

# Comparison of Freshman Baseline with First Year Seminar and Senior Exiting Assessment Results

## Academic Year 2016 – 2017

**Summer Assessment Workgroup Members:** Marie Archambault, Cam Brammer, Kim DeTardo-Bora, Robert Ellison, Victor Fet, Marty Laubach, Joan St. Germain, Anita Walz, Mary Welch, Mary Beth Reynolds and Tim Melvin (Office of Assessment), and Doug Nichols (Academic Affairs Technical Support).

### Executive Summary

#### *Background*

#### *Recommendations from 2016 Assessment Workgroup (with current status in red)*

##### *Recommendations regarding baseline and senior assessments*

1. The Assessment Workgroup recommended that baseline and senior exams include a preliminary check sheet asking students to rate each document for *accuracy, bias, and relevance*. We felt that this task, although not identical to the one asked of students during the FYS final exam due to the differing lengths of time allotted to the two assessments (90 minutes for baseline and senior assessments as compared to 120 minutes for FYS final exams) would provide greater equivalence between these baseline/senior assessments and FYS final exams. **This practice was not implemented last summer, primarily because of the short turn-around time between the end of the summer 2016 assessment and preparation for August 2016 freshman baseline assessments. Later, we were advised that there had been problems in FYS using check sheets; that information was more accurate when students explained their rationales for each answer.**

##### *Recommendations regarding FYS Exams*

1. The Assessment Workgroup continued to be concerned about the length of some of the documents accompanying the FYS final exams and, perhaps more pointedly, the variation in the length of these documents among the exams given. These documents ranged in length from 75 pages for the *Concealed Weapons Scenario* to 16 for the *Influenza Scenario*. That said, the page count was not a perfect predictor of difficulty because the density of print per page varied from document to document. Further, statistical analysis of the mean differences in student performance among the eight scenarios used during 2015-2016 on the eight traits of the rubric revealed only one scenario on which students scored significantly lower than on the others; that was the *Social Media Scenario*, which had a moderate number of document

pages (20) for students to read. The Assessment Workgroup recommended that the FYS Director and faculty review 2015 recommendations regarding the issue of page length and take the scenario comparison results from the Assessment Workgroup into consideration when deploying final exams. **The length of the documents for each scenario continued to vary somewhat (from 93 pages for Campus Speech to 17 Pages for Soda Ban) among those used in 2016-2017. However, our analysis for the 2016-2017 results showed no significant correlations between scenario page length and student performance on any trait except *Recommendations* and that correlation, although negative, was weak.**

2. The Assessment Workgroup recommended that FYS exams be reconfigured to ask students to discuss additional information they might need to make a final recommendation before they make the recommendation. This would bring the exam format more into line with what students are asked to do at baseline. **This change in ordering was implemented during academic year 2016-2017.**
3. The Assessment Workgroup recommended that students in FYS be explicitly asked to use information they provided regarding *bias*, *relevance*, and *accuracy* in items 1 – 7 of the final exam when composing their final recommendation. The Workgroup further noted that students should be told that the main part of the exam is the final recommendation and that this should be carefully considered and composed. **We are unsure of the status of this request.**
4. Workgroup members reiterated that all scenarios should include a sample of the format in which the final recommendation should be written. **We are unsure of the status of this request. However, further discussion among workgroup members in summer 2017 led to the conclusion that providing an explicit example was perhaps not necessary, or even desirable, as this would result in students simply copying the format.**

### ***Recommendations regarding Baseline/FYS/Senior Rubric***

The Assessment Workgroup recommended re-examining the *Communication Style* trait of the rubric again next year before beginning assessments. – **We normed the rubrics again this year, but did not change the wording of *communication style*.**

