Comparison of Freshman Baseline with First Year Seminar Assessment Results
Academic Year 2017 – 2018

Summer Assessment Workgroup Members: Marie Archambault, Cam Brammer, Kim DeTardo-Bora, Robert Ellison, Victor Fet, Marty Laubach, Joan St. Germain, Anita Walz, Mary Welch

Summer Assessment Support Staff: Mary Beth Reynolds and Tim Melvin (Office of Assessment), Doug Nichols (Academic Affairs Technical Support)

<u>Executive Summary</u>

Background

Recommendations from the 2017 Assessment Workgroup (with current status in red)

Recommendations regarding baseline and senior assessments

1. The Assessment Workgroup recommended that baseline and senior assessments include the rubric so that students would have a better idea of how we are assessing their work. The rubric was added to baseline assessments during the 2017 Week of Welcome. Senior assessments were replaced with an analysis of capstone artifacts in 2017-2018.
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 recommended 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. We recommended that these discussions 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 discussed in greater detail in the Blackboard Outcomes Assessment Report. We felt that 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. Mary Beth Reynolds and Kristen Huff met with chairs in all colleges except the College of Information Technology and Engineering during academic year 2017-2018 to ask that they encourage capstone instructors to align capstone projects to Marshall's Capstone Critical Thinking Outcome in Blackboard. This has resulted in programs from the Colleges of Arts

Recommendations regarding Baseline/FYS/Senior Rubric

Based on interrater reliability results, the Assessment Workgroup recommended re-examining the *Communication Style* trait of the rubric again before beginning the 2018 assessment process. The assessment workgroup modified the wording of Level 4 for this trait in May 2018.

Procedures for 2018 Assessment

General Procedures

In August 2017, 1,590 incoming freshmen at Marshall University completed baseline assessments. These 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*. Freshmen completing Marshall's mandatory First Year Seminar in Critical Thinking (FYS) completed assessments that were similar to those finished by incoming freshmen.

In May 2018 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 243 Marshall Freshman baseline assessments was drawn from the pool of 1,590 (15%) of the total number of assessments available.   One hundred eighty-four (184) of the 243 freshmen from our baseline sample (76%) completed (or partially completed) FYS assessments. Two of these students did not complete the recommendation part of the FYS assessment, reducing the number of FYS assessments that were scorable across all traits to 182. The reasons we had no FYS assessments from the remaining 59 students in the baseline sample were as follows: 2 withdrew from Marshall following Week of Welcome; 5 completed FYS, but did not complete the final exam; 33 did not take FYS and are no longer enrolled at Marshall; 7 are still enrolled at Marshall, but have not taken and are not registered for FYS; 8 are registered to take FYS in fall 2018; 1 is registered to take FYS in summer 2018; 3 are part of an Honors program that does not require FYS.

All assessments were de-identified and 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 Results at the End of FYS

Mean scores (on a scale of 1 – 4) for the 243 freshmen in the baseline sample were 2.31 for Information Literacy: *information needed*, 1.92 for Information Literacy: *source acknowledgment*, 2.23 for Inquiry-Based Thinking: *evidence*, 2.09 for Inquiry-Based Thinking: *viewpoints*, 2.2 for Inquiry-Based Thinking: *recommendations*, 2.11 for Communication Fluency: *development*, 1.6 for Communication Fluency: *convention/format*, and 2.37 for Communication Fluency: *communication style*. These means did not differ significantly from the mean baseline results for the 184 to 182 students whose baseline scores remained in the Baseline/FYS analysis.

As noted above, there were 182 freshmen who completed both a baseline assessment and an FYS final exam, with an additional two students completing a baseline and the part of the FYS final exam that aligned to Information Literacy: *information needed*. This resulted in paired sample comparisons for 184 matched pairs for Information Literacy*: information needed* and 182 matched pairs for all other traits. 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*: information needed and acknowledgment of sources*, for Inquiry-Based Thinking*: viewpoints and recommendations*, and for Communication Fluency*: convention/format*. We note that, for the past six 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 2.42. 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 higher performance on *Campus Speech* and *Flu Vaccine* than on *GMO for* IL: source acknowledgment. Additionally, performance on *Fracking* was significantly higher than performance on *Campus Speech* or *Social Media* for CF: communication style.

Recommendations from the 2018 Assessment Workgroup

Recommendations regarding baseline assessments

1. The Assessment Workgroup recommended that we ensure that all instructors are thoroughly familiar with the assignment instructions. Beginning fall 2018 incoming freshmen will complete baseline assessments online through the assignment module in Blackboard Learn during the first week of their UNI 101 courses. We will work with the baseline assessment creation team to ensure that instructions in Blackboard are clear. Additionally, we will communicate the instructions to the UNI 101 course director.

2. The Assessment Workgroup recommended that we clarify on the baseline/FYS rubric that the trait *Information needed* applies to Part A of the Assessment and that *all other traits* apply to Part B. This should add additional clarification for students regarding information needed in each section of the assessment.
3. The Assessment Workgroup recommended that both baseline and FYS assessments should make explicit to students the convention to be used for their recommendations.
4. The Assessment Workgroup also recommended that total page length of documents in document library should be similar across assessments.

Supporting Documentation

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

Academic Year 2017 - 2018

Review Procedures

- Two hundred forty-three (243) freshman baseline assessments were used for this evaluation. Freshman assessments represented approximately 15% of the 1,590 completed during the University's Week of Welcome in August 2017.

- One hundred eighty-four (184; 76%) of the 243 freshmen whose baseline assessments we sampled completed (or partially completed) similar assessments at the end of First Year Seminar (FYS). Two partial completers from FYS finished only the "Information Needed" section of the FYS assessment. This resulted in paired sample comparisons for 184 matched pairs for "Information Needed" and 182 matched pairs for all other rubric traits.

    - Assessments were de-identified.
    - 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 2018 – updated 5-14-2018

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 some sources of information from the DL.	Clearly acknowledges multiple 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 sophisticated language that skillfully communicates meaning to readers with clarity and fluency, and is virtually error-free.

Freshman Baseline Means

Mean Scores on a scale of 1 – 4, with 4 being the highest possible score
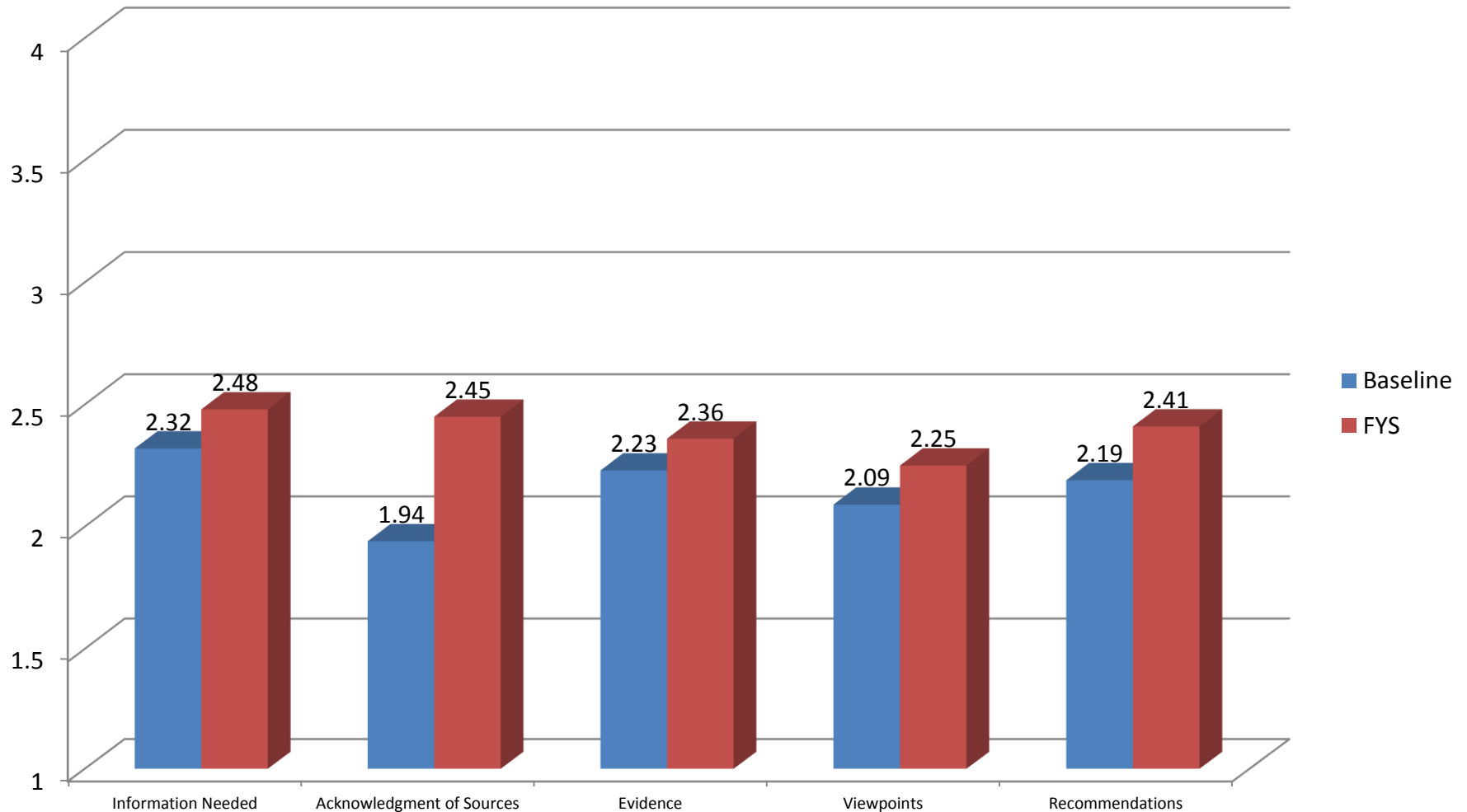
n = 243

