Comparison of Freshman Baseline with First Year Seminar Assessment Results

Academic Year 2021 – 2022

Summer Assessment Team Members: Marie Archambault, Clinton Brown, Kim DeTardo-Bora, Robert Ellison, Victor Fet, Marty Laubach, and Anita Walz

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

Executive Summary

Background

Recommendations from the 2021 Assessment Team (current status is in red)

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. No changes were made regarding assessment for FYS.
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. We will follow up with an additional analysis.
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. We did not include this information in the Assessment Newsletter this past year; we will make every effort to do so in academic year 2022-2023.

Procedures for the 2022 Assessment

General Procedures

In August 2021, 1,476 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*. As part of Marshall's mandatory First Year Seminar in Critical Thinking (FYS), students completed assessments that mirrored those they finished as incoming freshmen, with 949 FYS assessments uploaded into Blackboard. To obtain a sample of matched pairs of baseline and FYS assessments, we began by collecting a random sample of 500 FYS assessments. We then matched the students who completed these assessments with their baseline assessments. This process yielded a total of 363 matched pairs. From these matches, 188 were randomly discarded to yield a sample of 175 baseline and matching FYS assessments. Please note that our sample represented 12% of uploaded baseline and 18% of uploaded FYS assessments. During the Assessment Team's review, we discovered that thirteen baseline artifacts from our sample were blank, two would not open, and two more included only a response that aligned to the "Information Needed" rubric trait. An additional two FYS artifacts were blank. This reduced the usable number of matched pairs to 158 for "Information Needed" and 156 for all other rubric traits.

In May 2022, a group of seven 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 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 all traits (evidence, viewpoints, and recommendation/position) of *Inquiry-Based [Critical] Thinking*, and for all traits (development, convention/format, and communication style) of *Communication Fluency*. 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**	2.108 (0.5625)	2.351 (0.6666)	*t(157)* = -4.349, *p* < .001
	Source Acknowledgment	2.029 (0.8313)	2.279 (0.7476)	*t(155)* = -3.324, *p* = .001
Inquiry-Based (Critical) Thinking	**Evidence**	2.141 (0.6282)	2.365 (0.6282)	*t(155)* = -3.522, *p* = .001
	Viewpoints	1.962 (0.5780)	2.112 (0.5223)	*t(155)* = -2.701, *p* = .008
	Recommendation/Position	2.317 (0.6478)	2.519 (0.6114)	*t(155)* = -3.198, *p* = .002
Communication Fluency	**Development**	2.199 (0.7087)	2.452 (0.7552)	*t(155)* = -3.784, *p* < .001
	Convention/Format	2.407 (0.6619)	2.683 (0.8193)	*t(155)* = -3.986, *p* < .001
	Communication Style	2.587 (0.5272)	2.728 (0.4695)	*t(155)* = -2.960, *p* = .004

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**	10%
	Source Acknowledgment	17%
Inquiry-Based (Critical) Thinking	**Evidence**	18%
	Viewpoints	9%
	Recommendation/Position	8%
Communication Fluency	**Development**	15%
	Convention/Format	9%
	Communication Style	11%

This year's results showed a significant difference in performance based on scenario used for the FYS assessments for two traits (evidence and recommendation/position) of *Inquiry-Based [Critical] Thinking*, and for all traits (development, convention/format, and communication style) of *Communication Fluency*. For evidence, and viewpoints students scored significantly lower on GMO Foods than on the Online Gaming and on Social Media. On development and convention/format, students scored significantly lower on GMO Foods than on the other three scenarios (Online Gaming, Flu Vaccine, and Social Media). On communication style, students scored significantly lower on GMO Foods than on Online Gaming and Flu Vaccine. Also, gain scores between students in our sample who completed FYS in fall 2021 (*n* = 62) and those who completed FYS in spring 2022 (*n* = 95) differed significantly on only one outcome trait, *Communication Fluency* (convention/format), with students enrolled in the spring (mean gain = .432) outperforming students enrolled in the fall (mean gain = .025), *t* (110.183) = -2.806; *p* = .006. Again, we note that *Communication Fluency* is not an outcome of FYS. 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 (in this year's sample, in all) traits of *Information Literacy* and *Critical Thinking* skills between matriculation and the completion of First Year Seminar in Critical Thinking. As was the case this year, in 2019 and 2020 students' improvement reached statistical significance for all traits of both outcomes.

