference between the means, medians, and score distributions of transfer vs. freshman-start students for one dimension only—IL3, with scores for transfer students higher. The higher scores for transfer students might be explained by the fact that transfer students had, on average, more credit hours completed, with 66.7% of the transfer students having earned over 60 credits, while only 18.9% of the freshman-start students had earned over 60 credits (although the results on hours completed discussed below does not bolster this hypothesis). The samples of students with race/ethnicities other than white were too small to compare the groups.

To compare scores based on number of credit hours completed, two methods were used. First, students were grouped into four categories, having completed 0 – 30 credit hours (3% of the sample), those 31 – 60 credit hours (74% of the sample), 61 – 90 (16% of the sample), and over 90 credit hours (7% of the sample). 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 for any dimensions. However, calculating the Spearman rho correlation coefficient between scores and UNCW hours, transfer hours, and total hours showed a number of statistically significant correlations: UNCW hours with IL2 (-.343**) and IL5 (-.259*); transfer hours with IL2 (.276*); and total hours with IL3 (-.228*) and IL5 (-.230*).

IL2 was positively correlated with students’ GPA (.355**) and SAT-Verbal scores (.227*). Students’ GPA was also positively correlated with IL3 (.230*) and IL4 (.280**).
COMPARISON BETWEEN COURSES AND ASSIGNMENTS
Student work was only sampled from ENG 201, hence there are no course comparisons. Student work was sampled from online courses (80) and classroom-based courses (20). All sections were classroom-based, and all assignments were done out-of-class, so there is no comparison between assignment types in that regard.

THOUGHTFUL EXPRESSION (WRITTEN)
At the basic studies level, the UNCW Thoughtful Expression 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, 2009). The VALUE Written Communication rubric contains five dimensions that are aligned with the UNCW description of Thoughtful Expression.

SUMMARY OF SCORES BY DIMENSION
Twelve faculty scorers scored 80 work products from two courses from the Spring 2011 semester, ENG 201 and WMS210. Twenty-seven work products (33.8%) were scored by multiple scorers. All assignments were scored on each dimension (no dimension was considered not applicable for any assignment). Figure 4 provides illustration of the score distributions for each dimension.
RESULTS BY DIMENSION

WC1 Context of and Purpose for Writing

This dimension was one of the higher scoring Written Communication dimensions, along with WC4 and WC5. No work products demonstrated complete lack of attention to context, audience, purpose and to the assigned task (scores of 0). Just less than a quarter of the work products demonstrated minimal attention to context, audience, purpose, and to the assigned task (scores of 1). Over one third of work products demonstrated awareness of the context, audience, purpose, and assigned task (scores of 2). Slightly more 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). One in ten 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**

One out of 80 work products demonstrated no content development (scores of 0). Three in ten work products used appropriate and relevant content to develop simple ideas in some parts of the work (scores of 1). Four out of ten the work products used appropriate and relevant content to develop and explore ideas through the work (scores of 2). One quarter of the work products used appropriate, relevant and compelling content to explore ideas within the context of the discipline (scores of 3 and 4).

**WC3 Genre and Disciplinary Conventions**

Scores on this dimension were in the mid-range of scores. None of the work products failed to show an attempt to use a consistent system for organization and presentation (scores of 0). Slightly more than one in five work products demonstrated an attempt to use a consistent system for basic organization and presentation (scores of 1). Almost half of the work products followed expectations appropriate to the specific writing task for basic organization, content, and presentation (scores of 2). Three out of ten work products demonstrated consistent use of important conventions particular to the writing task, including stylistic choices (scores of 3 and 4).

**WC4 Sources and Evidence**

The scores on this dimension were generally high; this dimension received the highest total number of scores in the 3 and 4 category. However, this dimension also had the highest number of “0” scores: about one in eleven demonstrated no attempt to use sources to support ideas. Two out of eleven work products demonstrated an attempt to use sources to support ideas (scores of 1). One in three work products demonstrated an attempt to use credible and/or relevant sources to support ideas that were appropriate to the task (scores of 2). More than four in ten work products demonstrated consistent (score or 3) or skillful (score of 4) use of credible, relevant sources to support ideas.

**WC5 Control of Syntax and Mechanics**

No work products failed to meet the level 1 benchmark (score of 0). Approximately two in twelve work products used language that sometimes impeded meaning because of errors in usage (score of 1). Over four in ten work products used language that generally conveyed meaning with clarity, although writing included some errors (score of 2). Over one third of work products used straightforward language that generally conveyed meaning, with few errors (score of 3). One in 25 work products used graceful language that skillfully communicated meaning with clarity and fluency, with virtually no errors (score of 4).
**Correlation Between Dimensions**

All dimension scores were correlated with each other at the .01 level of significance, with a range of correlations from .398 to .729. See Appendix table C2 for the complete presentation of correlation coefficients for the Written Communication dimensions. The large and statistically significant correlations between the scores on each dimension of the rubric may suggest a lack of independent scoring on the part of the scorers; however, they may simply represent the interdependence among all aspects of written communication.

**Demographic and Preparedness Findings**

There were no statistically significant differences between the score distributions of transfer vs. freshman-start students, honors vs. non-honors students, or male vs. female students. The samples of students with race/ethnicities other than white were too small to compare the groups.

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 (10.0% of the sample), 31 – 60 credit hours (63.8% of the sample), 61 – 90 (16.3% of the sample), and over 90 credit hours (10.0% of the sample). Comparison of means (using ANOVA), medians (using Independent Samples Test of Medians), and distributions (using the Mann-Whitney U statistic) showed no statistical differences. Calculating the Spearman rho correlation coefficient between scores and total as well as transfer hours completed showed no statistically significant correlations. However, UNCW hours was negatively correlated with WC1 (-.224*), WC2 (-.289**), and WC4 (-.237*).

SAT-Verbal scores were positively correlated with all dimensions: WC1 (.287*), WC2 (.271*), WC3 (.282*), WC4 (.377**), and WC5 (.337**). Students’ GPA was positively correlated with WC1 (.269*), WC2 (.294**), WC4 (.405**), and WC5 (.275*). No dimensions were correlated with ACT or SAT-Math scores.

**Comparison Between Courses and Assignments**

Student work was sampled from ENG 201 (64) and WMS 210 (16). No statistically significant differences in the means, medians, or distributions of the two groups were found. All sections were classroom-based and all assignments were done out-of-class, hence there is no comparison between assignment types in that regard.
RESULTS FOR UPPER-DIVISION COURSES

PILOT 1 INFORMATION LITERACY AND THOUGHTFUL EXPRESSION (WRITTEN)

In the first pilot, student work products from one section of NSG 327 Clinical Reasoning/Scientific Inquiry were scored by 12 scorers using the Information Literacy and Written Communication rubrics.

DESCRIPTION OF SAMPLE
The sample of 45 students consisted of: 88.9% females, which is above the percent of females at the institution (60%); 93.3% Caucasian, which is greater than that the institution (83.9%); 37.8% transfer students, which is greater than the percent of transfer students at the institution (28.0%); and 6.7% honors students, is also greater than the percentage of honors students at the institution (4.0%). It is to be expected that the proportion of transfer students is higher in the junior than the average across all four years. The average SAT Verbal was 553, which was lower than that the Fall 2010 incoming freshmen (570), and the average SAT Math was 575, also lower than that of the Fall 2010 incoming freshmen (590).

INFORMATION LITERACY RESULTS
Forty work products were scored on the Information Literacy rubric by seven scorers. Twelve products were double, triple, or quadruple scored. IL3 Evaluate Information and Its Sources Critically and IL4 Use Information to Accomplish a Specific Purpose were left blank for the 5 and 14 work products, respectively. Figure 5 provides the score distributions for all applicable (non-blank) scores.
**PILOT 1 INFORMATION LITERACY RESULTS BY DIMENSION FOR APPLICABLE SCORES ONLY**

![Bar chart showing distribution of scores for Pilot 1 Information Literacy](image)

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2.5%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>1</td>
<td>17.5%</td>
<td>2.5%</td>
<td>11.4%</td>
<td>0.0%</td>
<td>20.0%</td>
</tr>
<tr>
<td>2</td>
<td>27.5%</td>
<td>27.5%</td>
<td>65.7%</td>
<td>23.1%</td>
<td>37.5%</td>
</tr>
<tr>
<td>3</td>
<td>32.5%</td>
<td>52.5%</td>
<td>22.9%</td>
<td>65.4%</td>
<td>35.0%</td>
</tr>
<tr>
<td>4</td>
<td>20.0%</td>
<td>17.5%</td>
<td>0.0%</td>
<td>11.5%</td>
<td>7.5%</td>
</tr>
<tr>
<td>25th%tile</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>50th%tile</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>75th%tile</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Mode</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

Figure 5 Distribution of Scores for Pilot 1 Information Literacy, Applicable Scores Only

**Results by Dimension**

**IL1 Extent of Information Needed**

Only one of the 40 work products was scored a zero on this dimension. Around two in ten student work products indicated difficulty defining the scope of the research question or thesis (score of 1). Slightly more than one in four work products incompletely defined the scope of the research question or thesis but did determine some key concepts (score of 2). Three out of ten work products defined the scope of the research question or thesis completely and determined key concepts (score of 3). Two in ten effectively defined the scope of the research question and effectively determined key concepts (score of 4).
IL2 Access Needed Information
No student work products received a score of zero for this dimension. Just under one in three in work products accessed information randomly or retrieved information that lacked relevance and quality (score of 1). Slightly more than one in four student work products accessed information using simple search strategies and retrieved information from some relevant, though limited and similar sources (score of 2). Over half of the work products accessed information using a variety of search strategies and from relevant sources, while also demonstrating the ability to refine the search strategy (score of 3). Just fewer than two in ten work products accessed information using effective, well-designed search strategies and from the most appropriate information sources (scores of 4).

IL3 Evaluate Information and Sources
Scores for this dimension of Information Literacy were not assigned for five student work products and tended to be lower than those for the other dimensions. No work products failed to show any evidence of evaluating information and it sources critically (score of 0). One in ten work products indicated that information was taken from sources without any interpretation or evaluation of the materials and that the viewpoints of the authors were not subject to questioning (score of 1). Over half of the work products indicated that information was taken from source(s) with some interpretation/evaluation, but did not include a coherent analysis of the material and viewpoints of the authors were taken mostly as fact (score of 2). Slightly less than two out of ten work products demonstrated that information was taken from source(s) with enough interpretation or evaluation to develop a coherent analysis of the material, with viewpoints of authors questioned (score of 3). No student work products received scores of four for this dimension.

IL4 Use Information Effectively
This dimension was scored for 26 of the 40 work products (those scorers determined the assignment did not require the student to use the information for a specific purpose). No student work products received a score of zero or one for this dimension. One out of seven papers communicated and organized information from sources, but that information was not fully synthesized (score of 2). Four in ten papers communicated, organized, and synthesized information from sources, achieving the intended purpose (score of 3). One out of fourteen papers communicated, organized, and synthesized information from sources, achieving the intended purpose with clarity and depth (score of 4).

IL5 Access and Use Information Ethically
This dimension deals with ethical use of information. No work products received a score 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. Two in ten work products received scores of 1, indicating the work 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). Over one third of the work products demonstrated two types of information-use strategies (score of 2). One third of the work products showed evidence of three types of information-use strategies (score of 3). One out of fourteen work products showed evidence of all types of information-use strategies (score of 4).

**Correlation between Dimensions**

All dimensions of Information Literacy were significantly correlated with each other except IL2 with IL3, IL2 with IL 5, and IL4 with IL5. The significant correlation coefficients range in magnitude from .337 to .658. Appendix table C3 contains all the IL correlation coefficients.

**Demographic and Preparedness Findings**

There were no statistically significant differences between the means, median, or score distributions of transfer vs. freshman-start students, honors vs. non-honors students, or male vs. female students. The samples of students with race/ethnicities other than white were too small to compare the groups.

To compare scores based on number of credit hours completed, two methods were used. First, students were grouped into two categories, those having completed 61 – 90 credit hours (60.0% of the sample), and over 90 credit hours (40.0% of the sample). Comparison of means (using ANOVA), medians (using Independent Samples Test of Medians), and distributions (using the Mann-Whitney U statistic) showed statistically significant differences in the medians between the groups only for IL4. Calculating the Spearman rho correlation coefficient between scores and UNCW hours, transfer hours, and total hours showed some statistically significant correlations. Total hours completed was correlated with IL4 (.503**), and UNCW hours completed was correlated with IL1 (.334*), IL2 (.328*), and IL4 (.400*). No dimensions were correlated with SAT-Verbal scores, SAT-Math scores, ACT scores, or students’ GPA.

**THOUGHTFUL EXPRESSION (WRITTEN) RESULTS**

Thirty-three work products were scored on the Written Communication rubric by five scorers. Eleven work products were double or triple scored using the Written Communication rubric. All Written Communication dimensions were evaluated as applicable for the assignment. Figure 6 provides the score distributions.
Results by Dimension

**WC1 Context of and Purpose for Writing**

Scores on this dimension were in the mid-range. No work products failed to demonstrate attention to context, audience, purpose, and to the assigned task (scores of 0). Two in five work products demonstrated minimal attention to context, audience, purpose, and to the assigned task (score of 1). Three in ten work products demonstrated awareness of the context, audience, purpose, and assigned task (score of 2). Almost half of the work products demonstrated adequate consideration of context, audience, and purpose, and a clear focus on the assigned task (score of 3). No 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 (score of 4).
WC2 Content Development
Scores on this dimension were the lowest. No work products demonstrated a lack of content development (score of 0). One fourth of the work products used appropriate and relevant content to develop simple ideas in some parts of the work (score of 1). Over four out of ten work products used appropriate and relevant content to develop and explore ideas through the work (score of 2). Three in ten work products used appropriate, relevant, and compelling content to explore ideas within the context of the discipline (score of 3), with one in ten demonstrating mastery of the subject (score of 4).

WC3 Genre and Disciplinary Conventions
Scores on this dimension were in the mid-range of scores. None of the work products lacked an attempt to use a consistent system for organization and presentation (score of 0). Slightly more than one in ten work products demonstrated an attempt to use a consistent system for basic organization and presentation (score of 1). Almost half of the work products followed expectations appropriate to the specific writing task for basic organization, content, and presentation (score of 2). Four in ten work products demonstrated consistent use of important conventions particular to the writing task, including stylistic choices (score of 3). No work products demonstrated detailed attention to and successful execution of a wide range of conventions particular to the discipline and/or writing task (score of 4).