Baseline
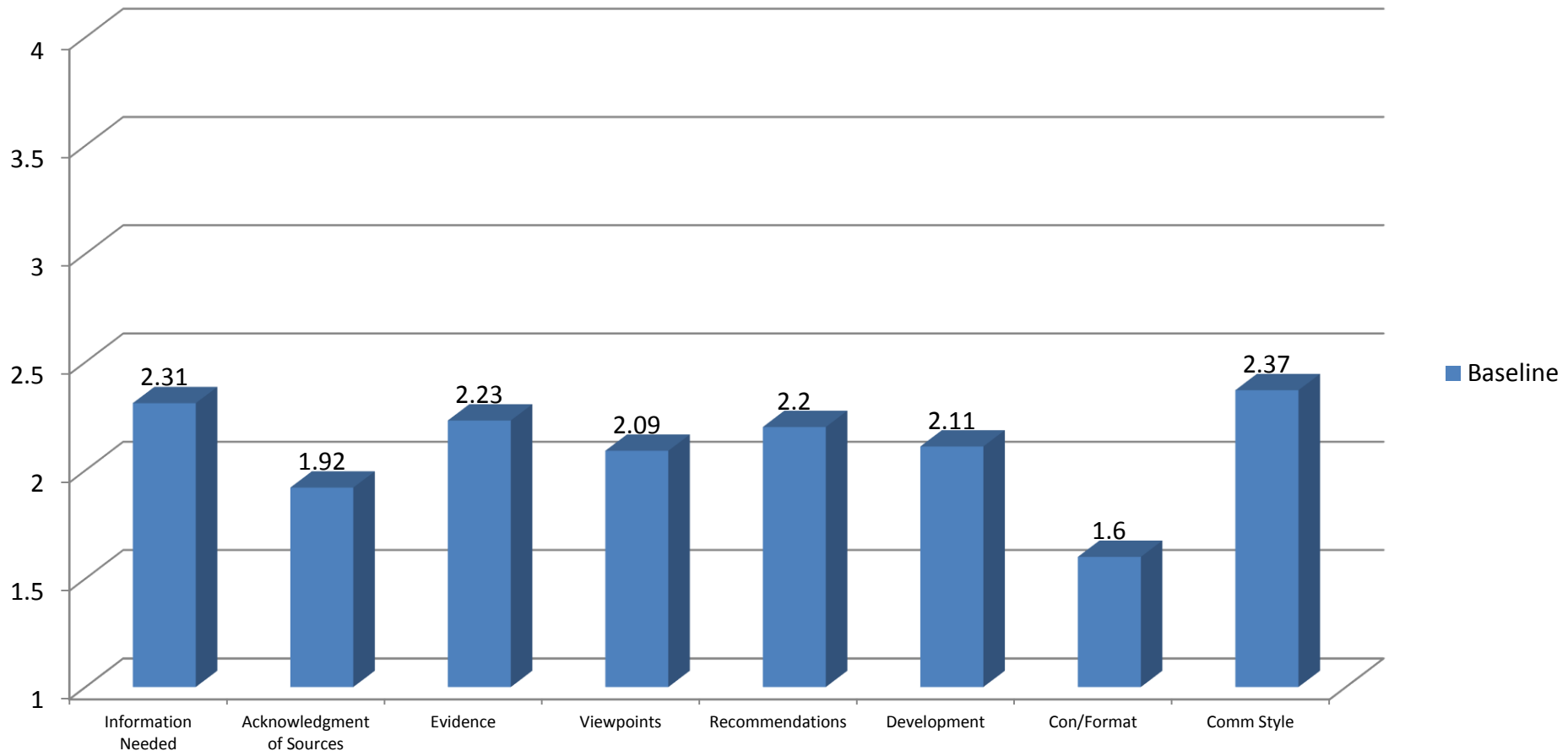
Freshman Baseline/FYS Comparisons

Mean Scores on a scale of 1 – 4, with 4 being the highest possible score
n = 184 for information needed; 182 for other traits
Mean differences are statistically significant for all traits except *Evidence*
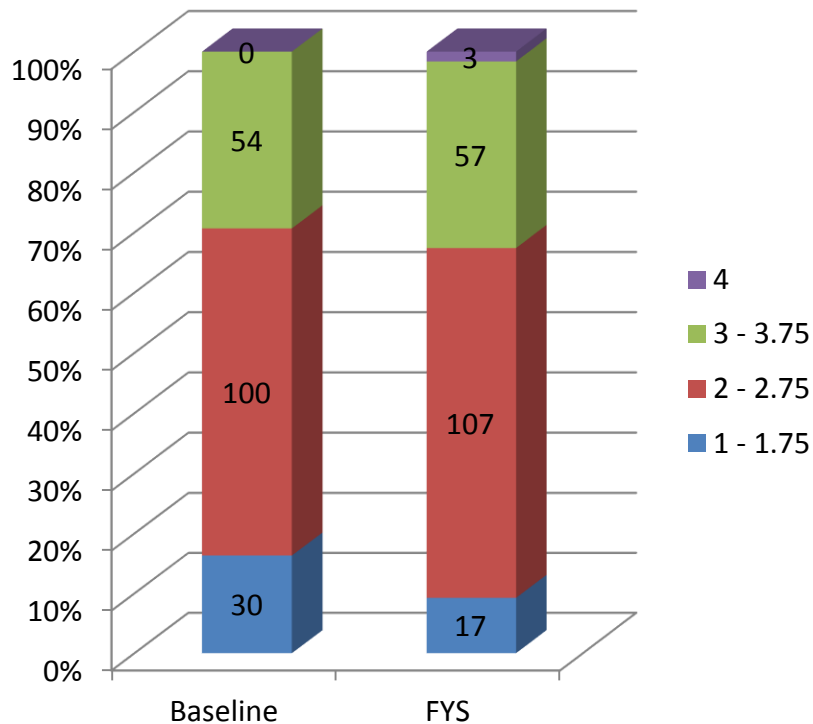
Freshman Baseline/FYS Comparisons

n = 184 (Information Needed); 182 (All Other Traits)

| Trait/ Performance Level	Info Needed	Acknowledgment of Sources	Evidence	Viewpoints	Recommendations
1 – 1.75 Baseline	30 (16%)	81 (45%	39 (21%)	45 (25%	44 (24%)
1 – 1.75 FYS	17 (9%)	40 (22%)	18 (10%)	19 (10%)	21 (12%)
2 – 2.75 Baseline	100 (54%)	72 (40%)	104 (57%)	113 (62%)	98 (54%)
2 – 2.75 FYS	107 (58%)	67 (37%)	124 (68%)	140 (77%)	102 (56%)
3 – 3.75 Baseline	54 (29%)	28 (15%)	37 (20%)	24 (13%)	40 (22%)
3 – 3.75 FYS	57 (31%)	69 (38%)	37 (20%)	21 (12%)	59 (32%)
4 Baseline	0	1 (1%)	2 (1%)	0	0
4 FYS	3 (2%)	6 (3%)	3 (2%)	2 (1%)	0
Grand Total Baseline	184 (100%)	182 (100%)	182 (100%)	182 (100%)	182 (100%)
Grand Total FYS	184 (100%)	182 (100%)	182 (100%)	182 (100%)	182 (100%)

Freshman Baseline/FYS Comparisons

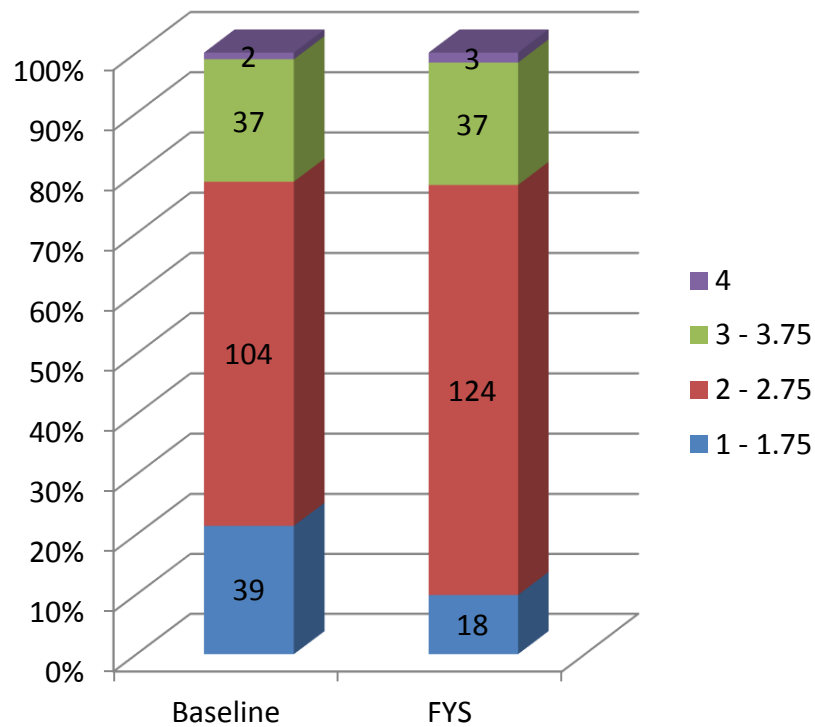n = 184 (Information Needed); 182 (Acknowledgment of Sources)