Recommendations from the 2022 Assessment Team

The Summer Assessment Team made the following recommendations:

1. That evidence documents attached to FYS scenarios be evaluated for equivalence of type, length, and complexity across scenarios.
2. That creators of baseline and FYS scenarios consider validated scales that students could use when assessing documents for creditability and relevance.
3. That students be asked to provide a two-sentence summary regarding why they have judged the credibility and relevance of each document as they have.
4. That the Summer Assessment Team review the current rubric before starting the assessment in summer 2023.
5. That we include a more comprehensive evaluation of information literacy in our ratings, e.g., if students say they're using peer-reviewed journals, note where that would that fall on the rubric scale.
6. That FYS instructors consider having students use the same convention/format to make their recommendations. This would place all students on the same playing field for achievement on *Communication Fluency*: convention/format; however, it is worth noting that *Communication Fluency* is not one of the outcomes of FYS, and we realize that FYS instructors might have a pedagogical reason for using different convention/formats for FYS scenarios.
7. That we consider using the same rubrics for baseline, FYS, and senior capstone projects.
8. As was recommended last year, 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.

MARSHALL UNIVERSITY®

Supporting Documentation

MARSHALL UNIVERSITY®

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

Academic Year 2021 - 2022

Review Procedures

- One hundred seventy-five (175) FYS critical thinking artifacts were matched with 175 baseline critical thinking artifacts. This number represented 18% of the 949 FYS artifacts and 12% of the 1,476 baseline artifacts uploaded to Blackboard.

- During the evaluation we discovered that thirteen baseline artifacts from our sample were blank, two would not open, and two more included only a response that aligned to the information needed rubric trait. An additional two FYS artifacts were blank. This reduced the usable number of matched pairs to 158 for information needed and 156 for all other rubric traits.

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

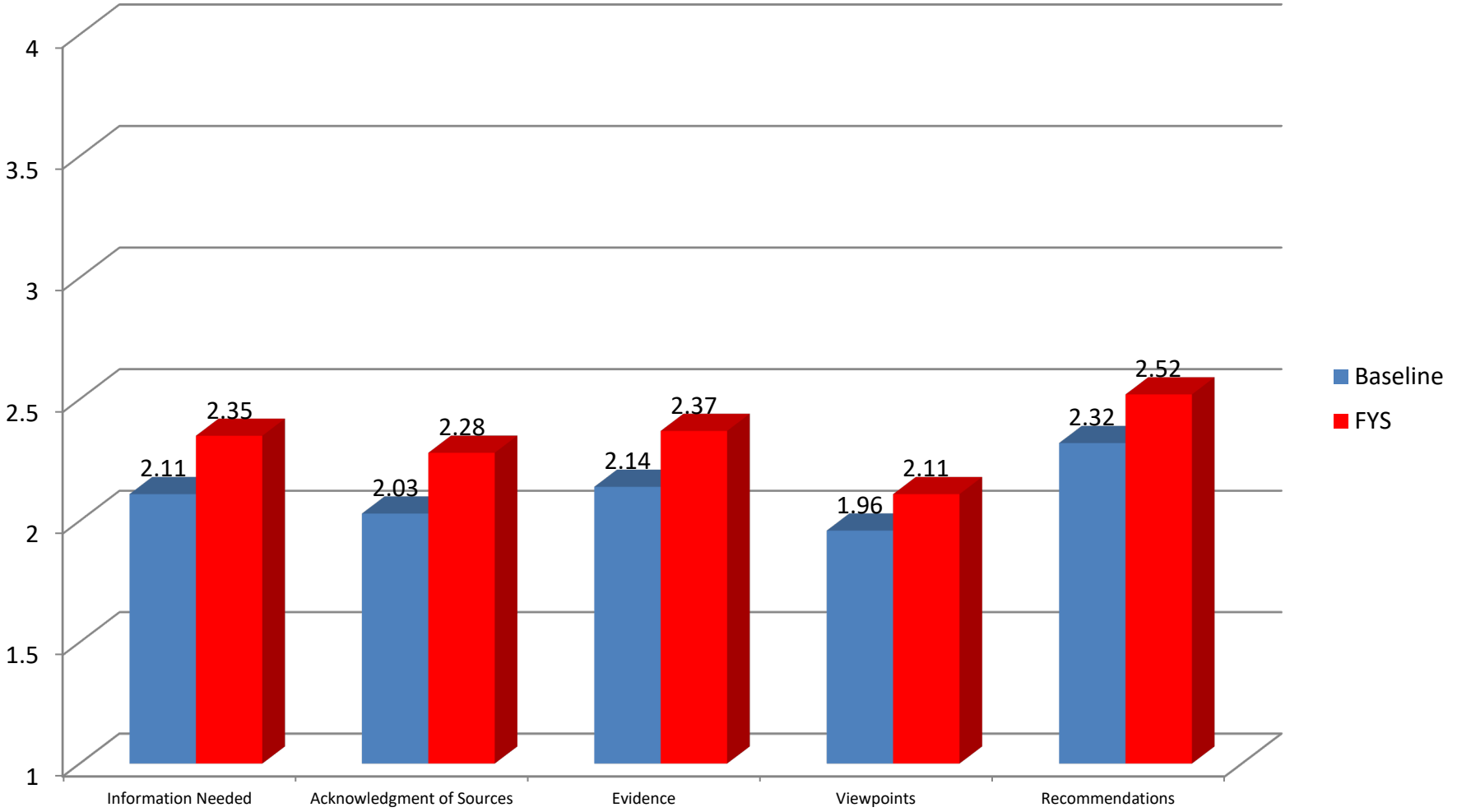
Outcomes	Traits	Performance Levels			
		1	2	3	4
Information Literacy	**Information Needed**	Does not acknowledge or assess the need for more information.	Acknowledges the need for more information but does not identify research methods/sources (or those identified are not feasible) that would address 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 = 158 (Information Needed); *n* = 156 (All other Traits)

