

General Education Assessment 2015-2016

DATA ANALYSIS

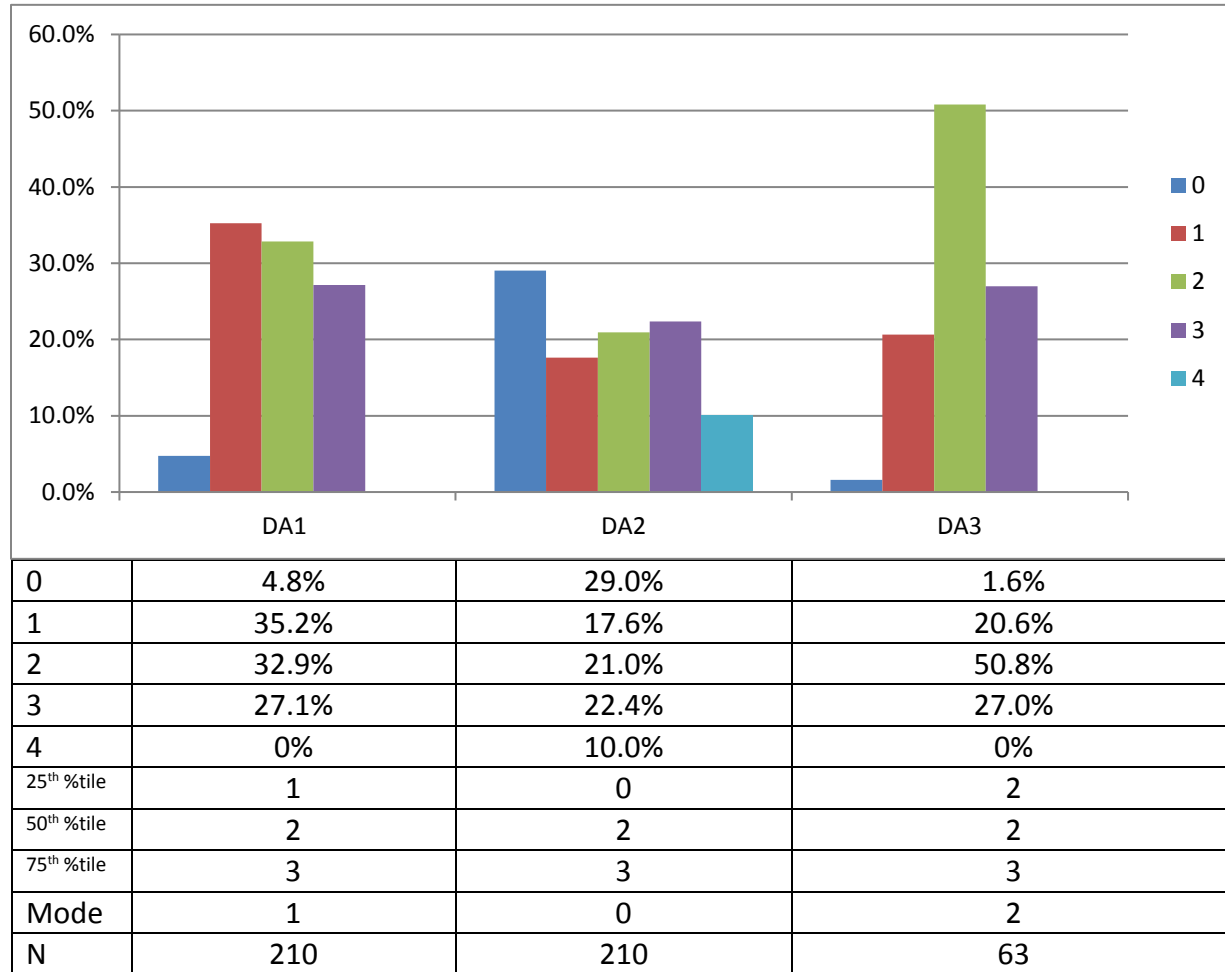
Overview Information for Data Analysis Assessment

- Total Number of Courses: 2 lower division courses with a total of 7 sections
 - MAT151: 5 sections
 - STT215: 2 sections
- All work products were collected during the Fall semester of 2015
- Types of work products collected: multi-part math question
- Learning goal assessed: Inquiry: Data Analysis
- University Studies components represented: Mathematics and Statistics, Quantitative and Logical Reasoning
- Rubric used: locally-developed Data Analysis rubric
- Total number of faculty scorers: 4
- Total number of work products sampled and scored: 210 (scores were not necessarily given in all dimensions for every work product)
- All work products were multiple-scored

Findings

The below table shows the applicable Data Analysis scores for work sampled during Fall 2015.