Information Needed
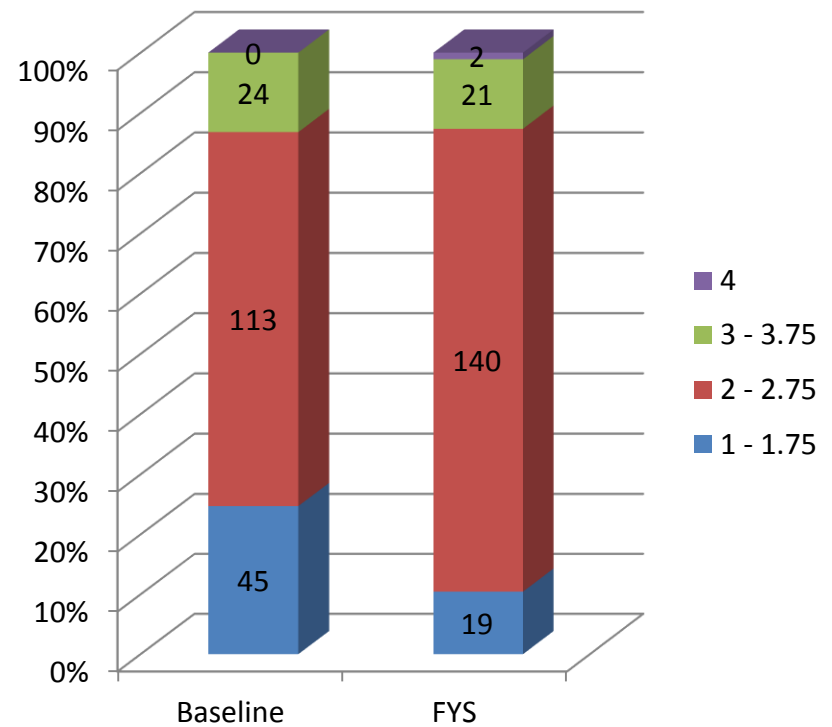
Acknowledgment of Sources
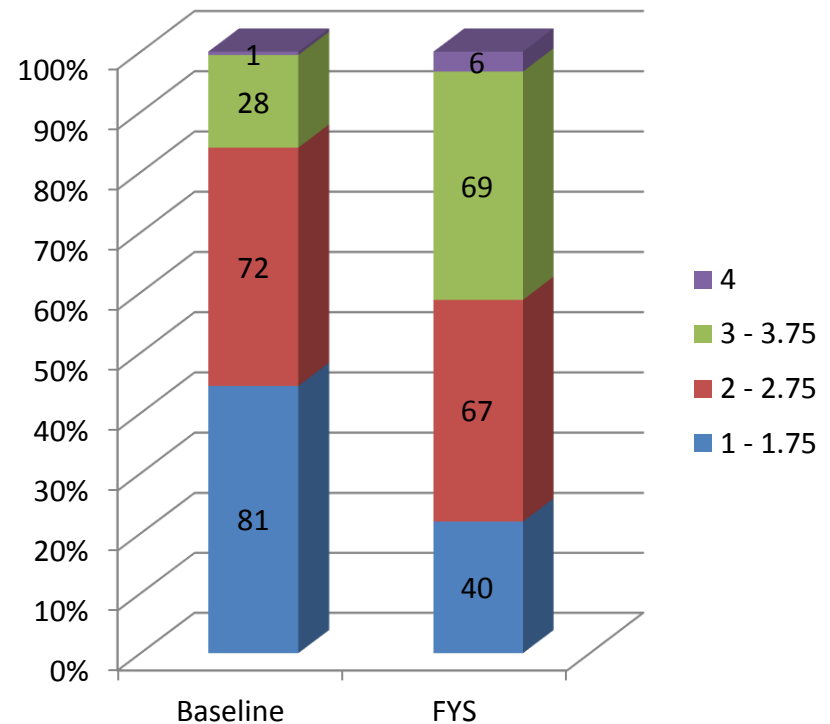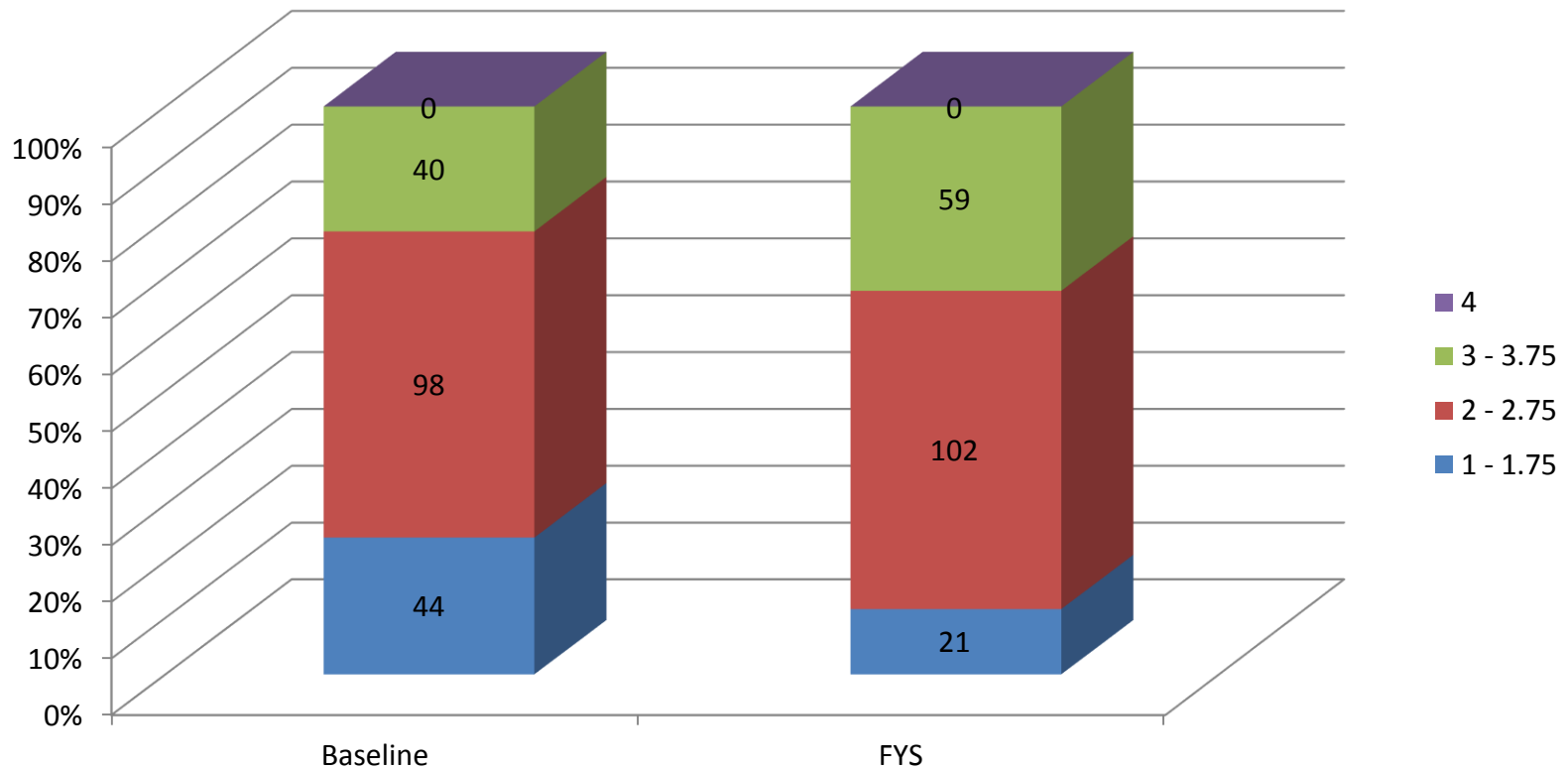
Freshman Baseline/FYS Comparisons
$n = 182$

Freshman Baseline/FYS Comparisons
n = 182

Recommendations

Baseline Inter-Rater Agreement Results

| Trait/ Agreement	Info Needed (Conservative Kappa = .385; Liberal Kappa = .969)	Acknowledgment of Sources (Conservative Kappa = .312; Liberal Kappa = .944)	Evidence (Conservative Kappa = .240; Liberal Kappa = .958)	Viewpoints (Conservative Kappa = .261; Liberal Kappa = .957)	Recommendations (Conservative Kappa = .246; Liberal Kappa = .948)
Agree	146 (60%)	132 (54%)	127 (52%)	135 (56%)	125 (51%)
Difference = 1 point or less	91 (37%)	100 (41%)	108 (44%)	100 (41%)	108 (44%)
Difference = 1.5 to 2 points	6 (2%)	11 (5%)	8 (3%)	8 (3%)	10 (4%)
Difference = 2.5 to 3 points	0	0	0	0	0
Total	243 (100%)	243 (100%)	243 (100%)	243 (100%)	243 (100%)

FYS Inter-Rater Agreement Results

Trait/ Agreement	Info Needed (Conservative Kappa = .226; Liberal Kappa = .905)	Acknowledgment of Sources (Conservative Kappa = .350; Liberal Kappa = .993)	Evidence (Conservative Kappa = .193; Liberal Kappa = .955)	Viewpoints (Conservative Kappa = .136; Liberal Kappa = .976)	Recommendations (Conservative Kappa = .265; Liberal Kappa = .964)
Agree	91 (49%)	99 (54%)	97 (53%)	102 (56%)	101 (55%)
Difference = 1 point or less	79 (43%)	82 (45%)	79 (43%)	77 (42%)	76 (42%)
Difference = 1.5 to 2 points	13 (7%)	1 (1%)	6 (3%)	3 (2%)	5 (3%)
Difference = 2.5 to 3 points	1 (1%)	0	0	0	0
Total	184 (100%)	182 (100%)	182 (100%)	182 (100%)	182 (100%)