Mean differences were statistically significant for *all traits*



Freshman Baseline/FYS Comparisons

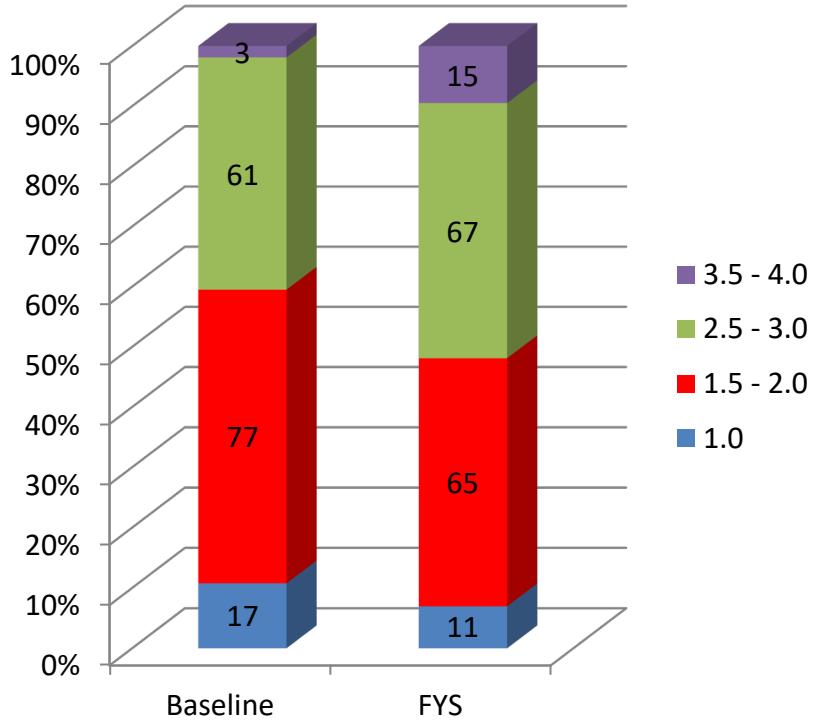
n = 158 (Information Needed); *n* = 156 (All Other Traits)