WC4 Sources and Evidence
The scores on this dimension were relatively high. No student work products lacked any attempt to use sources to support ideas (score of 0). About one in twenty work products demonstrated an attempt to use sources to support ideas (score of 1). Over half the work products demonstrated an attempt to use credible and/or relevant sources to support ideas that were appropriate to the task (score of 2). More than four in ten work products demonstrated consistent use of credible, relevant sources to support ideas (score of 3). No work products demonstrated skillful use of high-quality, credible, relevant sources to develop ideas appropriate to the discipline (score of 4).

WC5 Control of Syntax and Mechanics
The scores on this dimension were the highest. No work products failed to meet the level 1 benchmark (score of 0). Only one work product used language that sometimes impeded meaning because of errors in usage (score of 1). One third of the work products used language that generally conveyed meaning with clarity, although writing included some errors (score of 2). Almost two thirds of work products used straightforward language that generally conveyed meaning, with few errors (score of 3). No work products used graceful language that skillfully communicated meaning with clarity and fluency, with virtually no errors (score of 4).
Correlation between Dimensions
All dimensions of Written Communication were significantly correlated with each other. These correlation coefficients range from .421 to .705. Appendix table C3 contains all the WC correlation coefficients.

Demographic and Preparedness Findings
There were no statistically significant differences between means, medians, or the score distributions of transfer vs. freshman-start students, honors vs. non-honors students, or male vs. female students. The samples of students with race/ethnicities other than white were too small to compare the groups.

To compare scores based on number of credit hours completed, two methods were used. First, students were grouped into three categories, those having completed 31 – 60 credit hours (3.0% of the sample), 61 – 90 (57.6% of the sample), and over 90 credit hours (39.4% of the sample). Comparison of means (using ANOVA), medians (using Independent Samples Test of Medians), and distributions (using the Mann-Whitney U statistic) showed no statistical differences. Calculating the Spearman rho correlation coefficient between scores and UNCW hours, transfer hours, and total hours showed some statistically significant correlations for total hours only, which was correlated with WC1 (.361*), WC3 (.415*), and WC4 (.305*). There were no significant correlations between scores and SAT-Verbal scores, SAT-Math scores, ACT scores, or students’ GPA.

Pilot 2 Inquiry

Description of Sample
In Spring 2010, research reports from PLS 401 were scored by three faculty members on a department rubric which holistically assesses each of four UNCW Learning Goals—Critical Thinking, Information Literacy, Inquiry, and Thoughtful Expression. To examine student performance on Inquiry in more detail, the same 14 student work products were scored by the same scorers using the AAC&U VALUE Inquiry rubric, which divides inquiry into six dimensions. Scorers attended a norming and discussion session. All scorers then scored all work products individually. At a subsequent meeting, the scorers discussed each work product and their individual scores, and produced a final committee score. Demographic information was not available for this study.

Results
Figure 7 provides summary of the final committee scores.
### Pilot 2 Inquiry Results by Dimension for Applicable Scores Only

**Figure 7 Distribution of Scores for Pilot 2 Inquiry, Applicable Scores Only**

<table>
<thead>
<tr>
<th>IN1</th>
<th>IN2</th>
<th>IN3</th>
<th>IN4</th>
<th>IN5</th>
<th>IN6</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>71.4%</td>
</tr>
<tr>
<td>14.3%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>14.3%</td>
</tr>
<tr>
<td>50.0%</td>
<td>50.0%</td>
<td>50.0%</td>
<td>50.0%</td>
<td>50.0%</td>
<td>7.1%</td>
</tr>
<tr>
<td>50.0%</td>
<td>50.0%</td>
<td>50.0%</td>
<td>50.0%</td>
<td>50.0%</td>
<td>7.1%</td>
</tr>
<tr>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2.5</td>
<td>2.5</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

**Results by Dimension**

**IN1 Topic Selection**

Scores on this dimension were one of the two highest, although still lower than expectations. Over one-third of work products identified a manageable topic that left out relevant aspects (score of 2). Slightly less than two-thirds of the work products identified a focused and manageable topic that addressed relevant aspects of the topic (score of 3). No work products identified a creative, focused, and manageable topic that addressed potentially significant aspects of the topic (score of 4).
IN2 Existing Knowledge, Research, and/or Views
Scores on this dimension were one of the two highest, although still lower than expectations. Six of the 14 work products (42%) presented information from relevant sources representing limited points of view (score of 2). Half of the work products presented in-depth information from relevant sources representing various points of view (score of 3). One work product synthesized in-depth information from relevant sources representing various points of view (score of 4).

IN3 Design Process
Scores on this dimension were the second lowest. Two work products demonstrated some misunderstanding of the methodology or theoretical framework (score of 1). Half the work products utilized a process, but parts of the process were missing or incorrectly developed (score of 2). Slightly over a third of the work products demonstrated the ability to utilize a methodology that was appropriately developed, even though more subtle elements were not there (score of 3). No work products skillfully developed all elements of the methodology or theoretical framework (score of 4).

IN4 Analysis
Scores on this dimension were in the mid-range. Half the work products contained organized evidence, although the organization was not considered effective in revealing patterns, differences, or similarities (score of 2). Half the work products contained evidence organized effectively to reveal patterns, differences, or similarities related to the focus of the inquiry (score of 3). No work products organized and synthesized evidence to reveal insightful patterns, differences, and similarities related to the focus (score of 4).

IN5 Conclusions
Scores on this dimension were in the mid-range. Half the work products stated general conclusions that went beyond the scope of the inquiry (score of 2). Half the work products stated conclusions focused solely on and arising specifically from the inquiry finding (score of 3). No work products stated a conclusion that was a logical extrapolation from the inquiry findings (score of 4).

IN6 Limitations and Implications
Scores on this dimension were the lowest. Over 70% of the work products presented no limitations or implications (score of 0). Two work products presented limitations and implications that were irrelevant and unsupported (score of 1). One work product presented relevant and supported limitations and assumptions (score of 2). One work product discussed relevant and supported limitations and implication (score of 3). No work products insightfully discussed in detail relevant and supported limitations and implications (score of 4).
DEMOGRAPHIC AND PREPAREDNESS FINDINGS

Student names, and hence demographic and preparedness information, were not available for this sample.

PILOT 3 THOUGHTFUL EXPRESSION (ORAL)

DESCRIPTION OF SAMPLE

In Spring 2011, all three sections of PSY 410 Cognitive Psychology participated in piloting the AAC&U VALUE Oral Communication Rubric to assess student presentations of their research projects. The presentations were given in two open poster sessions. Students made their presentations one-on-one as faculty and other students approached their poster area.

There were four scorers, two course faculty and two scorers from the General Education Assessment office. The scorers met prior to the presentations to discuss the rubric. However, a typical norming session could not be done since there were no prior student work products for practice scoring and consensus building.

There were 73 students in the three sections. Fifty-seven student presentations were assessed with the rubric. Of those, 22 were assessed by one scorer, and the others were assessed by multiple (2, 3, or 4) scorers. Interrater reliability was calculated on all multiply-scored presentations. For the purposes of assessing individual performance on the five dimensions of Oral Communication, when scorers differed on the score assigned, the lowest score was used for this analysis.

The final sample of 57 students consisted of: 77.2% females, which closely aligns with the percent of females in the three sections (79.2%), but above the percent of females at the institution (60%); 87.7% Caucasian, which closely aligns with the percent of Caucasian students in the three sections (83.3%) and at the institution (83.9%); 40.4% transfer students, which closely aligns with the percent of transfer students in the three sections (37.5%), but above the percent at the institution (28.0%); and 5.3% honors students, which closely aligns with the honors students in the three sections (4.2%) and at the institution (4.9%). It is to be expected that the proportion of transfer students is higher in the senior year than across all four years. Also, the percent of females closely matches the 78.2% for the psychology major. The average GPA for the sample was 3.29 (compared to 3.24 for all students in the three sections), and the average SAT combined was 1149 (compared to 1141 for all students in the three sections).

RESULTS

Figure 8 provides a summary of the scores.
**PILOT 3 ORAL COMMUNICATION RESULTS BY DIMENSION FOR APPLICABLE SCORES ONLY**

<table>
<thead>
<tr>
<th></th>
<th>OC1</th>
<th>OC2</th>
<th>OC3</th>
<th>OC4</th>
<th>OC5</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>1</td>
<td>1.8%</td>
<td>1.8%</td>
<td>5.3%</td>
<td>3.5%</td>
<td>3.5%</td>
</tr>
<tr>
<td>2</td>
<td>31.6%</td>
<td>28.1%</td>
<td>29.8%</td>
<td>26.3%</td>
<td>24.6%</td>
</tr>
<tr>
<td>3</td>
<td>47.4%</td>
<td>61.4%</td>
<td>50.9%</td>
<td>59.6%</td>
<td>54.5%</td>
</tr>
<tr>
<td>4</td>
<td>21.2%</td>
<td>8.8%</td>
<td>14.0%</td>
<td>10.5%</td>
<td>17.5%</td>
</tr>
<tr>
<td>25&lt;sup&gt;th&lt;/sup&gt; %ile</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>50&lt;sup&gt;th&lt;/sup&gt; %ile</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>75&lt;sup&gt;th&lt;/sup&gt; %ile</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>mode</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

Figure 8 Distribution of Scores for Pilot 3 Oral Communication, Applicable Scores Only

**RESULTS BY DIMENSION**

**OC1 Organization**

Only one presentation did not have an observable organization pattern (score of 1). Less than one third of the presentations had organizational patterns that were intermittently observable (score of 2). Almost one half of the presentations had clearly and consistently observable organizational patterns (score of 3). One in five presentations had clearly and consistently observable organizational patterns that were skillful and made the content cohesive (score of 4).
**OC2 Language**
Only one presenter used language that was not appropriate to the audience (unclear and minimally supported effectiveness) (score of 1). Almost three in ten presenters made mundane and commonplace language choices that partially support the effectiveness of the presentation (score of 2). Over three in five presenters made thoughtful language choices that generally supported the effectiveness of the presentation (score of 3). One in ten presenters made imaginative, memorable, and compelling language choices that enhanced the effectiveness of the presentation (score of 4).

**OC3 Delivery**
Three presenters used delivery techniques that detracted from the understandability of the presentation (score of 1). Three in ten presenters used delivery techniques that made the presentation understandable, although the presenter appeared tentative (score of 2). One half of the presenters used delivery techniques that made the presentation interesting (score of 3). One in seven presenters used delivery techniques that made the presentation compelling, and the presenter appeared polished and confident (scores of 4).

**OC4 Supporting Materials**
Two presentations displayed insufficient supporting materials that only minimally supported the presentation or established the presenter’s credibility/authority on the topic (score of 1). One fourth of the presentations displayed supporting materials that made appropriate reference to the information/analysis and that partially supported the presentation or established the presenter’s credibility/authority on the topic (score of 2). Three in five presentations displayed supporting materials that made appropriate reference to the information/analysis and that generally supported the presentation or established the presenter’s credibility/authority on the topic (scores of 3). One in ten presentations displayed a variety of types of supporting materials that significantly supported the presentation or established the presenter’s credibility/authority on the topic (score of 4).

**OC5 Central Message**
Two presentations contained central messages that could be deduced, but that were not explicitly stated (score of 1). One fourth of the presentations contained central messages that were understandable, but were not often repeated and not memorable (score of 2). Over half of the presentations contained central messages that were clear, consistent, and supported (score of 3). Slightly less that one in five presentations contained central messages that were compelling, precisely stated, repeated, memorable, and strongly supported (score of 4).

**DEMOGRAPHIC AND PREPAREDNESS FINDINGS**
There were no significant differences between the means, medians, or score distributions between male and females, or between white, non-Hispanic students and students of all other
races/ethnicities (only 7 out of 57 self-identified as a race or ethnicity other than white, non-Hispanic, therefore analysis could not be done on individual designations). There were only 3 honors students, so no comparison could be made between scores for honors and all other students. For transfer students, there were significant (at the .05 level) negative differences between the means and score distributions for OC1, OC4, and OC5.

Individual scores on the five dimensions of Oral Communication were all positively correlated with the number of UNCW Hours completed (OC1 = .352**, OC2 = .285*, OC3 = .277*, OC4 = .441**, OC5 = .444**), and negatively correlated with Transfer Hours (OC1 = -.291*, OC2 = -.179, OC3 = -.143, OC4 = -.444**, OC5 = -.327*). There were no significant correlations with Total Hours completed. GPA was significantly correlated with two dimensions (OC3 = .310* and OC4 = .294*). There were no significant correlations between SAT Verbal or Math.¹

¹ * indicates significance at the .05 level; ** indicates significance at the .01 level.
PROCESS RESULTS

RELIABILITY OF SCORES

A NOTE ABOUT VALIDITY
Reliability is a necessary, but not sufficient, condition for validity. According to standards published jointly by the American Education Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME), validity is the degree to which evidence and theory support the interpretation of test scores for the proposed use (AERA, 1999). The VALUE rubrics were recommended as valid means to assess specific UNCW Learning Goals because (1) they align directly to the definitions of Critical Thinking, Information Literacy, and Written Communication adopted by UNCW, and (2) according to the AAC&U developers,

“The rubric development teams relied on existing campus rubrics when available, other organizational statements on outcomes, experts in the respective fields and faculty feedback from campuses throughout the process. Each VALUE rubric contains the most common and broadly shared criteria or core characteristics considered critical for judging the quality of student work in that outcome area” (AAC&U, 2010).

The Diversity rubric also aligns with the definitions within the UNCW Learning Goals, as well as the diversity component student learning outcomes. The rubric does, however, require additional vetting by faculty before it can meet validity standards.

MEASURING RELIABILITY
To ensure that a means of assessment, such as a rubric, that is considered to be valid produces reliable scores, we must also look at reliability, in this case interrater reliability. The Fall 2010 scoring event was the second instance of using the Critical Thinking rubric and the first time using the Diversity rubric. The Spring 2011 scoring event was the first time the Information Literacy and Oral Communication rubrics were used and the second time the Written Communication and Inquiry rubrics were used. Details about how scorers were normed are given in the Methodology chapter of this report. Briefly, scorer norming consisted of two stages. First, each scorer attended a two-hour workshop at which the rubric was reviewed and two to four student work products were scored and discussed. Second, on the day of scoring, scorers worked in groups of 2, 3, or 4. They began the scoring process for each assignment packet by scoring and discussing one common work product from their packets, and created additional scoring guidelines specific to that assignment, if necessary. There were a number of additional common student work products in each packet so that interrater reliability could be assessed. Only the independently scored work products were used to measure interrater reliability.
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.

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).

