Comparison of Freshman Baseline with First Year Seminar Assessment Results

Academic Year 2020 – 2021

Summer Assessment Team Members: Marie Archambault, Cam Brammer, Kim DeTardo-Bora, Robert Ellison, Victor Fet, Marty Laubach, Anita Walz, Mary Welch

Summer Assessment Support Staff: Mary Beth Reynolds, Adam Russell, and Chris Sochor

Executive Summary

Background

Recommendations from the 2020 Assessment Team (current status is in red)

The Summer Assessment Team made the following recommendations:

1. That, given that both baseline and FYS assessments will be delivered via the assignment module in Blackboard Learn in fall 2020, both groups be allotted one day dedicated to completing this assessment. For baseline assessments, this day will be during the first two weeks of the term. For FYS assessments, the task will continue to be the course's final exam, given the last week of the term. This recommendation was implemented.

2. That we provide greater clarity to the directions that align with the *Information Literacy*: information needed part of the exam/rubric. Directions will be modified to ask students to outline additional information they need to make recommendations regarding the issues posed in their scenarios and to suggest methods as to how they will acquire this information. This recommendation was implemented.

Procedures for the 2021 Assessment

General Procedures

In August 2020, 1,373 incoming freshmen at Marshall University uploaded baseline assessments into Blackboard as part of their assignments for Freshman First Class (UNI 100). 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 (Critical) Thinking*, and *Communication Fluency*. Freshmen completing Marshall's mandatory First Year Seminar in Critical Thinking (FYS) completed assessments that mirrored those they finished as incoming freshmen, with 1,105 FYS assessments uploaded into Blackboard. To obtain 200 matched pairs of baseline/FYS assessments, we began by collecting a random sample of 450 FYS assessments. We then matched the students who completed these assessments with their baseline assessments and checked to see that all matches had correct and completed uploads. This process yielded a total of 280 matched pairs. From these matches, eighty were randomly discarded to yield a sample of 200 baseline and matching FYS assessments. During the Assessment Team's review, we discovered that one of the 200 baseline assessments was not complete. This left us with a usable sample of 199 matched pairs, representing 14.5% of the baseline population and 18% of the FYS population.

In May 2021, a group of eight faculty representing several academic colleges from across the university evaluated the baseline/FYS sample using a rubric that allowed them to score each artifact across eight criteria (traits). These traits included information needed and source acknowledgment (*Information Literacy*), evidence, viewpoints, and recommendation/position (I*nquiry-Based [Critical] Thinking*), and development, convention/format, and communication style (*Communication Fluency*). This project was coordinated by the Office of Assessment and Quality Initiatives.

Each assessment had two independent raters. Please see the supporting documentation 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

The baseline and FYS means (and standard deviations) for the 199 students in the sample with scorable baseline and FYS exams are reported below. Please note that, for students with scorable baseline and FYS (i.e., pre-post) assessments, *paired-samples t-tests* using adjusted alpha levels to control for Type I error (.025 for I*nformation literacy*), (.017 for *Inquiry-Based [Critical] Thinking*), and (.017 for *Communication Fluency*) showed significant mean differences between freshman baseline and FYS results for both traits (information needed and source acknowledgment) of *Information Literacy*, for one trait (evidence) of *Inquiry-Based [Critical] Thinking*, and for two traits (development and convention/format) of *Communication Fluency*. For the comparisons that reached statistical significance, students performed significantly better at the end of FYS than they had on their baseline assessments. We further note that *Communication Fluency* is not an outcome of FYS.

Outcome	Trait	Baseline Mean (*SD*)	FYS Mean (*SD*)	Statistical Significance
Information Literacy	**Information Needed**	2.377 (0.6286)	2.525 (0.6067)	*t(198)* = -2.705, *p* = .007
	Source Acknowledgment	2.241 (0.9071)	2.457 (0.8657)	*t(198)* = -3.122, *p* = .002
Inquiry-Based (Critical) Thinking	**Evidence**	2.322 (0.6906)	2.503 (0.6714)	*t(198)* = -3.120, *p* = .002
	Viewpoints	2.048 (0.4757)	2.106 (0.4511)	*t(198)* = -1.406, *p* = .161
	Recommendation/Position	2.462 (0.6944)	2.550 (0.6276)	*t(198)* = -1.589, *p* = .114
Communication Fluency	**Development**	2.317 (0.8011)	2.513 (0.7342)	*t(198)* = -3.344, *p* = .001
	Convention/Format	2.513 (0.7841)	2.774 (0.7665)	*t(198)* = -3.830, *p* < .001
	Communication Style	2.663 (0.5771)	2.663 (0.5793)	*t(198)* = 0.000, *p* = 1.00

A frequency analysis also showed the following increases in students scoring between 2.5 and 4.0 on the rubric between baseline and FYS. Please see the supporting documentation following this summary for additional information.

Outcome	Trait	Percentage Gain in Students Scoring 2.5 to 4.0 from Baseline to FYS
Information Literacy	**Information Needed**	15%
	Source Acknowledgment	18%
Inquiry-Based (Critical) Thinking	**Evidence**	16%
	Viewpoints	2%
	Recommendation/Position	17%
Communication Fluency	**Development**	16%
	Convention/Format	18%
	Communication Style	-2%

This year's results showed a significant difference in performance based on scenario used for the FYS assessments for one trait (source acknowledgment) of *Information Literacy*, for one trait (viewpoints) of *Inquiry-Based [Critical] Thinking*, and for one trait (convention/format) of *Communication Fluency*. For source acknowledgment and viewpoints students scored significantly higher on GMO Foods than on the other three scenarios. However, on convention/format, students scored significantly higher on the Online Gaming and Social Media scenarios than on GMO Foods, and higher on the Social Media scenario than on the Flu Vaccine. Also, gain scores between students in our sample who completed FYS in fall 2020 (n = 88) and those who completed FYS in spring 2021 (n = 111) did not differ significantly on any outcome trait. Please refer to the supporting documentation for additional detail.

Conclusions

The conclusions reached from this year's analysis mirror those of every analysis this team has performed since 2013. Marshall's freshmen have shown significant improvement in at least some traits of information literacy and critical thinking skills between matriculation and the completion of First Year Seminar in Critical Thinking each year. In 2019 and 2020 students' improvement reached statistical significance for all traits of both outcomes. This year, students did not show significant gains on two traits (viewpoints and recommendation/position) of *Inquiry-Based [Critical] Thinking*. However, we note that this sample's baseline scores were higher than were the baseline scores for the past two years.

Recommendations from the 2021 Assessment Team

The Summer Assessment Team made the following recommendations:

1. That we follow-up with the Center for Teaching and Learning at the end of the summer to ask how the newly configured FYS course will be assessed.
2. That our assessment in summer 2022 include a comparison of student performance between large and small FYS sections. Note: We will need to control for any difference in student profiles between different sized sections.
3. That the Office of Assessment and Quality Initiatives continue to provide and distribute shorter reports in more digestible formats. We recommend that these reports be disseminated campus-wide through the Assessment Newsletter and shared with the Faculty Senate.

MARSHALL UNIVERSITY

Supporting Documentation

MARSHALL UNIVERSITY

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

Academic Year 2020 - 2021

Review Procedures

- Two hundred (200) FYS critical thinking artifacts were matched with 200 baseline critical thinking artifacts. During the evaluation we discovered that one baseline artifact was not complete, so discarded that student's FYS artifact as well. This reduced the total number of matched artifacts in our sample to 199. This represented 14.5% of the baseline of 1,373 and 18% of the FYS population of 1,105.

Review Procedures Continued

- 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, 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. (For this review, all raters were able to reconcile disagreements, so third raters were not needed).

Interrater Reliability

- We conducted interrater reliability analyses using the Cohen's Kappa statistical procedure. In so doing, we used the following rules, similar to those suggested by Stellmack, Kohneim-Kalkstein, Manor, Massey, & Schmitz (2009):
  - Since our scoring procedure was to average final scores between two raters when scores differed by only one point, we used that averaged score (e.g., 1.5) as the score for both raters, counting it as an agreement in the interrater reliability analysis.
  - For scores that were two or more points apart, the original score of each reviewer was used in the analysis. Therefore, these scores were counted as disagreements.

