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 <u>before</u> 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 <u>explicitly</u> asked to <u>use</u> 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 <u>and</u> 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 <u>except</u> 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 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.

Supporting Documentation

Comparison of Freshman Baseline and Senior Exiting Assessment Results

Academic Year 2016 – 2017

Review Procedures

- A total of 260 assessments (225 freshman and 35 senior) were used for this evaluation. Freshman assessments represented approximately 15% of the 1,500 completed (or partially completed) during the University's Week of Welcome in August 2016. Only 35 seniors completed the Marshall Developed Senior Assessment in spring 2017 (an additional 97 seniors completed the *Collegiate Learning Assessment [CLA+]*), so all 35 Marshall senior completers were included in this sample. 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 had a higher percentage of female students than that of the population.

  - Assessments were de-identified and raters did not know which were completed by freshmen and which by seniors.
  - Each assessment was scored across eight criteria.

- Each assessment had two independent raters and scores were determined in the following manner:
  - If raters assigned the same score, that became the score for the artifact.
  - If raters' scores differed by one point or less, e.g. Rater 1 assigned a score of 1 and Rater 2 a score of 2, the final score was the mean, i.e. 1.5.
  - If raters' scores differed by more than one point, e.g. Rater 1 assigned a score of 1 and Rater 2 a score of 3, the raters met to discuss the rationale for their scores to see if they could agree on a score or, at minimum, scores that differed by no more than one point.
  - If raters' scores differed by more than one point and, after discussion, they were not able to resolve the differences, a third rater was assigned to review the assessment.

Rules for Arriving at Final Scores when there were Three Raters: These rules were followed for all assessments conducted.

1. If the third rater's score agreed with one of the first two, the score with the two agreements was used.

2. If the first two raters' scores were two points apart, e.g. 1 and 3 and the third rater's score was in the middle, e.g. 2, the third rater's score was used.

3. If the first two raters' scores were two points apart, e.g. 1 and 3, and the third rater's score was between them, but a decimal, e.g. 1.5 or 2.5, the third rater's score was used.

4. If the first two raters' scores were two points apart, e.g. 1 and 3, and the third rater's score was a "4", the two scores closer together were averaged, e.g. 3.5.

5. IF the first two raters' scores were three points apart, e.g. 1 and 4, the third rater's score was averaged with the closest other rater; e.g. if the third rater's score was 3, the final score was 3.5; if the third rater's score was 2, the final score was 1.5.

Rubric Used for Scoring

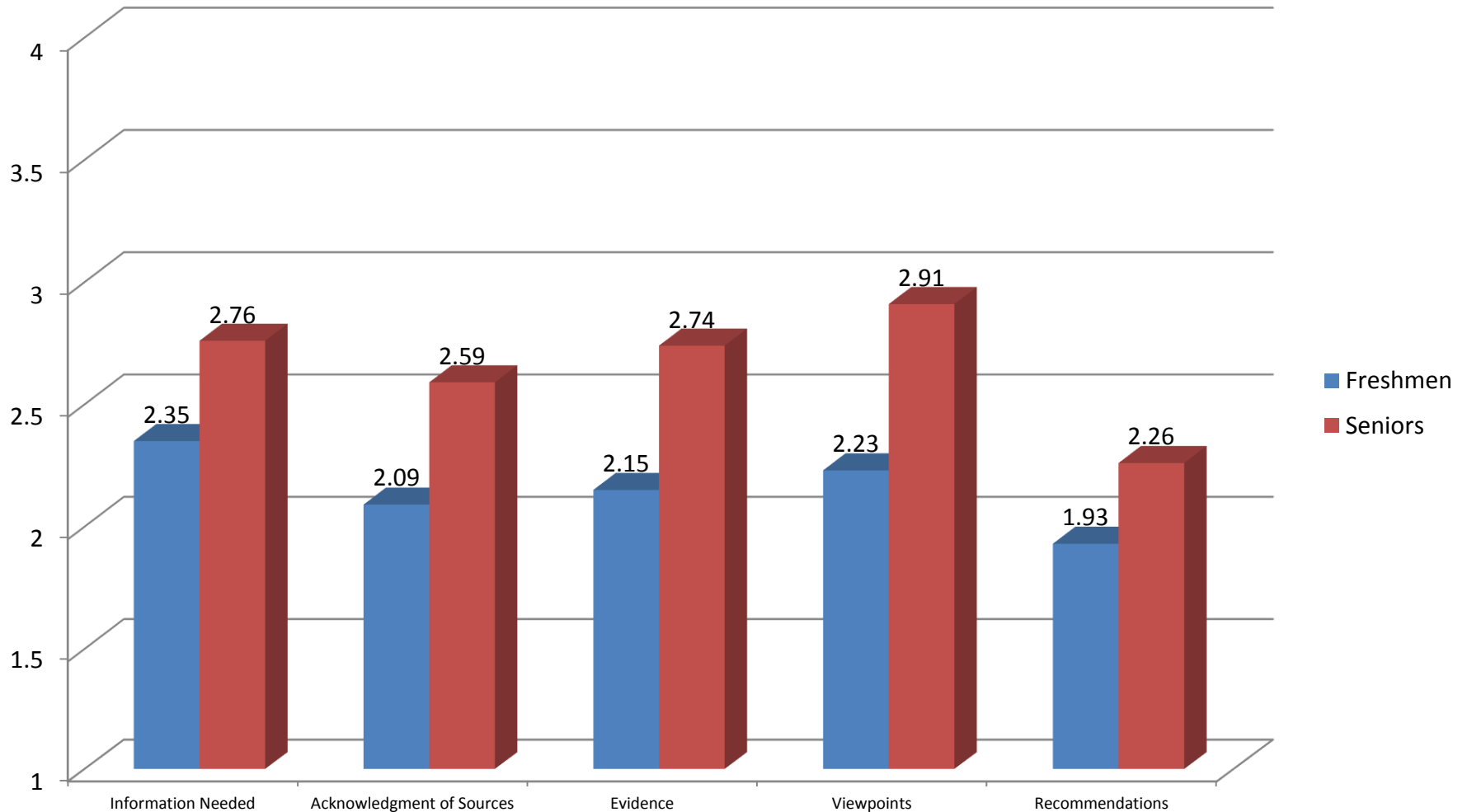Baseline/Senior Assessment Rubric – Summer 2016 – updated 5-18-2015

Outcomes	Traits	Performance Levels			
		1	2	3	4
Information Literacy	Information Needed	Does not acknowledge or assess the need for more information.	Acknowledges the need for more information but does not identify research methods/sources (or those identified are not feasible) that would address unanswered questions.	Assesses the need for more information and recommends general research methods/sources (that are feasible) that would address some unanswered questions.	Assesses the need for more information and recommends specific research methods/sources (that are feasible) that would address most unanswered questions.
	Source Acknowledgment	Fails to acknowledge sources from the DL.	Indirectly/vaguely acknowledges sources of information from the DL.	Clearly acknowledges relevant sources of information from the DL.	Integrates relevant information from the DL. Acknowledges sources used.
Inquiry-Based Thinking	Evidence	Disregards or misunderstands evidence from the DL.	Insufficient evidence is taken from sources in the DL or evidence is used without appropriate interpretation/evaluation (i.e. poor job).	Evidence is taken from relevant and valid sources in the DL with some interpretation/evaluation, but not enough to develop a coherent analysis or synthesis (i.e. adequate job).	Evidence is taken from relevant and valid sources in the DL with enough interpretation/evaluation to develop a coherent analysis or synthesis (i.e. good/excellent job).
	Viewpoints	Ignores viewpoints expressed in the DL.	Viewpoints expressed in the DL are taken as mostly fact, with little or no question.	Questions some viewpoints expressed in the DL.	Thoroughly questions and evaluates viewpoints expressed in the DL.
	Recommendation/Position	Either does not make a recommendation or makes a recommendation, but does not justify it in any way.	Recommendation is justified, but does not acknowledge different sides of the issue.	Recommendation is justified and takes into account different sides/complexities of the issue.	Recommendation takes into account the complexities of the issue. Any limits to the recommendation are acknowledged.
Communication Fluency	Development	Shows little or no evidence of developing his/her ideas.	Shows some development of ideas.	Shows a strong, but perhaps somewhat incomplete, development of ideas.	Produces a document in which the ideas have been fully developed.
	Convention/Format	Demonstrates minimal attention to basic organization and presentation and stylistic conventions.	Demonstrates some awareness of basic organization, content, and presentation and stylistic conventions.	Demonstrates consistent use of important conventions particular to a specific writing task, including organization, content, presentation, and stylistic choices.	Demonstrates detailed attention to and successful execution of a wide range of conventions particular to a specific writing task including organization, content, presentation, formatting, and stylistic choices.
	Communication Style	Uses language that impedes meaning because of errors in usage/mechanics.	Uses language that generally conveys meaning to readers with clarity, although writing may include some errors.	Uses straightforward language that generally conveys meaning to readers. The language in the document has few errors.	Uses graceful language that skillfully communicates meaning to readers with clarity and fluency, and is virtually error-free.

Freshman Baseline/Senior Exiting Comparisons
Mean Scores on a scale of 1 – 4, with 4 being the highest possible score
Freshman *n* = 224 (225 for Information Needed); Senior *n* = 35
All mean differences statistically significant

Legend:
- ■ Freshmen
- ■ Seniors

Category	Freshmen	Seniors
Information Needed	2.35	2.76
Acknowledgment of Sources	2.09	2.59
Evidence	2.15	2.74
Viewpoints	2.23	2.91
Recommendations	1.93	2.26

Freshman Baseline/Senior Exiting Comparisons

Freshman *n* = 225 for Information Needed; 224 for all other traits; Senior *n* = 35

Trait/ Performance Level	Info Needed	Acknowledgment of Sources	Evidence	Viewpoints	Recommendations
1 – 1.75 Freshmen	33 (15%)	89 (40%)	56 (25%)	46 (21%)	87 (39%)
1 – 1.75 Seniors	0 (0%)	7 (20%)	2 (6%)	2 (6%)	10 (29%)
2 – 2.75 Freshmen	124 (55%)	77 (34%)	126 (56%)	123 (55%)	104 (46%)
2 – 2.75 Seniors	20 (57%)	12 (34%)	13 (37%)	9 (26%)	12 (34%)
3 – 3.75 Freshmen	65 (29%)	54 (24%)	38 (17%)	53 (24%)	32 (14%)
3 – 3.75 Seniors	12 (34%)	12 (34%)	19 (54%)	22 (63%)	13 (37%)
4 Freshmen	3 (1%)	4 (2%)	4 (2%)	2 (1%)	1 (0%)
4 Seniors	3 (9%)	4 (11%)	1 (3%)	2 (6%)	0 (0%)
Grand Total Freshmen	225 (100%)	224 (100%)	224 (100%)	224 (100%)	224 (100%)
Grand Total Seniors	35 (100%)	35 (100%)	35 (100%)	35 (100%)	35 (100%)

Freshman Baseline/Senior Exiting Comparisons
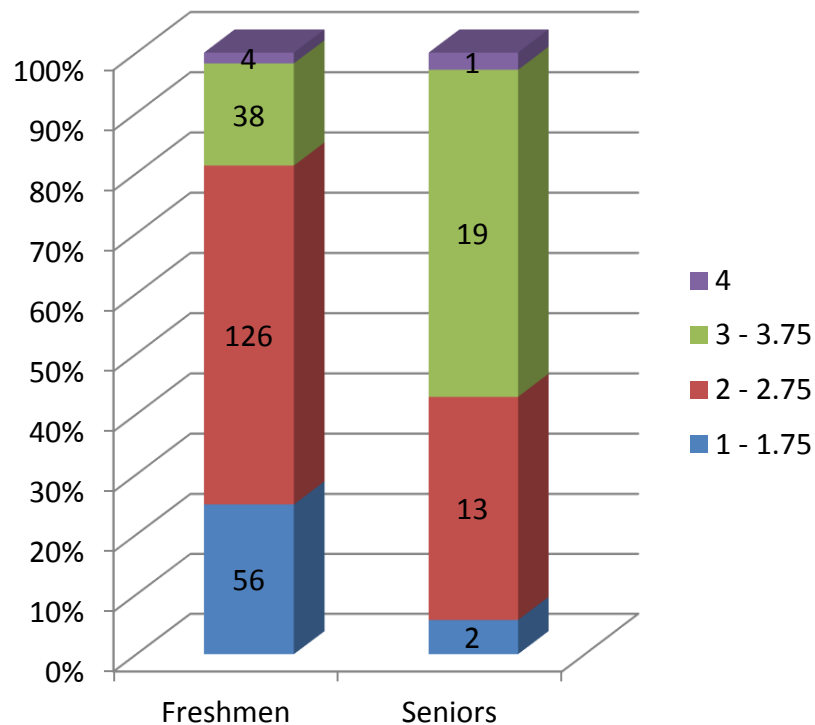## Freshman *n* = 225(IN); 224 (AS); Senior *n* = 35