**Sample Size**

**Basic Studies**

For Critical Thinking, 36 work products were double scored. Nine of those work products were discussed, leaving a sample of 28 (24.6%) for testing interrater reliability. For Diversity, 35 work products were double scored. Eight of those work products were discussed, leaving a sample of 27 (23.9%) for testing interrater reliability. For Information Literacy, 16 work products were double or triple scored. Six of those work products were discussed, leaving a sample of 10 (10%) for testing interrater reliability. For Written Communication, 27 work products were double or triple scored. Eight of those were discussed, leaving a sample of 19 work products (23.8%).
Pilot 1
Interrater reliability was examined on the scores from the seven scorers. For Information Literacy, 12 work products were double scored. For Written Communication, 11 work products were double scored.

Pilot 2
Interrater reliability was examined on the scores from the three scorers. All 14 student work products were scored by all three scorers.

Pilot 3
Interrater reliability was examined on the scores from the four scorers. Thirty-five student work products were scored by 2, 3 or 4 scorers.

IRR Results
Table 1 for Basic Studies and Table 2 for the three pilot studies provide results of commonly reported measures of interrater reliability. Percent agreement (the percent of pairwise matches), Percent agreement plus adjacent (the percent of pairwise exact matches plus cases where the scores differed by only one score level), and Krippendorff’s Alpha (a holistic measure across all four scorers) all provide measures of agreement. Average Spearman rho correlation coefficient (the average of the pairwise coefficients) provides a measure of consistency.
<table>
<thead>
<tr>
<th></th>
<th>Percent Agreement</th>
<th>Plus Percent Adjacent</th>
<th>Krippendorff’s Alpha</th>
<th>Spearman Rho</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Critical Thinking</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CT1 Explanation of Issues</td>
<td>42.9%</td>
<td>89.3%</td>
<td>0.412</td>
<td>0.402*</td>
</tr>
<tr>
<td>CT2.1 Evidence; Selecting and Using</td>
<td>17.9%</td>
<td>78.6%</td>
<td>0.067</td>
<td>0.045</td>
</tr>
<tr>
<td>CT2.2 Evidence; Critically Examining</td>
<td>50.0%</td>
<td>81.3%</td>
<td>-0.008</td>
<td>0.006</td>
</tr>
<tr>
<td>CT3 Influence of Context and Assumptions</td>
<td>28.6%</td>
<td>66.7%</td>
<td>0.100</td>
<td>0.104</td>
</tr>
<tr>
<td>CT4 Student’s Position</td>
<td>38.1%</td>
<td>81.0%</td>
<td>0.283</td>
<td>0.258</td>
</tr>
<tr>
<td>CT5 Conclusions and Related Outcomes</td>
<td>25.0%</td>
<td>60.7%</td>
<td>0.003</td>
<td>-0.010</td>
</tr>
<tr>
<td><strong>Diversity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DIV1 Knowledge of Diverse Perspectives</td>
<td>18.5%</td>
<td>66.7%</td>
<td>0.093</td>
<td>0.119</td>
</tr>
<tr>
<td>DIV2 Contextual Importance and Implications</td>
<td>34.6%</td>
<td>69.2%</td>
<td>0.144</td>
<td>0.123</td>
</tr>
<tr>
<td><strong>Information Literacy</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IL1 Determine extent of information needed</td>
<td>20.0%</td>
<td>60.0%</td>
<td>-0.577</td>
<td>-0.589</td>
</tr>
<tr>
<td>IL2 Access needed information</td>
<td>20.0%</td>
<td>70.0%</td>
<td>-0.288</td>
<td>-0.320</td>
</tr>
<tr>
<td>IL3 Evaluate information and sources critically</td>
<td>40.0%</td>
<td>100.0%</td>
<td>0.348</td>
<td>0.334</td>
</tr>
<tr>
<td>IL4 Use information to accomplish a purpose</td>
<td>40.0%</td>
<td>90.0%</td>
<td>0.248</td>
<td>0.192</td>
</tr>
<tr>
<td>IL5 Access and use information ethically</td>
<td>20.0%</td>
<td>70.0%</td>
<td>0.060</td>
<td>0.259</td>
</tr>
<tr>
<td><strong>Written Communication</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WC1 Context of and Purpose for Writing</td>
<td>15.8%</td>
<td>78.9%</td>
<td>0.073</td>
<td>0.089</td>
</tr>
<tr>
<td>WC2 Content Development</td>
<td>31.6%</td>
<td>94.7%</td>
<td>0.297</td>
<td>0.324</td>
</tr>
<tr>
<td>WC3 Genre and Disciplinary Conventions</td>
<td>26.3%</td>
<td>78.9%</td>
<td>-0.263</td>
<td>-0.258</td>
</tr>
<tr>
<td>WC4 Sources and Evidence</td>
<td>42.1%</td>
<td>94.7%</td>
<td>0.575</td>
<td>0.600**</td>
</tr>
<tr>
<td>WC5 Control of Syntax and Mechanics</td>
<td>52.6%</td>
<td>100.0%</td>
<td>0.605</td>
<td>0.605**</td>
</tr>
</tbody>
</table>

*Statistically significant at the .05 level
**Statistically significant at the .01 level
Table 2 Interrater Reliability for All Pilot Studies

<table>
<thead>
<tr>
<th>Information Literacy (Pilot 1)</th>
<th>Percent Agreement</th>
<th>Plus Percent Adjacent</th>
<th>Krippendorff’s Alpha</th>
<th>Spearman Rho</th>
</tr>
</thead>
<tbody>
<tr>
<td>IL1 Determine extent of information needed</td>
<td>10.0%</td>
<td>70.0%</td>
<td>0.144</td>
<td>.082</td>
</tr>
<tr>
<td>IL2 Access needed information</td>
<td>50.0%</td>
<td>80.0%</td>
<td>0.193</td>
<td>.138</td>
</tr>
<tr>
<td>IL3 Evaluate information and sources critically</td>
<td>0.0%</td>
<td>57.1%</td>
<td>-0.694</td>
<td>-.845</td>
</tr>
<tr>
<td>IL4 Use information to accomplish a purpose</td>
<td>16.7%</td>
<td>100.0%</td>
<td>0.411</td>
<td>.435</td>
</tr>
<tr>
<td>IL5 Access and use information ethically</td>
<td>80.0%</td>
<td>80.0%</td>
<td>0.534</td>
<td>.492</td>
</tr>
<tr>
<td>Written Communication (Pilot 1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WC1 Context of and Purpose for Writing</td>
<td>45.5%</td>
<td>81.8%</td>
<td>0.084</td>
<td>.138</td>
</tr>
<tr>
<td>WC2 Content Development</td>
<td>45.5%</td>
<td>90.9%</td>
<td>-0.037</td>
<td>.271</td>
</tr>
<tr>
<td>WC3 Genre and Disciplinary Conventions</td>
<td>36.4%</td>
<td>90.9%</td>
<td>0.013</td>
<td>.000</td>
</tr>
<tr>
<td>WC4 Sources and Evidence</td>
<td>45.5%</td>
<td>100.0%</td>
<td>0.265</td>
<td>.302</td>
</tr>
<tr>
<td>WC5 Control of Syntax and Mechanics</td>
<td>63.6%</td>
<td>90.9%</td>
<td>0.222</td>
<td>.395</td>
</tr>
<tr>
<td>Inquiry (Pilot 2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IN1 Topic Selection</td>
<td>38.1%</td>
<td>90.5%</td>
<td>0.184</td>
<td>.357</td>
</tr>
<tr>
<td>IN2 Existing Knowledge, Research, Views</td>
<td>28.6%</td>
<td>83.3%</td>
<td>0.026</td>
<td>.305</td>
</tr>
<tr>
<td>IN3 Design Process</td>
<td>33.3%</td>
<td>88.1%</td>
<td>0.354</td>
<td>.487</td>
</tr>
<tr>
<td>IN4 Analysis</td>
<td>50.0%</td>
<td>97.6%</td>
<td>0.291</td>
<td>.361</td>
</tr>
<tr>
<td>IN5 Conclusions</td>
<td>35.7%</td>
<td>97.6%</td>
<td>0.067</td>
<td>.036</td>
</tr>
<tr>
<td>IN6 Limitations and Implications</td>
<td>21.4%</td>
<td>57.1%</td>
<td>0.159</td>
<td>.543</td>
</tr>
<tr>
<td>Oral Communication (Pilot 3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OC1 Organization</td>
<td>38.4%</td>
<td>95.9%</td>
<td>0.3739</td>
<td>.407</td>
</tr>
<tr>
<td>OC2 Language</td>
<td>49.3%</td>
<td>93.2%</td>
<td>0.3497</td>
<td>.323</td>
</tr>
<tr>
<td>OC3 Delivery</td>
<td>42.5%</td>
<td>91.8%</td>
<td>0.4472</td>
<td>.476</td>
</tr>
<tr>
<td>OC4 Supporting Material</td>
<td>58.9%</td>
<td>97.3%</td>
<td>0.3520</td>
<td>.338</td>
</tr>
<tr>
<td>OC5 Central Message</td>
<td>57.5%</td>
<td>94.5%</td>
<td>0.3657</td>
<td>.316</td>
</tr>
</tbody>
</table>

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.

Comparing the results of the reliability indices for this study to the benchmark of .67 for Krippendorff’s alpha and .7 for Spearman’s rho, there are no dimensions in any rubric that meet these standards. Two of the Written Communication dimensions for basic studies (WC4 and WC5) come close. The percent agreement was greater than or equal to 50.0% in eight places. Looking at percent adjacent (that is, the scores that were within one level of each other), we find that most dimensions had greater than 70% of scores within one level of each other.

Overall, these various measures of reliability illustrate randomness in agreement and indicate that additional norming activities are required.

**PROCESS FEEDBACK**

For the Basic Studies and Pilot 1 assessment process, all scorers filled out two types of feedback forms. At the end of the day, each scorer filled out a process feedback survey. This survey asked for their opinions about how well each step of the process had gone, and for any recommendations for improvement. During the day, after completing each packet of student work products, each scorer filled out a rubric feedback form. This 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.

For Pilots 2 and 3, scorers provided feedback through conversations after the project completions.

**Basic Studies and Pilot 1 Scorer Feedback on Process**

Student work products for Pilot 1 were scored at the same time as the Basic Studies work products and by the same scorers. As a result, scorers completed a feedback survey about the process in general (since it was the same for both instances). Therefore, these are reported together here. Table 3 provides the results on the selected responses items on the survey.
<table>
<thead>
<tr>
<th>Table 3 Basic Studies, Pilot 1 Scorer Feedback on Process</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>The invitation to volunteer accurately described the experience.</td>
</tr>
<tr>
<td>FA2010 N=10</td>
</tr>
<tr>
<td>SP2011 N=24</td>
</tr>
<tr>
<td>The timing of the invitation gave adequate opportunity for attending workshops and scoring.</td>
</tr>
<tr>
<td>FA2010 N=10</td>
</tr>
<tr>
<td>SP2011 N=24</td>
</tr>
<tr>
<td>The 2-hour norming session adequately prepared me for what was expected of me during the scoring session.</td>
</tr>
<tr>
<td>FA2010 N=10</td>
</tr>
<tr>
<td>SP2011 N=24</td>
</tr>
<tr>
<td>The scoring session was well-organized.</td>
</tr>
<tr>
<td>FA2010 N=10</td>
</tr>
<tr>
<td>SP2011 N=24</td>
</tr>
<tr>
<td>The structure of the scoring made it reasonable to work for the full time.</td>
</tr>
<tr>
<td>FA2010 N=10</td>
</tr>
<tr>
<td>SP2011 N=24</td>
</tr>
<tr>
<td>When I had questions, one of the leaders was available to answer it.</td>
</tr>
<tr>
<td>FA2010 N=10</td>
</tr>
<tr>
<td>SP2011 N=24</td>
</tr>
<tr>
<td>When I had questions, the question was answered.</td>
</tr>
<tr>
<td>FA2010 N=10</td>
</tr>
<tr>
<td>SP2011 N=24</td>
</tr>
<tr>
<td>I was comfortable scoring student work products from outside my discipline on the broad Learning Goals.</td>
</tr>
<tr>
<td>FA2010 N=10</td>
</tr>
<tr>
<td>SP2011 N=24</td>
</tr>
<tr>
<td>The process is an appropriate way to assess students on the UNCW Learning Goals.</td>
</tr>
<tr>
<td>FA2010 N=10</td>
</tr>
<tr>
<td>SP2011 N=24</td>
</tr>
<tr>
<td>This process is valuable in improving student learning.</td>
</tr>
<tr>
<td>FA2010 N=10</td>
</tr>
<tr>
<td>SP2011 N=24</td>
</tr>
<tr>
<td>Survey Question</td>
</tr>
<tr>
<td>--------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>I would participate in this process again.</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>I would recommend participating in this process to my colleagues.</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

There were also three open-ended questions on the survey. The results of these are included in the summary below. The complete results of these questions are provided in Appendix D.

There was a high level of satisfaction with regard to most aspects of the process. The initial contact and explanations about the responsibilities of the volunteers was clear to all. Most scorers (66.7% for Fall and 76.2% for Spring) responded that the norming session adequately prepared them for the all-day scoring session. Positive comments were that it was invaluable, that it was helpful, and even fun. Although scorers felt prepared to score, they did provide recommendations for improving preparation. One scorer suggested that a longer training session would be useful at that finding inconsistencies among scorers concerning the rubric may need extended practice.

With regard to the all-day scoring session, scorers were generally satisfied. Three scorers from the Fall session and seven from the Spring session mentioned being pleased with the organization and flow of the process. Five scorers in the Spring 2010 scoring session commented that they enjoyed the opportunity to discuss student work with colleagues from other disciplines. With regard to the length of the session, seven scorers in the Spring session suggested that more time was needed to adequately score the student work, two mentions were made of fatigue and burnout, and one person suggested breaking the process into two sessions. Four scorers in that same session indicated that more quiet spaces for reading student work would be useful.

There were a number of comments related to the rubrics in the comments on the end-of-day survey. One person mentioned that the overall process helped her become more familiar and comfortable with the UNCW Learning Goals and with assessment in general. Another mentioned that it was a good way to learn more about the evaluation process across the university. One scorer indicated that having people from the discipline for which the papers were written as a scorers would help the scoring process. Another scorer questioned if groups or pairs for scoring were necessary, and a second person wondered if groups of three were more successful in reaching consensus about jointly-scored papers than were pairs. Most comments, though, concerned the match between the assignments and rubrics. Seven scorers (five from Fall and two from Spring) mentioned the issue of alignment of the assignment prompts to the rubric.
criteria. Most of these comments were concerned with the applicability of the rubric to the assignments as some dimensions did not seem to fit all assignments.