Rubric Used for Scoring

Baseline/FYS Assessment Rubric – Summer 2020 – updated 5-11-2020

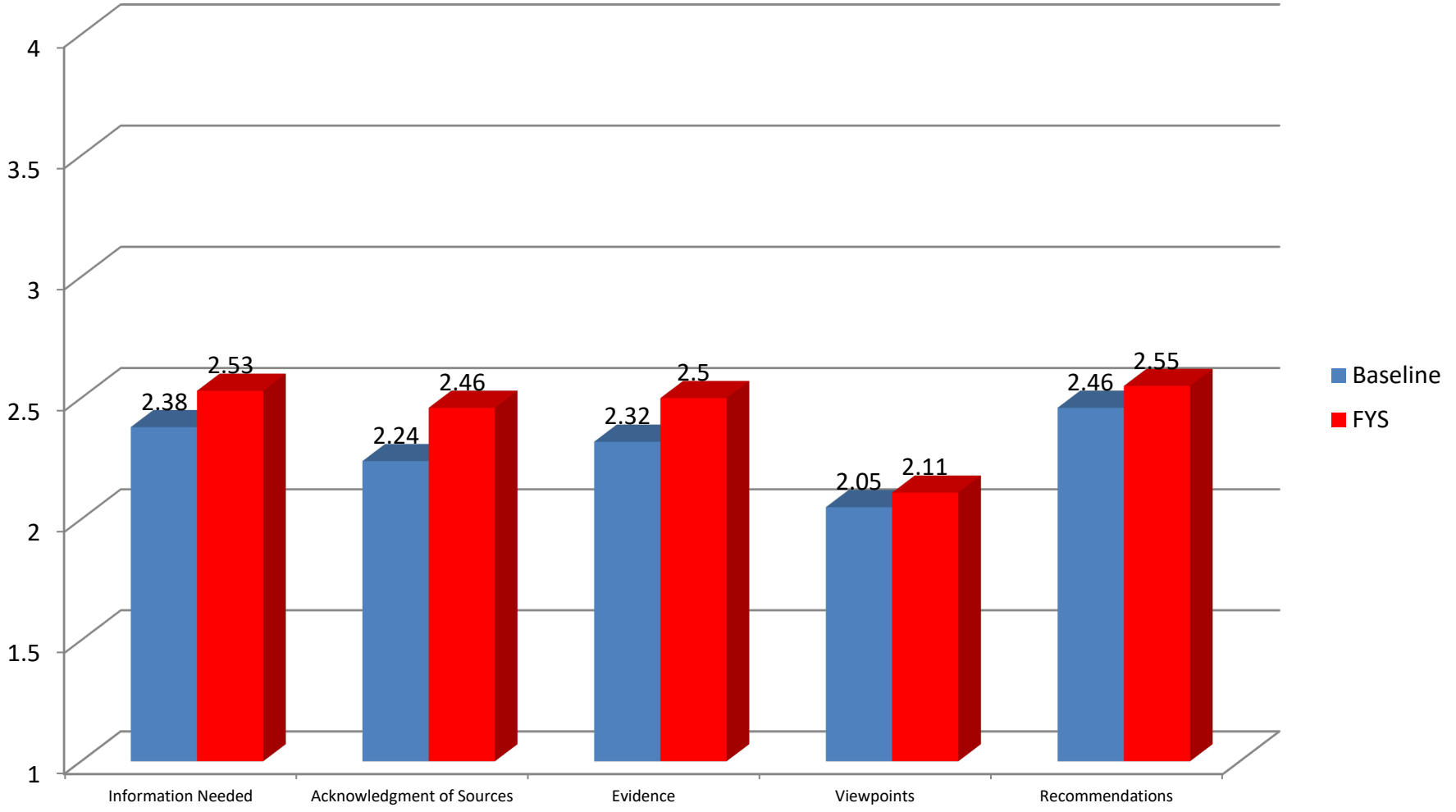
Outcomes	Traits	Performance Levels			
		1	2	3	4
Information Literacy	Information Needed	Does not acknowledge or assess the need for more information.	Acknowledges the need for more information but does not identify research methods/sources (or those identified are not feasible) that would address unanswered 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 (take a position) or makes a recommendation (takes a position), but does not justify it in any way.	Recommendation/position is justified, but does not acknowledge different sides of the issue.	Recommendation/position is justified and takes into account different sides/complexities of the issue.	Recommendation/position 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, 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/FYS Comparisons

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

n = 199

Mean differences are statistically significant for *Information Needed*, *Acknowledgment of Sources*, and *Evidence.*



Freshman Baseline/FYS Comparisons

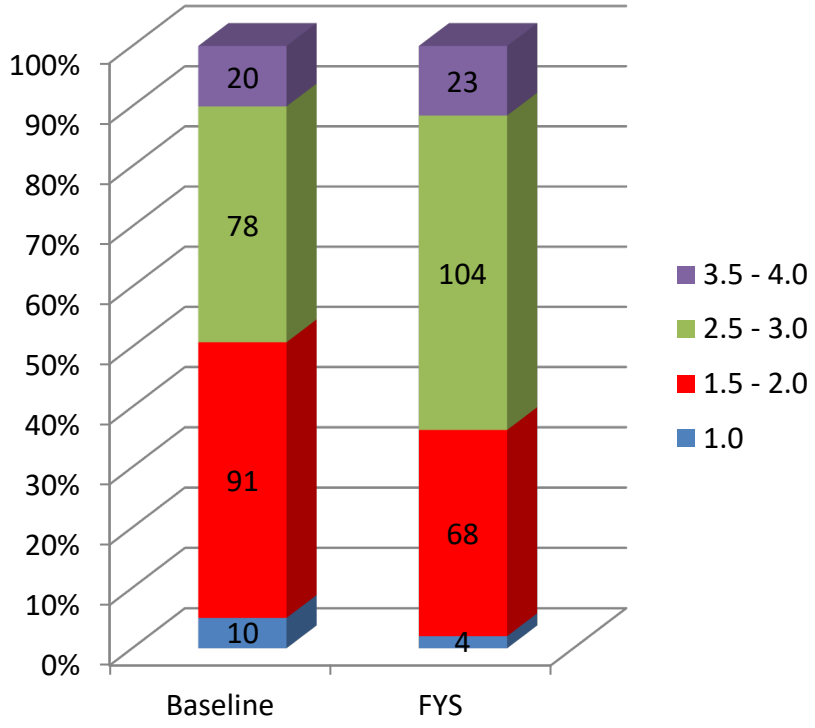
n = 199

Trait/ Performance Level	Info Needed	Acknowledgment of Sources	Evidence	Viewpoints	Recommendations
1.0 Baseline	10 (5%)	34 (17%)	14 (7%)	12 (6%)	11 (6%)
1.0 FYS	4 (2%)	24 (12%)	10 (5%)	9 (5%)	9 (5%)
1.5 – 2.0 Baseline	91 (46%)	82 (41%)	80 (40%)	125 (63%)	78 (39%)
1.5 – 2.0 FYS	68 (34%)	57 (29%)	53 (27%)	126 (63%)	45 (23%)
2.5 – 3.0 Baseline	78 (39%)	47 (24%)	88 (44%)	60 (30%)	87 (44%)
2.5 – 3.0 FYS	104 (52%)	85 (43%)	111 (56%)	61 (31%)	125 (63%)
3.5 – 4.0 Baseline	20 (10%)	36 (18%)	17 (9%)	2 (1%)	23 (12%)
3.5 – 4.0 FYS	23 (12%)	33 (17%)	25 (13%)	3 (2%)	20 (10%)
Grand Total Baseline	**199 (100%)**	**199 (100%)**	**199 (100%)**	**199 (100%)**	**199 (100%)**
Grand Total FYS	**199 (100%)**	**199 (100%)**	**199 (100%)**	**199 (100%)**	**199 (100%)**