Freshman Baseline/FYS Comparisons

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

n = 182

Mean differences are statistically significant for *Convention/Format*
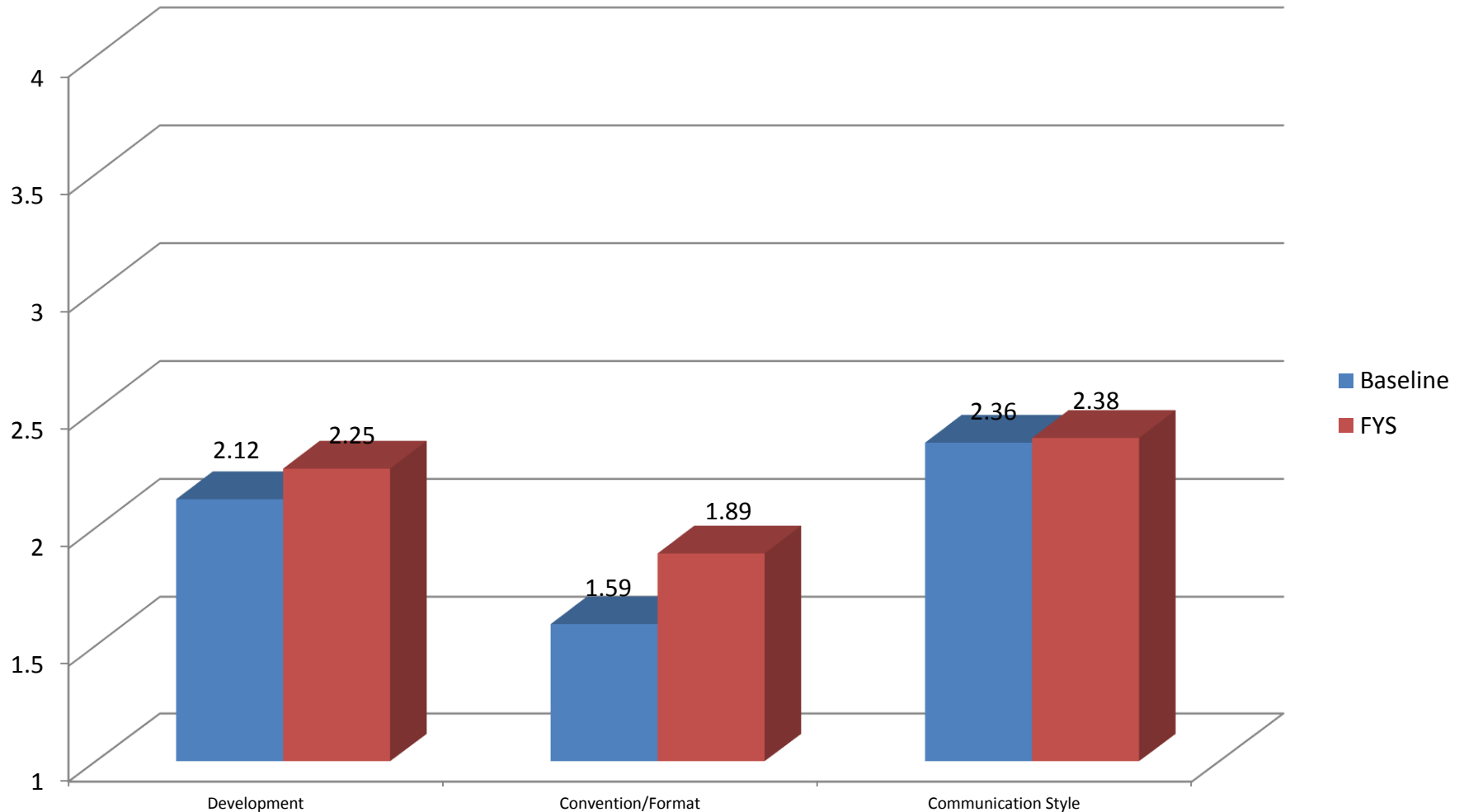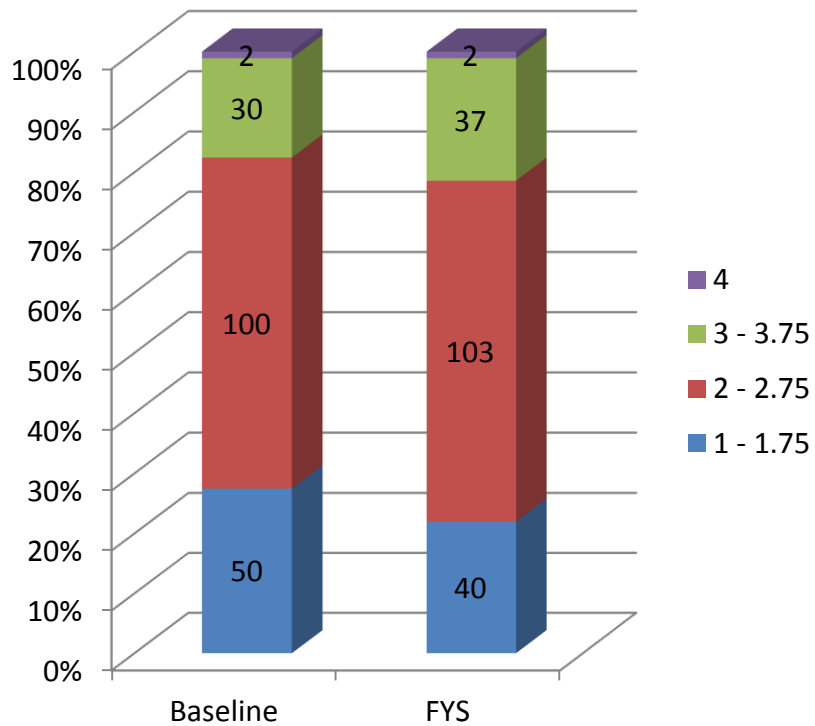
Freshman Baseline/FYS Comparisons
$n = 182$

| Trait/ Performance Level	Development	Convention/Format	Communication Style
1 – 1.75 Baseline	50 (27%)	128 (70%)	21 (12%)
1 – 1.75 FYS	40 (22%)	88 (48%)	33 (18%)
2 – 2.75 Baseline	100 (55%)	36 (20%)	113 (62%)
2 – 2.75 FYS	103 (57)	70 (38%)	95 (52%)
3 – 3.75 Baseline	30 (16%)	18 (10%)	48 (26%)
3 – 3.75 FYS	37 (20%)	22 (12%)	51 (28%)
4 Baseline	2 (1%)	0	0
4 FYS	2 (1%)	2 (1%)	3 (2%)
Grand Total Baseline	182 (100%)	182 (100%)	182 (100%)
Grand Total FYS	182 (100%)	182 (100%)	182 (100%)

Freshman Baseline/FYS Comparisons

n = 182

Development

Convention/Format
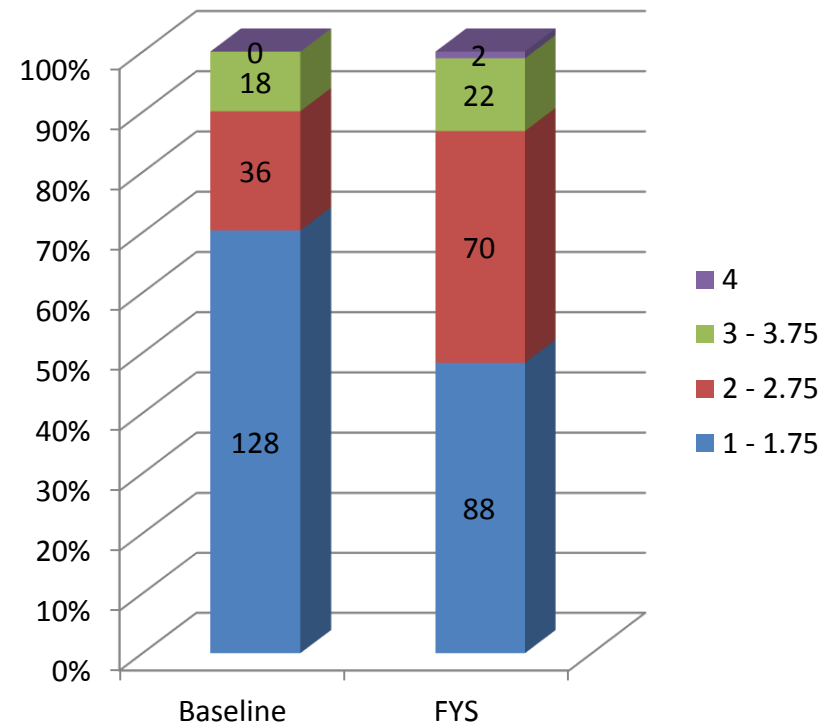
Freshman Baseline/FYS Comparisons
n = 182

Communication Style
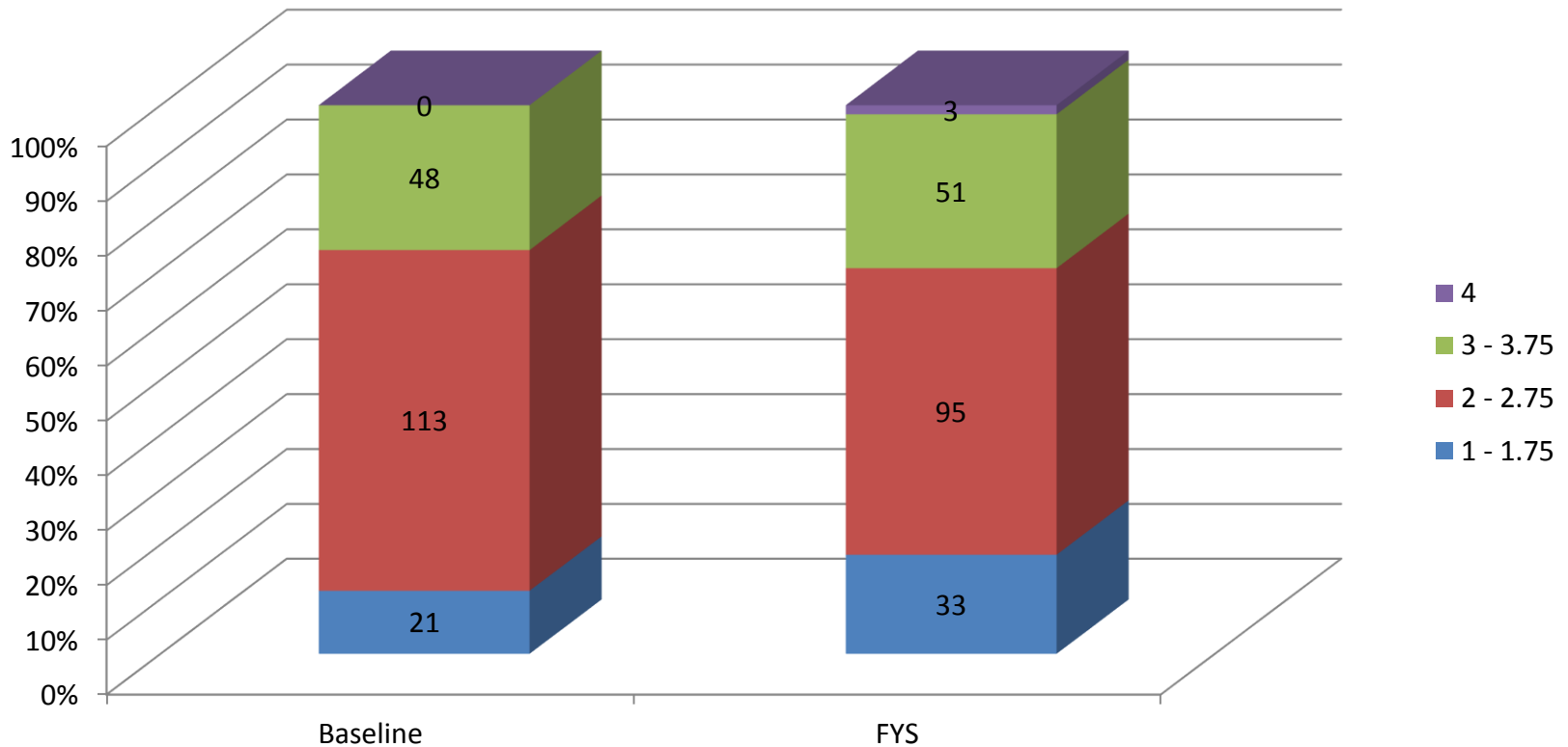
Baseline Inter-Rater Agreement Results