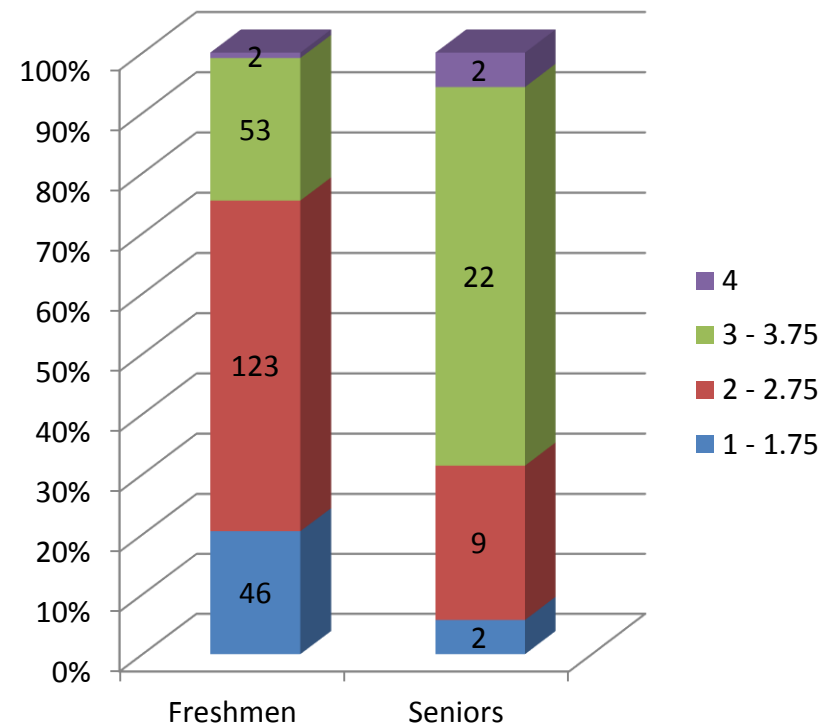**Information Needed**
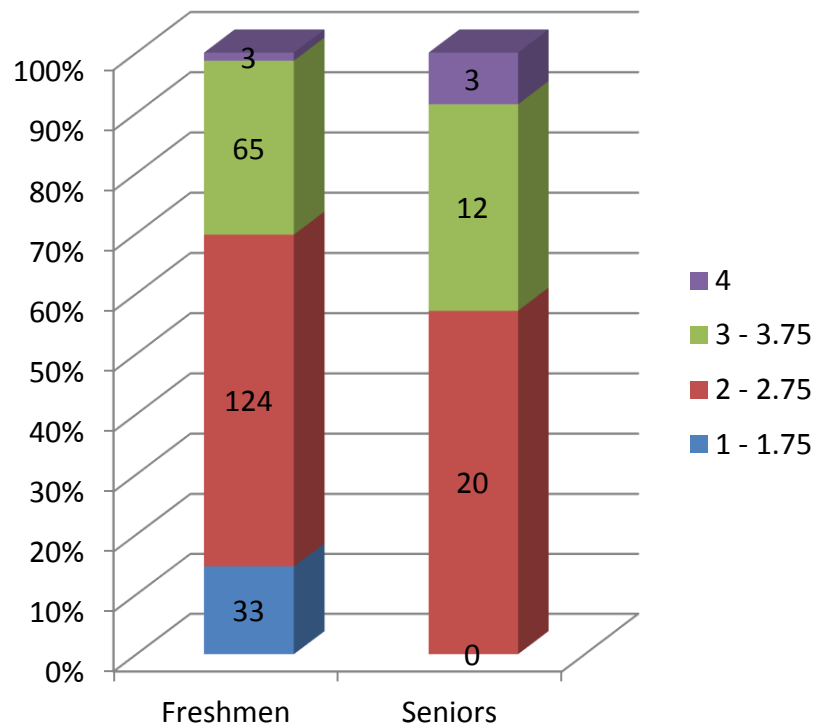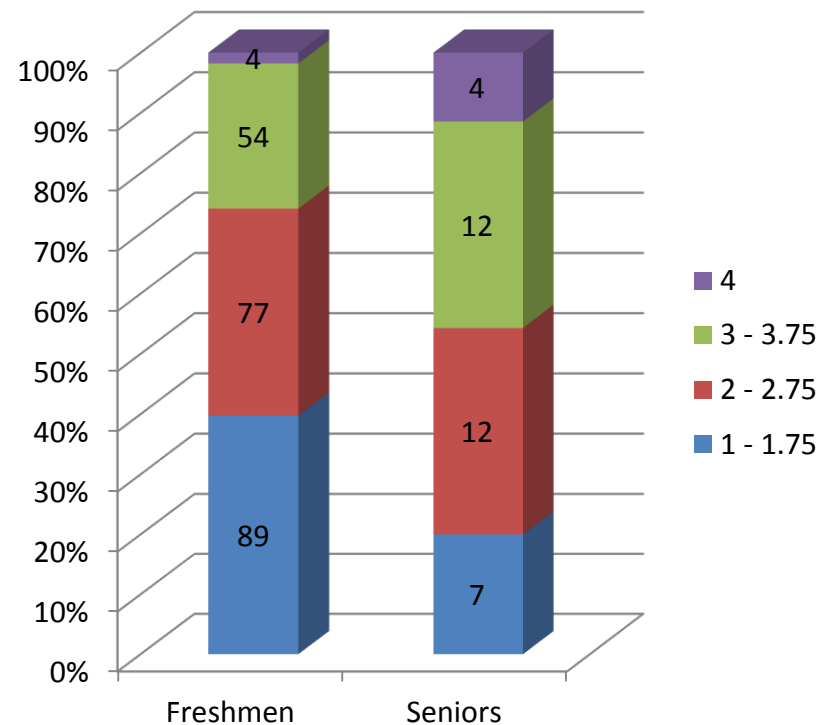
Acknowledgment of Sources

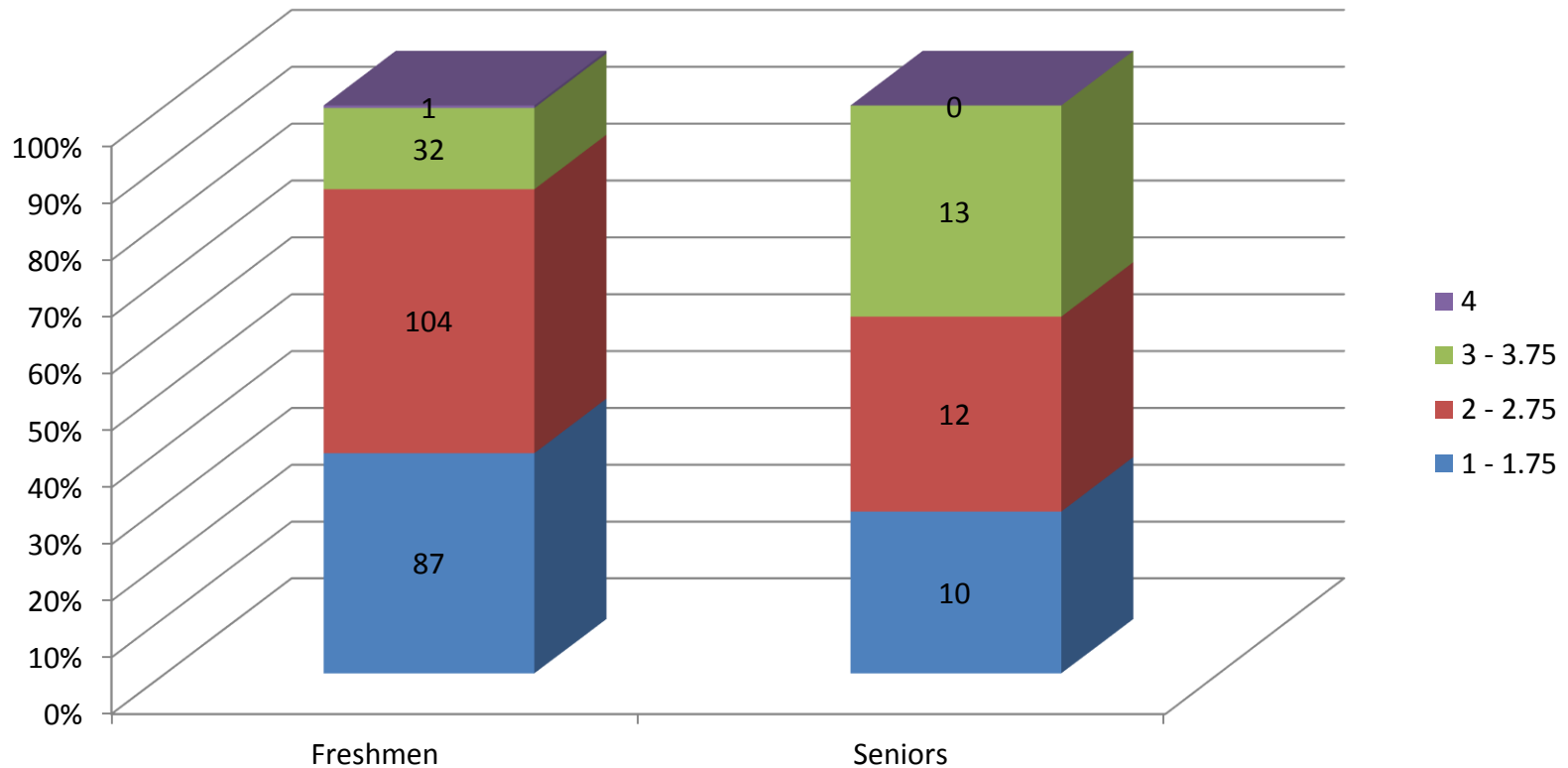# Freshman Baseline/Senior Exiting Comparisons
Freshman *n* = 224; Senior *n* = 35

Freshman Baseline/Senior Exiting Comparisons
Freshman *n* = 224; Senior *n* = 35

Recommendations

Freshman Baseline/Senior Exiting Assessment Inter-Rater Agreement Results

| Trait/ Agreement	Info Needed (Conservative Kappa = .378; Liberal Kappa = .932)	Acknowledgment of Sources (Conservative Kappa = .355; Liberal Kappa = .958)	Evidence (Conservative Kappa = .373; Liberal Kappa = .971)	Viewpoints (Conservative Kappa = .386; Liberal Kappa = .947)	Recommendations (Conservative Kappa = .375; Liberal Kappa = .876)
Agree	155 (60%)	141 (54%)	155 (60%)	155 (60%)	149 (58%)
Difference = 1 point or less	91 (35%)	109 (42%)	98 (38%)	93 (36%)	84 (32%)
Difference = 1.5 to 2 points	14 (5%)	9 (3%)	6 (2%)	11 (4%)	25 (10%)
Difference = 2.5 to 3 points	0	0	0	0	1 (0%)
Total	260 (100%)	259 (100%)	259 (100%)	259 (100%)	259 (100%)

