

# General Education Assessment 2015-2016

## INQUIRY and ANALYSIS

### Overview Information for Inquiry and Analysis Assessment

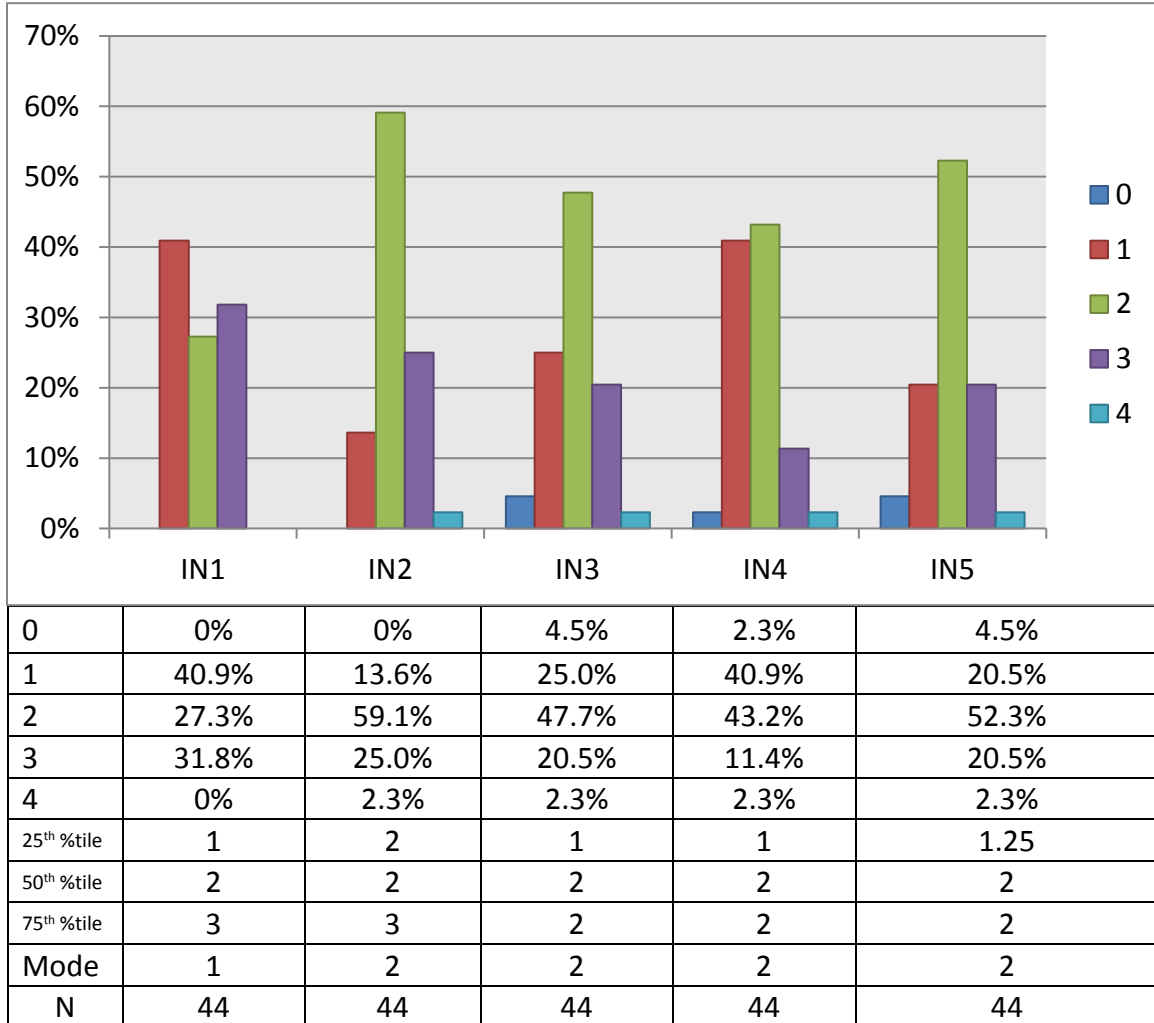
- Total number of courses: 4 with a total of 13 sections
  - LOWER: 1 course, EVSL 195, with 10 sections
  - UPPER: 3 courses: ARH 476, INT 490, and SOC 490, with 1 section each
- All work products were collected during the Fall Semester of 2015 and Spring Semester of 2016
- Types of work products collected: group papers, individual research papers, and term papers
- Learning goal assessed: Inquiry
- University Studies components represented: Capstone Course, Scientific Approaches to the Natural World
- Rubric used: AAC&U Inquiry & Analysis rubric
- Total number of faculty scorers: 8
- Total number of work products sampled: 89
- Total number of work products scored: 88 (one work product was not scored because the scorer felt that the work product was not applicable to the rubric)
  - Number lower division work products scored: 44
  - Number of upper division work products scored: 44
- Number of students contributing to each work product (most of the lower-division work was group work)
  - LOWER:
    - 1 student contributing = 2
    - 3 students contributing = 14
    - 4 students contributing = 24
    - 5 students contributing = 3
    - 6 students contributing = 1
  - UPPER:
    - 1 student contributing = 44
- Total number of multiple-scored work products: 30
  - LOWER: 12
  - UPPER: 18

## Findings

### Lower Division Results

The below table shows the applicable Inquiry and Analysis scores for work sampled during Fall 2015 and Spring 2016.