### ***Procedures for 2017 Assessment***

#### ***General Procedures***

In August 2016, 1,500 incoming freshmen at Marshall University completed baseline assessments (an additional 106 students completed the *Collegiate Learning Assessment [CLA+]*). Both assessments required students to analyze and evaluate information, solve problems, and write effectively. These skills are aligned to three of Marshall University’s outcomes; *Information Literacy*, *Inquiry-Based Thinking*, and *Communication Fluency*. In the spring semester of 2017, 132 graduating seniors completed the same assessments (35 the Marshall assessment and 97 the *CLA+*). The 132 seniors who completed either the *CLA+* or Marshall’s senior assessment did not differ significantly from the senior population in terms of entering academic ability based on ACT or SAT performance. However, the sample had a significantly higher mean college GPA (3.3) than the senior population (3.1) and the sample included a higher proportion of female students than did the population. Freshmen completing

Marshall's mandatory First Year Seminar in Critical Thinking (FYS) completed assessments that were similar to those finished by incoming freshmen and graduating seniors.

In May 2017 a group of nine faculty representing several academic colleges from across the university evaluated a sample of Marshall's assessment artifacts using a rubric that allowed them to score each artifact across eight criteria (traits). These included *information needed* and *source acknowledgment* (Information Literacy), *evidence*, *viewpoints*, and *recommendation/position* (Inquiry-Based Thinking), and *development*, *convention/format*, and *communication style* (Communication Fluency). This project was coordinated by the Office of Assessment.

A random sample of 225 Marshall Freshman baseline assessments was drawn from the pool of 1,500 (15%) of the total number of assessments available. However, we note that one of these assessments had results for only one trait of the rubric (Information Literacy: *information needed*), reducing the scorable number of baseline assessments for the other seven traits to 224. Since only 35 seniors completed the Marshall senior exiting assessment, we included all in our analysis, giving us a total of 260 assessment artifacts in our sample.

One hundred seventy-two (172) of the 225 freshmen from our baseline sample (76%) completed FYS assessments. The reasons we had no FYS assessments from 53 of the students in the baseline sample were as follows: 12 were enrolled in, and received credit for FYS, but did not complete the final exam (the instructor for seven of these students did not administer the FYS final exam to any students in the class); 8 were enrolled in, but did not receive credit for FYS; 18 were not enrolled in FYS during academic year 2016-2017; 2 completed FYS during summer 2016, so their scores could not be used as a "post baseline" measure; and 13 students withdrew from Marshall without completing FYS.

All assessments were de-identified and, for the freshman baseline/senior comparisons, raters did not know which were completed by freshmen and which by seniors. Each assessment had two independent raters. Please see the supporting information that follows this summary for a detailed explanation of scoring procedures.

## ***Results and Analysis***

### ***Comparison of Freshman Baseline to Senior Exiting Results and to Results at the End of FYS***

Mean scores (on a scale of 1 – 4) for seniors were significantly higher than freshman baseline measures on all criteria (traits). However, mean performance for seniors ranged from a low of 2.26 (*Inquiry-Based Thinking: recommendations and Communication Fluency: convention/format*) to a high of 2.76 (*Inquiry-Based Thinking: information needed*), indicating, as has been the case for the past four years, that there is room for improvement among Marshall's graduating seniors. Mean differences between freshman baseline performance and senior exiting performance ranged from a low of 0.33 for *Inquiry-Based Thinking: recommendations* to a high of 0.68 for *Inquiry-Based Thinking: viewpoints*. We note that, for the past five years, the difference between the mean scores of freshmen and seniors has averaged about one-half of a point (ranging from

0.27 to 0.96). Mean scores for seniors have never exceeded 3.04 (*Inquiry-Based Thinking: recommendations*) in 2013, with the average being about 2.6.