Trait/ Agreement	Development (Conservative Kappa = .296; Liberal Kappa = .954)	Convention/Format (Conservative Kappa = .250; Liberal Kappa = .949)	Communication Style (Conservative Kappa = .237; Liberal Kappa = .978)
Agree	132 (54%)	135 (56%)	136 (56%)
Difference = 1 point or less	102 (42%)	99 (41%)	103 (42%)
Difference = 1.5 to 2 points	9 (4%)	9 (4%)	4 (2%)
Difference = 2.5 to 3 points	0	0	0
Total	243 (100%)	243 (100%)	243 (100%)

FYS Inter-Rater Agreement Results

Trait/ Agreement	Development (Conservative Kappa = .233; Liberal Kappa = .945)	Convention/Format (Conservative Kappa = .255; Liberal Kappa = .911)	Communication Style (Conservative Kappa = .193; Liberal Kappa = .923)
Agree	90 (49%)	90 (49%)	86 (47%)
Difference = 1 point or less	84 (46%)	79 (43%)	85 (47%)
Difference = 1.5 to 2 points	7 (4%)	12 (7%)	11 (6%)
Difference = 2.5 to 3 points	1 (1%)	1 (1%)	0
Total	182 (100%)	182 (100%)	182 (100%)

Comparison of FYS Results for Each Trait by Scenario

Academic Year 2017 - 2018

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 = 14
- Fracking; n = 10
- GMO; n = 37
- Flu Vaccine; n = 25
- Trigger Warnings; n = 27
- College Costs; n = 42
- Social Media; n = 17
- Curriculum Coordinator; n = 12

Bar values: 2.79, 2.8, 2.51, 2.5, 2.39, 2.36, 2.4, 2.5
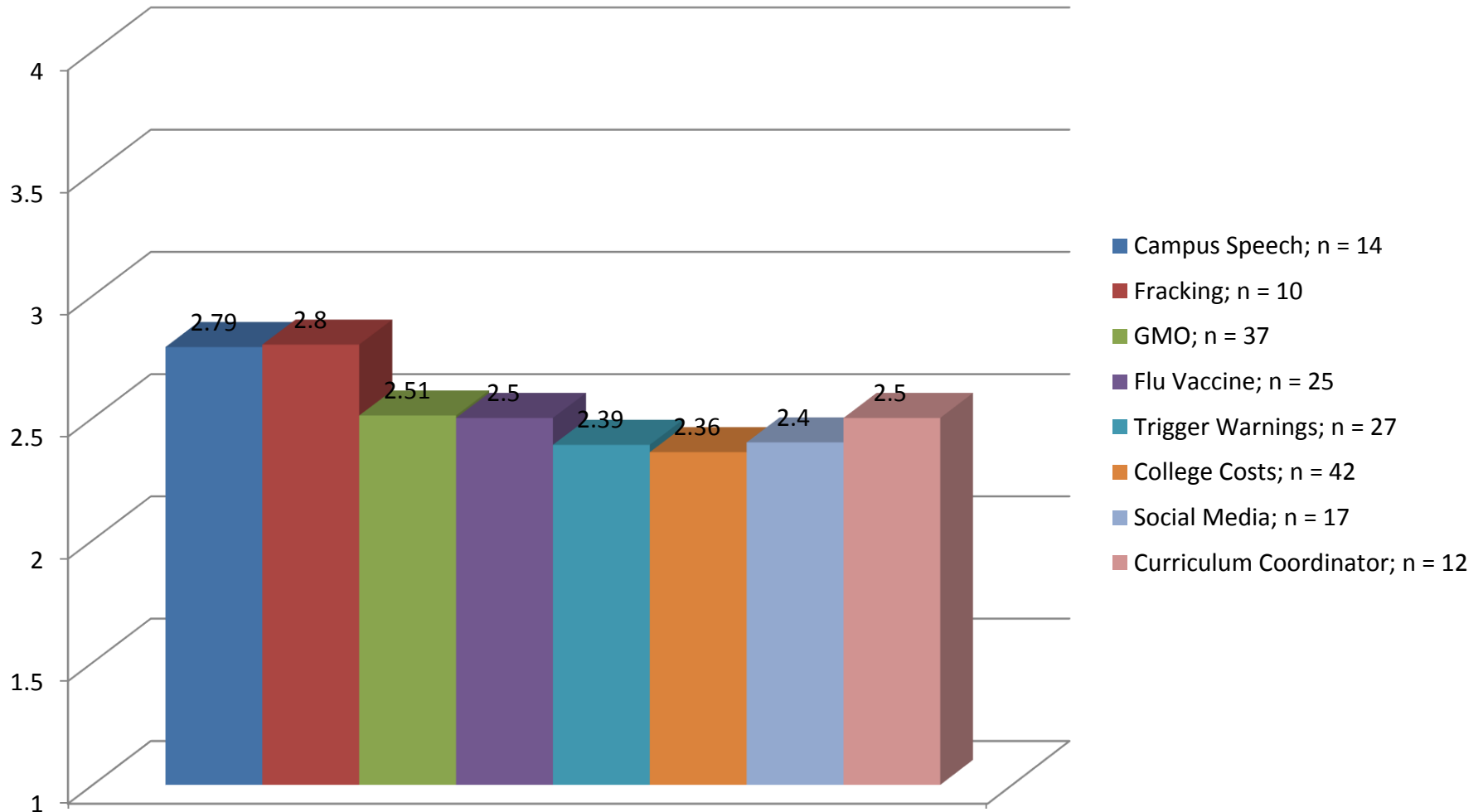
FYS Comparisons by Scenario for IL: Source Acknowledgment
Mean Scores on a scale of 1 – 4, with 4 being the highest possible score

A One-Way ANOVA revealed statistical significance; a Tukey Post-Hoc Analysis revealed that student performance on Campus Speech and Flu Vaccine was significantly higher than performance on GMO.

Legend:
- Campus Speech; n = 14
- Fracking; n = 10
- GMO; n = 37
- Flu Vaccine; n = 25
- Trigger Warnings; n = 27
- College Costs; n = 41
- Social Media; n = 17
- Curriculum Coordinator; n = 11

Values: 2.71, 2.8, 1.95, 2.64, 2.52, 2.49, 2.46, 2.68
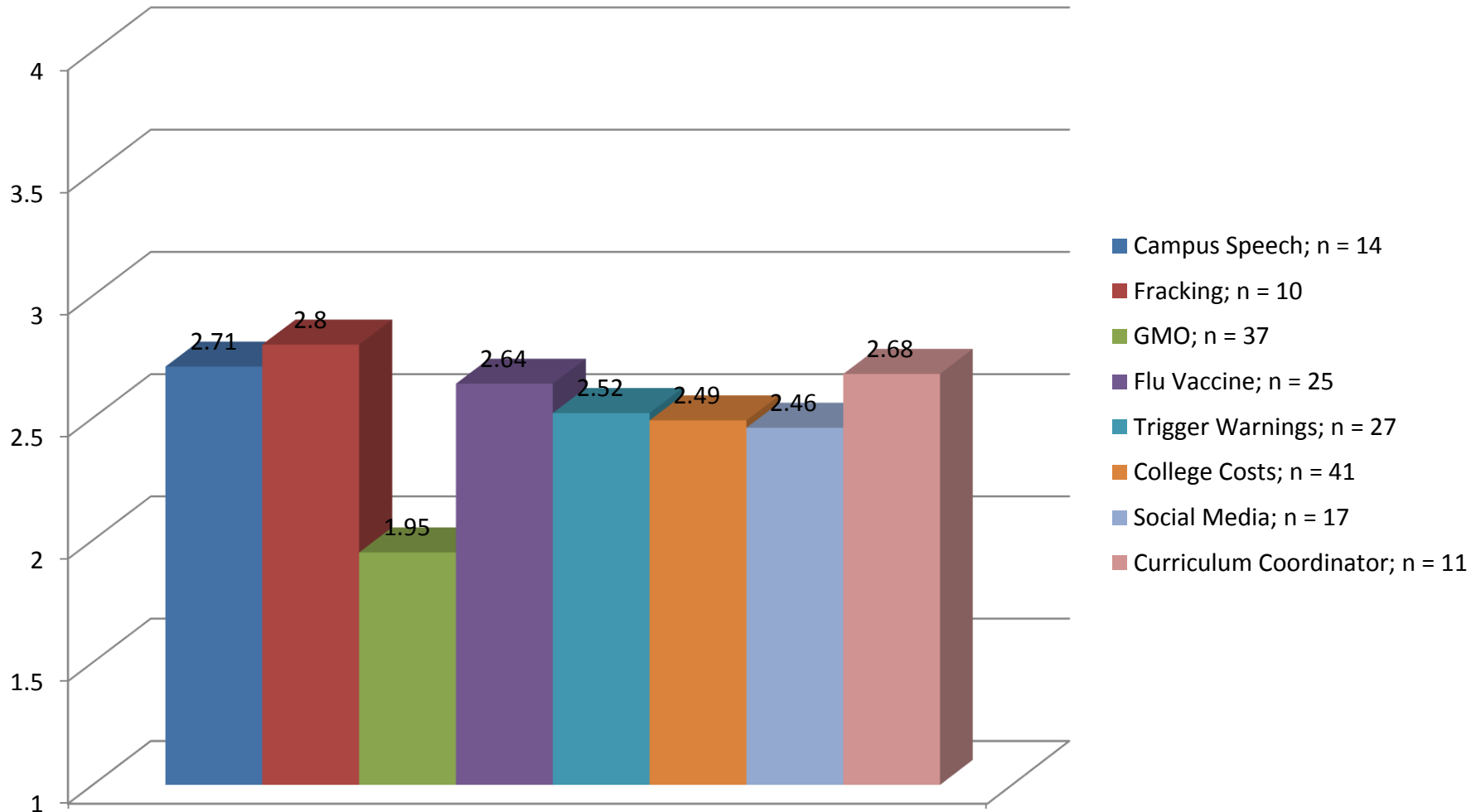
FYS Comparisons by Scenario for IBT: Evidence
Mean Scores on a scale of 1 – 4, with 4 being the highest possible score