Freshman Baseline/Senior Exiting Comparisons
Mean Scores on a scale of 1 – 4, with 4 being the highest possible score
Freshman *n* = 224 (225 for Information Needed); Senior *n* = 35
All mean differences statistically significant
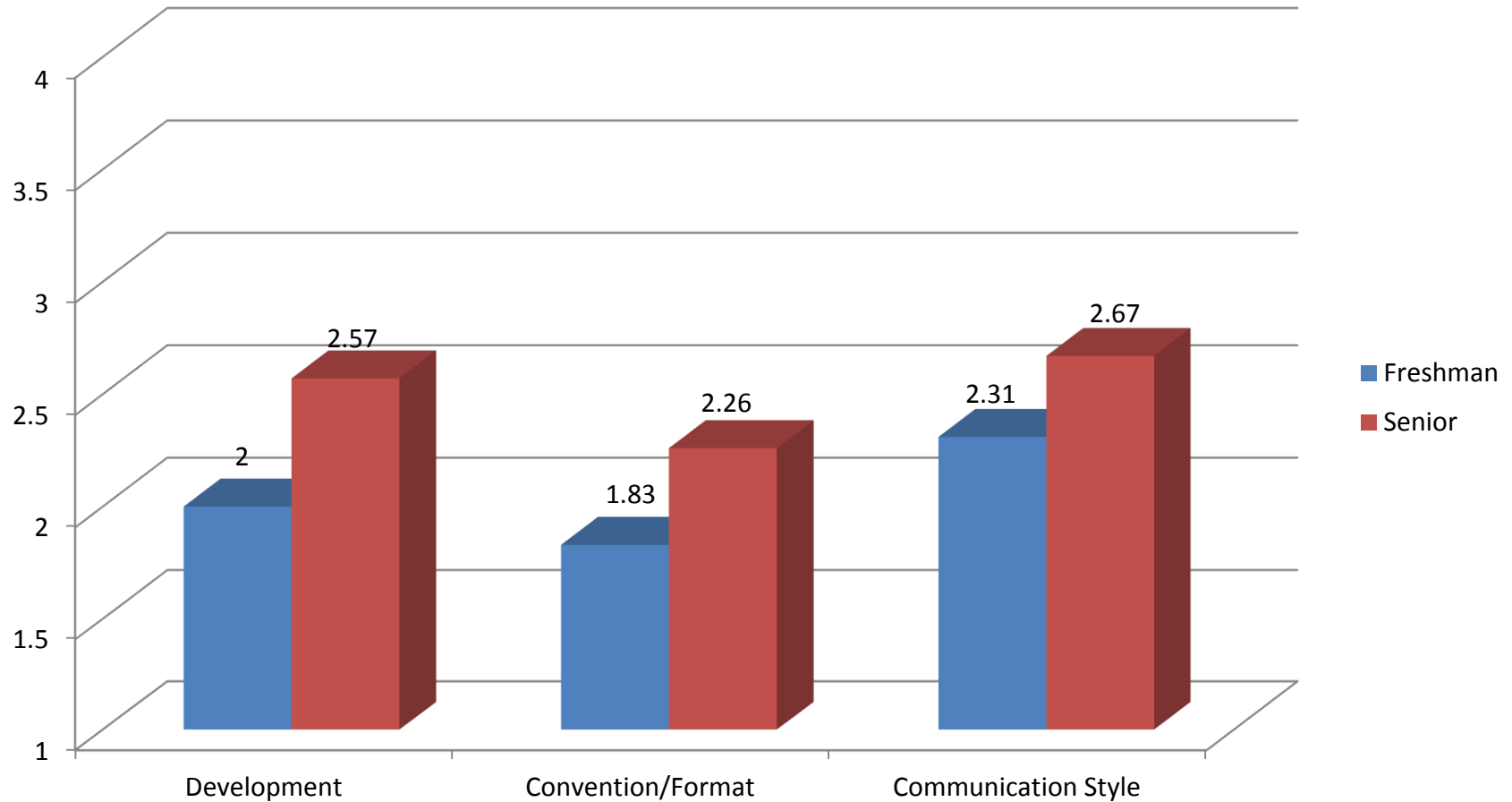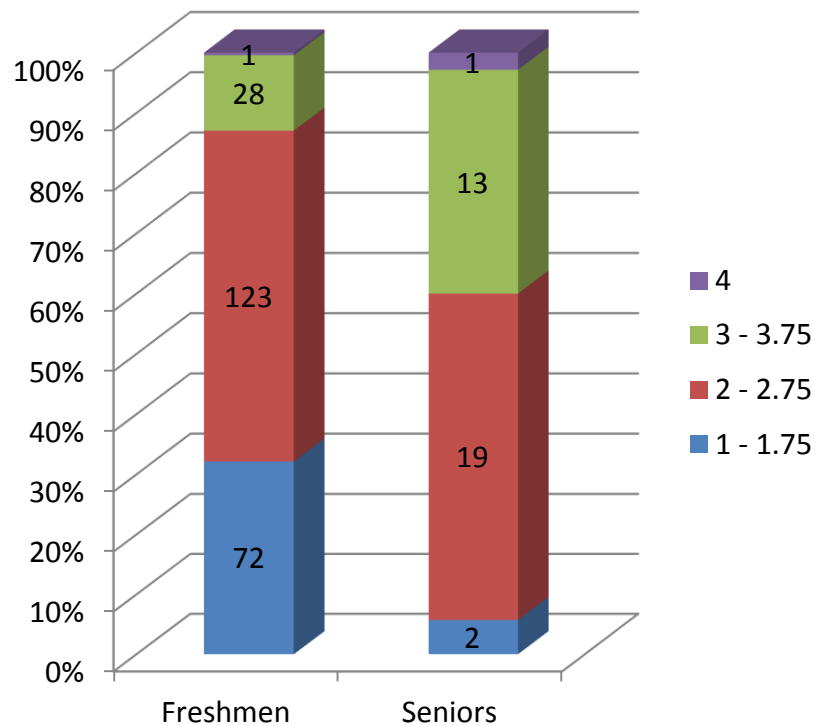
Freshman Baseline/Senior Exiting Comparisons
Freshman n = 224; Senior n = 35

Trait/ Performance Level	Development	Convention/Format	Communication Style
1 – 1.75 Freshmen	72 (32%)	109 (49%)	33 (15%)
1 – 1.75 Seniors	2 (6%)	10 (29%)	3 (9%)
2 – 2.75 Freshmen	123 (55%)	97 (43%)	146 (65%)
2 – 2.75 Seniors	19 (54%)	17 (49%)	16 (46%)
3 – 3.75 Freshmen	28 (13%)	17 (8%)	44 (20%)
3 – 3.75 Seniors	13 (37%)	5 (14%)	14 (40%)
4 Freshmen	1 (0%)	1 (0%)	1 (0%)
4 Seniors	1 (3%)	3 (9%)	2 (6%)
Grand Total Freshmen	224 (100%)	224 (100%)	224 (100%)
Grand Total Seniors	35 (100%)	35 (100%)	35 (100%)

Freshman Baseline/Senior Exiting Comparisons
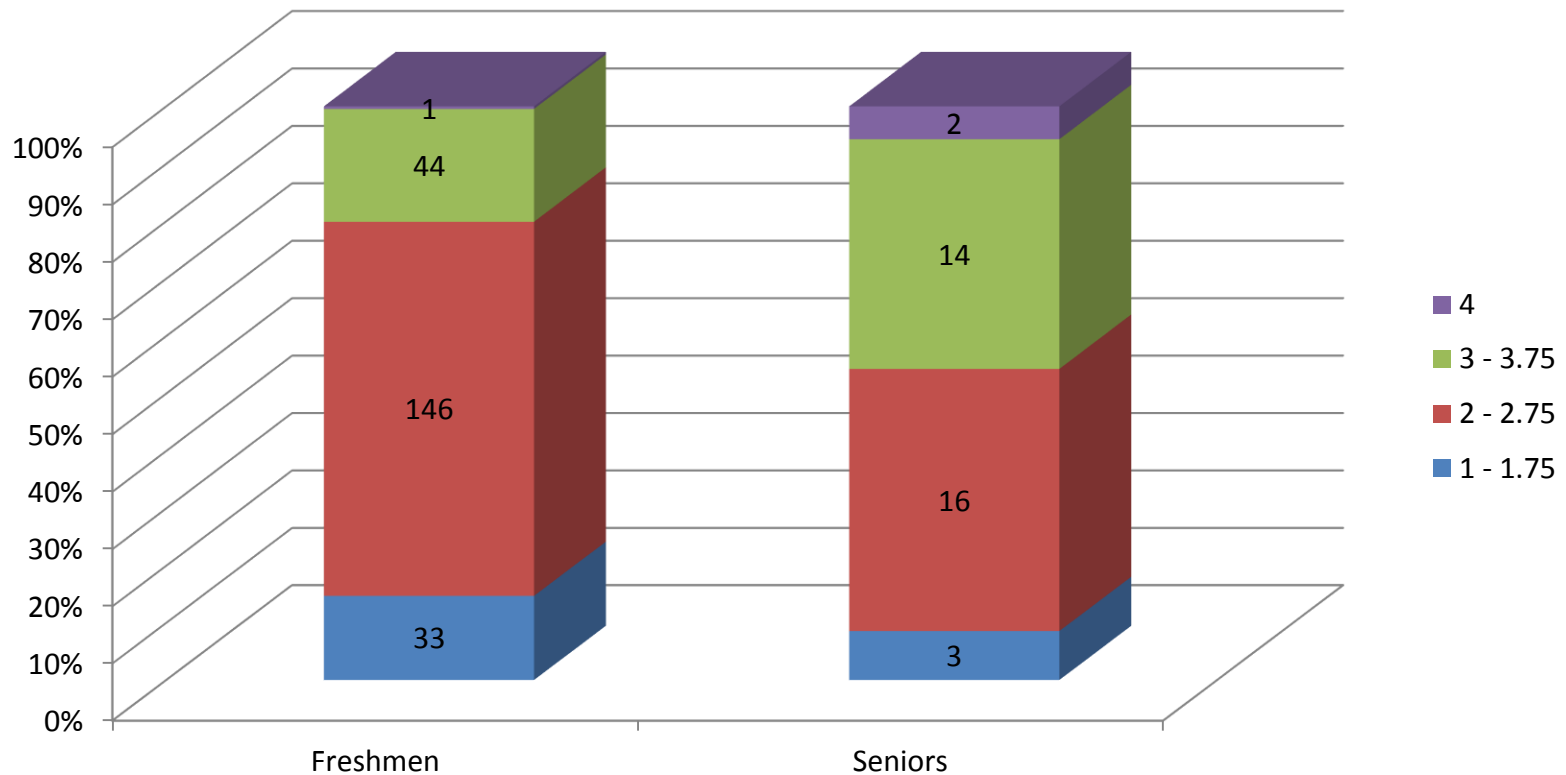Freshman *n* = 224; Senior *n* = 35
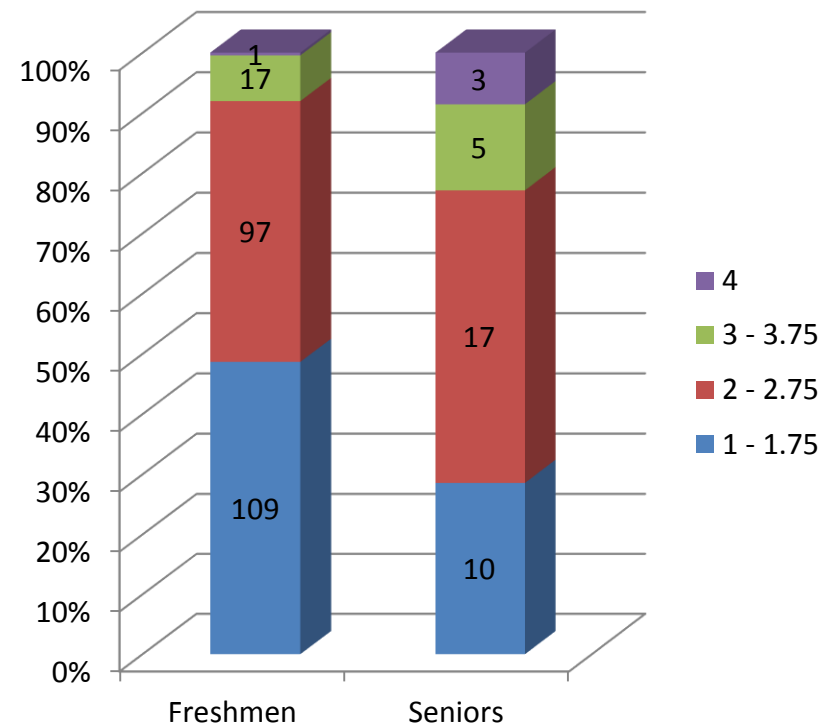
Freshman Baseline/Senior Exiting Comparisons
Freshman n = 224; Senior n = 35

Communication Style

Freshman Baseline/Senior Exiting Assessment Inter-Rater Agreement Results

Trait/ Agreement	Development (Conservative Kappa = .347; Liberal Kappa = .961)	Convention/Format (Conservative Kappa = .225; Liberal Kappa = .951)	Communication Style (Conservative Kappa = .122; Liberal Kappa = .969)
Agree	153 (59%)	133 (51%)	123 (47%)
Difference = 1 point or less	98 (38%)	116 (45%)	130 (50%)
Difference = 1.5 to 2 points	8 (3%)	10 (4%)	6 (2%)
Difference = 2.5 to 3 points	0	0	0
Total	259 (100%)	259 (100%)	259 (100%)

Comparison of Freshman Baseline and First-Year Seminar (FYS) Assessments

Academic Year 2016 - 2017

Review Procedures

- One hundred seventy-two (172; 76%) of the 225 freshmen who had completed (or partially completed) baseline assessments during Week of Welcome completed (or partially completed) similar assessments at the end of First Year Seminar (FYS). However, the baseline (WOW) partial completer completed only the "Information Needed" section, whereas the partial completer from FYS completed all sections <u>except</u> the "Information Needed" section. This resulted in paired sample comparisons for 171 matched pairs. FYS assessments were evaluated across the same eight criteria (traits) used to score freshman baseline assessments. Scoring methodology also was the same.