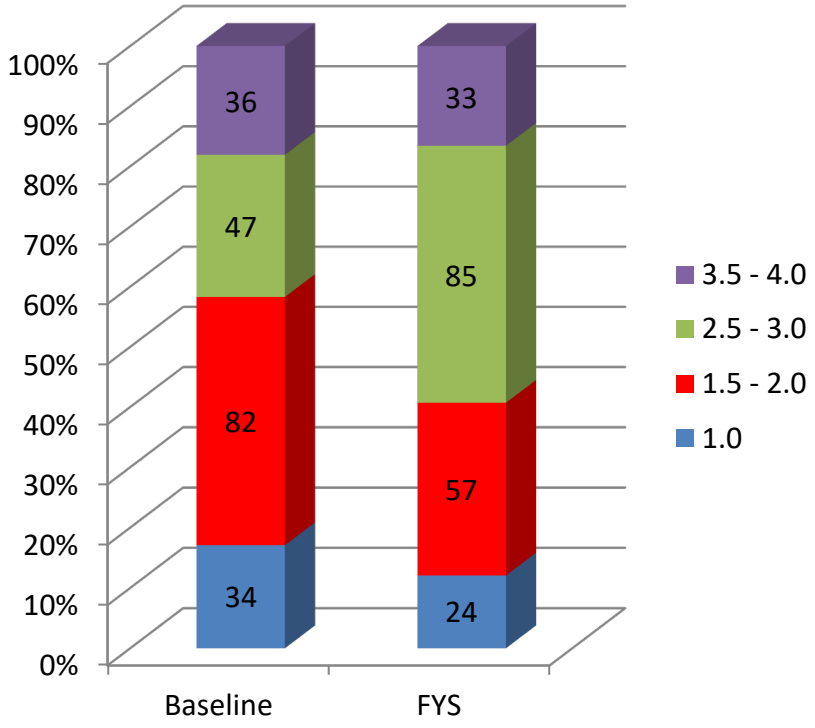
Freshman Baseline/FYS Comparisons

n = 199

Information Needed



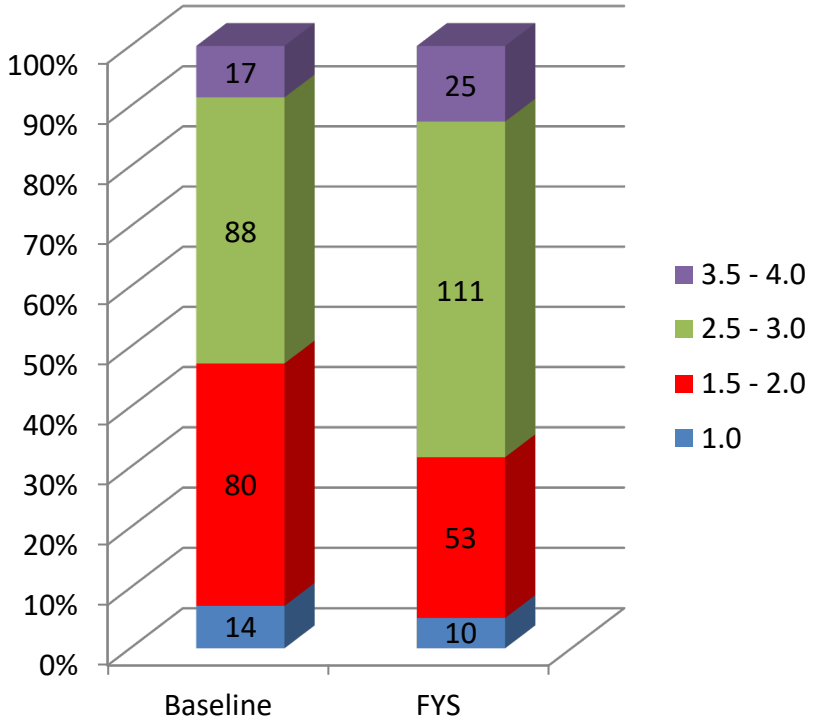
Acknowledgment of Sources



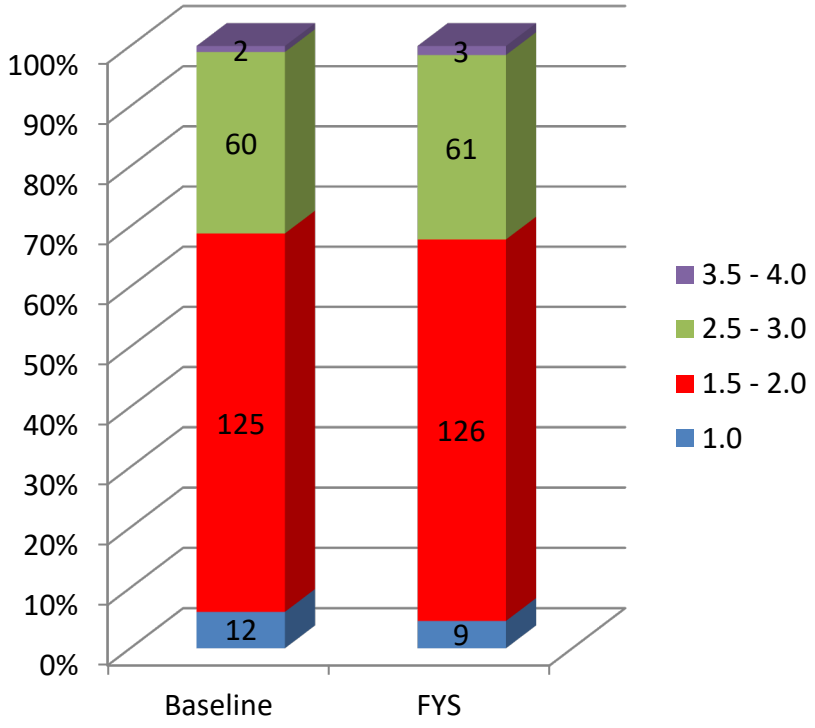
Freshman Baseline/FYS Comparisons

n = 199

Evidence



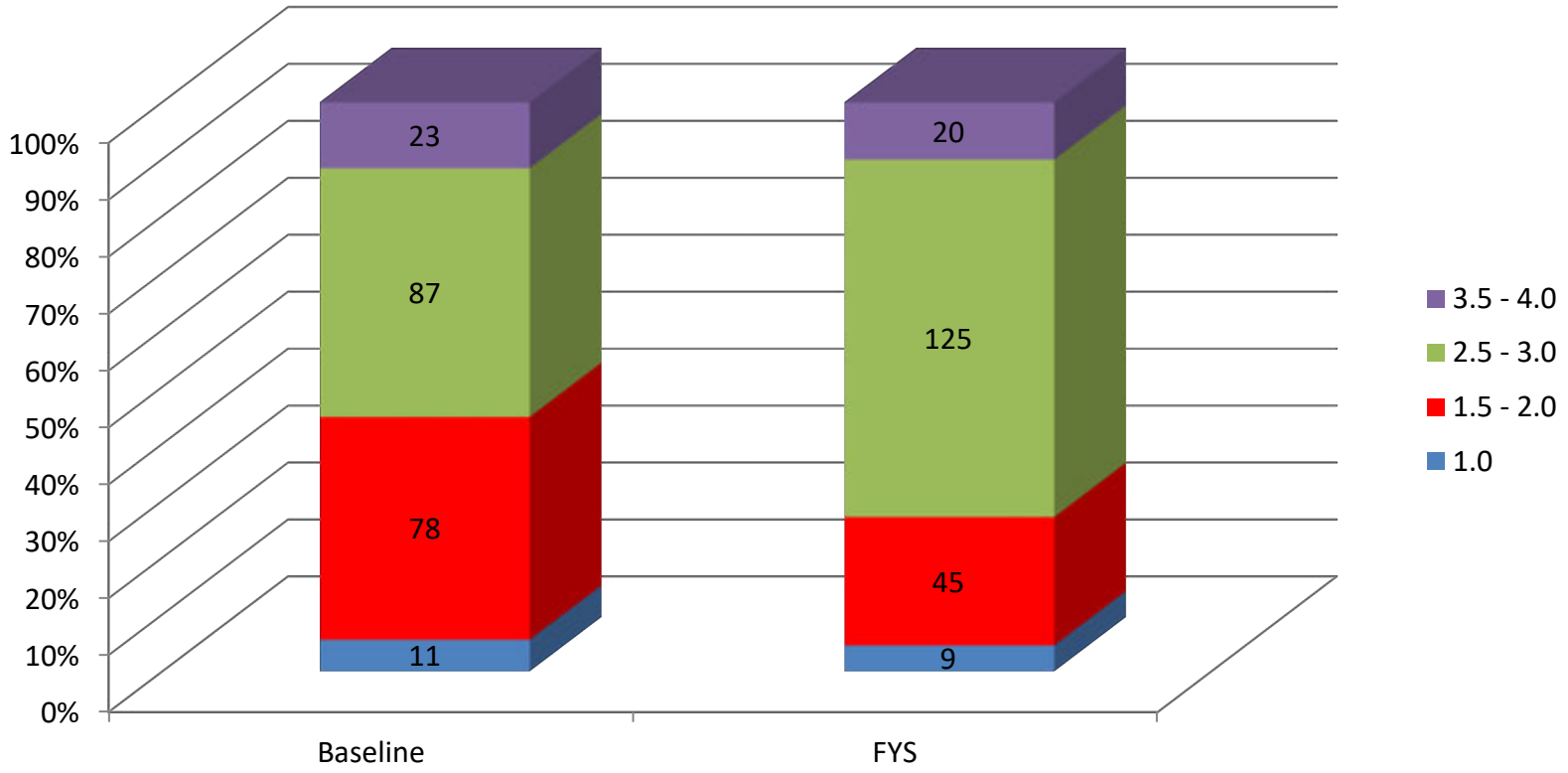
Viewpoints



Freshman Baseline/FYS Comparisons

n = 199

Recommendations



Baseline Inter-Rater Agreement Results

Trait/ Agreement	Info Needed : Cohen's Liberal Kappa = .837	Acknowledgment of Sources: Cohen's Liberal Kappa = .976	Evidence: Cohen's Liberal Kappa = .924	Viewpoints: Cohen's Liberal Kappa = .934	Recommendations: Cohen's Liberal Kappa = .935
Agree on score	119 (60%)	124 (62%)	108 (54%)	111 (56%)	132 (66%)
Difference = 1 point	57 (29%)	71 (36%)	79 (40%)	79 (40%)	57 (29%)
Difference = 2 points	23 (12%)	4 (2%)	12 (6%)	9 (5%)	9 (5%)
Difference = 3 points	0	0	0	0	1 (1%)
Total	**199 (100%)**	**199 (100%)**	**199 (100%)**	**199 (100%)**	**199 (100%)**

FYS Inter-Rater Agreement Results

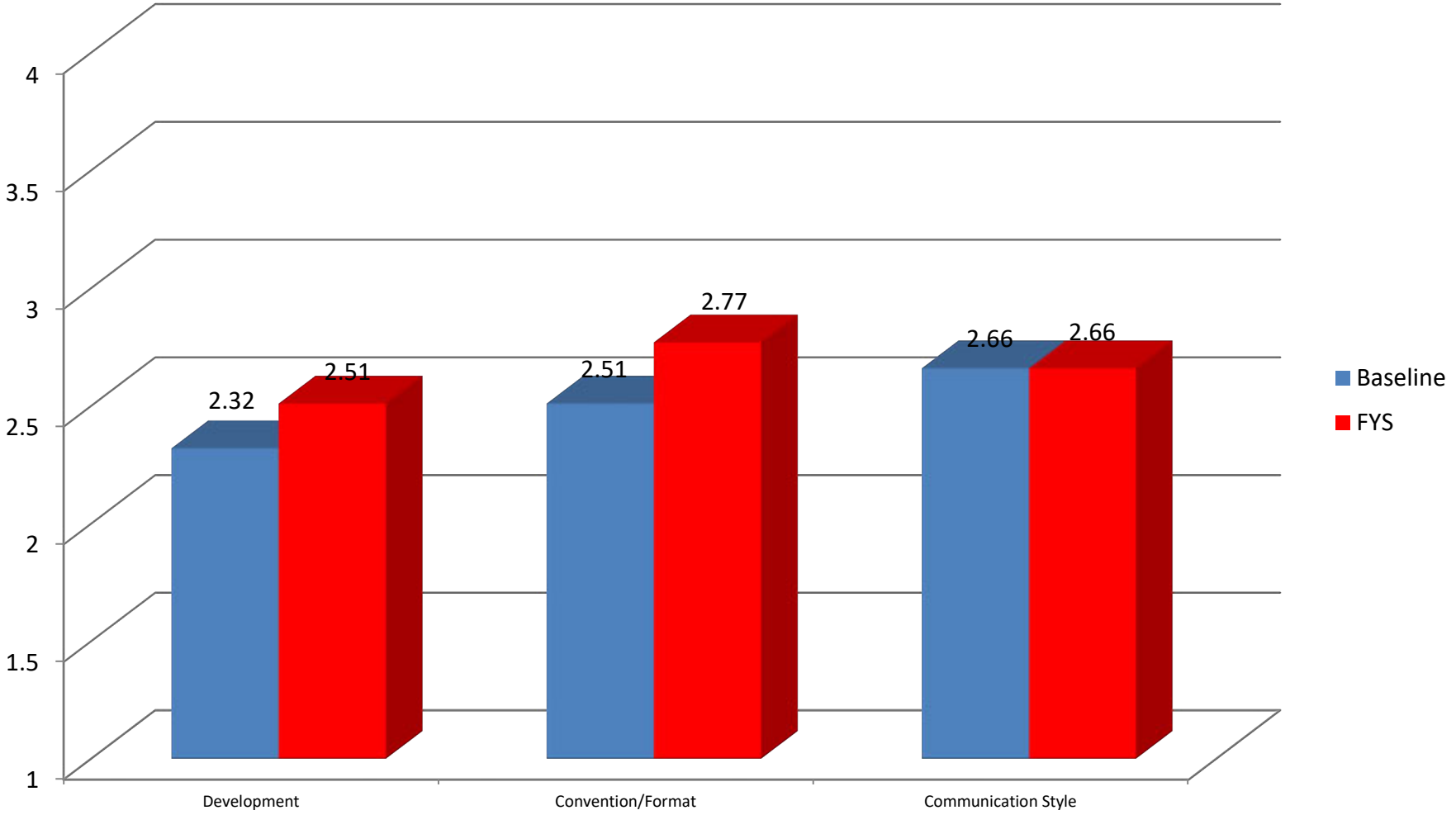
Trait/ Agreement	Info Needed : Cohen's Liberal Kappa = .916	Acknowledgment of Sources: Cohen's Liberal Kappa = .982	Evidence: Cohen's Liberal Kappa = .943	Viewpoints: Cohen's Liberal Kappa = .984	Recommendations: Cohen's Liberal Kappa = .948
Agree on score	111 (55.5%)	128 (64%)	92 (46%)	125 (62.5%)	101 (50.5%)
Difference = 1 point	76 (38%)	69 (34.5%)	99 (49.5%)	73 (36.5%)	91 (45.5%)
Difference = 2 points	13 (6.5%)	3 (1.5%)	8 (4%)	2 (1%)	8 (4%)
Difference = 3 points	0	0	1 (0.5%)	0	0
Total	**200 (100%)**	**200 (100%)**	**200 (100%)**	**200 (100%)**	**200 (100%)**

Freshman Baseline/FYS Comparisons

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

n = 199

Mean differences are statistically significant for *Development* and Convention/Format.