In 2015 the workgroup discussed the two-pronged approach that Marshall uses to compare student performance in *Information Literacy*, *Inquiry-Based Thinking* (aka *Critical Thinking*), and *Communication Fluency* between freshman baseline and senior exiting assessments, namely that some students take the nationally standardized *Collegiate Learning Assessment (CLA+)*, while the rest take a similar assessment developed by Marshall University faculty. This process works well for freshmen and, although having representative senior samples that are large enough to draw meaningful conclusions remains problematic, the cooperation of Marshall's senior capstone instructors who ask their students to participate has helped in this regard. We also note that for the past several years the *CLA+* and Marshall Assessment results have mirrored each other. Mean senior performance on the *CLA+* for spring 2015 and 2016 was 1112 ( $n = 99$ ) and 1100 ( $n = 108$ ) respectively. Both of these mean scores placed Marshall's mean level of senior performance at the *proficient* level. However, Marshall's senior mean performance for spring 2017 was 1091 ( $n = 97$ ), placing Marshall's mean level of senior performance at the *basic* level. Although the differences among the three mean scores for these years were not statistically significant, we are concerned about the gradual decline in our means and we are concerned that our seniors' overall mean for 2017 fell into the *basic* level of performance. We note that categorical levels of performance are *below basic*, *basic*, *proficient*, *accomplished*, and *advanced*. As with our university created assessments, these results strongly suggest a continued need to work to help our students improve their ability to analyze issues and problems, evaluate evidence that might help them to arrive at solutions or to make recommendations concerning issues, while being aware of their own assumptions and considering the potential consequence of proposed solutions and/or recommendations.

As noted above, there were 172 freshmen who completed (or partially completed) both a baseline assessment and an FYS final exam. However, the baseline partial completer completed only the *Information Literacy: information needed* section, whereas the partial completer from FYS completed all sections except the *Information Literacy: information needed* section. This resulted in paired sample comparisons for 171 matched pairs. For these students, *paired-samples t-tests* using adjusted alpha levels to control for Type I error (.025 for information literacy), (.017 for *Inquiry-Based Thinking*), and (.017 for *Communication Fluency*) showed significant mean differences between freshman baseline and FYS results for *Information Literacy: acknowledgment of sources*, for *Inquiry-Based Thinking: recommendations*, and for *Communication Fluency: development and convention/format*. We note that, for the past five years, the difference between the mean scores of FYS and baseline performance has averaged about three-tenths (.32) of a point (ranging from 0.01 to 1.29). Mean scores for FYS final exams have never exceeded 3.18 (*Communication Fluency: cohesion* – a trait that has since been revised) in 2013, with the average being about 2.45. This year's results showed that, for most traits, there were no significant differences in student performance between any pairs of scenarios. Exceptions to this overall pattern were significantly lower performance on *Campus Speech* than on *Music*, *Social Media* and *Soda Ban* and significantly lower performance on *Genetically Modified Foods (GMO)* than on *Music* and *Soda Ban* on the outcome *Inquiry-Based Thinking: evidence*. Performance was also significantly lower on *Campus Speech* than on *Music* and *Social Media* and significantly lower on *Open Carry* than on *Music* for the outcome *Inquiry-Based Thinking: viewpoints*.

## ***Recommendations from the 2017 Assessment Workgroup***

### ***Recommendations regarding baseline and senior assessments***

1. The Assessment Workgroup recommended that baseline and senior assessments include the rubric so that students have a better idea of how we are assessing their work.
2. The Assessment Workgroup also conducted a pilot in which they scored a very small sample of capstone project artifacts using the AAC&U's *Critical Thinking* and *Written Communication* Value rubrics. The group found these rubrics easy to use and their scoring resulted in very few scores of "not applicable" (N/A). Given this result and the difficulty we have experienced over the years in drawing truly representative samples of seniors to complete either the *CLA+* or Marshall's Senior Assessment, we recommend that staff from the Assessment Office encourage degree programs to use the Blackboard Assignment Module to align their senior capstone assignments with the AAC&U's *Critical Thinking* and *Written Communication* Value rubrics. These discussions can be incorporated into larger discussions regarding the process of creating assignments in Blackboard and aligning them to appropriate outcomes of Marshall's Baccalaureate Degree Profile (BDP), which we discuss in greater detail in the Blackboard Outcomes Assessment Report. This has the potential to allow us to evaluate a truly random sample of artifacts from multiple degree programs and apply validated rubrics to assess work that students complete as part of their degree programs.

### ***Recommendations regarding Baseline/FYS/Senior Rubric***

Based on interrater reliability results, the Assessment Workgroup recommends re-examining the *Communication Style* trait of the rubric again next year before beginning the 2018 assessment process.