Freshman Baseline/FYS Comparisons

Mean Scores on a scale of 1 – 4, with 4 being the highest possible score

n = 171

Mean differences are statistically significant for *Acknowledgment of Sources* and for *Recommendations*
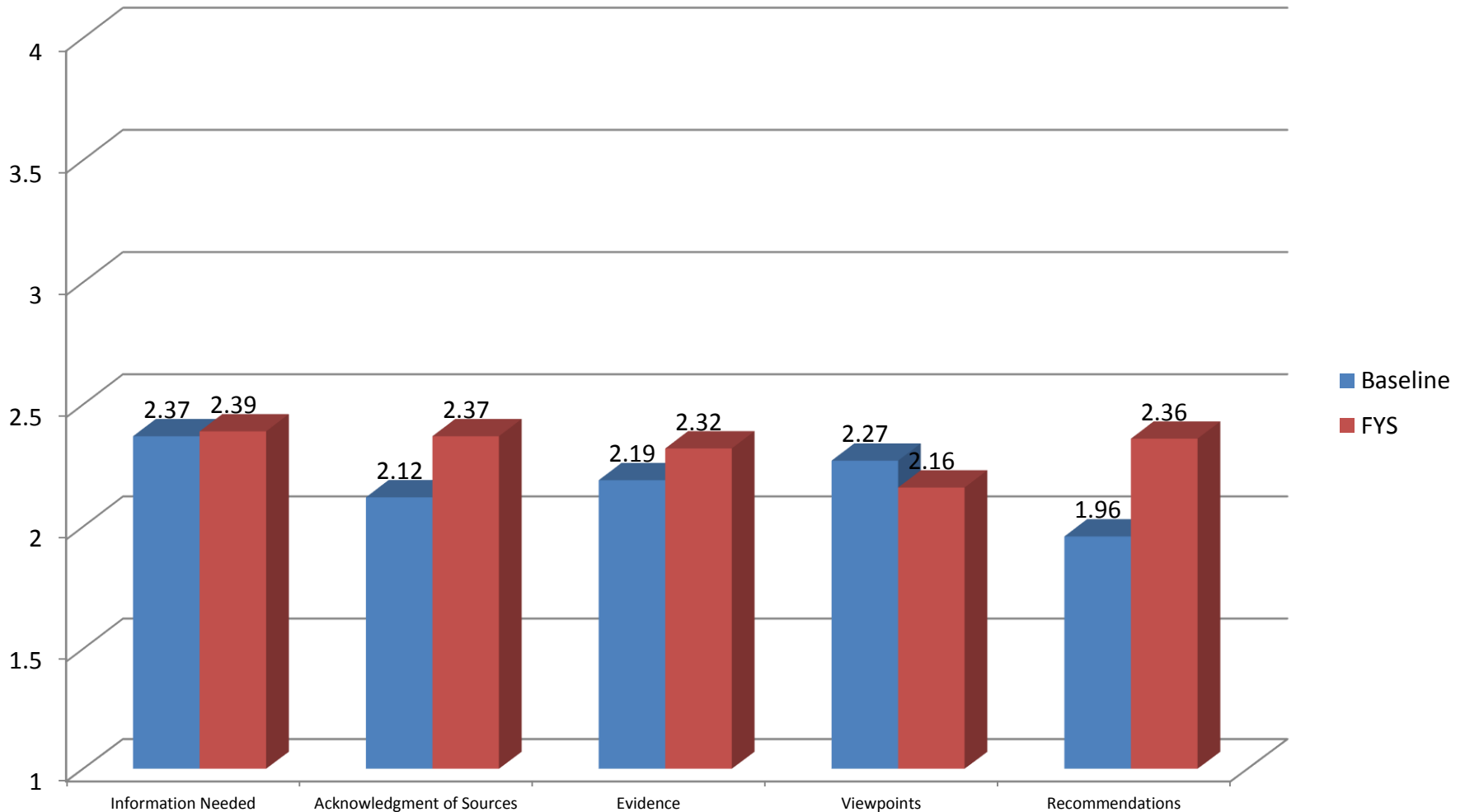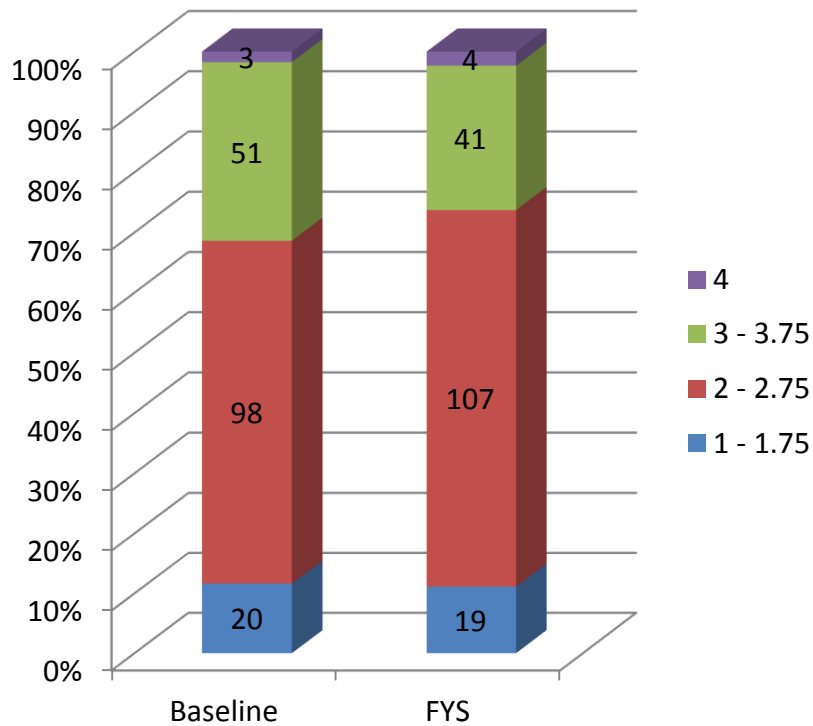
Freshman Baseline/FYS Comparisons

n = 171 or 172 (see below)

| Trait/ Performance Level	Info Needed	Acknowledgment of Sources	Evidence	Viewpoints	Recommendations
1 – 1.75 Baseline	20 (12%)	65 (38%)	37 (22%)	34 (20%)	63 (37%)
1 – 1.75 FYS	19 (11%)	52 (30%)	41 (24%)	41 (24%)	26 (15%)
2 – 2.75 Baseline	98 (57%)	60 (35%)	99 (58%)	93 (54%)	80 (47%)
2 – 2.75 FYS	107 (63%)	40 (23%)	84 (49%)	103 (60%)	100 (58%)
3 – 3.75 Baseline	51 (30%)	43 (25%)	31 (18%)	42 (25%)	27 (16%)
3 – 3.75 FYS	41 (24%)	77 (45%)	42 (24%)	25 (15%)	43 (25%)
4 Baseline	3 (2%)	3 (2%)	4 (2%)	2 (1%)	1 (1%)
4 FYS	4 (2%)	3 (2%)	5 (3%)	3 (2%)	3 (2%)
Grand Total Baseline	172 (100%)	171 (100%)	171 (100%)	171 (100%)	171 (100%)
Grand Total FYS	171 (100%)	172 (100%)	172 (100%)	172 (100%)	172 (100%)

Freshman Baseline/FYS Comparisons
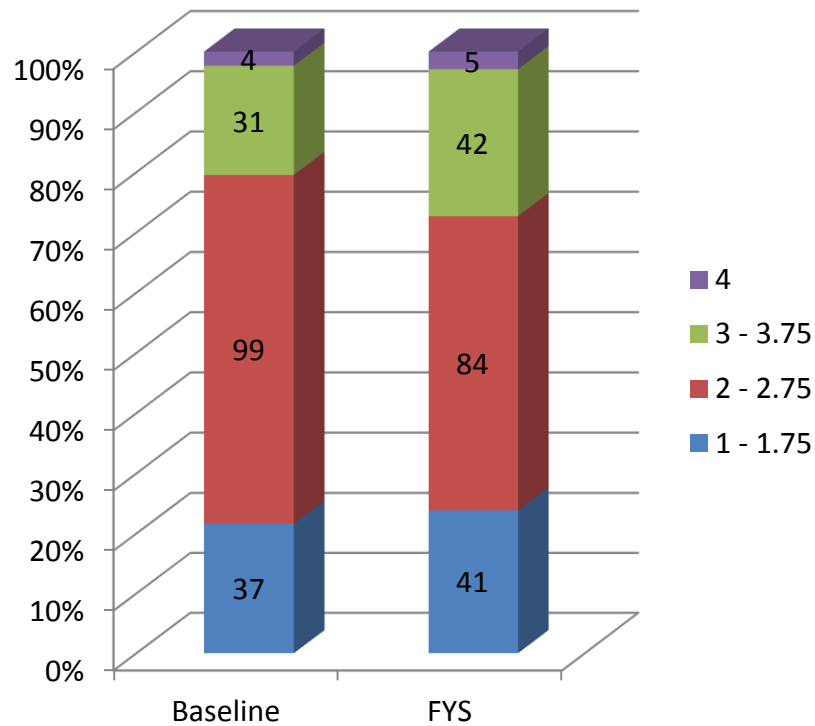n = 171/172
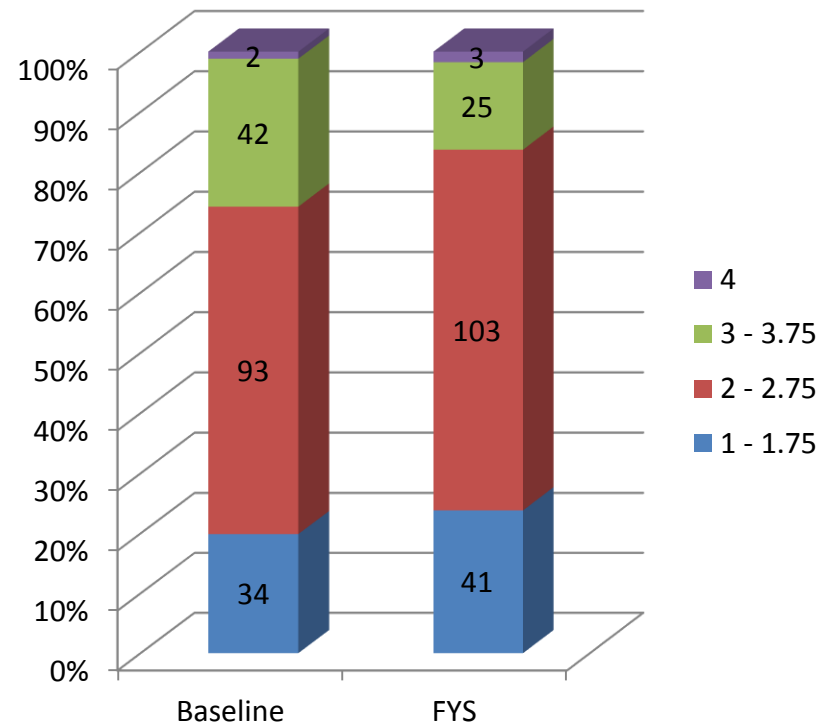
Information Needed
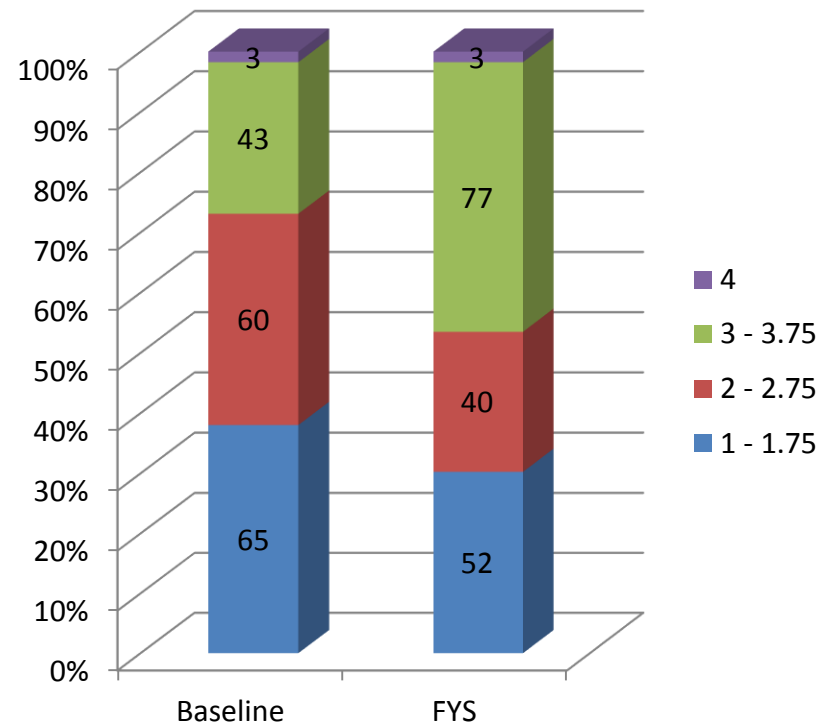
Acknowledgment of Sources