Freshman Baseline/FYS Comparisons

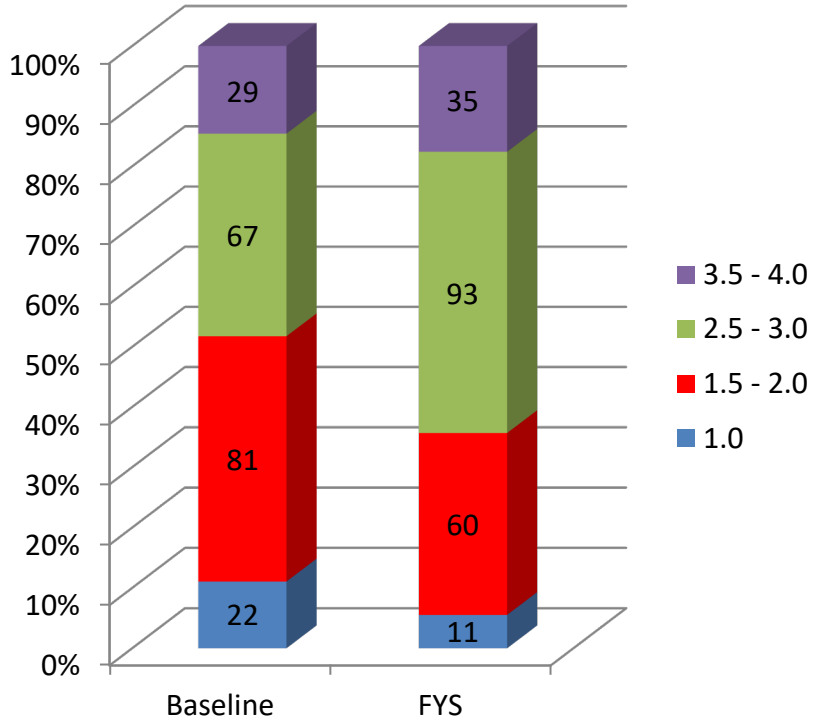
n = 199

Trait/ Performance Level	Development	Convention/Format	Communication Style
1.0 Baseline	22 (11%)	12 (6%)	2 (1%)
1.0 FYS	11 (6%)	12 (6%)	2 (1%)
1.5 – 2.0 Baseline	81 (41%)	68 (34%)	38 (19%)
1.5 – 2.0 FYS	60 (30%)	32 (16%)	42 (21%)
2.5 – 3.0 Baseline	67 (34%)	86 (43%)	141 (71%)
2.5 – 3.0 FYS	93 (47%)	93 (47%)	134 (67%)
3.5 – 4.0 Baseline	29 (15%)	33 (17%)	18 (9%)
3.5 – 4.0 FYS	35 (18%)	62 (31%)	21 (11%)
Grand Total Baseline	**199 (100%)**	**199 (100%)**	**199 (100%)**
Grand Total FYS	**199 (100%)**	**199 (100%)**	**199 (100%)**

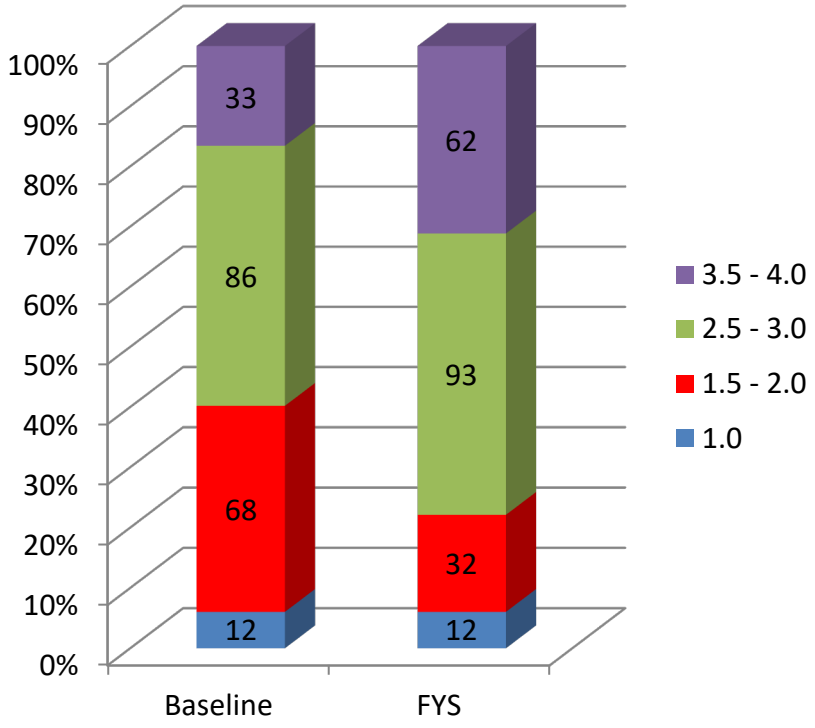
Freshman Baseline/FYS Comparisons

n = 199

Development



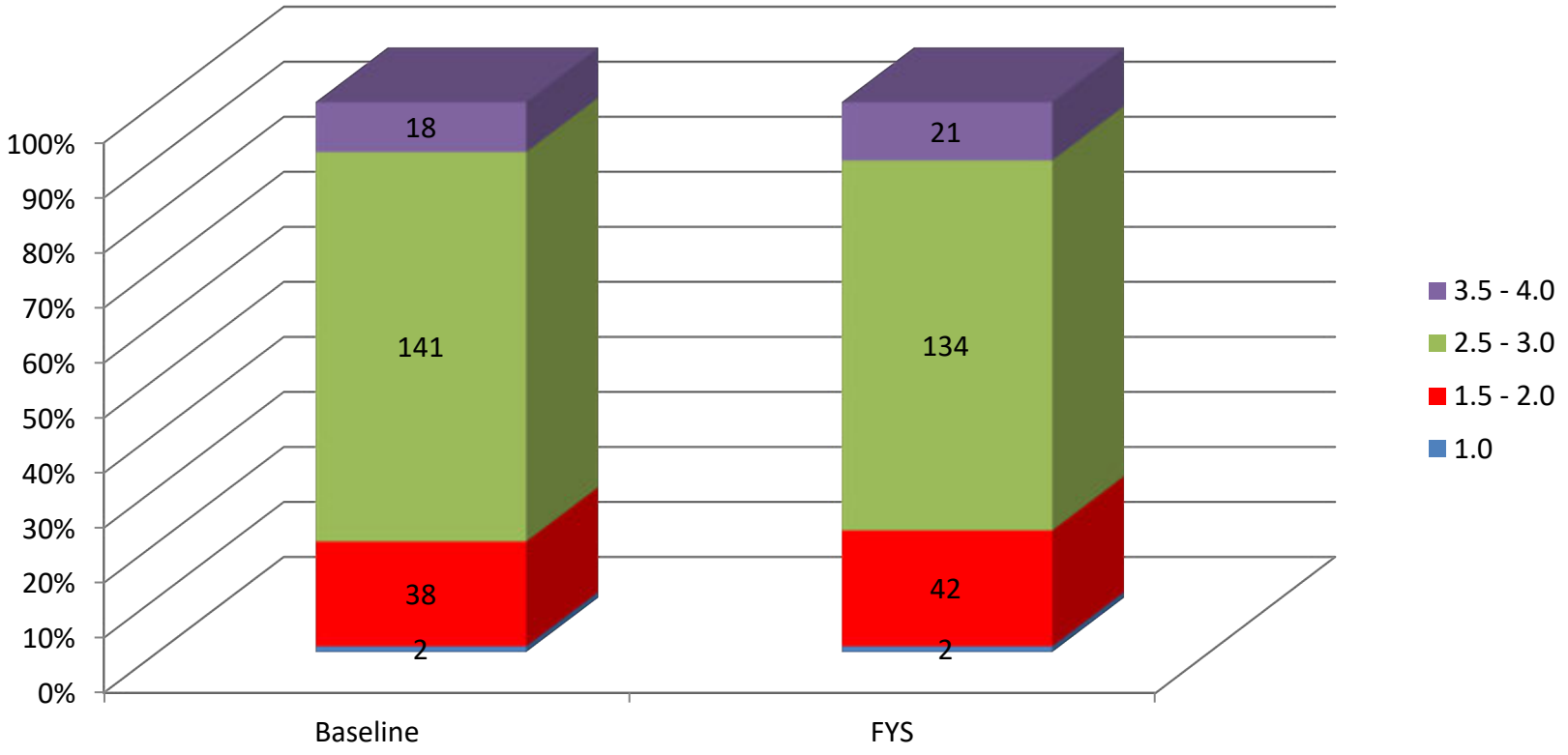
Convention/Format



Freshman Baseline/FYS Comparisons

n = 199

Communication Style



Baseline Inter-Rater Agreement Results

Trait/ Agreement	Development: Cohen's Liberal Kappa = .951	Convention/Format: Cohen's Liberal Kappa = .969	Communication Style: Cohen's Liberal Kappa = .937
Agree on score	114 (57%)	122 (61%)	111 (56%)
Difference = 1 point	77 (39%)	72 (36%)	79 (40%)
Difference = 2 points	8 (4%)	5 (3%)	9 (5%)
Difference = 3 points	0	0	0
Total	**199 (100%)**	**199 (100%)**	**199 (100%)**