**Basic Studies and Pilot 1 Scorer Feedback on Rubric and Required Assumptions**

**Critical Thinking**

Of the five rubric dimensions for Critical Thinking, only “Explanation of issues” was perceived as a good fit for the assignment by a majority of scorers (70%). For the remaining dimensions, most scorers felt that the dimension fit the work products, although many indicated that some assumptions were necessary to achieve the fit. However, it is important to note that some (20%-30%) of the scorers indicated that each of the remaining four dimensions did not fit the assignment.

In response to the open-ended questions about the Critical Thinking rubric, all scorers described some issue about the fit between the assignment prompt and the rubric dimensions. A close look at the individual comments reveals that most comments about issues of fit between the rubric and assignment were about the “critically examining viewpoints of experts” dimension (9 instances). Scorers also had issues with the fit between the “conclusions and outcomes” dimension and the assignments scored (5 occurrences), as well as the “student’s position” dimension (4 occurrences), the “evidence” dimension (3 occurrences), and the “context and assumptions” dimension (3 occurrences). Many participants (60%) also commented about student achievement of the criteria expectations in more general terms, without explicitly linking that achievement to the limits of the assignment.

Several scorers mentioned criteria with issues unrelated to the fit of the dimension to the assignment. For example, “explanation of issues” and “student’s position” were mentioned as potentially overlapping and difficult to separate by one scorer. Another scorer noted that “synthesis” and “analysis” need to be distinguished from one another more clearly.

Finally, some scorers offered specific suggestions about improvements to the rubric. One suggested adding “ability to express ideas succinctly and cogently” as a dimension. Another scorer mentioned the need to address instances of incorrect evidence being used by a student. Addressing the problem of rubric-assignment fit, one scorer suggested adding “if applicable” to each of the dimensions, and another suggested developing criteria to be used in selecting an assignment for the process.

**Diversity**

Of the two Diversity rubric dimensions, no scorers felt that the dimensions did not fit. They were, however, divided in their opinions about the degree of fit: 62.5% felt that the “knowledge of diverse perspectives” dimension fit well with the assignment, while 37.5% felt that it fit with
assumptions. For the “contextual importance” dimension, half of the scorers felt the dimension was a good fit with the assignment while half thought it fit with assumptions.

Scorer response was sparse for the open-ended questions concerning the Diversity rubric. Two scorers did not respond at all, and only four scorers provided substantive feedback. Of these four, three participants indicated some concern about the limits of the assignment prompt in allowing students to provide products meeting the dimensions. Three scores also indicated some concern about student ability to perform to levels of the dimensions. Finally, one scorer indicated that “awareness of racism” might be included on the rubric, another mentioned “being able to define diversity” as a possible improvement, and a third indicated that some parts of the rubric were not as applicable to cultural diversity as to human.

Information Literacy
For the first two dimensions of the Information Literacy rubric, “determine the extent of the information needed” and “access the needed information”, scorers thought the dimensions either fit well with the assignment (IL1: 61.1%, IL2: 67.7%) or fit with assumptions (22.2%, 16.7%). However, for the other three dimensions, scorers were divided on the fit of the dimensions. The most dramatic division occurred with the “use information effectively to accomplish a specific purpose” dimension. While about most of the scorers felt this dimension fit well or with assumptions, about one quarter thought it did not fit due to the narrow requirements of the assignment.

Responding to the open-ended questions, about half of the scorers (46.2%) cited issues with the alignment of the assignment with the rubric dimensions. Several comments also indicated some issues with applying the rubric (8 occurrences). In particular, scorers mentioned that not having access to the original sources made it difficult to know if the information was used in ways true to the original context or had been paraphrased correctly (3 occurrences). One scorer mentioned that he found it difficult to distinguish between common knowledge and ideas requiring attribution. Another indicated that it was “not always easy to discern if [the] student had synthesized information so that the intended purpose was achieved” and suggested established minimum criteria for this purpose. A third scorer wrote that it was not possible to determine fully the extent of information needed.

Some suggestions were given in the scorer feedback for improving the Information Literacy rubric and scoring process. These included separating the first dimension, “determine the extent of information needed” into two statements, one dealing with defining the scope of the thesis and the other dealing with the type of information selected. Another scorer suggested developing some anchor papers to use as benchmark samples of student work falling at each level of the rubric.
Written Communication
For the Written Communication rubric, scorers indicated that most dimensions fit, either as written or with assumptions. Only for the “content development” and “genre and disciplinary conventions” dimensions did one scorer indicate a complete lack of fit. While the majority of scorers did not mention a need for additional assumptions, between 13.6% and 36.4% did for the other three dimensions, with the most accommodations needed for the “sources and evidence” dimension. The reasons given by the scorers for these responses are discussed below.

For the open-ended questions about the rubric dimensions, 73.7% of scorers provided comments about the connection between the assignment prompt and the student work products. There were a number of statements indicating that the assignment instructions explicitly limited the level of achievement possible when using the rubric (9 occurrences). Additionally, many scorers mentioned that, because one of the assignments specified a two-part student work product (an annotated bibliography and a reflective paper) that it necessitated scoring both parts together, though one portion was often of higher quality than the other (6 occurrences).

There were several comments that fell into thematic categories concerning the rubric in general. Four scorers indicated that they found the “genre and disciplinary conventions” dimension difficult to score as they were unfamiliar with the conventions, or because of the multidisciplinary nature of the work. Four scorers commented that particular language in the rubric needed clarification or to be defined; for example, “some errors” vs. “few errors” and “skillful use” vs. “consistent use”. Two scorers mentioned finding it difficult to tease apart the dimensions of written communication, as they are all closely related (i.e., purpose and content development). Finally, two scorers mentioned that they found instances of students’ failure to follow instructions or problems with grammar and syntax and that these were problematic quality criteria; while these statements were not expanded upon, it may point to a need to separate these conventions from others in the “genre and disciplinary conventions” dimension.

Several suggestions for changes to the rubric were provided. One scorer suggested the addition of dimensions such as “values, beliefs, problem-solving, and analysis”. Another suggested adding a component of “application to global thinking.” Two scorers suggested that adding a dimension to the rubric that attends to the instructor’s directions—such as number of sources, formatting requirements, etc. that may not be detailed in the assignment prompt—would be useful. Finally, one scorer suggested that, if planning-level work such as annotated bibliographies is used, that an additional dimension of “potential to develop” could be added.
Inquiry
Pilot 2 scorers met multiple times throughout the process, including the final day when scorers were compared and committee scores assigned. One scorer summarized feedback from all scorers.

The scoring process helped the scorers come to a more common understanding of the rubric dimensions in relation to their discipline and the assignment. The team of scorers made a number of observations from their work. Regarding the rubric, they concluded that there is not much difference between levels 2 and 3 of the rubric for most dimensions. Regarding the assignment, they determined that faculty need to more clearly explain to students that, as part of the conclusion of their papers, it is important to discuss their study and the limitations of their study (which 71% of the students had not done at all). Finally, the inquiry process for the political science discipline takes many forms, both quantitative and qualitative. Students are introduced to the various methods employed by different sub-groups, which provides broad exposure. But it also means that they don’t get extensive practice in one methodology. In addition, the scorers felt that students are missing guidance in determining when a particular method is most useful. The scorers felt that this may partly explain the fact that scores were lower overall than expected.

Thoughtful Expression (Oral)
The scorers met after the poster sessions to compare and discuss scores. The most significant general finding was that each scorer experienced a different presentation. It was determined that the order for the student was important, as their first presentation may be the most tentative, with presentations improving with experience up to a point, and potentially becoming less precise after multiple times through the presentation. It was noted by one faculty member that many students were more nervous when presenting to their instructor, who was responsible for their course grade. Despite these observations, the scorers were pleased with the level of agreement across scorers for most students. Of 175 incidents of multiple scored dimensions, in only 12 (5.7%) incidents were there scores more than one level different.

It was determined that the rubric is unclear in a few places. For example, one scorer assumed that the poster should be referred to by the student during the presentation, and hence scored as part of OC3 Delivery. Others scored the poster only under OC4 Supporting Materials. Another discovery was that, while all scorers were influenced by speaking issues such as fillers and pauses, some scorers included this in OC2 Language, and others in OC3 Delivery. Another potential problem comes from subjective adjectives such as “compelling.”

INSTRUCTOR FEEDBACK
A brief survey was sent to the 12 instructors who provided the student work products. Five responded. This is similar to the four out of thirteen response rate last year. Instructors were
asked to comment on the process of selecting an assignment, score levels they would expect on the assignment, the collection process, and any other parts of the process.

All five respondents said that the assignment selection process was not difficult, and that they felt their assignment fit the rubric well. One mentioned that the workshop gave them good exposure to the rubric. Another mentioned that they worked hard to make sure the assignment did fit. Instructors were asked to indicate which level of the rubric scale they would choose as the goal for the students in their class. For the Diversity rubric, all respondents (who taught 100-level courses) indicated Level 2 for all dimensions. For the Critical Thinking rubric (also used in 100-level courses this year), one respondent indicated Level 2 and the other Level 3 for all dimensions. There was only one response regarding Information Literacy and Written Communication, and that response (from a 200-level Composition instructor) was Level 4 for all dimensions.

Regarding the collection process, three respondents said that the process went very smoothly. Within these responses were instructors who collected the work products in hard copy as well as one instructor who collected work products electronically. The other two respondents, both using some version of electronic submission, had encountered problems. In one class, some students had difficulty making sure that the Student Work Product Cover Sheet was incorporated into the assignment. The other stated that the collection process “was a mess,” but no specifics were given.

When asked for additional comments on the experience, one instructor said the process “went fairly well.” Another stated that it “was a great learning experience.” Another stated that while “my expectations are Level 2, I find that an increasing number of students are entering the university will Level 1 skills, and find it difficult within a semester to progress to Level 2 without additional support beyond the classroom.”
DISCUSSION, LIMITATIONS AND RECOMMENDATIONS

Returning to the purpose of the general education assessment activities in academic year 2010-2011, what evidence was obtained for the following questions?

Within the Basic Studies courses, the following questions were examined:

- What are the overall abilities of students taking basic studies courses with regard to the UNCW Learning Goals of Critical Thinking, Diversity, Information Literacy, and Written Communication?
- What are the relative strengths and weaknesses within the subskills of those goals?
- Are there any differences in performance based on 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?

Within the upper division courses, the following questions were examined:

- What are the abilities of students taking higher-level courses with regard to the UNCW Learning Goals of Written Communication, Information Literacy, Inquiry, and Thoughtful Expression (Oral Communication)?
- What are the relative strengths and weaknesses within the subskills of those goals?
- Are there any differences in performance based on 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?

CRITICAL THINKING

The median score on for all dimensions except CT2.2 was 2, which means that at least half of student work products sampled demonstrate performance at the first Milestone level, level 2. The median score for CT2.2 was 1, indicating that less than half of student work products demonstrated performance at the first Milestone level (level 2). Table 4 shows the percent of work products scored at a level 2 or higher and the percent of work products scored at a level 3 or higher for each dimension.
Table 4 Critical Thinking Percent of Sample Scored at Least 2 and at Least 3

<table>
<thead>
<tr>
<th>Dimension</th>
<th>% of Work Products Scored 2 or higher</th>
<th>% of Work Products Scored 3 or Higher</th>
</tr>
</thead>
<tbody>
<tr>
<td>CT1 Explanation of Issues</td>
<td>73.7%</td>
<td>36.0%</td>
</tr>
<tr>
<td>CT2.1 Evidence: Selecting and Using</td>
<td>72.8%</td>
<td>38.6%</td>
</tr>
<tr>
<td>CT2.2 Evidence: Critically Examining</td>
<td>40.9%</td>
<td>13.6%</td>
</tr>
<tr>
<td>CT3 Influence of Context and Assumptions</td>
<td>59.2%</td>
<td>33.7%</td>
</tr>
<tr>
<td>CT4 Student’s Position</td>
<td>58.2%</td>
<td>30.6%</td>
</tr>
<tr>
<td>CT5 Conclusions and Related Outcomes</td>
<td>59.6%</td>
<td>23.7%</td>
</tr>
</tbody>
</table>

It should be no surprise that, just as last year’s results showed, students performed higher in explaining the issue (CT1) and in presenting evidence from sources (now CT2.1) than they did on the other dimensions. But even the CT2.1 scores show that only 38.6% of students provided a coherent analysis based on evaluation of the sources.

Important to the discussion is how well the assignments matched up to the dimensions of the rubric. Two of the five assignments were scored on all dimensions. Two of the assignments were not scored on CT2.2 Evidence: Critically Examining, and one assignment was not scored on both CT3 Influence of Context and Assumptions and CT4 Student’s Position. CT3 requires the student to demonstrate the ability to identify their own and others’ assumptions and to examine various relevant contexts surrounding the issue. This is an important critical thinking skill in which students need more practice.

Overall, the evidence shows that critical thinking skills increased with credit hours completed. All dimensions except for CT2.2 Critically Examining Evidence are positively correlated with the number of hours completed, which suggests that student skills on most areas of critical thinking improve over the first several semesters of courses at UNCW. While the university is still in the process of determining an expected level of attainment on this rubric for basic studies students, a score of 2 will certainly be the minimum.

It bears mentioning here that there were statistically significant differences between classroom-based courses and online courses for two dimensions, CT1 Explanation of Issues and CT2.2 Critically Examining Evidence, with the classroom-based work product means and medians slightly higher. Additionally, there were statistically significant differences between four dimensions (CT1, CT2.1, CT3, and CT4) for in-class essay assignments and at-home term papers, with the in-class essays scoring lower.

This is the second study using the Critical Thinking rubric. The results for all dimensions except CT2, which cannot be directly compared due to rubric changes, were higher in this study. It is important to mention that we made a small change to the Critical Thinking rubric between years.
1 and 2. According to feedback we received from faculty scorers after the first round of using the VALUE Critical Thinking Rubric, the second dimension, Evidence, was difficult to apply. This dimension contains two statements, one addressing the level of interpretation and development of analysis, and the other focused on questioning the viewpoints of experts. Based on this feedback, we piloted a change to the rubric in which the two statements were applied independently. When we did this, the scores on the first part, interpreting the evidence and developing an analysis (CT2.1), are the highest of all dimensions, and the scores on the second part, questioning the viewpoints of the experts (CT 2.2), are the lowest of all dimensions. The information found from dissecting the dimension is quite important, as it suggests that students need to be instructed on the importance of including an author’s viewpoint in critical analysis.

**DIVERSITY**

The median score on both DV1 and DV2 was 2, which means that more than half the work products received scores of 2 or higher. Table 5 below shows the percent of work products scored at a level 2 or higher and the percent of work products scored at a level 3 or higher for each dimension.