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

Legend:
- Campus Speech; n = 14
- Fracking; n = 10
- GMO; n = 37
- Flu Vaccine; n = 25
- Trigger Warnings; n = 27
- College Costs; n = 41
- Social Media; n = 17
- Curriculum Coordinator; n = 11

Values: 2.39, 2.6, 2.12, 2.6, 2.39, 2.35, 2.29, 2.45
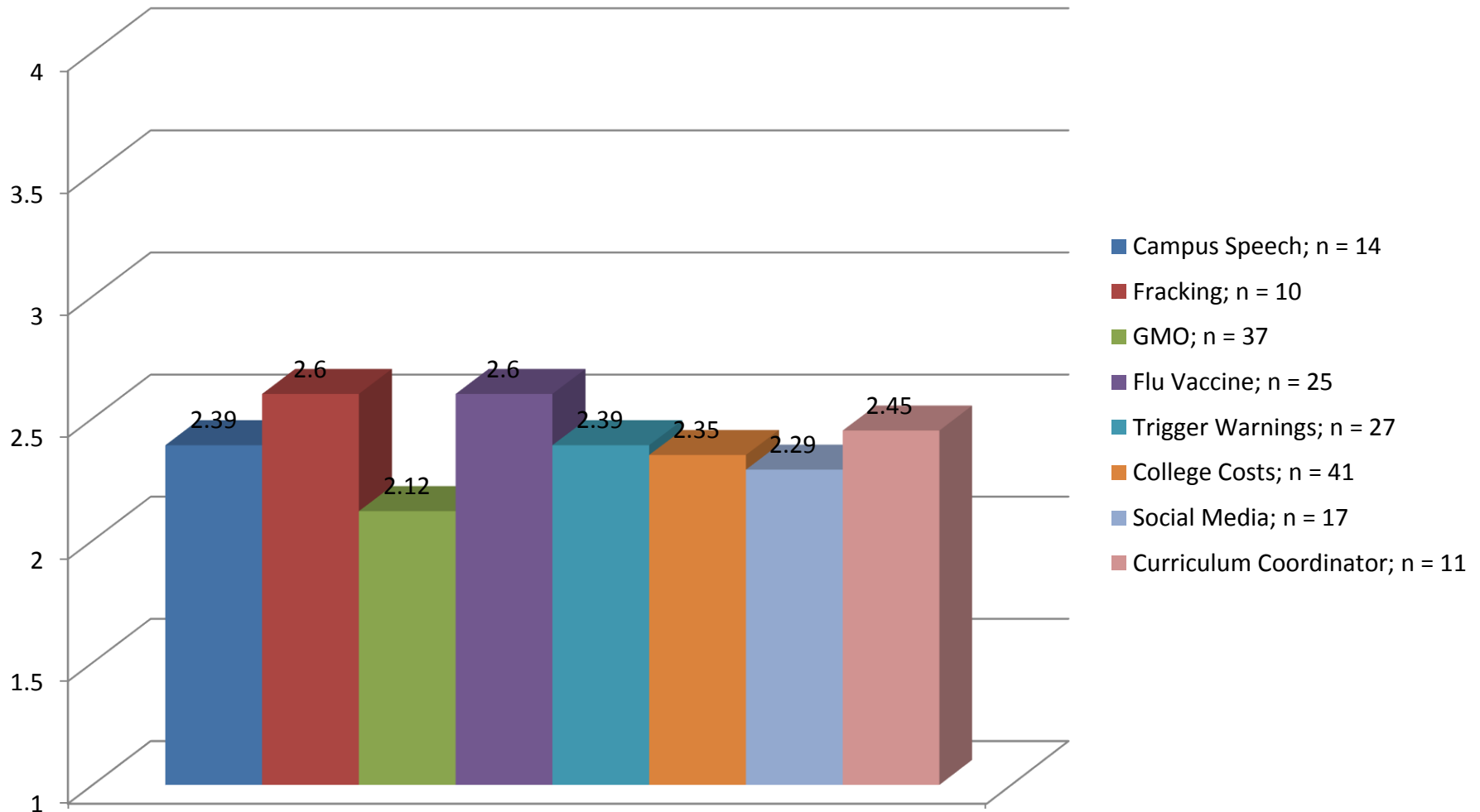
FYS Comparisons by Scenario for IBT: Viewpoints
Mean Scores on a scale of 1 – 4, with 4 being the highest possible score

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

Legend:
- Campus Speech; n = 14
- Fracking; n = 10
- GMO; n = 37
- Flu Vaccine; n = 25
- Trigger Warnings; n = 27
- College Costs; n = 41
- Social Media; n = 17
- Curriculum Coordinator; n = 11

Values: 2.43, 2.5, 2.05, 2.34, 2.3, 2.21, 2.26, 2.23
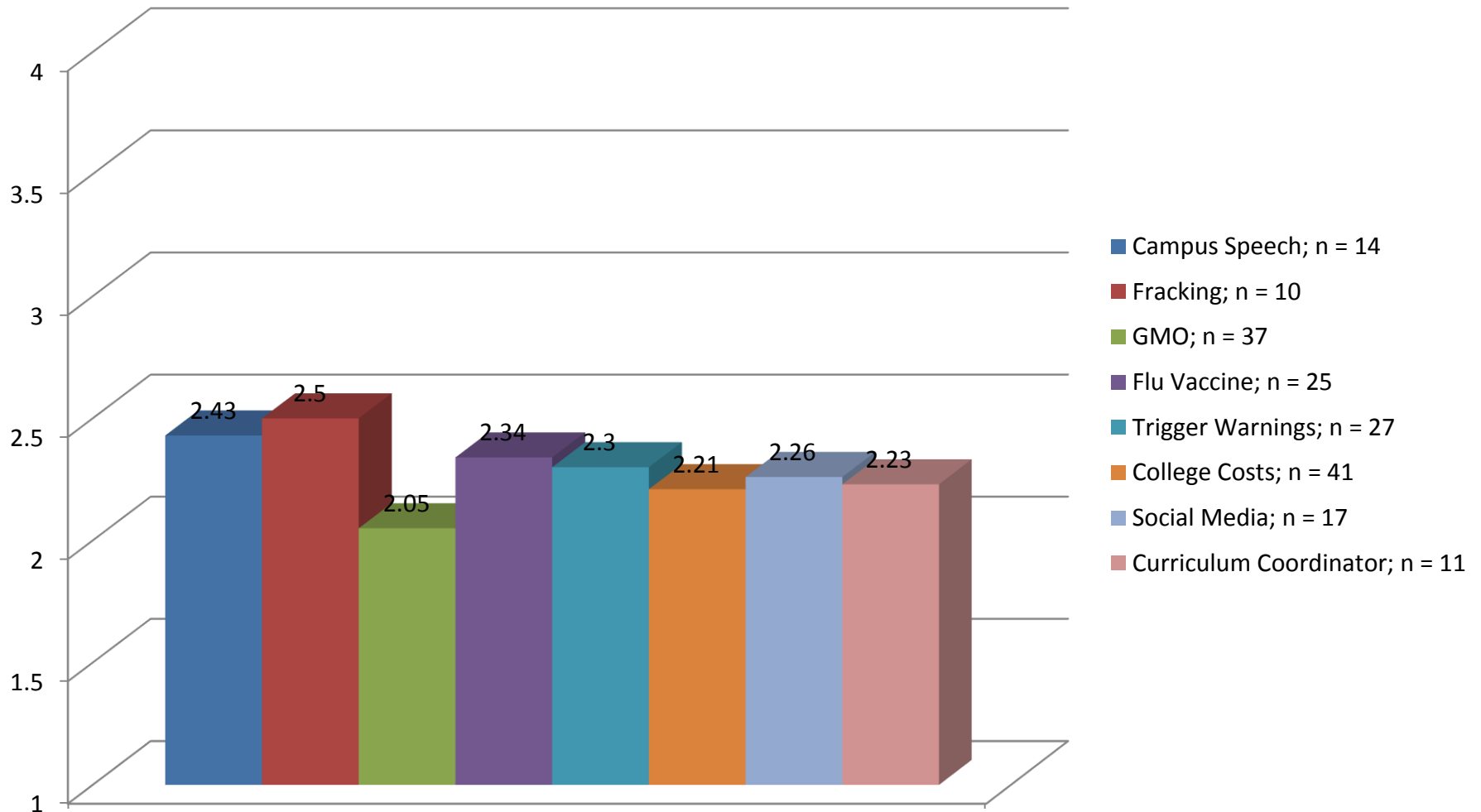
FYS Comparisons by Scenario for IBT: Recommendations
Mean Scores on a scale of 1 – 4, with 4 being the highest possible score

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

- ■ Campus Speech; n = 14
- ■ Fracking; n = 10
- ■ GMO; n = 37
- ■ Flu Vaccine; n = 25
- ■ Trigger Warnings; n = 27
- ■ College Costs; n = 41
- ■ Social Media; n = 17
- ■ Curriculum Coordinator; n = 11