FYS Inter-Rater Agreement Results

Trait/ Agreement	Development: Cohen's Liberal Kappa = .957	Convention/Format: Cohen's Liberal Kappa = .861	Communication Style: Cohen's Liberal Kappa = .903
Agree on score	95 (47.5%)	84 (42%)	107 (43.5%)
Difference = 1 point	98 (49%)	93 (46.5%)	79 (39.5%)
Difference = 2 points	6 (3%)	21 (10.5%)	13 (6.5%)
Difference = 3 points	1 (0.5%)	2 (1%)	1 (0.5%)
Total	**200 (100%)**	**200 (100%)**	**200 (100%)**

MARSHALL UNIVERSITY

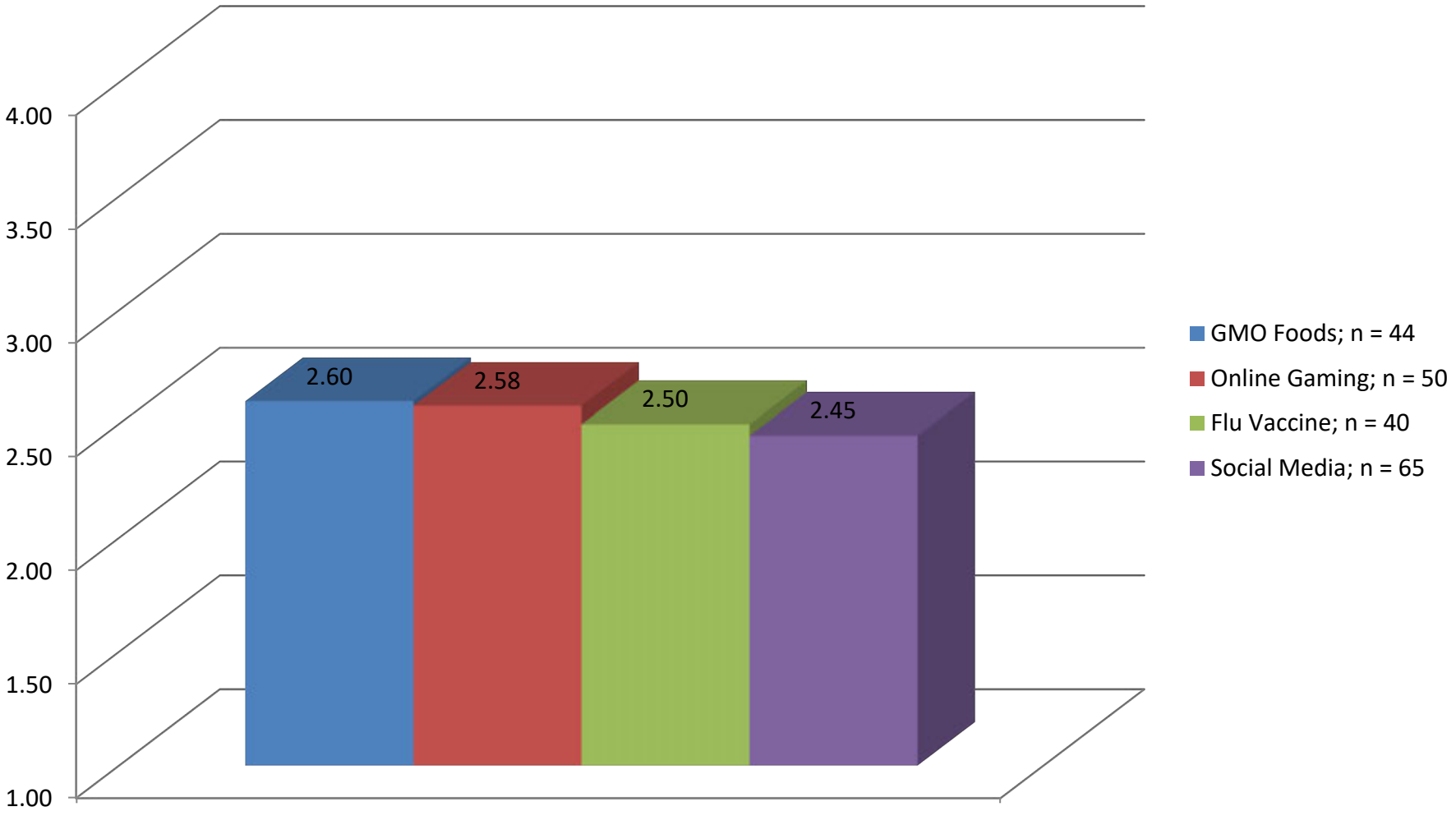
Comparison of FYS Results for Each Trait by Scenario

Academic Year 2020 - 2021

FYS Comparisons by Scenario for IL: Information Needed

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

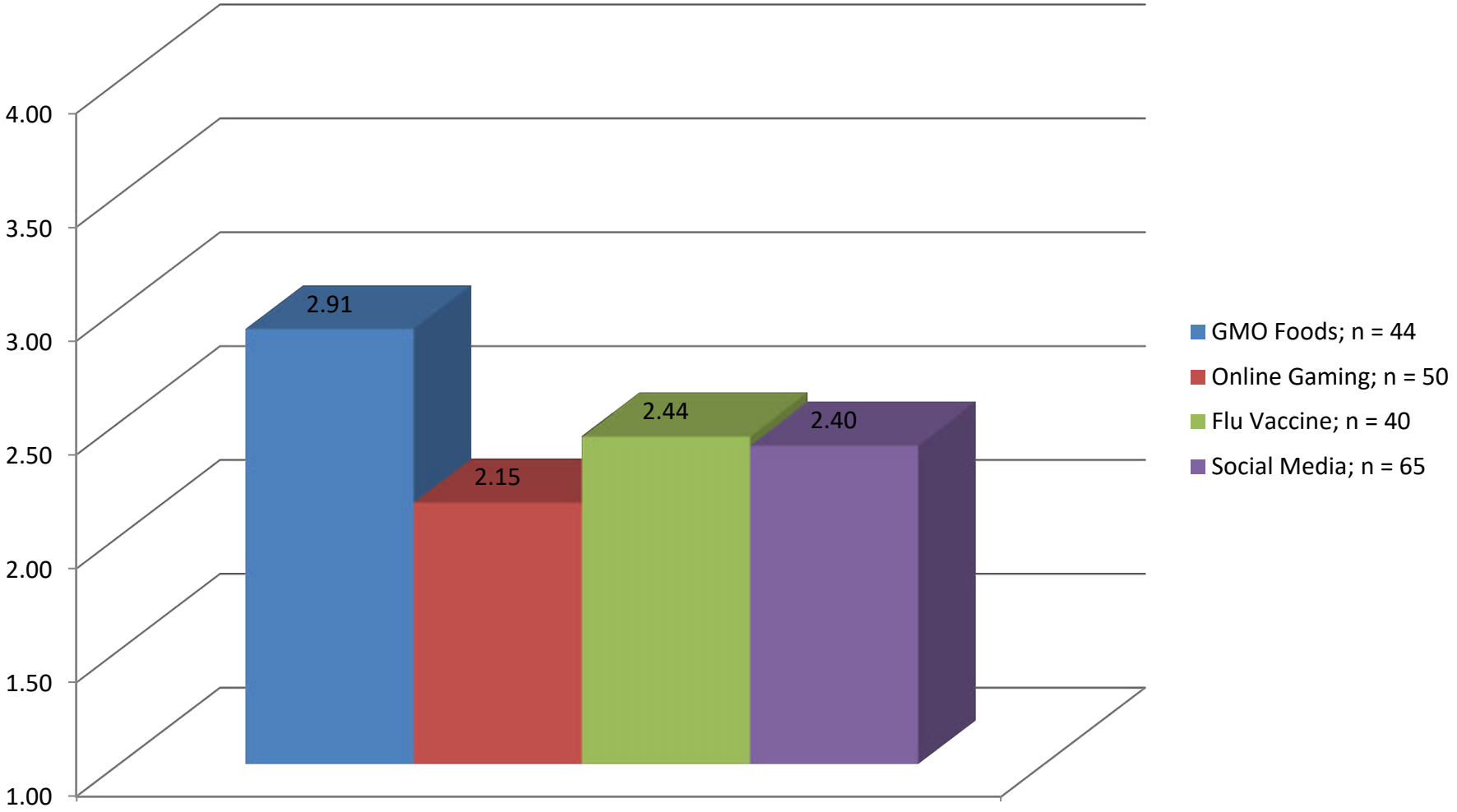
A One-Way ANOVA did not reveal any statistically significant differences in means across the scenarios.