<table>
<thead>
<tr>
<th>Dimension</th>
<th>% of Work Products Scored 2 or higher</th>
<th>% of Work Products Scored 3 or Higher</th>
</tr>
</thead>
<tbody>
<tr>
<td>DV1 Knowledge of Diverse Perspectives</td>
<td>64.6%</td>
<td>30.1%</td>
</tr>
<tr>
<td>DV2 Contextual Importance and Implications</td>
<td>57.7%</td>
<td>30.6%</td>
</tr>
</tbody>
</table>

This was the first pilot study for the Diversity rubric and evidence comes from only one discipline (history). Students’ performances on Diversity are comparable to CT3 Influence of Context and Assumptions and CT4 Student’s Position, which were in the mid-range of critical thinking scores. This result is a favorable indication of the rubric’s validity, as the dimensions are strongly related. In fact, the correlation coefficients between the two diversity and two critical thinking dimensions were statistically significant at the .01 level, and ranged from .778 to .813. Scorer feedback was positive regarding the fit between the rubric and the five assignments, with all scorers indicating that each dimension either fit well or fit with assumptions. There were three suggestions for changes to the rubric, which will be considered before the next study.

There were no statistically significant differences between the means, medians, and distributions of credit-hour groups (i.e. freshman, sophomores, juniors), suggesting that achievement in dimensions of the Diversity rubric do not change as students progress through their coursework at UNCW.
**INFORMATION LITERACY**

Results were statistically significantly higher for 300 level for two of five dimensions—IL2 and IL4—and lower for IL5. IL1 was not well suited for the 300-level assignment, as students were only required to find one article. Table 6 below shows, for both the Basic Studies courses and the Pilot 1 course, the percent of work products scored at a level 2 or higher and the percent of work products scored at a level 3 or higher for each dimension.

Table 6 Information Literacy Percent of Sample Scored at Least 2 and at Least 3

<table>
<thead>
<tr>
<th>Dimension</th>
<th>% of Work Products Scored 2 or higher</th>
<th>% of Work Products Scored 3 or Higher</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>200 courses</td>
<td>300 course</td>
</tr>
<tr>
<td>IL1 Extent of Information Needed</td>
<td>89.0%</td>
<td>80.0%</td>
</tr>
<tr>
<td></td>
<td>51.0%</td>
<td>52.5%</td>
</tr>
<tr>
<td>IL2 Access Needed Information</td>
<td>91.0%</td>
<td>97.5%</td>
</tr>
<tr>
<td></td>
<td>49.0%</td>
<td>70.0%</td>
</tr>
<tr>
<td>IL3 Evaluate Information and Sources</td>
<td>88.0%</td>
<td>88.6%</td>
</tr>
<tr>
<td></td>
<td>41.0%</td>
<td>22.9%</td>
</tr>
<tr>
<td>IL4 Use Information for a Purpose</td>
<td>88.0%</td>
<td>100.0%</td>
</tr>
<tr>
<td></td>
<td>44.0%</td>
<td>76.9%</td>
</tr>
<tr>
<td>IL5 Access and Use Ethically and Legally</td>
<td>90.0%</td>
<td>80.0%</td>
</tr>
<tr>
<td></td>
<td>59.0%</td>
<td>42.5%</td>
</tr>
</tbody>
</table>

Table 6 indicates that the percentage of students scoring a “2” or above is lower for 300-level courses for IL1, IL3, and IL5. The percentage of students scoring “3” or higher is higher in the 300-level course for all dimensions except IL3 and IL5. Possible explanations for this are the limitations of the assignment (students only had to select a single article for the assignment) and the unfamiliarity of the content with regards to the scorers. In their feedback, some scorers mentioned that it was difficult to know what “good sources” for this assignment would look like, and question their ability to determine if students had selected the best sources to achieve their purpose.

Of interest is the fact that there were indications from a number of tests that transfer students taking ENG 201 scored significantly higher than freshman starter on IL3. Although this might be attributed to the fact that transfer students in general had completed more total hours, there were also findings that scores on IL3 and IL5 went down as UNCW hours increased. There were no significant difference found in the NSG 327 findings between transfer and freshman starters, and only positive correlations with UNCW and total hours completed.

**THOUGHTFUL EXPRESSION (WRITTEN)**

Means, medians and distributions were higher for the 300 level for all 5 dimensions, though only statistically significantly higher for WC5. Table 7 below shows the percent of work products scored at a level 2 or higher and the percent of work products scored at a level 3 or higher for each dimension.
Table 7 Thoughtful Expression (Written) Percent of Sample Scored at Least 2 and at Least 3

<table>
<thead>
<tr>
<th>Dimension</th>
<th>% of Work Products Scored 2 or higher</th>
<th>% of Work Products Scored 3 or Higher</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>200 courses</td>
<td>300 course</td>
</tr>
<tr>
<td>WC1 Context of and Purpose for Writing</td>
<td>76.3%</td>
<td>78.5%</td>
</tr>
<tr>
<td>WC2 Content Development</td>
<td>67.5%</td>
<td>75.8%</td>
</tr>
<tr>
<td>WC3 Genre and Disciplinary Conventions</td>
<td>67.6%</td>
<td>87.9%</td>
</tr>
<tr>
<td>WC4 Sources and Evidence</td>
<td>71.4%</td>
<td>93.9%</td>
</tr>
<tr>
<td>WC5 Control of Syntax and Mechanics</td>
<td>83.9%</td>
<td>97.0%</td>
</tr>
</tbody>
</table>

Student performance was strongest on WC5 Control of Syntax and Mechanics at both levels, and weakest on WC2 Content Development, again at both levels. In last year’s basic studies results, performance was strongest on WC1 Context and Purpose for Writing, which ranked tied for second this year at both levels.

As with Information Literacy, UNCW hours completed was negatively correlated with a number of dimensions of Written Communication (WC1, WC2, and WC4) for the ENG 201 results. However, this was not the case for the NSG 327 results, where there were significant positive correlations between total hours completed and WC1, WC3, and WC4.

It is also important to note that assignments were chosen for assessing Information Literacy as the main criterion. Therefore, a few of the assignments, such as the annotated bibliography for two sections of ENG 201, provided more difficult products with which to assess Written Communication, especially the Content Development dimension.

Pilot 2 Inquiry

At least 50% of students scored at a level 3 or higher for 4 of the 6 dimensions, although only one score of 4 was given by the committee across all six dimensions. Table 8 shows the percent of work products scored at a level 3 or higher for each dimension (percent of scores at a level of 2 or higher are not considered relevant for senior-level work).
The assessment committee members determined that the results for IN3 Design Process could be attributed to the fact that there is not one standard research methodology for Political Science. Students are exposed to the methodology of the faculty member teaching a particular section of the research methodology course. Students need exposure to multiple methodologies and for which types of problems they are most appropriate. The committee members also determined that it is important remind students of the importance of discussing the limitations of their studies.

**Pilot 3 Thoughtful Expression (oral)**

The median score on all dimensions was a 3, and no dimension had less that 64% of the scores fall below a 3. Table 9 shows the percent of work products scored at a level 3 or higher for each dimension (percent of scores at a level of 2 or higher are not considered relevant for senior-level work).

The most significant general finding was that each scorer experienced a different presentation. It was determined that the order for the student was important, as their first presentation may be the most tentative, with presentations improving with experience up to a point, and potentially becoming less precise after multiple times through the presentation. It was noted by one faculty member that many students were more nervous when presenting to their instructor, who was responsible for their course grade. Despite these observations, the scorers were pleased with the
level of agreement across scorers for most students. Of 175 incidents of multiple scored dimensions, in only 12 (5.7%) incidents were there scores more than one level different.

It was determined that the rubric is unclear in a few places. For example, one scorer assumed that the poster should be referred to by the student during the presentation, and hence scored as part of OC3 Delivery. Others scored the poster only under OC4 Supporting Materials. Another discovery was that, while all scorers were influenced by speaking issues such as fillers and pauses, some scorers included this in OC2 Language, and others in OC3 Delivery. Another potential problem comes from subjective adjectives such as “compelling.”

The results and scorer discussions lead to the following recommendations for future assessment strategies.

- Pilot the rubric with multiple scorers in a whole class presentation, where each scorer shares the same experience.
- Clarify language in the rubric and develop assumptions to govern the application of the rubric.
- Video record a sample of student presentations for use in scorer norming sessions.
- Compare scores between course instructors and other scorers to determine whether there are significant quality effects when presenting to the instructor. Use results to determine whether to include instructors as scorers for General Education Assessment purposes.

**Relative Strengths and Weaknesses across Rubrics**

Comparison of scores across rubrics should be done cautiously and should only serve as a starting place for further investigation. This is because the criteria for each level of the rubric cannot be assumed to be scaled the same. For example, Level 2 cannot be considered to be in the identical place on a scale of abilities for each of the rubrics. With this in mind, determination of university-wide expectations for performance in basic studies courses should be done on a rubric-by-rubric basis.

With this caveat in mind, it is helpful for the prioritization of effort to look at potential areas of strength and weakness. Looking at the results from all scoring, the following areas stand out.

**Areas of Relative Strengths for Basic Studies:**

- All of the dimensions of Information Literacy produced scores higher than dimensions from all other rubrics. Within Information Literacy, IL5 Access and use information ethically and IL 1 Determine extent of information needed produced the highest scores.
- Within Critical Thinking, CT1 Explanation of Issues and CT2.1 Evidence: selecting and using produced the highest scores.
Within Written Communication, WC1 Context of and Purpose for Writing and WC5 Control of Syntax and Mechanics produced the highest scores.

Areas of Relative Weaknesses for Basic Studies:

- Both dimensions of Diversity and four of the Critical Thinking dimensions produced the lowest across all rubrics.
- Within Information Literacy, IL3 Evaluate information and sources critically produced the lowest scores.
- Within Written Communication, WC2 Content Development produced the lowest scores.

Areas of Relative Strengths for 300- and 400-level Courses:

- IL2 Access needed information and IL4 Use information to accomplish a purpose produced the highest scores within Information Literacy.
- WC5 Control of Syntax and Mechanics produced the highest scores within Written Communication.
- IN1 Topic Selection and IN2 Existing Knowledge, Research, and/or Views produced the highest scores within Inquiry.
- OC1 Organization and OC5 Central Message produced the highest scores within Oral Communication.

Areas of Relative Weaknesses for 300- and 400-level Courses:

- IL3 Evaluate information and sources critically produced the lowest scores within Information Literacy.
- WC2 Content Development produced the lowest scores within Written Communication.
- IN6 Limitations and Implications produced the lowest scores within Inquiry.
- OC3 Delivery produced the lowest scores within Oral Communication.

In general, scores were higher for selecting and explaining information, and were lowest for examining information critically. The findings in this year continue to point towards the need to provide students more opportunities to practice higher-order thinking skills, starting with general education courses.

**Methodology and Process**

This assessment methodology ran fairly smoothly during this implementation. Feedback was generally good from both instructor and scorer participants. Based on the feedback from scorers and instructors, and the results presented in this report, there are some areas for further work.
PROCESS OF SELECTING ASSIGNMENTS

Most assignments selected for scoring with the Diversity, Information Literacy, and Written Communication rubrics matched their respected rubrics well. However, this was not the case for the Critical Thinking rubric. There were three Critical Thinking dimensions that scorers deemed applicable to all assignments, and three dimensions were deemed not applicable for some assignments (CT2.2 was not applicable for two assignments and CT3 and CT4 were not applicable for one assignment). This is similar to the assessment process during the 2009-2010 academic year, when the Critical Thinking rubric was deemed partly not applicable to some assignments. This indicates clearly the need for additional discussion of the dimensions of critical thinking in the initial workshop for instructors, in addition to follow up during the selection process. This is not meant to suggest, however, that all assignments selected for general education assessment purpose must align with all dimensions of the rubric. It would also be helpful for instructional purposes for there to be more dissemination of information about the UNCW Learning Goals, such as through Center for Teaching Excellence workshops, and inclusion of these goals as appropriate in course syllabi.

INTERRATER RELIABILITY AND THE PROCESS OF NORMING SCORERS

Interrater reliability was lower this year than last year. A major factor that may account for this is that last year, scoring was done in a day-long session. Scorers were able to discuss initial work products with partners and still have time to settle in to scoring work products from one assignment. This year, scoring was done in half day sessions, increasing the number of scorers, and decreasing the number of work products scored by each scorer. It’s hypothesized that scoring becomes more consistent as scorers become familiar with using the rubric, and with applying it to a particular assignment. Initial discussion within scoring teams took up a large amount of time. In the spring semester, with 25 scorers in the room at once, these discussions caused the noise level to increase, possibly distracting from conversations. In addition, the number of facilitators during the scoring session was smaller than last year. This reduced the ability for them to interact with the scoring teams as they discussed their initial work product and determined common assumptions. It was also evident from comparing some of the scores to the group assumptions that not all scorers carried through on the common assumptions. Reliability can be improved in three ways. The first is to go back to the all-day scoring session. The second is through improving the training process. Scorers themselves noted that additional training would be beneficial. The third means is to increase the number of facilitators.

LIMITATIONS

Faculty are still becoming acquainted with the rubrics used in general education, therefore student ability is not the only thing reflected in the results. The relative newness of the rubric to both the instructors selecting assignments and the faculty doing the scoring has implications about the reliability of the results.
The sample of student work products was created to be a random sample from representative courses in the Basic Studies curriculum. Still, it represents just five subject areas in one point in time. Therefore, the results in this report should be combined with other evidence to get a more complete picture of student abilities. We will continue to collect evidence from across the spectrum of courses to provide a clearer picture of UNCW student abilities. They should also be combined with finding from other instruments, such as the ETS Proficiency Profile. Another limitation of the results is that interrater reliability measures were lower than optimal.

Although the sample of students who contributed work products was representative of the UNCW undergraduate student body, the sample of students from all race/ethnicity classifications other than white was not large enough to test for differences between groups. Although the ratio is representative of that of the University as a whole, further studies will need to include ways to analyze potential differences so that they can be addressed if necessary.

RECOMMENDATIONS

Although there are limitations to this study, some recommendations that will only have positive effects on student learning can be made in light of these findings. Based on the analysis of the findings presented in the previous sections of this chapter, the Learning Assessment Council approved the following recommendations to improve both student learning and the General Education Assessment process on October 4, 2011.