*Lower Division Score Distribution without N/A or blanks*



Dimension IN 6 Limitations and Implications was deemed not applicable for all the lower-division assignments and so is not represented on the above chart and table. The other dimensions were deemed applicable for all the lower-division papers. Looking at the number of level three and level four scores, dimensions IN 1 Topic Selection and IN 2 Existing Knowledge had the highest percentage of scores at these higher achievement levels. However, IN 1 did not have any scores at the highest score level, level four. Dimension IN 4 Analysis had the largest percentage of scores at the zero and one score level. IN 3 Design Process and IN 5 Conclusions both had higher percentages of level-

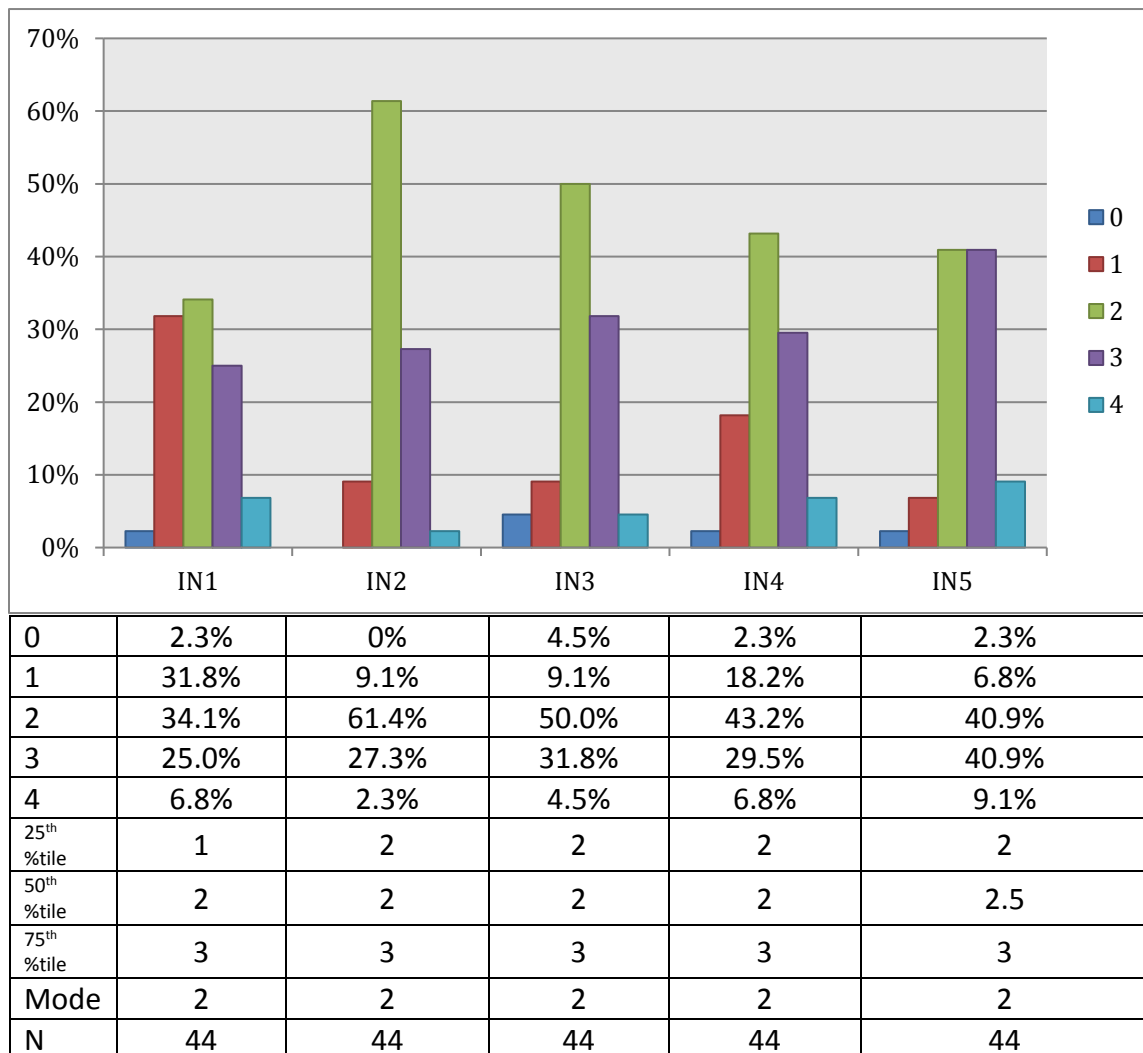
zero scores though both dimensions showed scores fairly normally distributed across the five score levels.

All dimension scores were significantly positively correlated with the other dimension scores at the 0.01 level of significance.