FYS Comparisons by Scenario for IL: Source Acknowledgment

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

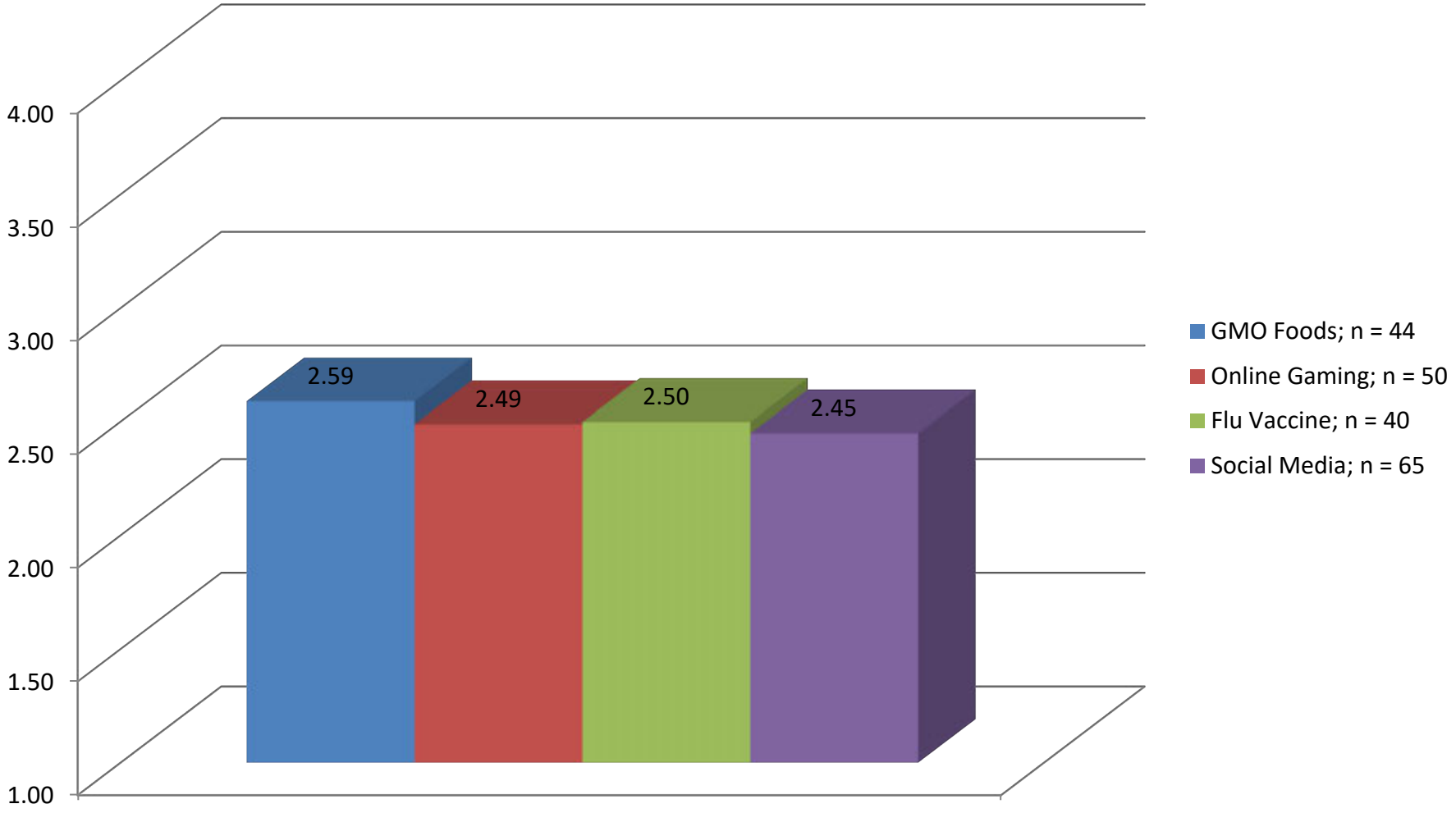
A One-Way ANOVA revealed statistically significant differences across scenarios. Post-Hoc analysis showed that students performed significantly better on GMO Foods than on Online Gaming, Flu Vaccine, and Social Media.



FYS Comparisons by Scenario for BT: Evidence

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

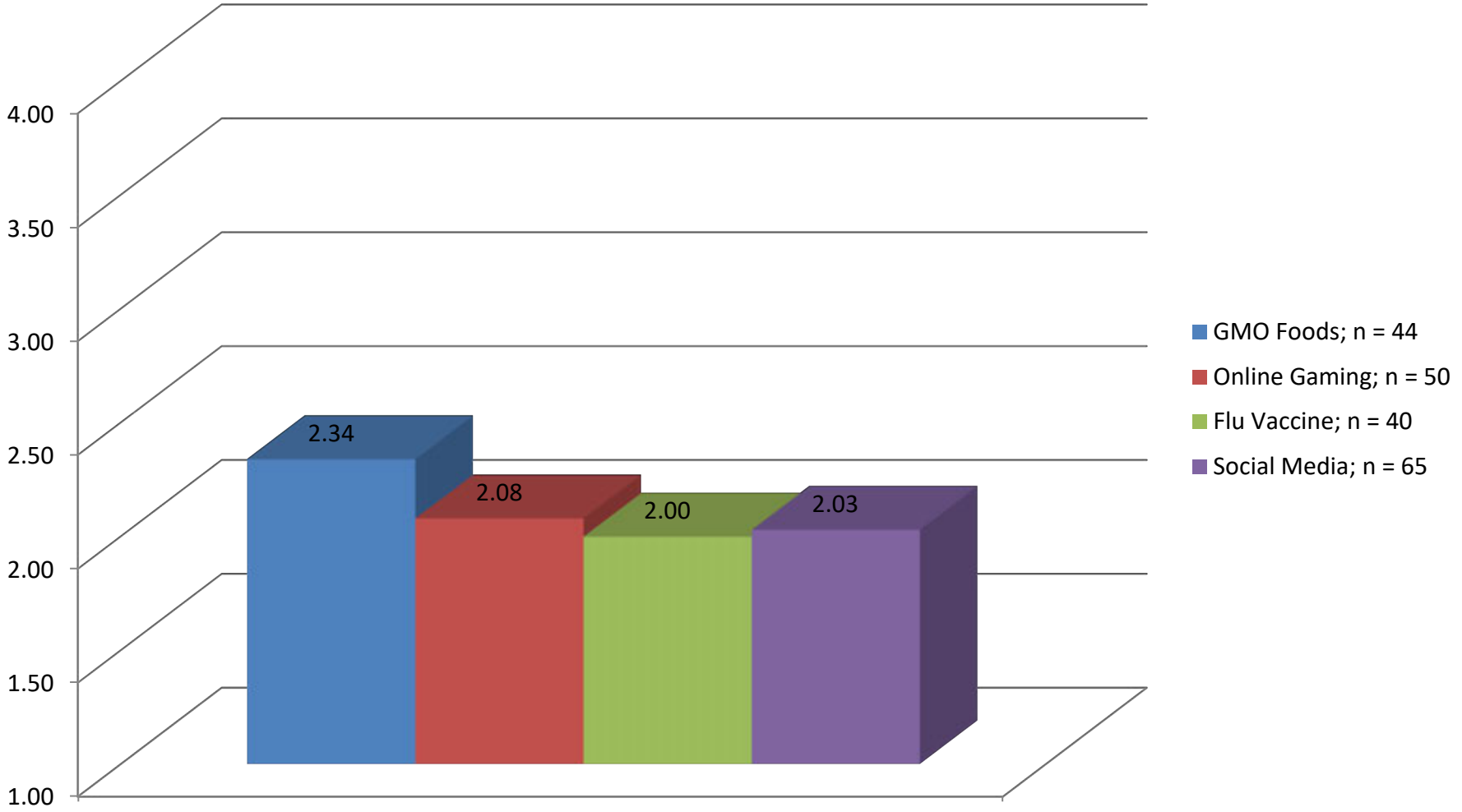
A One-Way ANOVA did not reveal any statistically significant differences in means across the scenarios.



FYS Comparisons by Scenario for IBT: Viewpoints

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

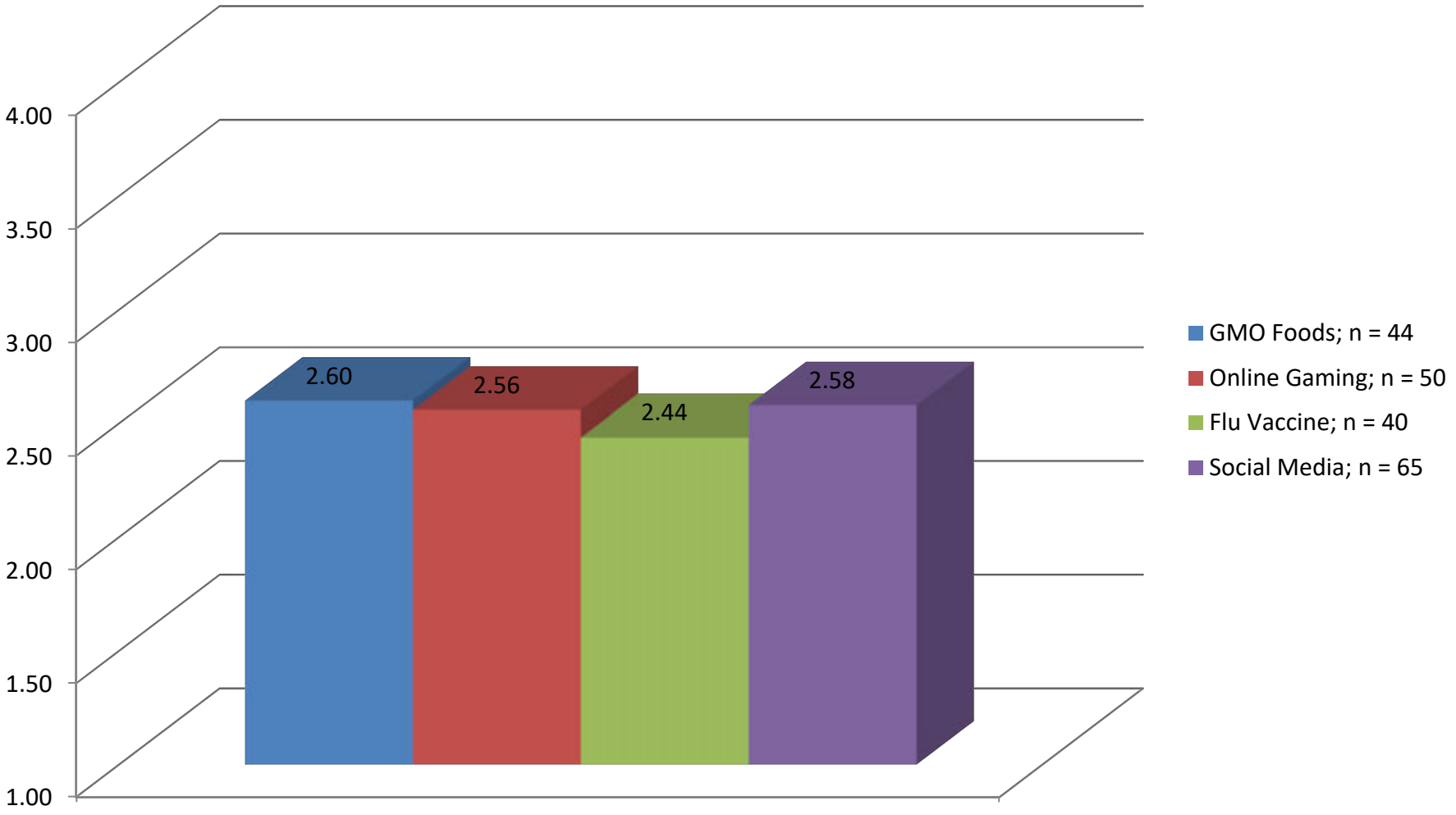
A One-Way ANOVA revealed statistically significant differences across scenarios. Post-Hoc analysis showed that students performed significantly better on GMO Foods than on Online Gaming, Flu Vaccine, and Social Media.