- Continue efforts to improve interrater reliability
  - Slightly longer workshops for faculty scorers
  - Have more facilitators present during scoring
  - Have longer scoring sessions
- Present targeted content workshops where we invite on-campus experts in various disciplines to share some of the specific ways they are introducing the UNCW Learning Goals into their courses
- Introduce Learning Outcomes and guidelines for Explorations Beyond the Classroom and Thematic Transdisciplinary Clusters which will include a synthesis of information and aspects related to critical thinking. This will increase exposure to these learning goals.
- Develop a central information site where we can share summaries of reports, best practices of our colleagues, relevant literature and information from other institutions, and other information than may be helpful for faculty seeking to improve how they present information related to our learning goals. For this year, the goal would be to start this effort, with expansion in future years.
- Reemphasize the desirability of including the relevant learning outcomes on syllabi of courses approved for the new University Studies curricula (or at least providing links to these goals on the syllabi).
FOLLOW UP ON PREVIOUS RECOMMENDATIONS

2010 RECOMMENDATION 1
Levels of expected performance at the basic studies, or lower division, level should be developed for each rubric.

**Actions**: Sub-committees were created to review the rubric scales against discipline-specific criteria. Instructors providing work products were asked to provide opinions on what they thought would be appropriate benchmark scores for their students on the assignments. No consensus has been reached, but the range has been narrow to 2 or 3, as expected. It may not be possible to narrow further. The fact that the rubric authors list both levels as “Milestone” levels forewarns of this outcome. We will continue to compare results from lower-division courses to both milestones, and we will compare upper division results to rubric levels 3 and 4.

2010 RECOMMENDATION 2
Additional exposure to the content of and rationale for the UNCW Learning Goals should be provided to increase faculty ownership and awareness of these Goals. The LAC will ask the Center for Teaching Excellence to provide a workshop series on these Goals. The LAC will ask the University Curriculum Committee to consider actions in this area.

**Actions**: During the 2010 – 2011 academic year, the Center for Teaching Excellence started a series called “UNCW Learning Goals: Teaching and Assessing.” In Fall 2010, Kim Sawrey and Linda Siefert offered a workshop on Critical Thinking, and in Spring 2011 Anne Pemberton and Linda Siefert offered a workshop on Information Literacy. The workshops, which have been well received, will continue through the eight Learning Goals.

2010 RECOMMENDATION 3
To increase student exposure to the writing process, the Writing Intensive component of University Studies should be implemented by Fall 2012.

**Actions**: The Writing Intensive component of University Studies is on schedule for implementation starting in Fall 2012. A workshop was presented through CTE on teaching and assessment practices for these courses by Cara Cilano and Linda Siefert.

2010 RECOMMENDATION 4
Modifications and improvements to the general education assessment process should be made as needed, including the following: modify rubrics based on feedback, develop benchmarks work products, and enhance instructor and scorer workshops.

**Actions**: The Critical Thinking rubric was modified to separate a dimension into two parts in response to scorer feedback. Subject-matter experts were added into the process of selecting benchmark papers for training scorers.
2010 RECOMMENDATION 5

Long term implementation schedule should provide flexibility for targeting additional sampling for specific learning goals that are characterized by ambiguous or unclear assessment results. For 2010 – 2011, Critical Thinking will be sampled for this purpose.

Actions: Critical Thinking was assessed again in Fall 2010. A four-year plan was created that assesses each learning goal at least once every three years, and provides room for follow-up assessment of any ambiguous or unclear findings.
REFERENCES AND RESOURCES


http://www.uncw.edu/assessment/uncwLearningGoals.html


APPENDIX A RUBRICS USED

AAC&U Critical Thinking Rubric (with modifications)
Locally created Diversity Rubric
AAC&U Information Literacy Rubric
AAC&U Written Communication Rubric
AAC&U Inquiry Rubric
AAC&U Oral Communication Rubric
**CRITICAL THINKING VALUE RUBRIC (AAC&U)**  
Modified January 2011 UNCW

**Definition:** Critical Thinking is a habit of mind characterized by the comprehensive exploration of issues, ideas, artifacts, and events before accepting or formulating an opinion or conclusion.

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

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>Milestones</th>
<th>Capstone</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. Explanation of Issues</strong></td>
<td>Issue/problem to be considered critically is stated without clarification or description.</td>
<td>Issue/problem to be considered critically is stated but description leaves some terms undefined, ambiguities unexplored, boundaries undetermined, and/or backgrounds unknown.</td>
<td>Issue/problem to be considered critically is stated, described, and clarified so that understanding is not seriously impeded by omissions.</td>
</tr>
<tr>
<td><strong>2.1 Evidence</strong></td>
<td>Information is taken from source(s) without any interpretation/evaluation.</td>
<td>Information is taken from source(s) with some interpretation/evaluation, but not enough to develop a coherent analysis or synthesis.</td>
<td>Information is taken from source(s) with enough interpretation/evaluation to develop a coherent analysis or synthesis.</td>
</tr>
<tr>
<td><strong>2.2 Critically examining viewpoints of experts</strong></td>
<td>Viewpoints of experts are taken as fact, without question.</td>
<td>Viewpoints of experts are taken as mostly fact, with little questioning.</td>
<td>Viewpoints of experts are subject to questioning.</td>
</tr>
<tr>
<td><strong>3. Influence of context and assumptions</strong></td>
<td>Shows an emerging awareness of present assumptions (sometimes labels assertions as assumptions). Begins to identify some contexts when presenting a position.</td>
<td>Questions some assumptions. Identifies several relevant contexts when presenting a position. May be more aware of others’ assumptions than one’s own (or vice versa).</td>
<td>Identifies own and others’ assumptions and several relevant contexts when presenting a position.</td>
</tr>
<tr>
<td><strong>4. Student’s position</strong></td>
<td>Specific position is stated, but is simplistic and obvious.</td>
<td>Specific position acknowledges different sides of an issue.</td>
<td>Specific position takes into account the complexities of an issue. Others’ points of view are acknowledged within position.</td>
</tr>
<tr>
<td><strong>5. Conclusions and related outcomes</strong></td>
<td>Conclusion is inconsistently tied to some of the information discussed; related outcomes are oversimplified.</td>
<td>Conclusion is logically tied to information (because information is chosen to fit the desired conclusion); some related outcomes are identified clearly.</td>
<td>Conclusion is logically tied to a range of information, including opposing viewpoints; related outcomes are identified clearly.</td>
</tr>
</tbody>
</table>

NOTES (CONTINUE ON OTHER SIDE):
**UNCW LEARNING GOALS**  
**DIVERSITY SCORING RUBRIC**

**Basic Studies:** Students will describe the importance and implications of human diversity.  
**UNCW Graduate:** Students will examine the importance and implications of human diversity.

**Diversity** constitutes the knowledge, skills and attitudes necessary to examine the importance and implications of cultural and ethnic human differences. Diversity examines the significance of historical, political, social, racial, ethnic and cultural realities through critical thinking to understand and explain their implications in human endeavors.

**Use of Terms in this Rubric:**  
**Human diversity** includes, but is not limited to, differences in race, ethnicity, socio-economic status, gender, age, sexual orientation, ability, and religious affiliation.  
**Disciplinary contexts** include, but are not limited to, historical, political, artistic, geographic, social, and economic.  
**Identify** means point out, and could be as simple as list, although it could also include brief description. **Discuss** means present in detail or investigate by reasoning or argument.  
**Examine** means inquire into carefully or test in order to determine the nature, condition, or quality of a thing.  
**Importance** can mean both the quality or state of being important AND an important aspect of or significance. In this latter regard, **Importance** and **Implications** can overlap.

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

<table>
<thead>
<tr>
<th>Dimension</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Knowledge of Diverse Perspectives</td>
<td>Begins to identify facets of the perspective(s) being studied, but connection may be vague.</td>
<td>Identifies some facets of the perspective(s) being studied.</td>
<td>Discusses many significant facets of the perspective(s) being studied.</td>
<td>Examines the complexity of facets of the perspective(s) being studied.</td>
<td></td>
</tr>
<tr>
<td>2. Contextual Importance and Implications</td>
<td>Shows a vague awareness of the importance and/or implications of human diversity within the present disciplinary context.</td>
<td>Identifies some factors that illustrate the importance and implications of human diversity within the present disciplinary context.</td>
<td>Discusses factors that establish the importance and implications of human diversity within the present disciplinary context.</td>
<td>Thoroughly examines and establishes the importance and implications of human diversity within the present disciplinary context.</td>
<td></td>
</tr>
</tbody>
</table>

NOTES (CONTINUE ON OTHER SIDE):
<table>
<thead>
<tr>
<th>Capstone</th>
<th>Milestones</th>
<th>Benchmark</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Determine the Extent of Information Needed</td>
<td>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.</td>
<td>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.</td>
<td>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.</td>
</tr>
<tr>
<td>2. Access the Needed Information</td>
<td>Accesses information using effective, well-designed search strategies and from most appropriate information sources.</td>
<td>Accesses information using variety of search strategies and from relevant information sources. Demonstrates ability to refine search.</td>
<td>Accesses information using simple search strategies and retrieves information from some relevant, though limited and similar, sources.</td>
</tr>
<tr>
<td>3. Evaluate Information and Its Sources Critically</td>
<td>Information is taken from source(s) with enough interpretation/evaluation to develop a comprehensive analysis of the material; viewpoints of authors are questioned thoroughly.</td>
<td>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.</td>
<td>Information is taken from source(s) with some interpretation/evaluation, but not a coherent analysis of the material; viewpoints of authors are taken as fact, without question.</td>
</tr>
<tr>
<td>4. Use Information Effectively to Accomplish a Specific Purpose</td>
<td>Communicates, organizes and synthesizes information from sources to fully achieve a specific purpose, with clarity and depth</td>
<td>Communicates, organizes and synthesizes information from sources. Intended purpose is achieved.</td>
<td>Communicates and organizes information from sources. The information is not yet synthesized, so the intended purpose is not fully achieved.</td>
</tr>
<tr>
<td>5. Access and Use Information Ethically and Legally</td>
<td>Consistently uses all of the following information use strategies: • use of citations and references, • choice of paraphrasing, summary, or quoting, • using information in ways that are true to original context, • distinguishing between common knowledge and ideas requiring attribution; AND demonstrates a full understanding of the ethical and legal restrictions on the use of published, confidential, and/or proprietary information.</td>
<td>Consistently uses three of the following information use strategies: • use of citations and references, • choice of paraphrasing, summary, or quoting, • using information in ways that are true to original context, • distinguishing between common knowledge and ideas requiring attribution; AND demonstrates a full understanding of the ethical and legal restrictions on the use of published, confidential, and/or proprietary information.</td>
<td>Consistently uses two of the following information use strategies: • use of citations and references, • choice of paraphrasing, summary, or quoting, • using information in ways that are true to original context, • distinguishing between common knowledge and ideas requiring attribution; AND demonstrates a full understanding of the ethical and legal restrictions on the use of published, confidential, and/or proprietary information.</td>
</tr>
</tbody>
</table>

NOTES:
**Written Communication VALUE Rubric**

_for more information, please contact 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.*

<table>
<thead>
<tr>
<th>Capstone</th>
<th>Milestones</th>
<th>Benchmark</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

1. **Context of and Purpose for Writing**

_Includes considerations of audience, purpose, and the circumstances surrounding the writing task(s)._  

- **Demonstrates a thorough understanding of context, audience, and purpose that is responsive to the assigned task(s) and focuses all elements of the work.**
- **Demonstrates adequate consideration of context, audience, and purpose and a clear focus on the assigned task(s) (e.g., the task aligns with audience, purpose, and context).**
- **Demonstrates awareness of context, audience, purpose, and to the assigned task(s) (e.g., begins to show awareness of audience's perceptions and assumptions).**
- **Demonstrates minimal attention to context, audience, purpose, and to the assigned task(s) (e.g., expectation of instructor or self as audience).**

2. **Content Development**

- Uses appropriate, relevant, and compelling content to illustrate mastery of the subject, conveying the writer's understanding, and shaping the whole work.
- Uses appropriate, relevant, and compelling content to explore ideas within the context of the discipline and shape the whole work.
- Uses appropriate and relevant content to develop and explore ideas through most of the work.
- Uses appropriate and relevant content to develop simple ideas in some parts of the work.

3. **Genre and Disciplinary Conventions**

_Formal and informal rules inherent in the expectations for writing in particular forms and/or academic fields._  

- Demonstrates detailed attention to and successful execution of a wide range of conventions particular to a specific discipline and/or writing task(s) including organization, content, presentation, formatting, and stylistic choices
- Demonstrates consistent use of important conventions particular to a specific discipline and/or writing task(s), including organization, content, presentation, and stylistic choices
- Follows expectations appropriate to a specific discipline and/or writing task(s) for basic organization, content, and presentation
- Attempts to use a consistent system for basic organization and presentation.

4. **Sources and Evidence**

- Demonstrates skillful use of high-quality, credible, relevant sources to develop ideas that are appropriate for the discipline and genre of the writing
- Demonstrates consistent use of credible, relevant sources to support ideas that are situated within the discipline and genre of the writing
- Demonstrates an attempt to use credible and/or relevant sources to support ideas that are appropriate for the discipline and genre of the writing
- Demonstrates an attempt to use sources to support ideas in the writing.

5. **Control of Syntax and Mechanics**

- Uses graceful language that skillfully communicates meaning to readers with clarity and fluency, and is virtually error-free.
- Uses straightforward language that generally conveys meaning to readers. The language in the portfolio has few errors.
- Uses language that generally conveys meaning to readers with clarity, although writing may include some errors.
- Uses language that sometimes impedes meaning because of errors in usage.