Data Analysis Score Distribution without N/A or blanks



Dimension DA2 Explanation of Results had both the highest percentage of four-level scores as well as the highest percentage of zero-level scores. Dimension DA1, Summarizing and Analyzing Data, and no level four scores. DA 3, Reasoning, was scored for the fewest number of papers; most scorers deemed DA 3 to be not applicable for the assignments scored.

Scores for DA 1 and DA 3 were both significantly correlated at the 0.01 level with scores for DA2. Scores for DA 1 and DA 3 were not significantly correlated, however.

Analysis Across Course Levels For Data Analysis

Demographic and Preparedness Findings

There were no statistically significant differences between the means, medians, and score distributions of gender, ethnicity groups, transfer students vs. non-transfer students, honors vs. non-honors students, or Isaac Bear vs. non-Isaac Bear students.

To compare scores based on number of credit hours completed, two methods were used. First, students were grouped into four categories, those having completed 0-30 credit hours, 31-60 credit hours, 61-90, and over 90 credit hours. Comparison of means (using ANOVA), medians (using Independent Samples test of medians) and distributions (using the Mann-Whitney U statistic) showed a statistically significant difference between the groups for DA 2 such that students with 91 credit hours or more scored the highest followed by students with 61-90 credit hours, followed by students with 31-60 credit hours, and finally students with 30 credit hours or less. Looking at Spearman rho correlation coefficients, the number of total hours completed was significantly positively correlated with DA 2 (.321**). Additionally, GPA was significantly positively correlated with DA 2 (.242**) as was ACT score (.202*).

Comparisons Between Criteria

Scores were compared across a number of criteria. There was a statistical difference between the scores on all dimensions for work completed in courses taught by tenure-line faculty vs. non-tenure-line faculty. Comparison of means (using ANOVA), medians (using Independent Samples test of medians) and distributions (using the Mann-Whitney U statistic) showed a statistically significant difference between the groups for DA 1 and DA 2 such that scores on work completed in courses taught by tenure-line faculty scored the highest. Additionally, there was a statistically significant difference between the groups for DA 3 such that scores on work completed in courses taught by tenure-line faculty scored the highest. Additionally, DA 3 was only found to be applicable in work from these same tenure-line faculty. All work products being collected from lower-division courses, so comparisons between scores from lower-division vs. upper-division courses could not be made. Finally, the sampled courses were in the same university studies components, so no comparisons regarding University Studies components could be made.

Interrater Reliability

All papers were duplicate-scored between each pair of faculty scorers so that interrater reliability could be assessed. The following table summarizes the reliability measures.

Dimension	N	Percent Agreement	Plus Percent Adjacent	Krippendorff's alpha
DA 1 Summarizing and Analyzing Data	210	75.7%	99.5%	.647
DA 2 Explanation of Results	210	35.2%	77.1%	.188
DA 3 Reasoning	63	38.1%	81.1%	-.125

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 [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 though DA 1 gets close. 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 75% of scores in agreement or within one level of each other.

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.

Inquiry Data Analysis		
Dimension	% of Work Products Scored Two or Higher	% of Work Products Scored Three or Higher
DA 1 Summarizing and Analyzing Data	60.0%	27.1%
DA 2 Explanation of Results	53.4%	32.4%
DA 3 Reasoning	77.8%	27.0%

Looking at the total percentage of work products, between one-quarter and one-third of student work products scored either a three or four (with a four being the expectation of what a UNCW graduate would score at graduation) on all three Inquiry Data Analysis dimensions. While one dimension, DA 2 Explanation of Results, had just under half of the scores at a level

zero or level one, it is important remember that such scores are the expectation of what an entering freshman might achieve, and that the courses sampled for this assessment were lower-division courses.