FYS Comparisons by Scenario for IBT: Recommendation/Position

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

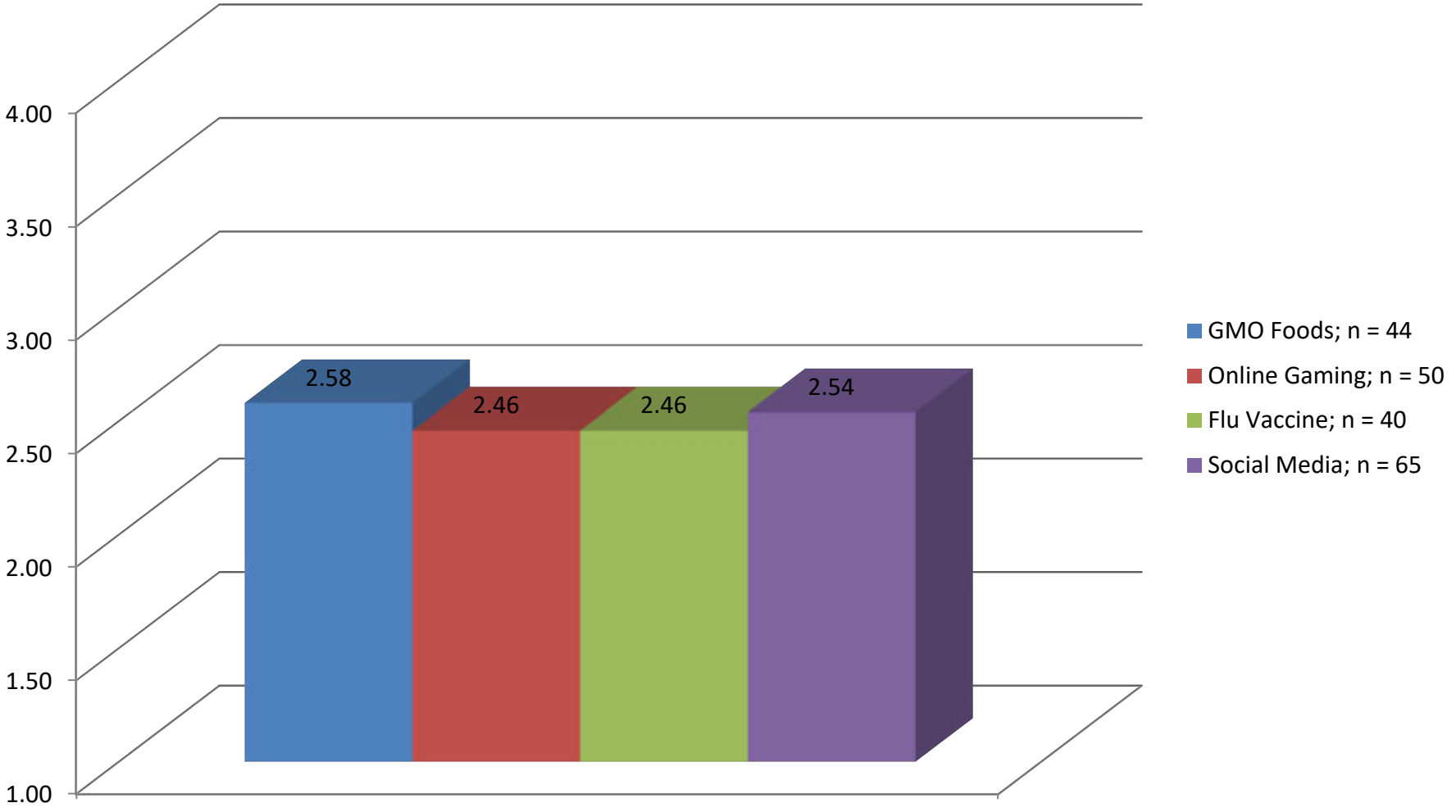
A One-Way ANOVA did not reveal any statistically significant differences in means across the scenarios.



FYS Comparisons by Scenario for CF: Development

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

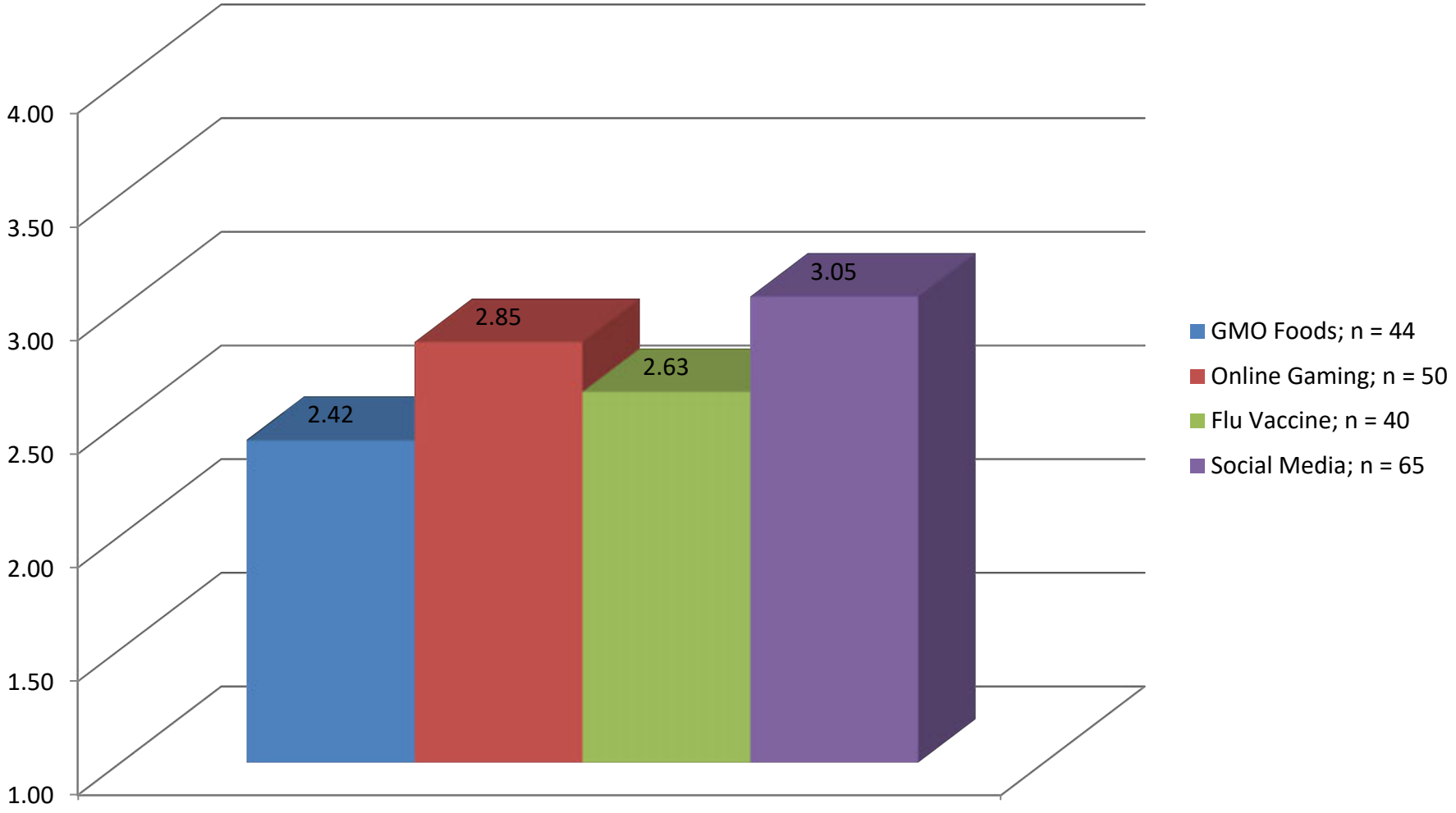
A One-Way ANOVA did not reveal any statistically significant differences in means across the scenarios.



FYS Comparisons by Scenario for CF: Convention/Format

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

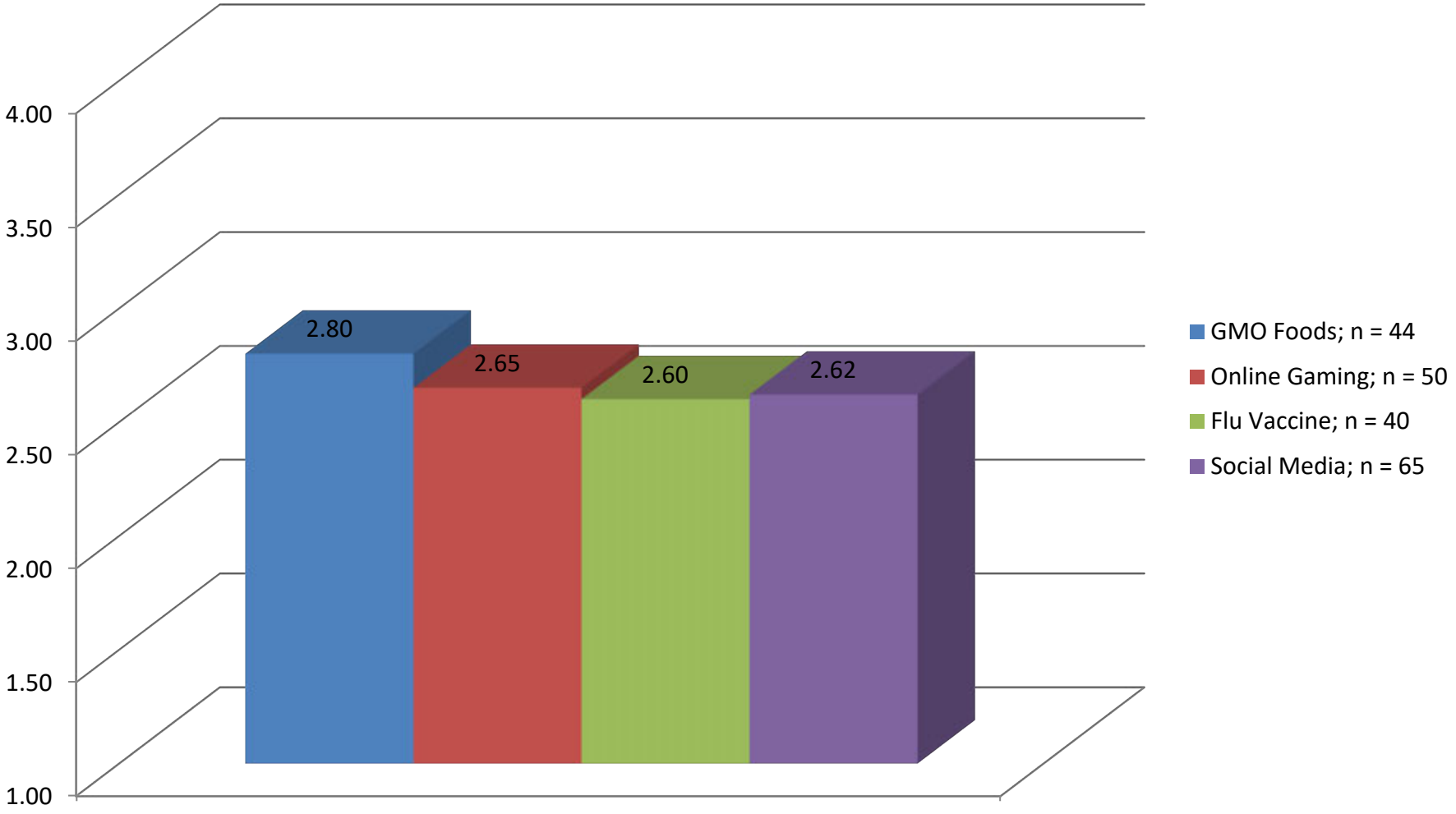
A One-Way ANOVA revealed statistically significant differences across scenarios. Post-Hoc analysis showed that students performed significantly better on Online Gaming and Social Media than on GMO Foods and performed significantly better on the Social Media than on the Flu Vaccine.