Trait/ Performance Level	Info Needed	Acknowledgment of Sources	Evidence	Viewpoints	Recommendations
1.0 Baseline	17 (11%)	37 (24%)	18 (12%)	19 (12%)	11 (7%)
1.0 FYS	11 (7%)	22 (14%)	7 (4%)	10 (6%)	4 (3%)
1.5 – 2.0 Baseline	77 (49%)	60 (38%)	70 (45%)	92 (59%)	49 (31%)
1.5 – 2.0 FYS	65 (41%)	48 (31%)	54 (35%)	87 (56%)	42 (27%)
2.5 – 3.0 Baseline	61 (39%)	50 (32%)	61 (39%)	43 (28%)	92 (59%)
2.5 – 3.0 FYS	67 (42%)	81 (52%)	86 (55%)	56 (36%)	99 (63%)
3.5 – 4.0 Baseline	3 (2%)	9 (6%)	7 (4%)	2 (1%)	4 (3%)
3.5 – 4.0 FYS	15 (9%)	5 (3%)	9 (6%)	3 (2%)	11 (7%)
Grand Total Baseline	**158 (100%)**	**156 (100%)**	**156 (100%)**	**156 (100%)**	**156 (100%)**
Grand Total FYS	**158 (100%)**	**156 (100%)**	**156 (100%)**	**156 (100%)**	**156 (100%)**

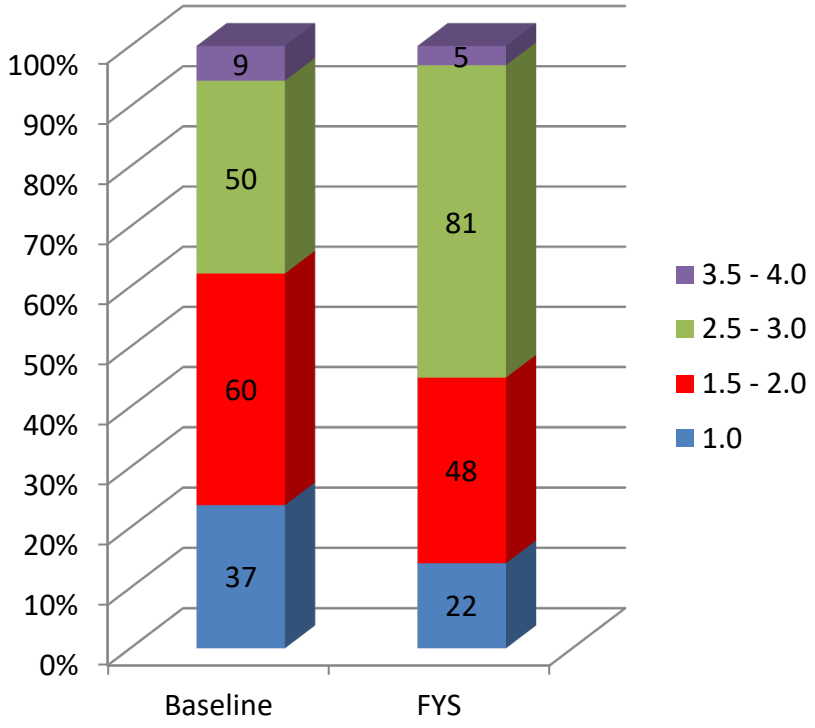
Freshman Baseline/FYS Comparisons

n = 158 (Information Needed); *n* = 156 (Acknowledgment of Sources)