**NOTES:**
**INQUIRY AND ANALYSIS VALUE RUBRIC**

*for more information, please contact 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._

<table>
<thead>
<tr>
<th></th>
<th>Capstone</th>
<th>Milestones</th>
<th>Benchmark</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Topic selection</strong></td>
<td>Identifies a creative, focused, and manageable topic that addresses potentially significant yet previously less-explored aspects of the topic.</td>
<td>Identifies a focused and manageable/doable topic that appropriately addresses relevant aspects of the topic.</td>
<td>Identifies a topic that while manageable/doable, is too narrowly focused and leaves out relevant aspects of the topic.</td>
<td>Evaluators are encouraged to assign a zero to any work sample or collection of work that does not meet benchmark (cell one) level performance.</td>
</tr>
<tr>
<td><strong>Existing Knowledge, Research, and/or Views</strong></td>
<td>Synthesizes in-depth information from relevant sources representing various points of view/approaches.</td>
<td>Presents in-depth information from relevant sources representing various points of view/approaches.</td>
<td>Presents information from relevant sources representing limited points of view/approaches.</td>
<td>Presents information from irrelevant sources representing limited points of view/approaches.</td>
</tr>
<tr>
<td><strong>Design Process</strong></td>
<td>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.</td>
<td>Critical elements of the methodology or theoretical framework are appropriately developed, however, more subtle elements are ignored or unaccounted for.</td>
<td>Critical elements of the methodology or theoretical framework are missing, incorrectly developed, or unfocused.</td>
<td>Inquiry design demonstrates a misunderstanding of the methodology or theoretical framework.</td>
</tr>
<tr>
<td><strong>Analysis</strong></td>
<td>Organizes and synthesizes evidence to reveal insightful patterns, differences, or similarities related to focus.</td>
<td>Organizes evidence to reveal important patterns, differences, or similarities related to focus.</td>
<td>Organizes evidence, but the organization is not effective in revealing important patterns, differences, or similarities.</td>
<td>Lists evidence, but it is not organized and/or is unrelated to focus.</td>
</tr>
<tr>
<td><strong>Conclusions</strong></td>
<td>States a conclusion that is a logical extrapolation from the inquiry findings.</td>
<td>States a conclusion focused solely on the inquiry findings. The conclusion arises specifically from and responds specifically to the inquiry findings.</td>
<td>States a general conclusion that, because it is so general, also applies beyond the scope of the inquiry findings.</td>
<td>States an ambiguous, illogical, or unsupported conclusion from inquiry findings.</td>
</tr>
<tr>
<td><strong>Limitations and Implications</strong></td>
<td>Insightfully discusses in detail relevant and supported limitations and implications.</td>
<td>Discusses relevant and supported limitations and implications.</td>
<td>Presents relevant and supported limitations and implications.</td>
<td>Presents limitations and implications, but they are possibly irrelevant and unsupported.</td>
</tr>
</tbody>
</table>
**ORAL COMMUNICATION VALUE RUBRIC**

*for more information, please contact 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.*

<table>
<thead>
<tr>
<th>Capstone 4</th>
<th>Milestones 3</th>
<th>Benchmark 1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Organization</strong></td>
<td>Organizational pattern (specific introduction and conclusion, sequenced material within the body, and transitions) is clearly and consistently observable within the presentation.</td>
<td>Organizational pattern (specific introduction and conclusion, sequenced material within the body, and transitions) is intermittently observable within the presentation.</td>
</tr>
<tr>
<td>Language</td>
<td>Language choices are imaginative, memorable, and compelling, and enhance the effectiveness of the presentation. Language in presentation is appropriate to audience.</td>
<td>Language choices are thoughtful and generally support the effectiveness of the presentation. Language in presentation is appropriate to audience.</td>
</tr>
<tr>
<td>Delivery</td>
<td>Delivery techniques (posture, gesture, eye contact, and vocal expressiveness) make the presentation compelling, and speaker appears polished and confident.</td>
<td>Delivery techniques (posture, gesture, eye contact, and vocal expressiveness) make the presentation interesting, and speaker appears comfortable.</td>
</tr>
<tr>
<td>Supporting Material</td>
<td>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.</td>
<td>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.</td>
</tr>
<tr>
<td>Central Message</td>
<td>Central message is compelling (precisely stated, appropriately repeated, memorable, and strongly supported.)</td>
<td>Central message is clear and consistent with the supporting material.</td>
</tr>
</tbody>
</table>
APPENDIX B DIMENSION MEANS AND STANDARD DEVIATIONS

Note of caution: Data from these rubrics cannot be assumed to be interval-level data. That is, although a level 2 is considered higher, or larger, than a level 1, it is not proper to assume that a student that scores at a level 2 is twice as knowledgeable as a student who scored at a level 1; nor can we assume that, whatever the difference is between these two categories, that it is exactly the same as the difference between levels 2 and 3. In addition, the scale of quality criteria may differ between the three rubrics. This table should be analyzed with extreme prudence, and no hypothesis should be made solely from this information.

Table B1 Basic Studies Means and Standard Deviations for Each Rubric Dimension

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>CT1 Explanation of Issues</td>
<td>2.18</td>
<td>1.01</td>
<td>114</td>
</tr>
<tr>
<td>CT2.1 Evidence; selecting and using</td>
<td>2.19</td>
<td>0.99</td>
<td>114</td>
</tr>
<tr>
<td>CT2.2 Evidence; critically examining</td>
<td>1.52</td>
<td>0.83</td>
<td>66</td>
</tr>
<tr>
<td>CT3 Influence of context and assumptions</td>
<td>1.92</td>
<td>1.21</td>
<td>98</td>
</tr>
<tr>
<td>CT4 Student’s position</td>
<td>1.95</td>
<td>1.08</td>
<td>98</td>
</tr>
<tr>
<td>CT5 Conclusions and related outcomes</td>
<td>1.82</td>
<td>1.08</td>
<td>114</td>
</tr>
<tr>
<td>DIV1 Knowledge of diverse perspectives</td>
<td>1.96</td>
<td>1.11</td>
<td>113</td>
</tr>
<tr>
<td>DIV2 Contextual importance and implications</td>
<td>1.82</td>
<td>1.18</td>
<td>111</td>
</tr>
<tr>
<td>IL1 Determine extent of information needed</td>
<td>2.59</td>
<td>0.95</td>
<td>100</td>
</tr>
<tr>
<td>IL2 Access needed information</td>
<td>2.48</td>
<td>0.79</td>
<td>99</td>
</tr>
<tr>
<td>IL3 Evaluate information and sources critically</td>
<td>2.35</td>
<td>0.81</td>
<td>100</td>
</tr>
<tr>
<td>IL4 Use information to accomplish a purpose</td>
<td>2.45</td>
<td>0.90</td>
<td>100</td>
</tr>
<tr>
<td>IL5 Access and use information ethically</td>
<td>2.69</td>
<td>0.98</td>
<td>100</td>
</tr>
<tr>
<td>WC1 Context of and Purpose for Writing</td>
<td>2.26</td>
<td>0.95</td>
<td>80</td>
</tr>
<tr>
<td>WC2 Content Development</td>
<td>1.99</td>
<td>0.92</td>
<td>80</td>
</tr>
<tr>
<td>WC3 Genre and Disciplinary Conventions</td>
<td>2.11</td>
<td>0.80</td>
<td>80</td>
</tr>
<tr>
<td>WC4 Sources and Evidence</td>
<td>2.10</td>
<td>1.07</td>
<td>79</td>
</tr>
<tr>
<td>WC5 Control of Syntax and Mechanics</td>
<td>2.25</td>
<td>0.78</td>
<td>80</td>
</tr>
</tbody>
</table>
Table B2 Pilot Studies (300 and 400 level courses) Means and Standard Deviations for Each Rubric Dimension

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>IL1 Determine extent of information needed</td>
<td>2.50</td>
<td>1.09</td>
<td>40</td>
</tr>
<tr>
<td>IL2 Access needed information</td>
<td>2.85</td>
<td>.736</td>
<td>40</td>
</tr>
<tr>
<td>IL3 Evaluate information and sources critically</td>
<td>2.11</td>
<td>.583</td>
<td>35</td>
</tr>
<tr>
<td>IL4 Use information to accomplish a purpose</td>
<td>2.88</td>
<td>.588</td>
<td>26</td>
</tr>
<tr>
<td>IL5 Access and use information ethically</td>
<td>2.30</td>
<td>.883</td>
<td>40</td>
</tr>
<tr>
<td>WC1 Context of and Purpose for Writing</td>
<td>2.27</td>
<td>.801</td>
<td>33</td>
</tr>
<tr>
<td>WC2 Content Development</td>
<td>2.15</td>
<td>.906</td>
<td>33</td>
</tr>
<tr>
<td>WC3 Genre and Disciplinary Conventions</td>
<td>2.27</td>
<td>.674</td>
<td>33</td>
</tr>
<tr>
<td>WC4 Sources and Evidence</td>
<td>2.36</td>
<td>.603</td>
<td>33</td>
</tr>
<tr>
<td>WC5 Control of Syntax and Mechanics</td>
<td>2.61</td>
<td>.556</td>
<td>33</td>
</tr>
<tr>
<td>IN1 Topic Selection</td>
<td>2.64</td>
<td>.497</td>
<td>14</td>
</tr>
<tr>
<td>IN2 Existing Knowledge, Research, and/or Views</td>
<td>2.64</td>
<td>.633</td>
<td>14</td>
</tr>
<tr>
<td>IN3 Design Process</td>
<td>2.21</td>
<td>.699</td>
<td>14</td>
</tr>
<tr>
<td>IN4 Analysis</td>
<td>2.50</td>
<td>.519</td>
<td>14</td>
</tr>
<tr>
<td>IN5 Conclusions</td>
<td>2.50</td>
<td>.519</td>
<td>14</td>
</tr>
<tr>
<td>IN6 Limitations and Implications</td>
<td>0.50</td>
<td>.941</td>
<td>14</td>
</tr>
<tr>
<td>OC1 Organization</td>
<td>2.88</td>
<td>.738</td>
<td>57</td>
</tr>
<tr>
<td>OC2 Language</td>
<td>2.77</td>
<td>.627</td>
<td>57</td>
</tr>
<tr>
<td>OC3 Delivery</td>
<td>2.74</td>
<td>.768</td>
<td>57</td>
</tr>
<tr>
<td>OC4 Supporting Material</td>
<td>2.77</td>
<td>.682</td>
<td>57</td>
</tr>
<tr>
<td>OC5 Central Message</td>
<td>2.86</td>
<td>.743</td>
<td>57</td>
</tr>
</tbody>
</table>
### Appendix C Correlations Between Rubric Dimensions

Table C1 Correlation between Dimensions Basic Studies Fall 2010—Spearman rho Rank Order Correlation Coefficients

<table>
<thead>
<tr>
<th></th>
<th>CT1</th>
<th>CT2.1</th>
<th>CT2.2</th>
<th>CT3</th>
<th>CT4</th>
<th>CT5</th>
<th>DV1</th>
<th>DV2</th>
</tr>
</thead>
<tbody>
<tr>
<td>CT1</td>
<td>1.000</td>
<td>.754</td>
<td>.543</td>
<td>.821</td>
<td>.763</td>
<td>.718</td>
<td>.696</td>
<td>.730</td>
</tr>
<tr>
<td>n</td>
<td>114</td>
<td>114</td>
<td>66</td>
<td>98</td>
<td>98</td>
<td>114</td>
<td>113</td>
<td>111</td>
</tr>
<tr>
<td>CT2</td>
<td>.754*</td>
<td>1.000</td>
<td>.443*</td>
<td>.836</td>
<td>.783</td>
<td>.703</td>
<td>.795</td>
<td>.803</td>
</tr>
<tr>
<td>n</td>
<td>114</td>
<td>114</td>
<td>66</td>
<td>98</td>
<td>98</td>
<td>114</td>
<td>113</td>
<td>111</td>
</tr>
<tr>
<td>CT3</td>
<td>.543*</td>
<td>.443*</td>
<td>1.000</td>
<td>.683</td>
<td>.517</td>
<td>.409</td>
<td>.217</td>
<td>.262</td>
</tr>
<tr>
<td>n</td>
<td>66</td>
<td>66</td>
<td>66</td>
<td>50</td>
<td>50</td>
<td>66</td>
<td>65</td>
<td>63</td>
</tr>
<tr>
<td>CT4</td>
<td>.821*</td>
<td>.836</td>
<td>.683*</td>
<td>.100</td>
<td>.833</td>
<td>.752</td>
<td>.809</td>
<td>.813</td>
</tr>
<tr>
<td>n</td>
<td>98</td>
<td>98</td>
<td>50</td>
<td>98</td>
<td>98</td>
<td>98</td>
<td>97</td>
<td>95</td>
</tr>
<tr>
<td>CT5</td>
<td>.763*</td>
<td>.783*</td>
<td>.517*</td>
<td>.833*</td>
<td>1.000</td>
<td>.728</td>
<td>.778</td>
<td>.811*</td>
</tr>
<tr>
<td>n</td>
<td>98</td>
<td>98</td>
<td>98</td>
<td>98</td>
<td>98</td>
<td>98</td>
<td>97</td>
<td>95</td>
</tr>
<tr>
<td>CT6</td>
<td>.718*</td>
<td>.703*</td>
<td>.409*</td>
<td>.752*</td>
<td>.728*</td>
<td>1.000</td>
<td>.704*</td>
<td>.700*</td>
</tr>
<tr>
<td>n</td>
<td>114</td>
<td>114</td>
<td>66</td>
<td>98</td>
<td>98</td>
<td>114</td>
<td>113</td>
<td>111</td>
</tr>
<tr>
<td>DIV1</td>
<td>.696*</td>
<td>.795*</td>
<td>.217</td>
<td>.809*</td>
<td>.778*</td>
<td>.704*</td>
<td>1.000</td>
<td>.859*</td>
</tr>
<tr>
<td>n</td>
<td>113</td>
<td>113</td>
<td>65</td>
<td>97</td>
<td>97</td>
<td>113</td>
<td>113</td>
<td>111</td>
</tr>
<tr>
<td>DIV2</td>
<td>.730*</td>
<td>.803*</td>
<td>.262</td>
<td>.813*</td>
<td>.811*</td>
<td>.700*</td>
<td>.859*</td>
<td>1.000</td>
</tr>
<tr>
<td>n</td>
<td>111</td>
<td>111</td>
<td>63</td>
<td>95</td>
<td>95</td>
<td>111</td>
<td>111</td>
<td>111</td>
</tr>
</tbody>
</table>

*Statistically significant at the .05 level  
**Statistically significant at the .01 level
### Table C2 Correlation between Dimensions Basic Studies Spring 2011—Spearman rho Rank Order Correlation Coefficients