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 did not reveal any statistically significant differences in means across the scenarios.



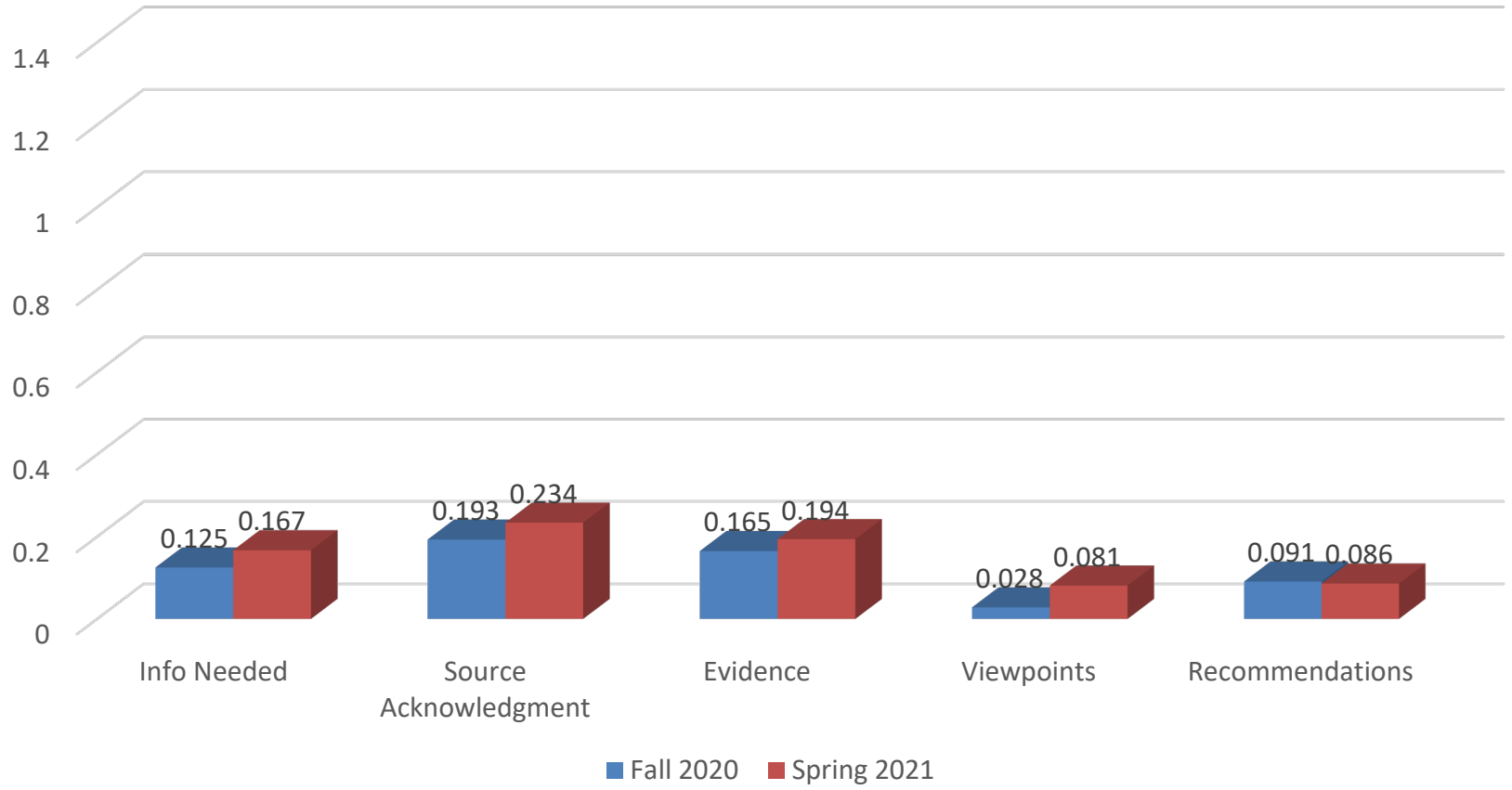
Comparison of Baseline to FYS Mean Gain Score for Each Trait by Semester of FYS

Academic Year 2020 - 2021

Baseline to FYS Mean Gain Scores for Each Trait

n = 88 in fall and 111 in spring

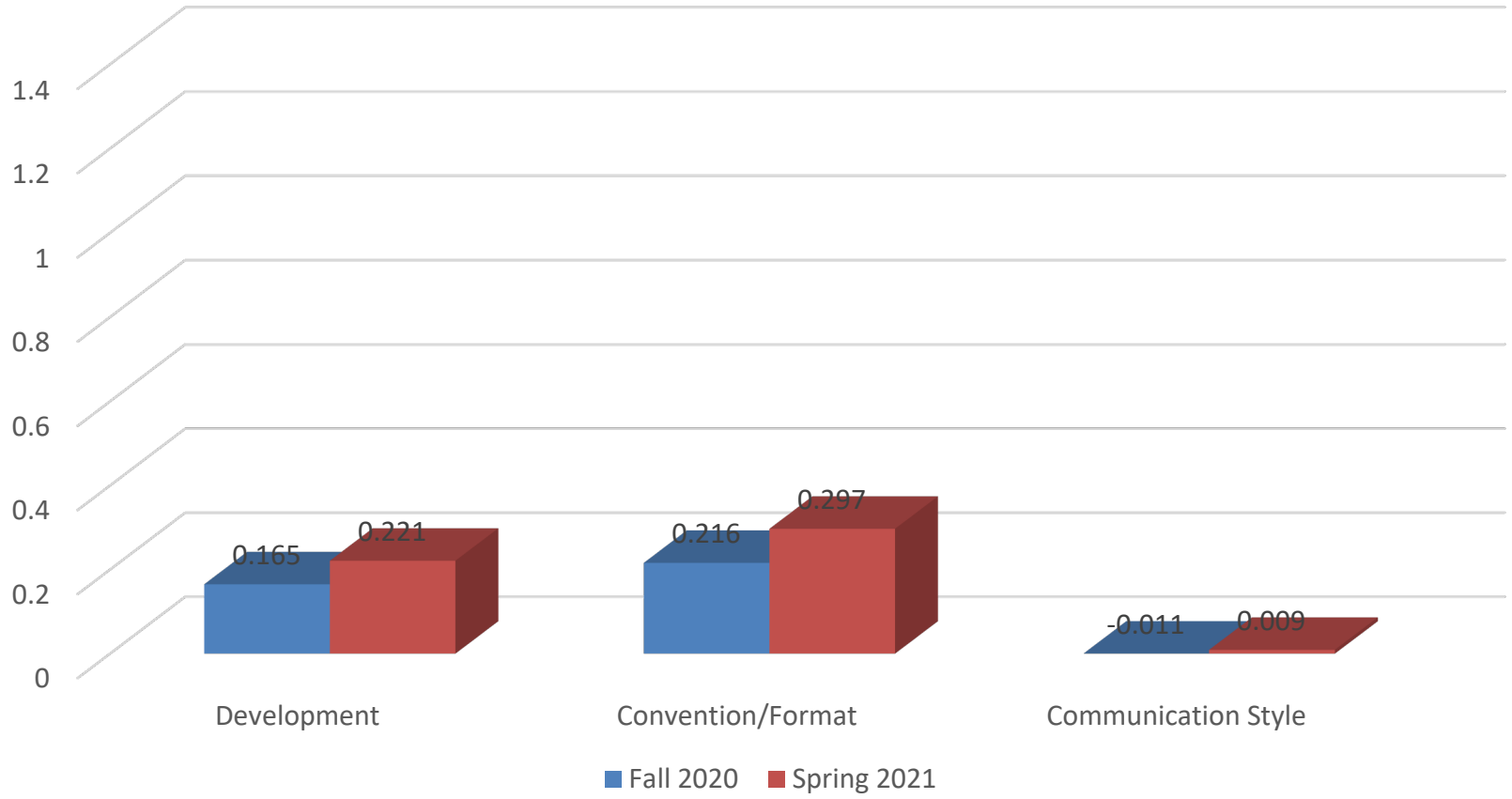
(Differences between fall and spring were not statistically significant)



Baseline to FYS Mean Gain Scores for Each Trait

n = 88 in fall and 111 in spring

(Differences between fall and spring were not statistically significant)



Reference

Stellmack, M.A., Kohneim-Kalkstein, Y. L, Manor, J. E., Massey, A. R., & Schmitz, J. A. P. (2009). An assessment of reliability and validity of a rubric for grading APA-style introductions. *Teaching of Psychology*, *36*, 102-107.