Information Needed



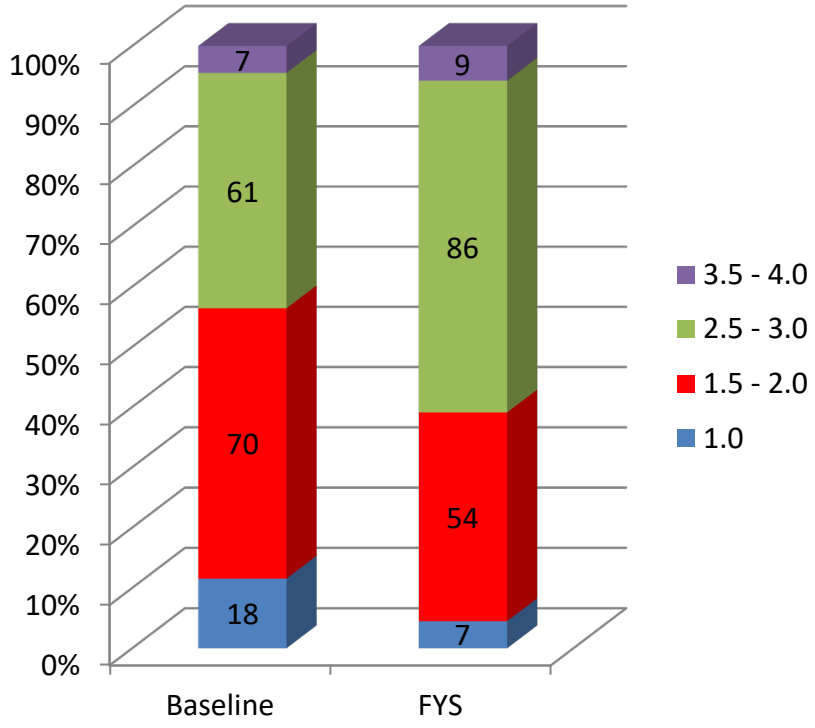
Acknowledgment of Sources



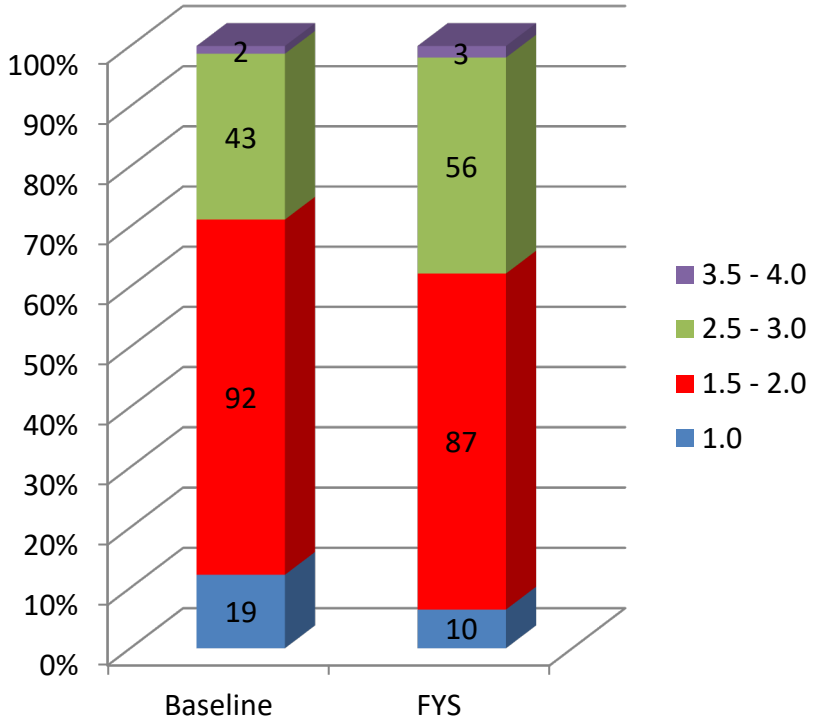
Freshman Baseline/FYS Comparisons

n = 156

Evidence



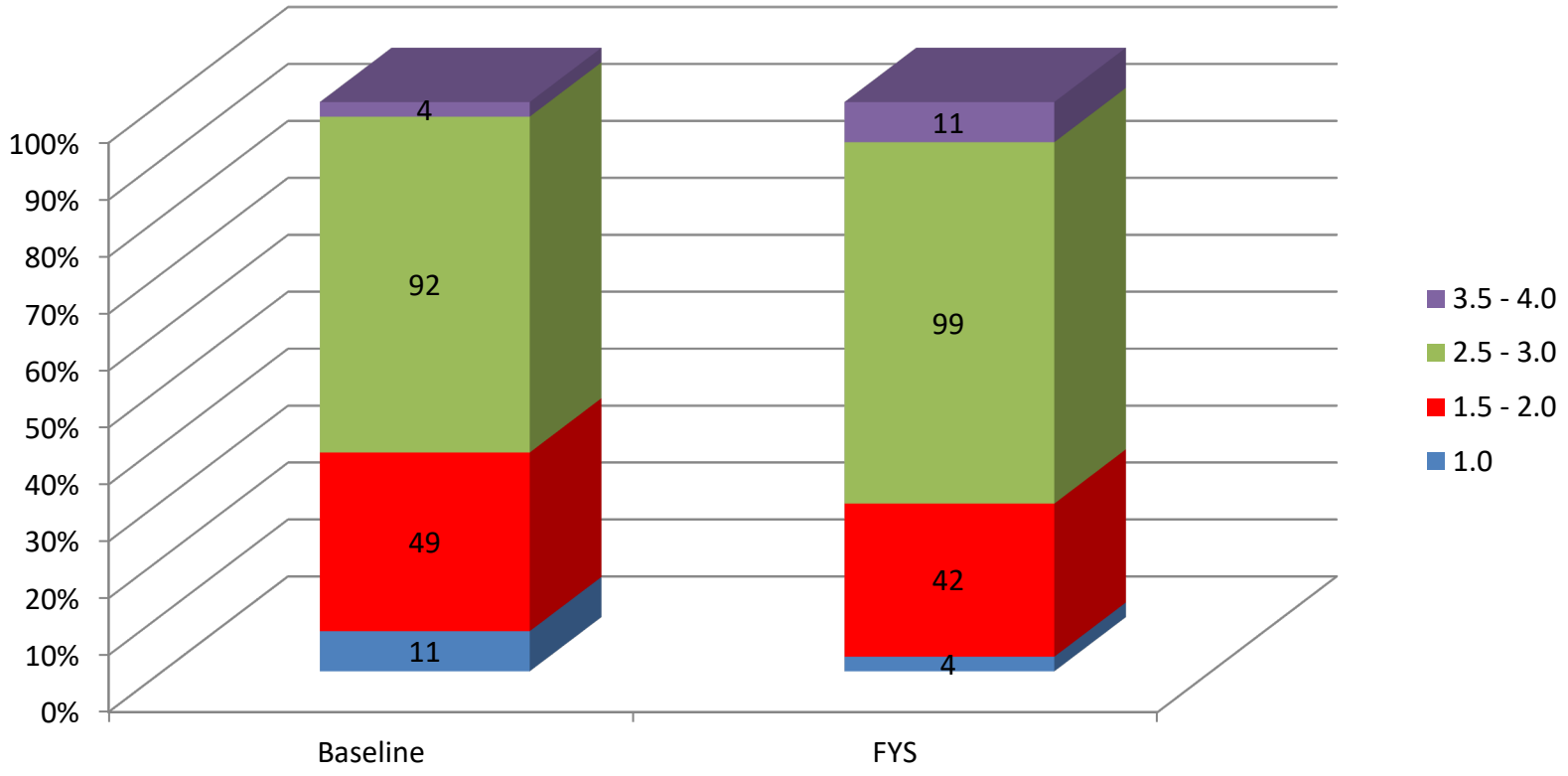
Viewpoints



Freshman Baseline/FYS Comparisons

n = 156

Recommendations



Baseline Inter-Rater Agreement Results

Trait/ Agreement	Info Needed : Cohen's Kappa (Liberal) = .958	Acknowledgment of Sources: Cohen's Kappa (Liberal) = .985	Evidence: Cohen's Kappa (Liberal) = .932	Viewpoints: Cohen's Kappa (Liberal) = .984	Recommendations: Cohen's Kappa (Liberal) = .921
Agree on score	89 (56%)	103 (65%)	83 (53%)	88 (56%)	74 (47%)
Difference = 1 point	66 (41%)	53 (34%)	66 (42%)	68 (43%)	74 (47%)
Difference = 2 points	5 (3%)	2 (1%)	9 (6%)	2 (1%)	9 (6%)
Difference = 3 points	0	0	0	0	1 (1%)
Total	**160 (100%)**	**158 (100%)**	**158 (100%)**	**158 (100%)**	**158 (100%)**