<table>
<thead>
<tr>
<th></th>
<th>IL1</th>
<th>IL2</th>
<th>IL3</th>
<th>IL4</th>
<th>IL5</th>
<th>WC1</th>
<th>WC2</th>
<th>WC3</th>
<th>WC4</th>
<th>WC5</th>
</tr>
</thead>
<tbody>
<tr>
<td>IL1</td>
<td>1.000</td>
<td>0.298*</td>
<td>0.420*</td>
<td>0.554*</td>
<td>0.284*</td>
<td>-0.090</td>
<td>-0.066</td>
<td>-0.032</td>
<td>0.049</td>
<td>0.085</td>
</tr>
<tr>
<td>n</td>
<td>100</td>
<td>99</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>63</td>
<td>63</td>
<td>63</td>
<td>63</td>
<td>63</td>
</tr>
<tr>
<td>IL2</td>
<td>0.298*</td>
<td>1.000</td>
<td>0.513*</td>
<td>0.537*</td>
<td>0.508*</td>
<td>0.004</td>
<td>0.088</td>
<td>0.039</td>
<td>0.146</td>
<td>0.226</td>
</tr>
<tr>
<td>IL3</td>
<td>0.420*</td>
<td>0.513*</td>
<td>1.000</td>
<td>0.676*</td>
<td>0.579*</td>
<td>0.085</td>
<td>0.121</td>
<td>0.048</td>
<td>0.213</td>
<td>0.096</td>
</tr>
<tr>
<td>n</td>
<td>100</td>
<td>99</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>63</td>
<td>63</td>
<td>63</td>
<td>62</td>
<td>63</td>
</tr>
<tr>
<td>IL4</td>
<td>0.554*</td>
<td>0.537*</td>
<td>0.676*</td>
<td>1.000</td>
<td>0.593*</td>
<td>0.056</td>
<td>0.065</td>
<td>0.152</td>
<td>0.148</td>
<td>0.175</td>
</tr>
<tr>
<td>n</td>
<td>100</td>
<td>99</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>63</td>
<td>63</td>
<td>63</td>
<td>62</td>
<td>63</td>
</tr>
<tr>
<td>IL5</td>
<td>0.284*</td>
<td>0.508*</td>
<td>0.579*</td>
<td>0.593*</td>
<td>1.000</td>
<td>0.251*</td>
<td>0.137</td>
<td>0.157</td>
<td>0.134</td>
<td>0.253*</td>
</tr>
<tr>
<td>n</td>
<td>100</td>
<td>99</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>63</td>
<td>63</td>
<td>63</td>
<td>62</td>
<td>63</td>
</tr>
<tr>
<td>WC1</td>
<td>-0.090</td>
<td>0.004</td>
<td>0.085</td>
<td>0.056</td>
<td>0.251*</td>
<td>1.000</td>
<td>0.729*</td>
<td>0.663**</td>
<td>0.539**</td>
<td>0.398**</td>
</tr>
<tr>
<td>n</td>
<td>63</td>
<td>62</td>
<td>63</td>
<td>63</td>
<td>63</td>
<td>80</td>
<td>80</td>
<td>80</td>
<td>79</td>
<td>80</td>
</tr>
<tr>
<td>WC2</td>
<td>-0.006</td>
<td>0.088</td>
<td>0.121</td>
<td>0.065</td>
<td>0.137</td>
<td>0.729*</td>
<td>1.000</td>
<td>0.719**</td>
<td>0.650**</td>
<td>0.564**</td>
</tr>
<tr>
<td>n</td>
<td>63</td>
<td>62</td>
<td>63</td>
<td>63</td>
<td>63</td>
<td>80</td>
<td>80</td>
<td>80</td>
<td>79</td>
<td>80</td>
</tr>
<tr>
<td>WC3</td>
<td>-0.032</td>
<td>0.039</td>
<td>0.048</td>
<td>0.152</td>
<td>0.157</td>
<td>0.663**</td>
<td>0.719**</td>
<td>1.000</td>
<td>0.585**</td>
<td>0.536**</td>
</tr>
<tr>
<td>n</td>
<td>63</td>
<td>62</td>
<td>63</td>
<td>63</td>
<td>63</td>
<td>80</td>
<td>80</td>
<td>80</td>
<td>79</td>
<td>80</td>
</tr>
<tr>
<td>WC4</td>
<td>0.049</td>
<td>0.146</td>
<td>0.213</td>
<td>0.148</td>
<td>0.134</td>
<td>0.539**</td>
<td>0.650**</td>
<td>0.585**</td>
<td>1.000</td>
<td>0.665**</td>
</tr>
<tr>
<td>n</td>
<td>62</td>
<td>61</td>
<td>62</td>
<td>62</td>
<td>62</td>
<td>79</td>
<td>79</td>
<td>79</td>
<td>79</td>
<td>79</td>
</tr>
<tr>
<td>WC5</td>
<td>0.085</td>
<td>0.226</td>
<td>0.096</td>
<td>0.175</td>
<td>0.253*</td>
<td>0.398**</td>
<td>0.564**</td>
<td>0.536**</td>
<td>0.665**</td>
<td>1.000</td>
</tr>
<tr>
<td>n</td>
<td>63</td>
<td>62</td>
<td>63</td>
<td>63</td>
<td>63</td>
<td>80</td>
<td>80</td>
<td>80</td>
<td>79</td>
<td>80</td>
</tr>
</tbody>
</table>

*Statistically significant at the .05 level
**Statistically significant at the .01 level
### Table C3 Correlation between Dimensions Pilot 1 Spring 2011—
Spearman rho Rank Order Correlation Coefficients

<table>
<thead>
<tr>
<th></th>
<th>IL1</th>
<th>IL2</th>
<th>IL3</th>
<th>IL4</th>
<th>IL5</th>
<th>WC1</th>
<th>WC2</th>
<th>WC3</th>
<th>WC4</th>
<th>WC5</th>
</tr>
</thead>
<tbody>
<tr>
<td>IL1</td>
<td>1.000</td>
<td>.598</td>
<td>.658</td>
<td>.610</td>
<td>.337</td>
<td>.302</td>
<td>.439</td>
<td>.501</td>
<td>.348</td>
<td>.190</td>
</tr>
<tr>
<td>n</td>
<td>40</td>
<td>40</td>
<td>35</td>
<td>26</td>
<td>40</td>
<td>28</td>
<td>28</td>
<td>28</td>
<td>28</td>
<td>28</td>
</tr>
<tr>
<td>IL2</td>
<td>.598</td>
<td>1.000</td>
<td>.253</td>
<td>.613</td>
<td>.217</td>
<td>.270</td>
<td>.382</td>
<td>.401</td>
<td>.333</td>
<td>.199</td>
</tr>
<tr>
<td>n</td>
<td>40</td>
<td>40</td>
<td>35</td>
<td>26</td>
<td>40</td>
<td>28</td>
<td>28</td>
<td>28</td>
<td>28</td>
<td>28</td>
</tr>
<tr>
<td>IL3</td>
<td>.658</td>
<td>.253</td>
<td>1.000</td>
<td>.443</td>
<td>.479</td>
<td>.213</td>
<td>.267</td>
<td>.455</td>
<td>.237</td>
<td>.074</td>
</tr>
<tr>
<td>n</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>21</td>
<td>35</td>
<td>23</td>
<td>23</td>
<td>23</td>
<td>23</td>
<td>23</td>
</tr>
<tr>
<td>IL4</td>
<td>.610</td>
<td>.613</td>
<td>.443</td>
<td>1.000</td>
<td>.262</td>
<td>.430</td>
<td>.309</td>
<td>.558</td>
<td>.595</td>
<td>.203</td>
</tr>
<tr>
<td>n</td>
<td>26</td>
<td>26</td>
<td>21</td>
<td>26</td>
<td>26</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
</tr>
<tr>
<td>IL5</td>
<td>.337</td>
<td>.217</td>
<td>.479</td>
<td>.262</td>
<td>1.000</td>
<td>.052</td>
<td>.014</td>
<td>.317</td>
<td>.458</td>
<td>-.007</td>
</tr>
<tr>
<td>n</td>
<td>40</td>
<td>40</td>
<td>35</td>
<td>26</td>
<td>40</td>
<td>28</td>
<td>28</td>
<td>28</td>
<td>28</td>
<td>28</td>
</tr>
<tr>
<td>WC1</td>
<td>.302</td>
<td>.270</td>
<td>.213</td>
<td>.430</td>
<td>.052</td>
<td>1.000</td>
<td>.703</td>
<td>.705</td>
<td>.657</td>
<td>.530</td>
</tr>
<tr>
<td>n</td>
<td>28</td>
<td>28</td>
<td>23</td>
<td>17</td>
<td>28</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>33</td>
</tr>
<tr>
<td>WC2</td>
<td>.439</td>
<td>.382</td>
<td>.267</td>
<td>.309</td>
<td>.014</td>
<td>.703</td>
<td>1.000</td>
<td>.600</td>
<td>.573</td>
<td>.581</td>
</tr>
<tr>
<td>n</td>
<td>28</td>
<td>28</td>
<td>23</td>
<td>17</td>
<td>28</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>33</td>
</tr>
<tr>
<td>WC3</td>
<td>.501</td>
<td>.401</td>
<td>.455</td>
<td>.558</td>
<td>.317</td>
<td>.705</td>
<td>.600</td>
<td>1.000</td>
<td>.563</td>
<td>.596</td>
</tr>
<tr>
<td>n</td>
<td>28</td>
<td>28</td>
<td>23</td>
<td>17</td>
<td>28</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>33</td>
</tr>
<tr>
<td>WC4</td>
<td>.348</td>
<td>.333</td>
<td>.237</td>
<td>.595</td>
<td>.458</td>
<td>.657</td>
<td>.573</td>
<td>.563</td>
<td>1.000</td>
<td>.421</td>
</tr>
<tr>
<td>n</td>
<td>28</td>
<td>28</td>
<td>23</td>
<td>17</td>
<td>28</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>33</td>
</tr>
<tr>
<td>WC5</td>
<td>.190</td>
<td>.199</td>
<td>.074</td>
<td>.203</td>
<td>-.007</td>
<td>.530</td>
<td>.581</td>
<td>.596</td>
<td>.421</td>
<td>1.000</td>
</tr>
<tr>
<td>n</td>
<td>28</td>
<td>28</td>
<td>23</td>
<td>17</td>
<td>28</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>33</td>
</tr>
</tbody>
</table>
APPENDIX D DETAILED SCORER FEEDBACK

Scorer Qualitative Comments.

What parts of the process worked the best?

Fall 2010 comments:
- The norming session and overall organization of the process. It was done well.
- The norming session was a very good part of the plan. Discussing the 1st paper with a partner gave me confidence to do the others.
- The norming session was very helpful. It made me feel more comfortable with my scoring practices.
- The norming session is really necessary prior to the scoring session—I would have felt really overwhelmed if I hadn't attended.
- I thought this process was explained well and was very organized. The relaxed environment was also very helpful.
- Overall the quality criteria worked well.
- The first scoring session where you compare your scoring results and feel confident scoring on your own
- Was organized very well
- Training was helpful

Spring 2011 comments:
- The norming exercise was very helpful prior to the scoring itself.
- I enjoyed learning how the scores would be analyzed in the norming session. And the organization of the entire event was wonderful.
- The discussion with peers to work out an issue
- Working in groups
- The training was very helpful especially recommending to underline the words most relative in the rubric. Made the review process flow easily.
- The norming process: thorough understanding of rubric (and what to do concerning assumptions)
- The norming session was extremely helpful and the guidance and discussion of how to score was excellent.
- I think giving feedback after scoring was the best part of the process.
- Organized
- Organizational structure
- Norming session discussion was extremely helpful.
- Workshop was very good preparation—very clear expectations
- Norming session
- Well-organized
- The norming session was invaluable to my participation in the scoring process.
- Norming session, networking outside of my area/school
- Very well organized! The norming session is particularly important/useful.
- The norming session was enormously helpful.
- All materials were organized and easy to use.
• Discussing the first student assignment with other colleagues was helpful.
• Pre workshop was very valuable. Linda is well organized and adequately explains the process.
• Getting trained first and working with other faculty from a different discipline.

In what ways could the scoring process be improved?

Fall 2010 comments:
• Perhaps more communication with instructor in which we are scoring. Not sure prompts addressed issues adequately-or perhaps they were not supposed to.
• Training, perhaps, could take longer. Finding consistencies among scorers concerning the rubrics may need more practice. I'm not convinced the process is working as desired unless the prompts/assignments are better.
• This is a good process--I have no suggestions at this time!
• Number the elements on the rubric. I basically worked alone and it was fine--not sure if a partner is necessary but perhaps the proof will be in the scoring.
• The only improvements I can think of are listed on the rubric feedback.
• I think maybe adding racism to diversity. I was amazed at how many papers did not use the term racism. Probably paper that fit the rubric a little better.
• It was great!
• Need criteria for prompts scored
• The prompts for the actual assignments to be scored could be analyzed more closely before selected.

Spring 2011 comments:
• Allocate more time for norming/workshop prior to scoring
• Noise level during discussion portion was a bit distracting
• The first part I didn’t get done because the assignment was 6-7 pages while the 2nd packet I finished early because the assignment was shorter. I don’t know if there is a way to balance that out.
• I think it all worked well.
• Perhaps have two 2-hour sessions. My sense was that some were getting burned out two hours into the scoring.
• I experienced fatigue near the end by working straight through lunch. When I typically grade papers for my own class I usually do not complete it in one sitting. I don’t feel my scoring was as well done as the morning.
• For scientific papers/research oriented things, it would be more effective to have people trained in that discipline to review those because of specialized knowledge of the sources/relevancy of the content.
• Less noise within the room
• None known
• Allow more time. I had less than 45 minutes to evaluate 10 papers.
• Discuss potential issues or problems in using the rubric.
• I would be interested to see if groups of 3 had better success at reaching consensus (as pairs seemed to go whichever way the most opinionated person went).
• More privacy/space to read quietly
• My papers were particularly thick and not possible to evaluate in the given time. Some of the assignments did not lend themselves well to this evaluation, such as the annotated bibliography where arguments cannot be adequately developed.
• Time became a crunch, but that’s because I work slowly.
• To decrease talking during the process. I was easily distracted from reading the assignments.
• Perhaps allow more time to score fewer student samples.
• It worked great! Do not change. Maybe snacks for training day.

Any other comments or suggestions.

Fall 2010 comments:
• Enjoyed and learned. Thanks.
• Until there is some consistency in how assignments or prompts to the writing are constructed, the outcomes may not be valid in some way.
• Thanks! This was a good experience!
• Thank you!
• I'd love to do this again! Thank you for the opportunity!
• Compared to last spring--much better to do it split over two days--it wasn't as draining.
• Thanks for the experience!
• Thanks again for lunch!

Spring 2011 comments:
• Great way to assess student writing learning goals at UNCW—keep it up!
• The discussion of the first essay was beneficial. Thank you!
• Coffee for (early) morning session would be greatly appreciated!
• Thanks for lunch!
• Do instructors who teach writing intensive courses know how these are being assess with relationship to overall university goals?
• The rubric at times was a challenge—the second assignment from Nursing seemed too basic and resulted in high scores for all dimensions except one. I did enjoy this very much! Thank you.
• Enjoyed contributing and would love to do it again!
• Great-enjoyed-would do it again!
• Well done!
• Very well done! Thank you!
• Great overall!
• The process has helped me to become more familiar with-and comfortable with-UNCW Learning Goals and overall assessment. Thank you!
• Thank you. I enjoyed it. Great way to interact with colleagues from other disciplines and to learn more about the evaluation process across the University.
• Perhaps more time (opportunities) for group/pair discussions.
• Job well done. It was fun and informational that makes me want to work with other disciplines outside nursing.