Freshman Baseline/FYS Comparisons
n = 171/172
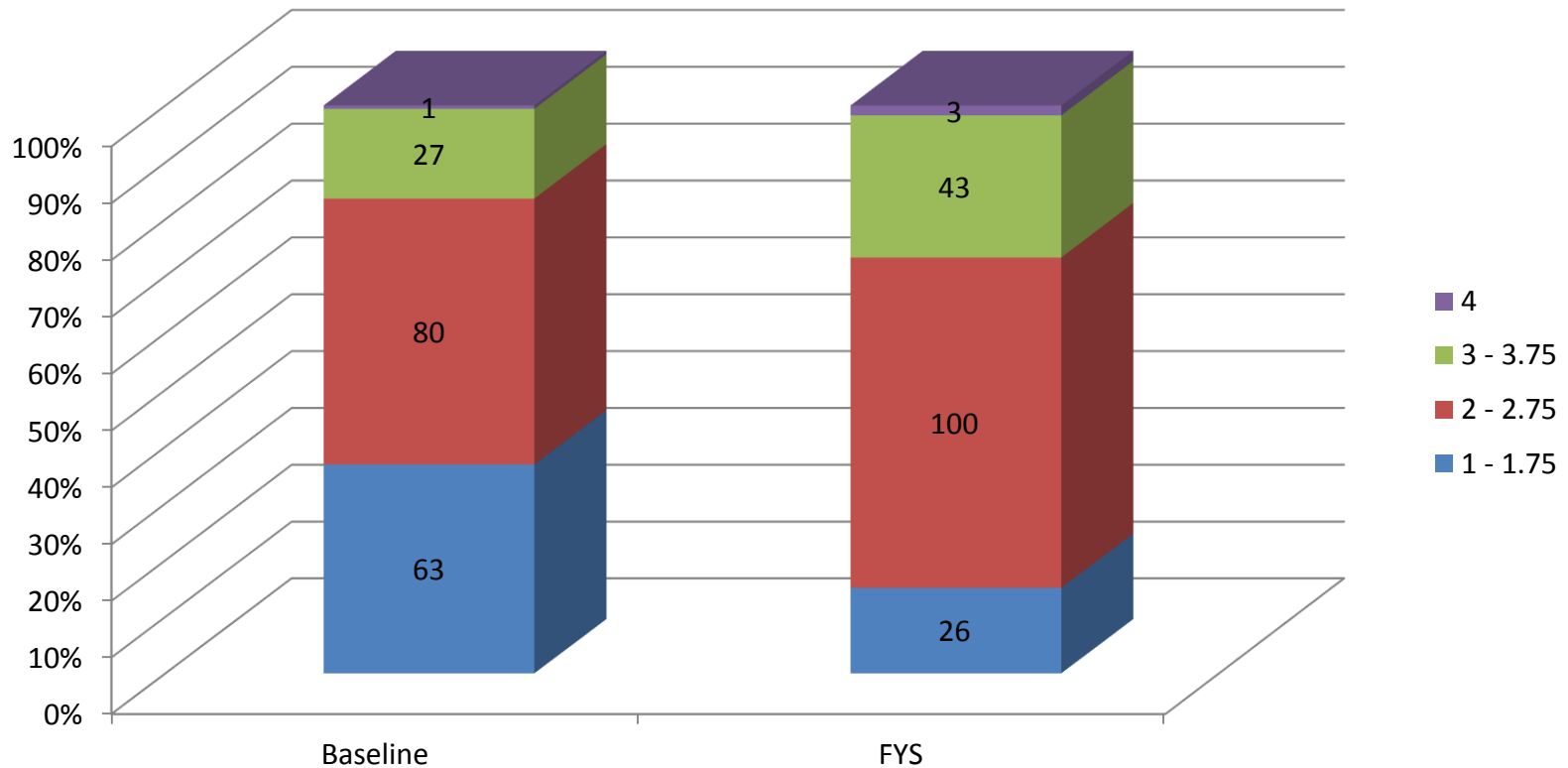
Freshman Baseline/FYS Comparisons
n = 171/172

Recommendations

FYS Inter-Rater Agreement Results

Trait/ Agreement	Info Needed (Conservative Kappa = .256; Liberal Kappa = .909)	Acknowledgment of Sources (Conservative Kappa = .379; Liberal Kappa = .943)	Evidence (Conservative Kappa = .199; Liberal Kappa = .942)	Viewpoints (Conservative Kappa = .205; Liberal Kappa = .910)	Recommendations (Conservative Kappa = .228; Liberal Kappa = .933)
Agree	90 (53%)	98 (57%)	80 (47%)	86 (50%)	89 (52%)
Difference = 1 point or less	69 (40%)	66 (38%)	84 (49%)	74 (43%)	74 (43%)
Difference = 1.5 to 2 points	12 (7%)	8 (5%)	8 (5%)	11 (6%)	8 (5%)
Difference = 2.5 to 3 points	0	0	0	1 (1%)	1 (1%)
Total	171 (100%)	172 (100%)	172 (100%)	172 (100%)	172 (100%)

Freshman Baseline/FYS Comparisons

Mean Scores on a scale of 1 – 4, with 4 being the highest possible score

n = 171

Mean differences are statistically significant for *Development* and for *Convention/Format*
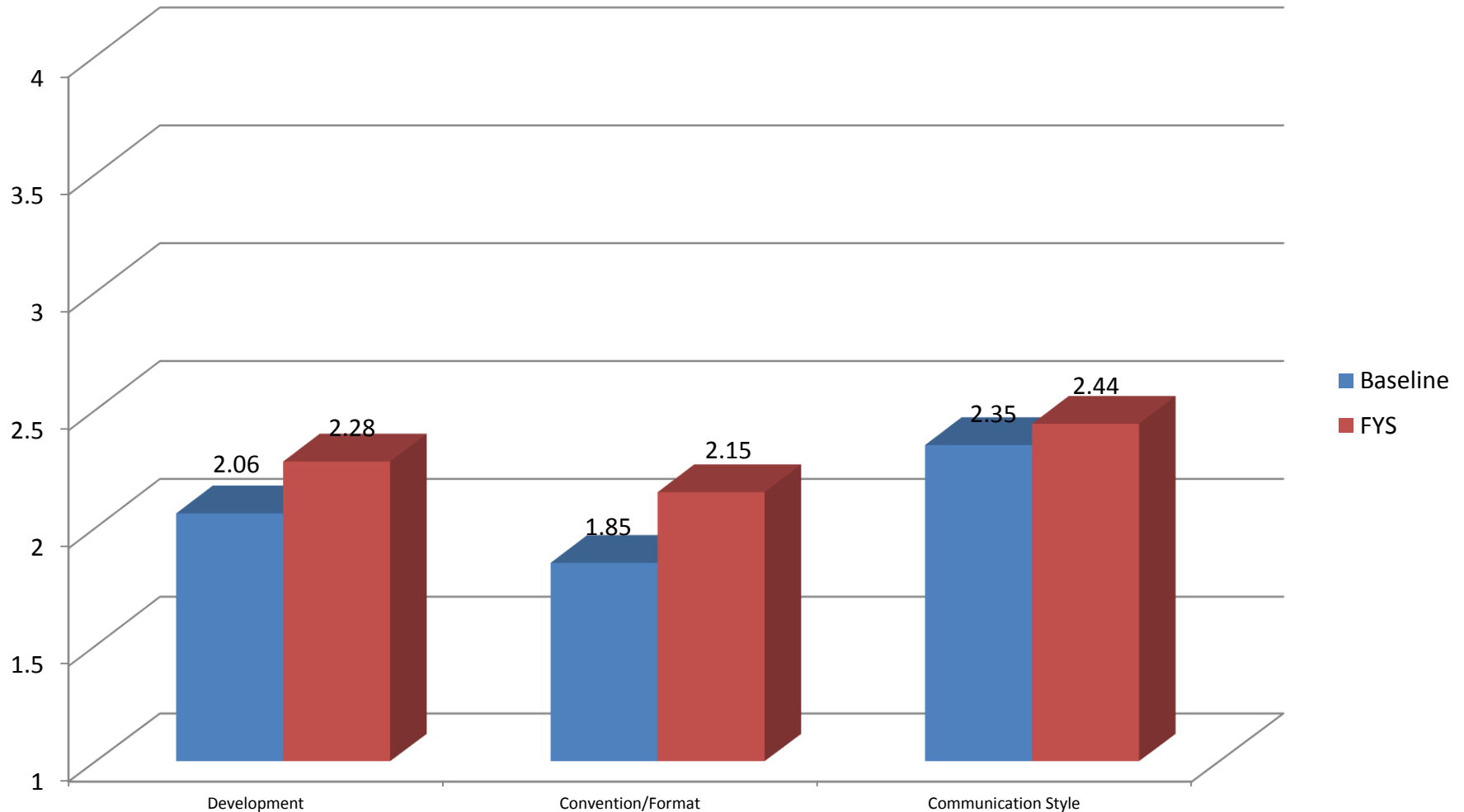
Freshman Baseline/FYS Comparisons
n = 171/172

Trait/ Performance Level	Development	Convention/Format	Communication Style
1 – 1.75 Baseline	47 (27%)	83 (49%)	21 (12%)
1 – 1.75 FYS	31 (18%)	40 (23%)	18 (10%)
2 – 2.75 Baseline	98 (57%)	72 (42%)	114 (67%)
2 – 2.75 FYS	100 (58%)	103 (60%)	102 (59%)
3 – 3.75 Baseline	25 (15%)	15 (9%)	35 (20%)
3 – 3.75 FYS	39 (23%)	28 (16%)	51 (30%)
4 Baseline	1 (1%)	1 (1%)	1 (1%)
4 FYS	2 (1%)	1 (1%)	1 (1%)
Grand Total Baseline	171 (100%)	171 (100%)	171 (100%)
Grand Total FYS	172 (100%)	172 (100%)	172 (100%)

Freshman Baseline/FYS Comparisons
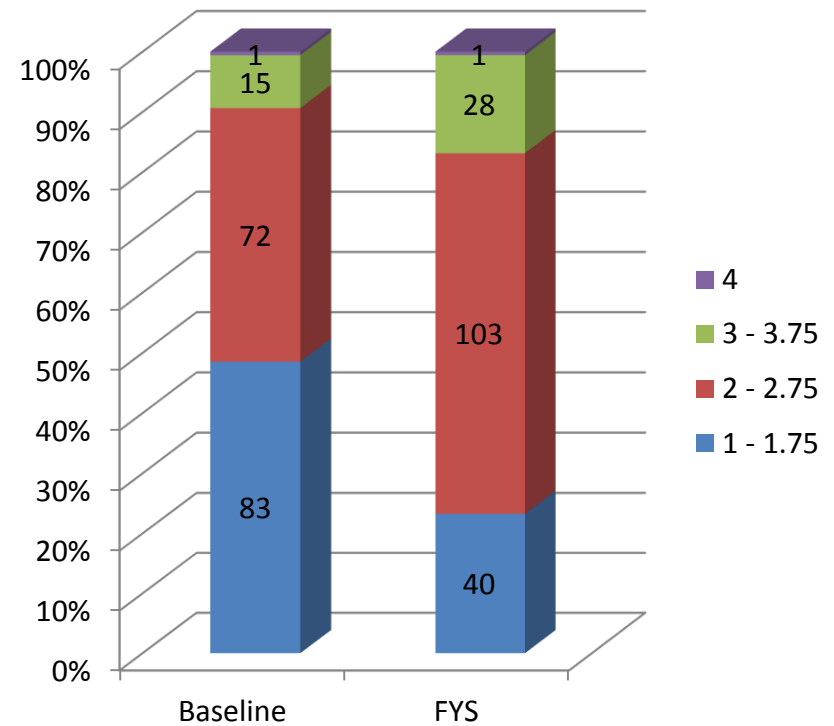n = 171/172