2.43   2.65   2.32   2.56   2.52   2.27   2.44   2.45
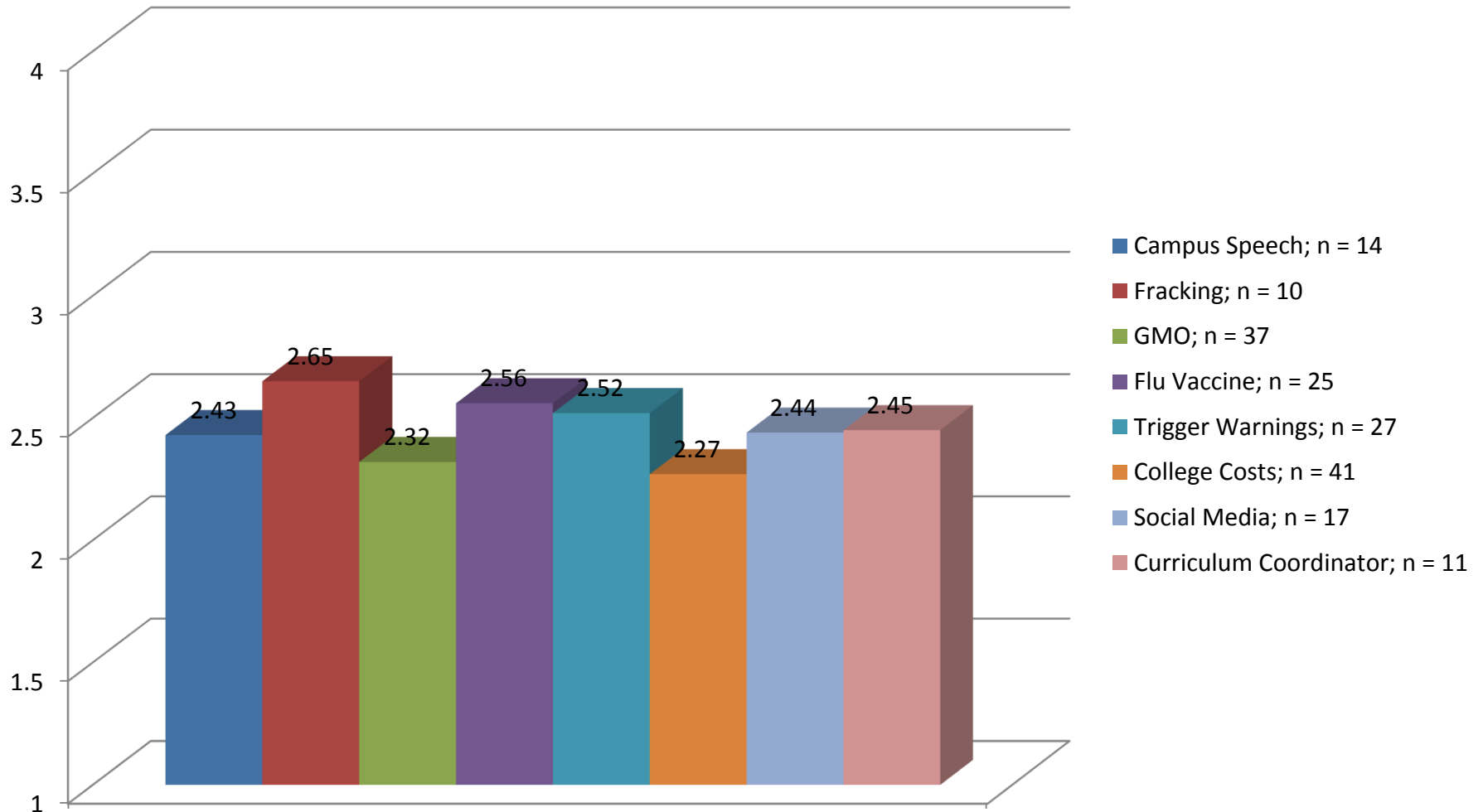
FYS Comparisons by Scenario for CF: Development
Mean Scores on a scale of 1 – 4, with 4 being the highest possible score

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

Legend:
- Campus Speech; n = 14
- Fracking; n = 10
- GMO; n = 37
- Flu Vaccine; n = 25
- Trigger Warnings; n = 27
- College Costs; n = 41
- Social Media; n = 17
- Curriculum Coordinator; n = 11

Values: 2.29, 2.5, 2.03, 2.34, 2.35, 2.26, 2.26, 2.18
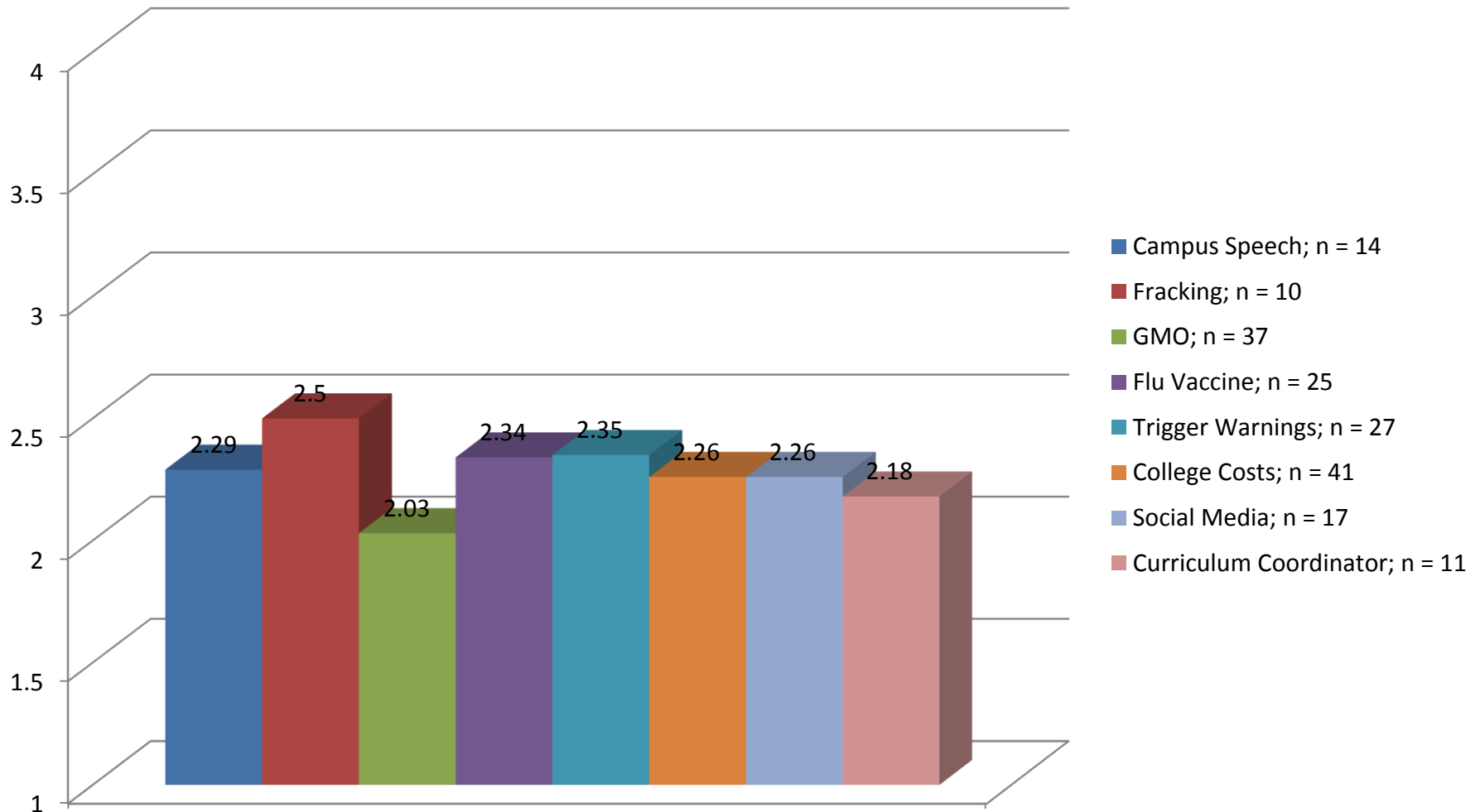
FYS Comparisons by Scenario for CF: Convention/Format
Mean Scores on a scale of 1 – 4, with 4 being the highest possible score

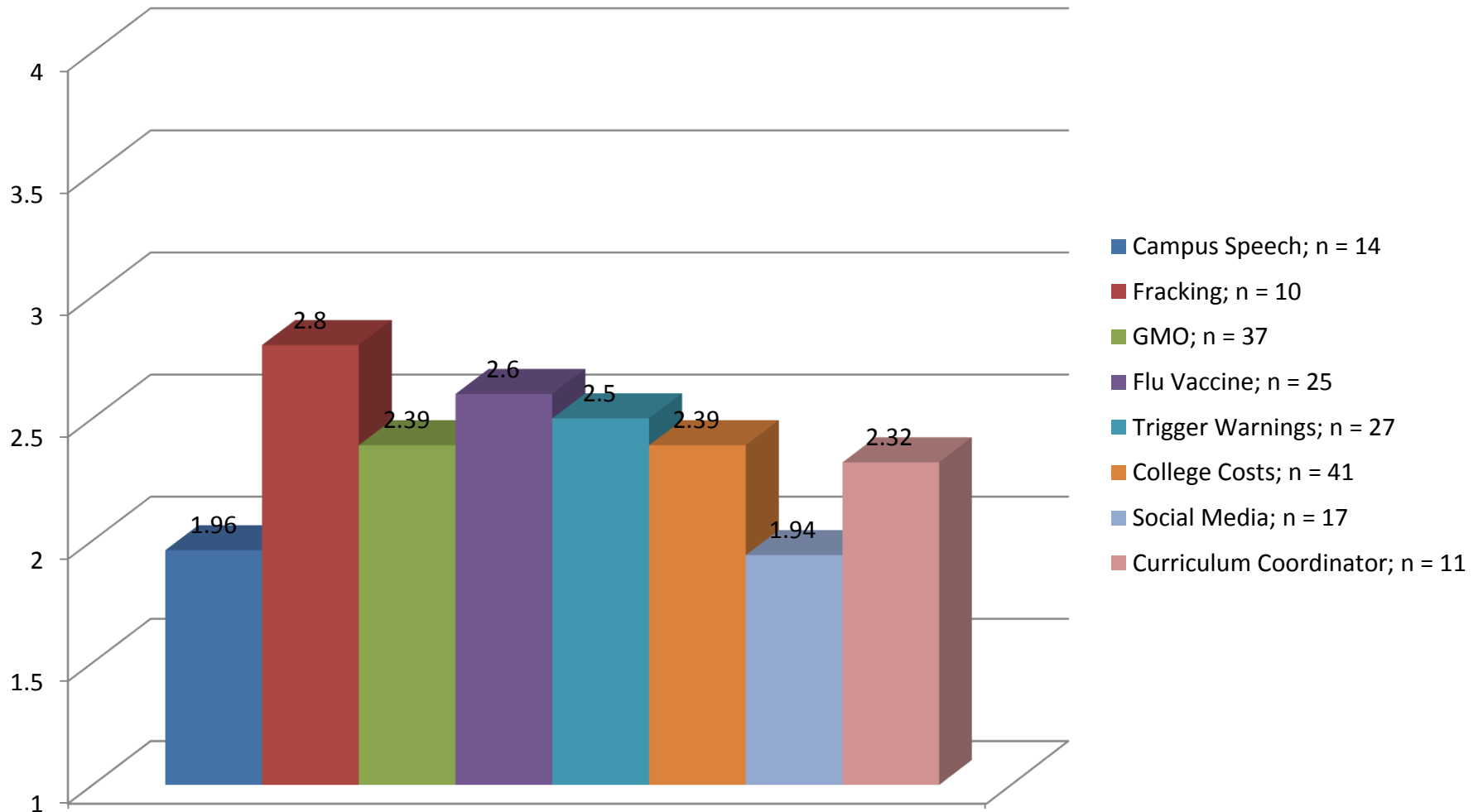A One-Way ANOVA did not reveal any statistically significant differences in means across the scenarios.