### Upper Division Results

The following table shows the applicable Inquiry scores for work sampled from upper-division courses during the 2015-2016 academic year.

*Upper Division Score Distribution without N/A or blanks*



Scorers determined dimensions IN 1 through IN 5 to be applicable for all assignments. IN 6 Limitations and Implications was deemed applicable for only five papers, but by only one scorer, and so was not included in the score analysis. Looking at the other five

dimensions, scores on IN 5 Conclusions were next highest, based on the percentage of scores and the three and four score level. Scores on IN 1 Analysis had the greatest percentage of scores at the zero score level and one score level. IN 2 Existing Knowledge and IN 3 Design Process were fairly similar in score distribution, though IN 3 had a larger spread with some work products scoring at all score levels. IN 1 Topic Selection shows a score distribution that is approaching normal.

Scores on rubric dimensions IN 1 through IN 5 were significantly positively correlated and the 0.01 level with scores on the other dimensions with the exception of IN 1 with IN 2, which were not significantly correlated.

### *Analysis Across Course Levels For Inquiry and Analysis*

#### *Demographic and Preparedness Findings*

A large number of work products collected were group assignments, which makes comparisons at the student level not possible for those work products. However, for the work products that were completed by only one student, there were no statistically significant differences between the means, medians, and score distributions of gender, ethnicity groups, or honors vs. non-honors students.

To compare scores on single-student work products based on number of credit hours completed, two methods were used. First, students were grouped into four categories, those having completed 0-30 credit hours, 31-60 credit hours, 61-90, and over 90 credit hours. Comparison of means (using ANOVA), medians (using Independent Samples test of medians) and distributions (using the Mann-Whitney U statistic) showed no statistically significant differences between the groups. Looking at Spearman rho correlation coefficients, ACT scores were significantly negatively correlated with IN2 (-.773\*). Additionally, SAT Verbal scores were significantly positively correlated with IN1 (.433\*) and SAT Math scores were significantly negatively correlated with IN2 (-.448\*).

#### *Comparisons Between Criteria*

Comparisons were made across a number of criteria. Work was collected from courses taught by both tenure-line faculty and part-time faculty; there was no significant difference in scores on work collected from the two types of faculty. Additionally, there was no difference in scores between upper- and lower-division courses across all dimensions. Likewise, there were no significant differences in scores across the University Studies components.

## Interrater Reliability for Inquiry & Analysis

There were a number of common papers scored between each pair of faculty scorers so that interrater reliability could be assessed (30 out of 88, or 34.09% of the total number of papers). The following table summarizes the reliability measures.

Dimension	N	Percent Agreement	Plus Percent Adjacent	Krippendorff's alpha
IN 1 Topic Selection	30	30.0%	63.3%	-.0869
IN 2 Existing Knowledge, Research and/or Views	30	23.3%	90.0%	.1290
IN 3 Design Process	30	40.0%	83.3%	.2381
IN 4 Analysis	30	46.7%	80.0%	.3717
IN 5 Conclusions	30	33.3%	86.7%	.2939

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

Comparing the results of the reliability indices for this study to the benchmark of .67 for Krippendorff's Alpha, there are no dimension of the rubric that meets these standards. Looking at percent agreement plus adjacent (that is, the scores that were within one level of each other), we find that all dimensions had at least 80% of scores in agreement or within one level of each other for all dimensions except IN 1, which had a lower percent agreement plus percent adjacent percentage of 63.3%.

## Discussion

The table below shows the percent of work products scored at a level two or higher and the percent of work products scored at a level three or higher for each dimension for lower-division courses and then for upper-division courses.

Inquiry: Lower-Division Courses		
Dimension	% of Work Products Scored Two or Higher	% of Work Products Scored Three or Higher
IN 1 Topic Selection	60.0%	31.8%
IN 2 Existing Knowledge, Research and/or Views	86.4%	27.3%
IN 3 Design Process	70.5%	22.7%
IN 4 Analysis	56.8%	13.6%
IN 5 Conclusions	75.0%	22.7%
IN 6 Limitations and Implications	--	--
Inquiry: Upper-Division Courses		
Dimension	% of Work Products Scored Two or Higher	% of Work Products Scored Three or Higher
IN 1 Topic Selection	65.9%	31.8%
IN 2 Existing Knowledge, Research and/or Views	90.9%	29.5%
IN 3 Design Process	86.4%	36.4%
IN 4 Analysis	79.5%	36.4%
IN 5 Conclusions	90.9%	50.0%
IN 6 Limitations and Implications	--	--

Both lower-division and upper-division courses had a minimum of 56% of papers scoring at least a two across all dimensions, though upper-division courses had higher percentages of papers scoring above a two for all dimensions. Looking at the percentage of work products scoring three or higher, lower- and upper-division courses showed fairly similar percentages for IN 1 and IN 2. However, looking at the other dimensions, IN 3, IN 4, and IN 5, the upper-division courses had a higher number of work products scoring at a three or higher than did the lower-division courses.