Development
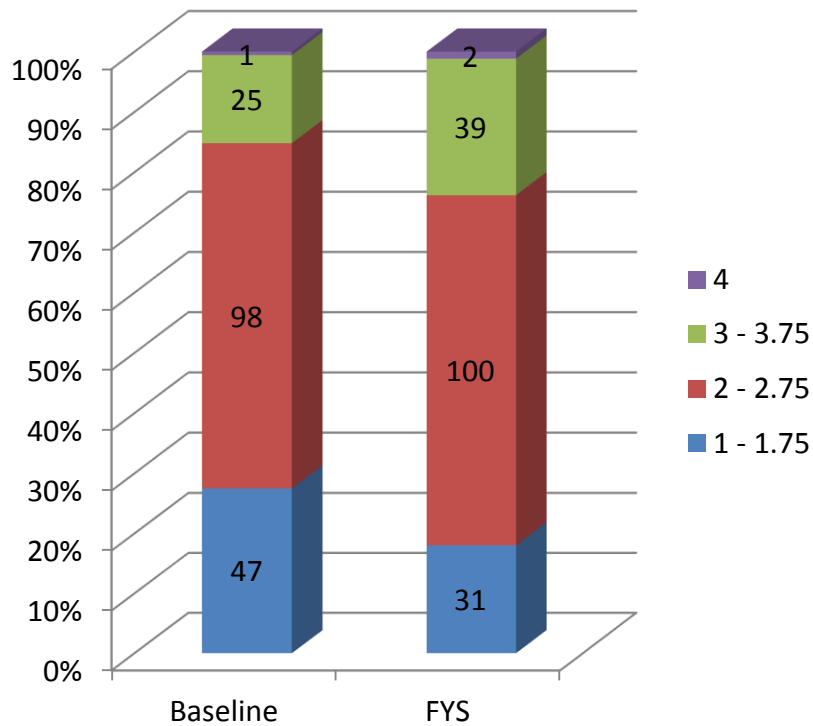
Convention/Format

Freshman Baseline/FYS Comparisons
n = 171/172

Communication Style
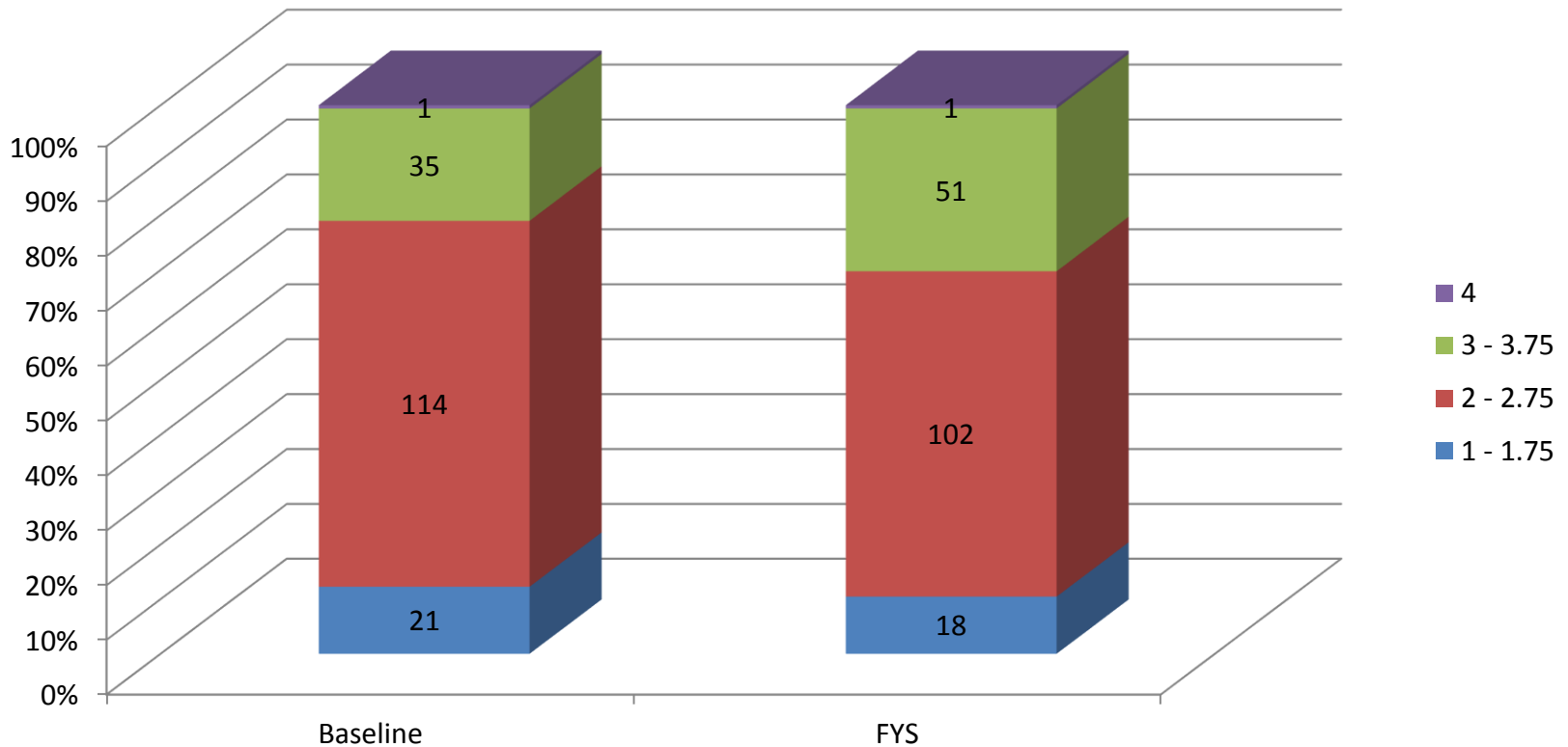
FYS Inter-Rater Agreement Results

Trait/ Agreement	Development (Conservative Kappa = .304; Liberal Kappa = .963)	Convention/Format (Conservative Kappa = .266; Liberal Kappa = .940)	Communication Style (Conservative Kappa = .204; Liberal Kappa = .932)
Agree	95 (55%)	94 (55%)	87 (51%)
Difference = 1 point or less	72 (42%)	70 (41%)	76 (44%)
Difference = 1.5 to 2 points	5 (3%)	8 (5%)	9 (5%)
Difference = 2.5 to 3 points	0	0	0
Total	172 (100%)	172 (100%)	172 (100%)

Comparison of FYS Results for Each Trait by Scenario

Academic Year 2016 - 2017

FYS Comparisons by Scenario for IL: Information Needed
Mean Scores on a scale of 1 – 4, with 4 being the highest possible score

A One-Way ANOVA did not reveal any statistically significant differences in means across the scenarios.

Legend:
- Campus Speech; n = 21
- Fracking; n = 14
- GMO; n = 36
- Music; n = 17
- Open Carry; n = 17
- Religious Freedom; n = 21
- Social Media; n = 27
- Soda Ban; n = 18

Values: 2.33, 2.29, 2.6, 2.49, 2.35, 2.12, 2.46, 2.22
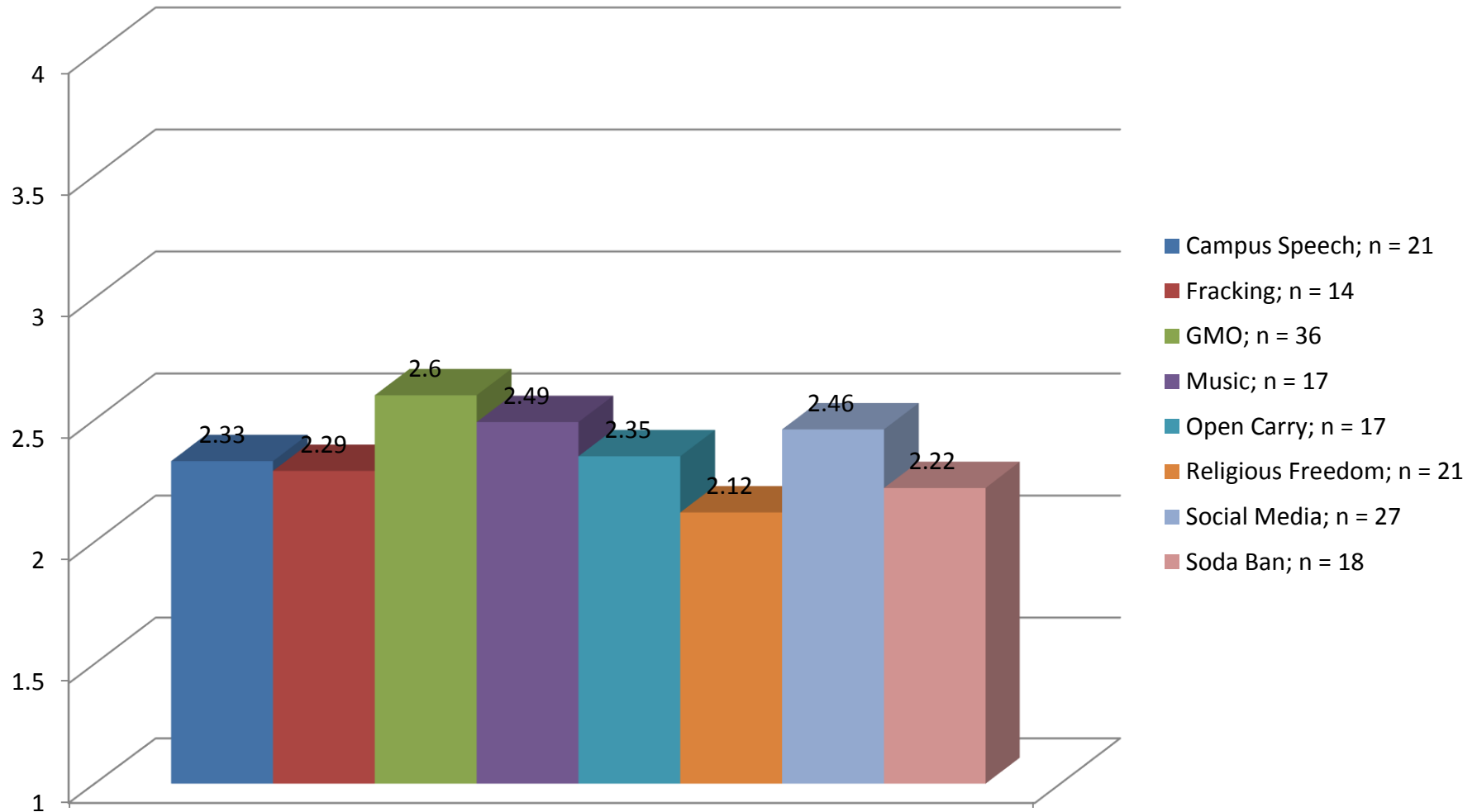
FYS Comparisons by Scenario for IL: Source Acknowledgment
Mean Scores on a scale of 1 – 4, with 4 being the highest possible score

A One-Way ANOVA revealed statistical significance; however, a Tukey Post-Hoc Analysis revealed no significant differences among individual pairs of scenarios.