Legend:
- Campus Speech; n = 14
- Fracking; n = 10
- GMO; n = 37
- Flu Vaccine; n = 25
- Trigger Warnings; n = 27
- College Costs; n = 41
- Social Media; n = 17
- Curriculum Coordinator; n = 11

Values: 2, 2.4, 1.73, 2, 1.95, 1.78, 1.85, 1.91
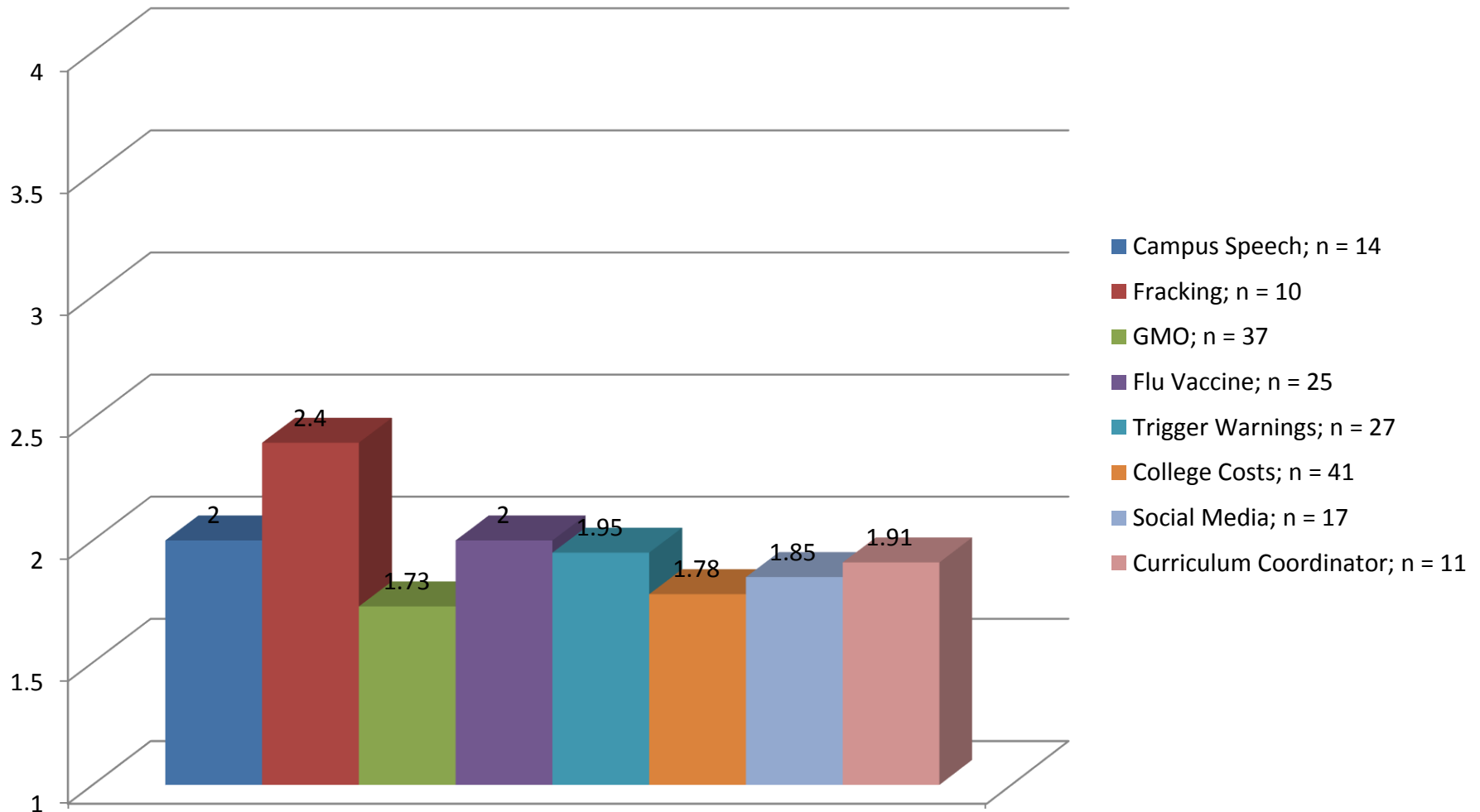
FYS Comparisons by Scenario for CF: Communication Style
Mean Scores on a scale of 1 – 4, with 4 being the highest possible score

A One-Way ANOVA revealed statistical significance; a Tukey Post-Hoc Analysis revealed that student performance on Fracking was significantly higher than performance on Campus Speech and on Social Media. Performance on Flu Vaccine also was significantly higher than performance on Social Media.

.

Legend:
- Campus Speech; n = 14
- Fracking; n = 10
- GMO; n = 37
- Flu Vaccine; n = 25
- Trigger Warnings; n = 27
- College Costs; n = 41
- Social Media; n = 17
- Curriculum Coordinator; n = 11

Values:
- Campus Speech: 1.96
- Fracking: 2.8
- GMO: 2.39
- Flu Vaccine: 2.6
- Trigger Warnings: 2.5
- College Costs: 2.39
- Social Media: 1.94
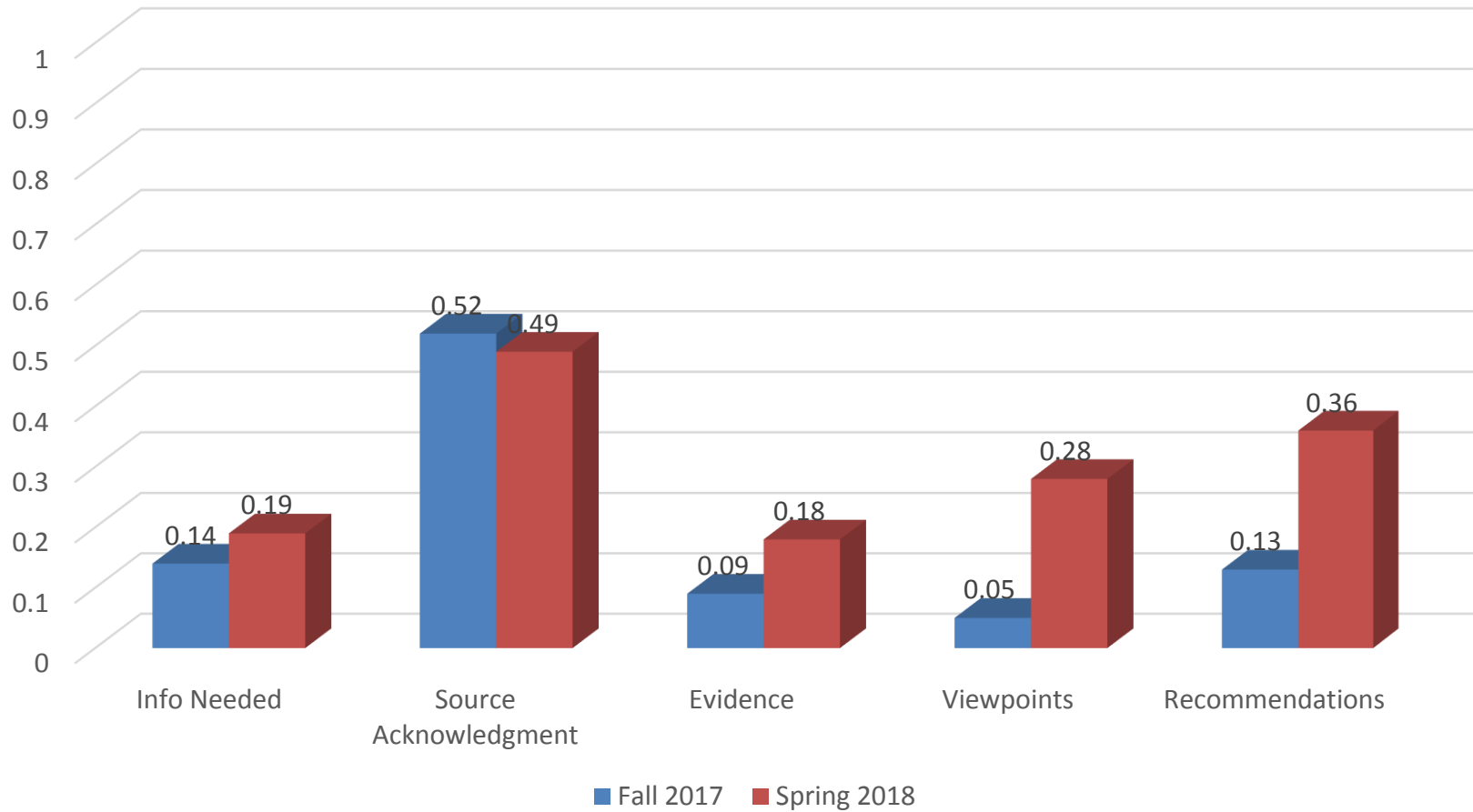- Curriculum Coordinator: 2.32

Comparison of Baseline to FYS Mean Gain Score for Each Trait by Semester of FYS

Academic Year 2017 - 2018

Baseline to FYS Mean Gain Scores for Each Trait
n = 104 in fall and 78 in spring (except for Fall Info Needed, where n = 106)

Baseline to FYS Mean Gain Scores for Each Trait

n = 104 in fall and 78 in spring