Legend:
- Campus Speech; n = 22
- Fracking; n = 14
- GMO; n = 36
- Music; n = 17
- Open Carry; n = 17
- Religious Freedom; n = 21
- Social Media; n = 27
- Soda Ban; n = 18

Values: 2, 2.57, 2.08, 2.56, 2.24, 2.48, 2.54, 2.83
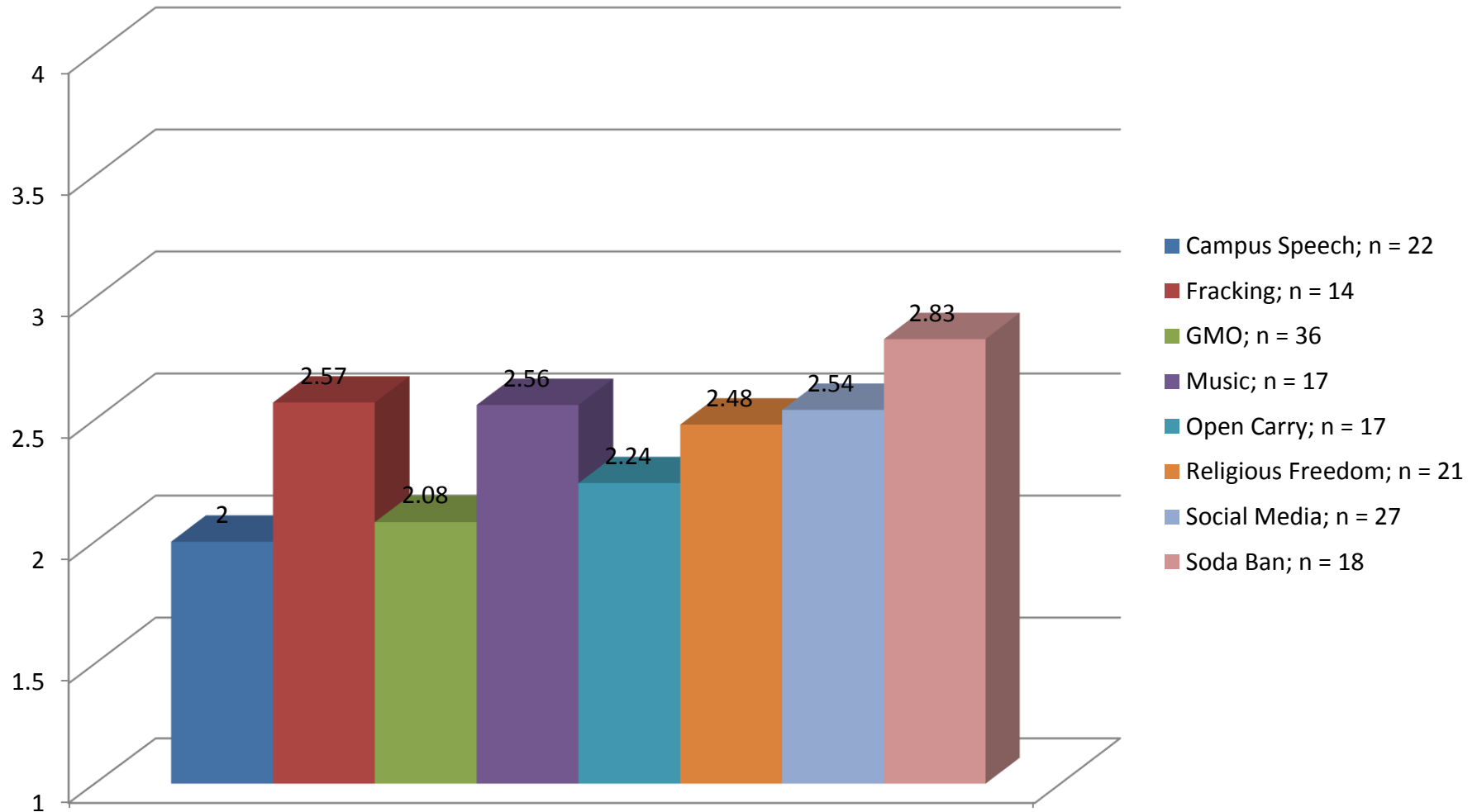
FYS Comparisons by Scenario for IBT: Evidence
Mean Scores on a scale of 1 – 4, with 4 being the highest possible score

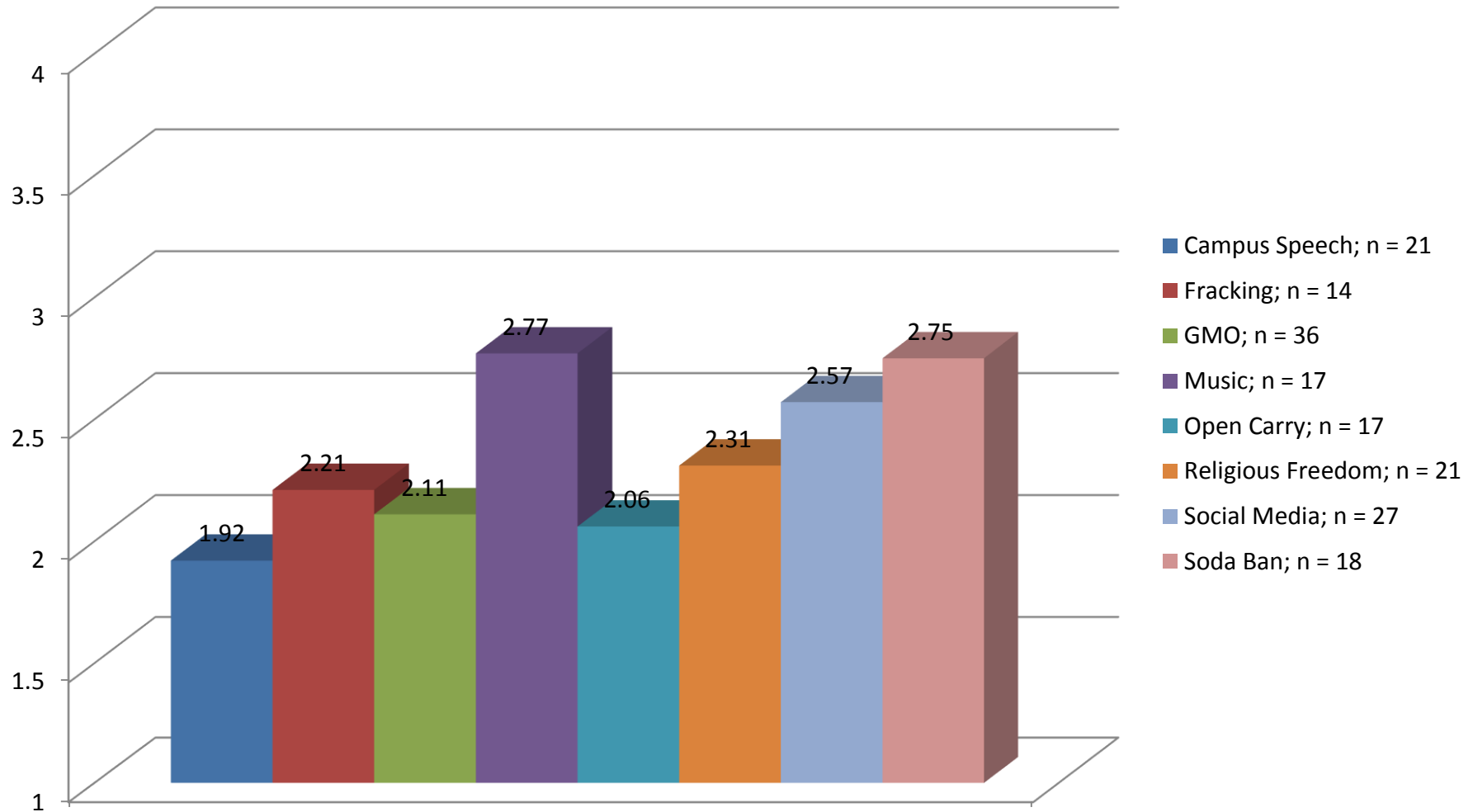A One-Way ANOVA revealed statistical significance; a Tukey Post-Hoc Analysis revealed that student performance on Campus Speech was significantly lower than their performance on Music, Social Media, and Soda Ban; performance on GMO was significantly lower than performance on Music and Soda Ban.

Legend:
- Campus Speech; n = 21
- Fracking; n = 14
- GMO; n = 36
- Music; n = 17
- Open Carry; n = 17
- Religious Freedom; n = 21
- Social Media; n = 27
- Soda Ban; n = 18

Values: 1.92, 2.21, 2.11, 2.77, 2.06, 2.31, 2.57, 2.75

FYS Comparisons by Scenario for IBT: Viewpoints
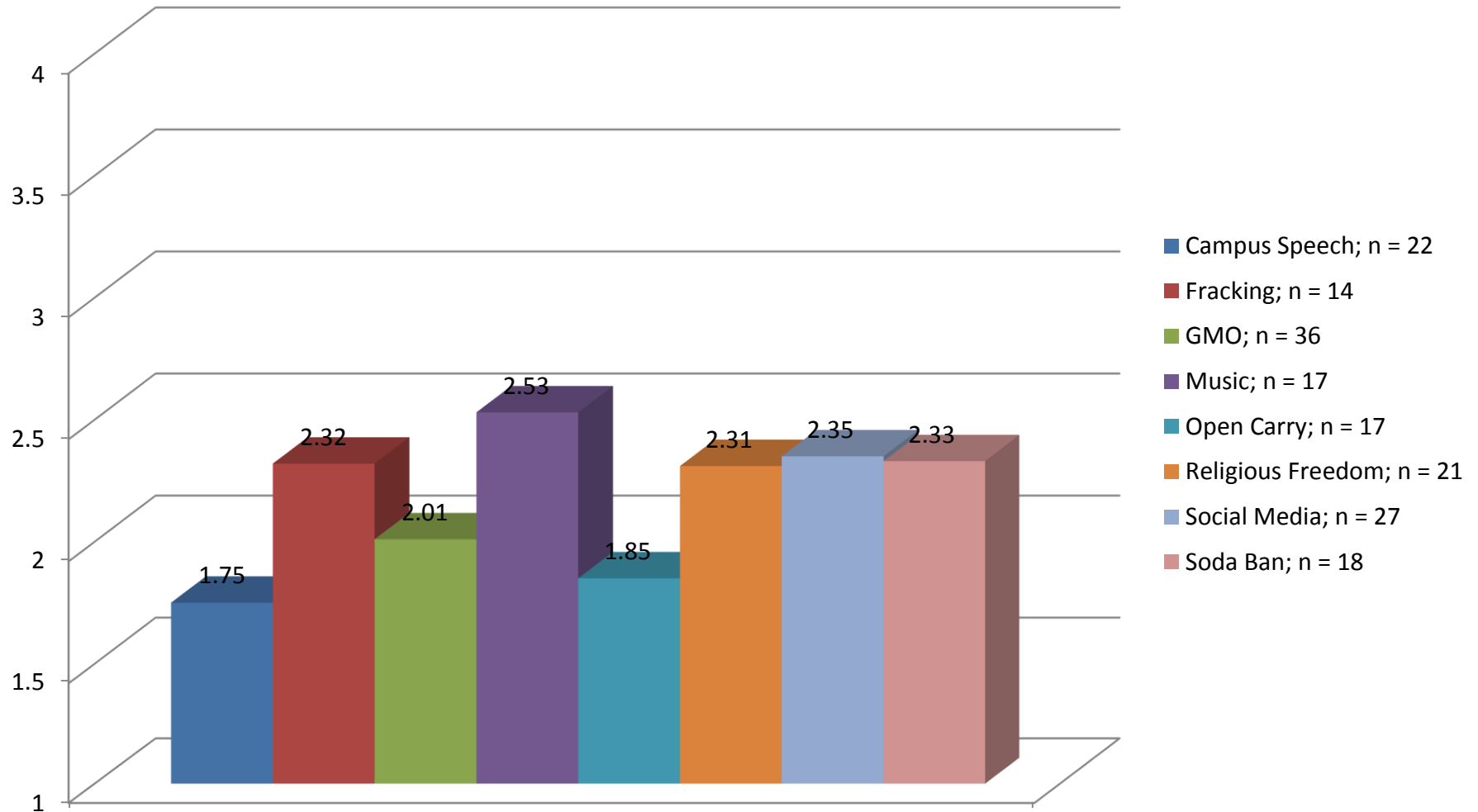## Mean Scores on a scale of 1 – 4, with 4 being the highest possible score

A One-Way ANOVA revealed statistical significance; a Tukey Post-Hoc Analysis revealed that student performance on Campus Speech was significantly lower than their performance on Music and Social Media; performance on Open Carry was significantly lower than performance on Music.

Legend:
- Campus Speech; n = 22
- Fracking; n = 14
- GMO; n = 36
- Music; n = 17
- Open Carry; n = 17
- Religious Freedom; n = 21
- Social Media; n = 27
- Soda Ban; n = 18

Values:
- Campus Speech: 1.75
- Fracking: 2.32
- GMO: 2.01
- Music: 2.53
- Open Carry: 1.85
- Religious Freedom: 2.31
- Social Media: 2.35
- Soda Ban: 2.33

FYS Comparisons by Scenario for IBT: Recommendations
Mean Scores on a scale of 1 – 4, with 4 being the highest possible score

A One-Way ANOVA revealed statistical significance; however, a Tukey Post-Hoc Analysis revealed no significant differences among individual pairs of scenarios.

Legend:
- Campus Speech; n = 22
- Fracking; n = 14
- GMO; n = 36
- Music; n = 17
- Open Carry; n = 17
- Religious Freedom; n = 19
- Social Media; n = 27
- Soda Ban; n = 18

Values: 2.18, 2.46, 2.26, 2.65, 2.09, 2.14, 2.67, 2.46
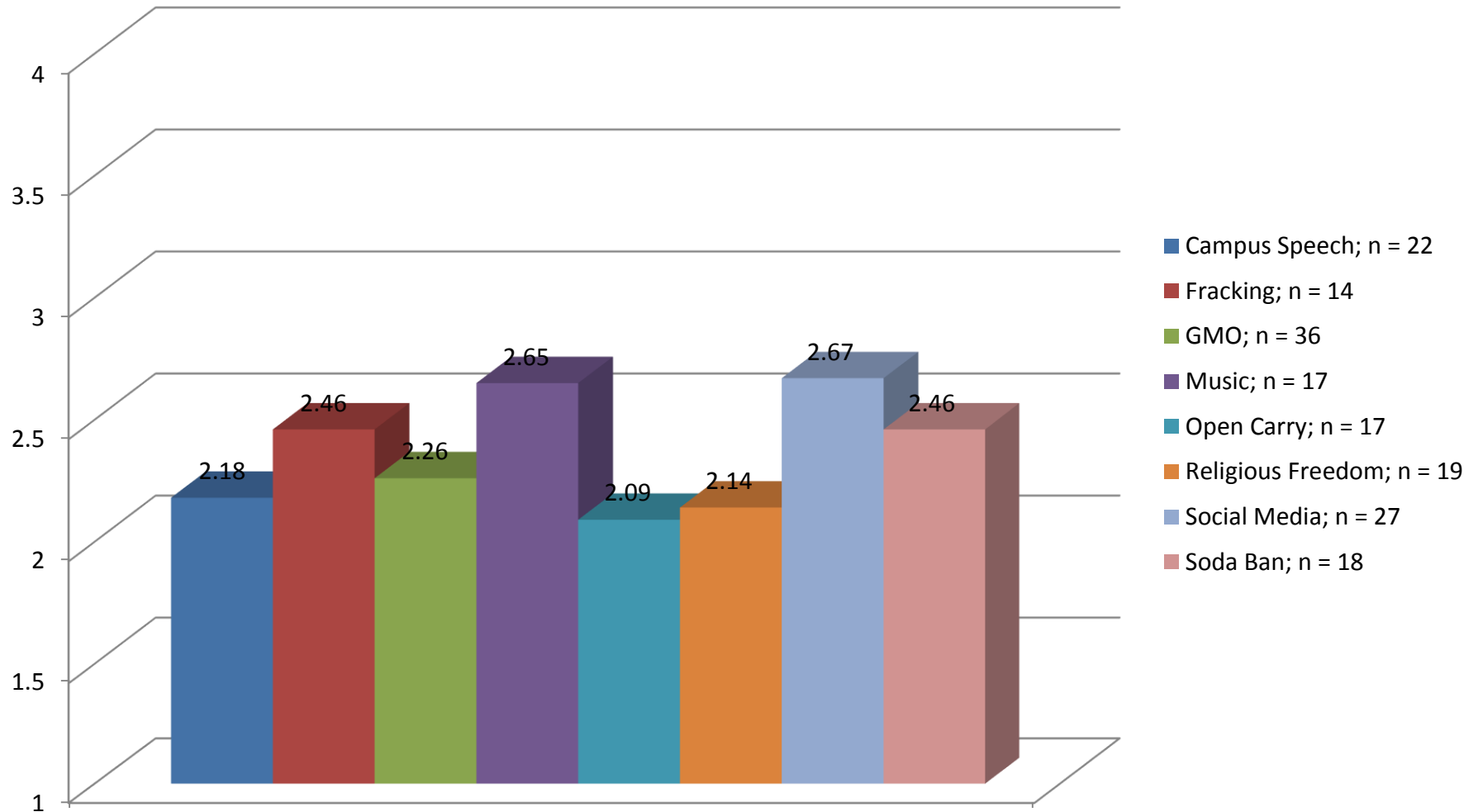
FYS Comparisons by Scenario for CF: Development
Mean Scores on a scale of 1 – 4, with 4 being the highest possible score

A One-Way ANOVA revealed statistical significance; however, a Tukey Post-Hoc Analysis revealed no significant differences among individual pairs of scenarios.
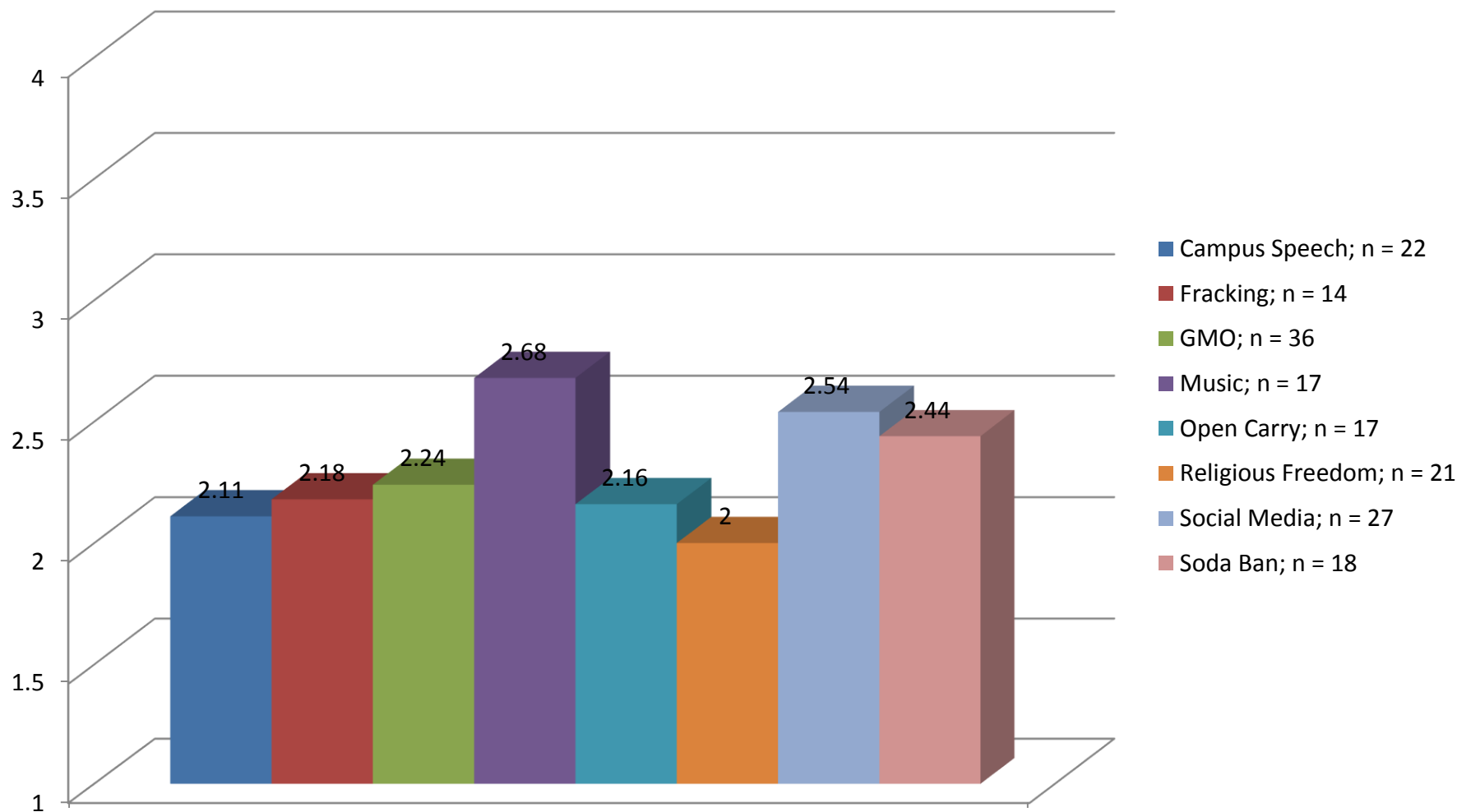

Legend:
- Campus Speech; n = 22
- Fracking; n = 14
- GMO; n = 36
- Music; n = 17
- Open Carry; n = 17
- Religious Freedom; n = 21
- Social Media; n = 27
- Soda Ban; n = 18

Values: 2.11, 2.18, 2.24, 2.68, 2.16, 2, 2.54, 2.44

FYS Comparisons by Scenario for CF: Convention/Format
Mean Scores on a scale of 1 – 4, with 4 being the highest possible score

A One-Way ANOVA did not reveal any statistically significant differences in means across the scenarios.

Legend:
- Campus Speech; n = 22
- Fracking; n = 14
- GMO; n = 36
- Music; n = 17
- Open Carry; n = 17
- Religious Freedom; n = 21
- Social Media; n = 27
- Soda Ban; n = 18

Values: 2.11, 2.18, 2.06, 2.24, 2, 2, 2.3, 2.39

FYS Comparisons by Scenario for CF: Communication Style
Mean Scores on a scale of 1 – 4, with 4 being the highest possible score
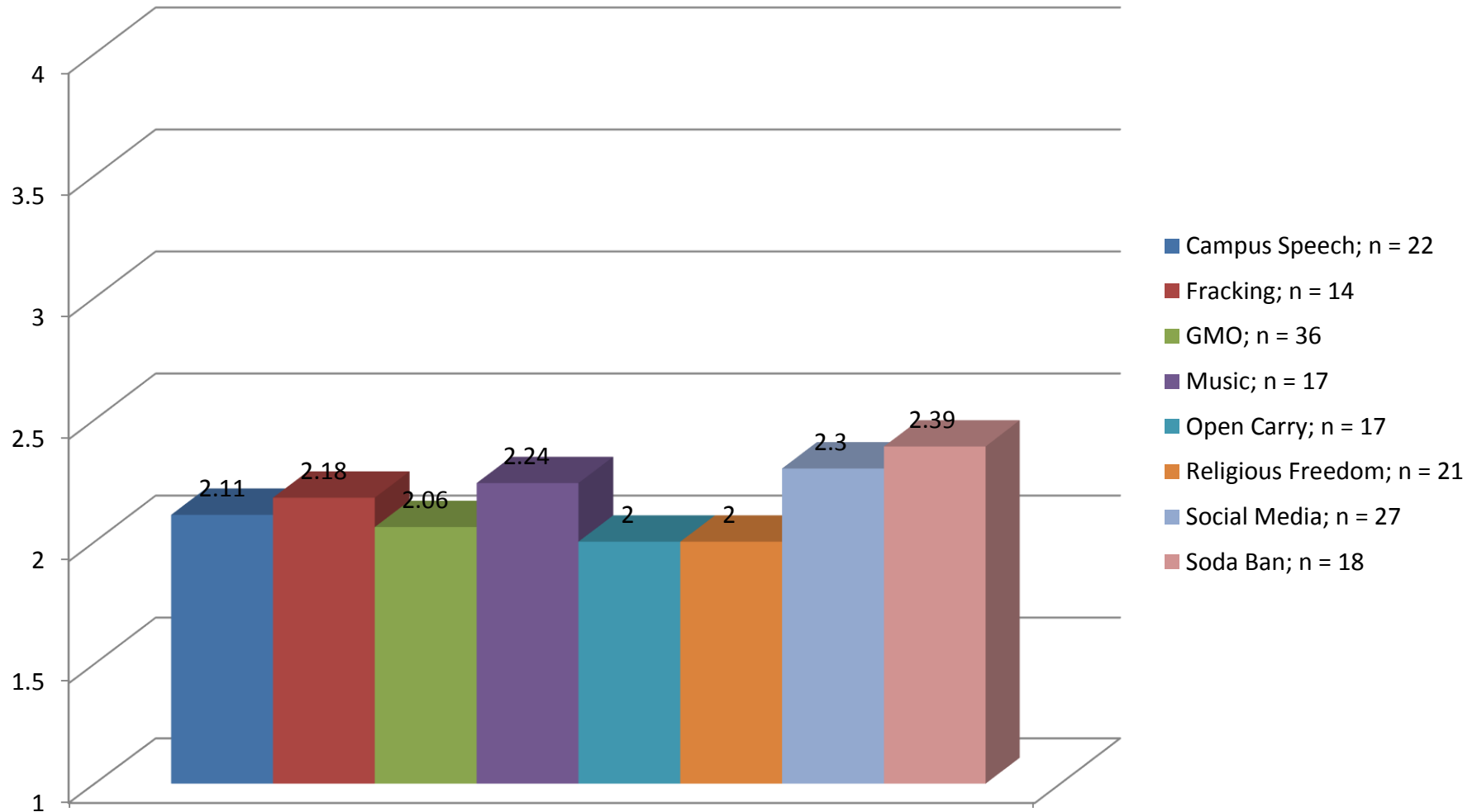
A One-Way ANOVA revealed statistical significance; however, a Tukey Post-Hoc Analysis revealed no significant differences among individual pairs of scenarios.

Legend:
- Campus Speech; n = 22 — 2.48
- Fracking; n = 14 — 2.21
- GMO; n = 36 — 2.42
- Music; n = 17 — 2.62
- Open Carry; n = 17 — 2.09
- Religious Freedom; n = 21 — 2.36
- Social Media; n = 27 — 2.54
- Soda Ban; n = 18 — 2.69