FYS Inter-Rater Agreement Results

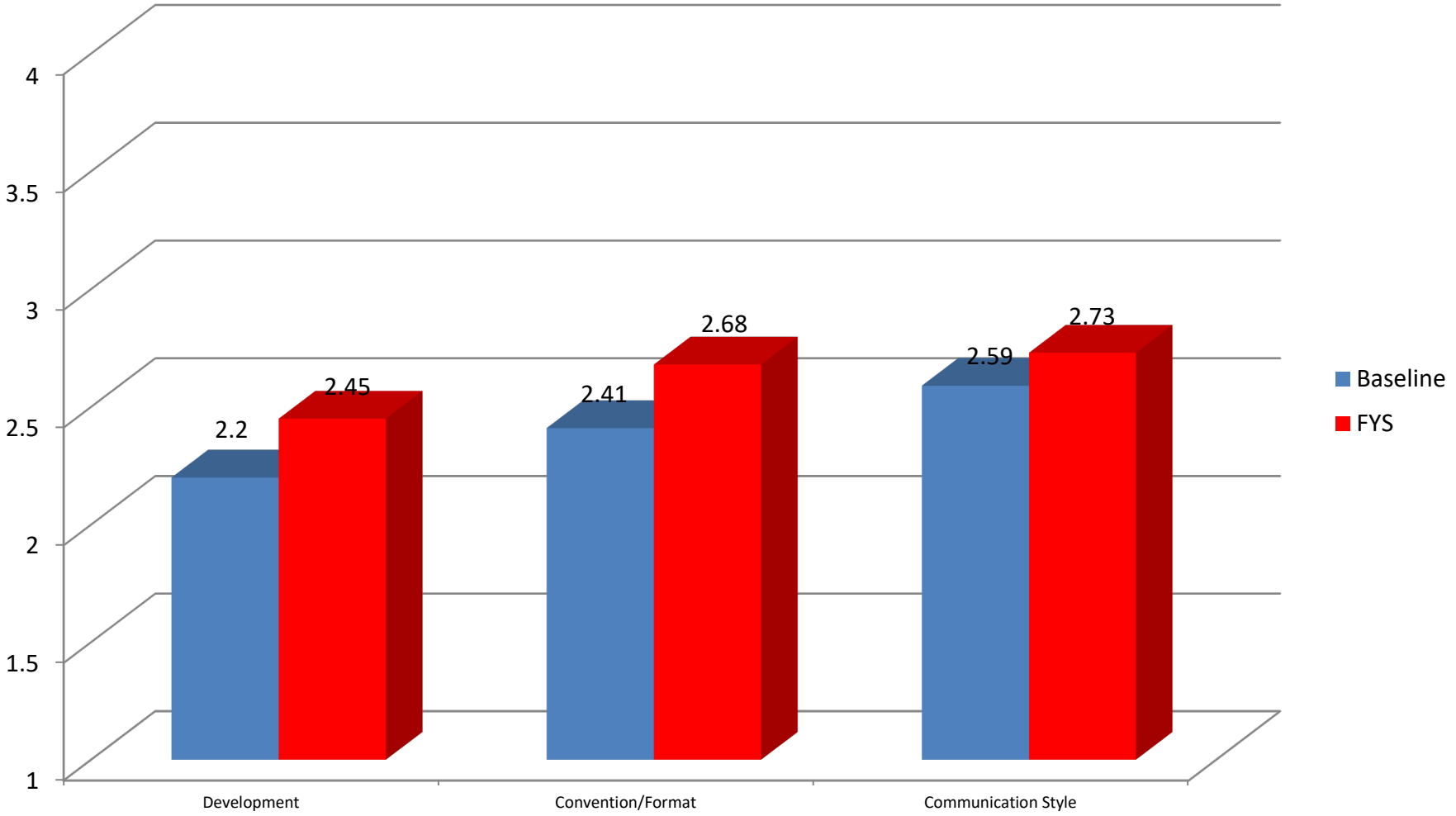
Trait/ Agreement	Info Needed : Cohen's Kappa (Liberal) = .963	Acknowledgment of Sources: Cohen's Kappa (Liberal) = .985	Evidence: Cohen's Kappa (Liberal) = .912	Viewpoints: Cohen's Kappa (Liberal) = .952	Recommendations: Cohen's Kappa (Liberal) = .931
Agree on score	101 (58%)	113 (65%)	77 (45%)	89 (51%)	86 (50%)
Difference = 1 point	67 (39%)	58 (34%)	84 (49%)	78 (45%)	78 (45%)
Difference = 2 points	5 (3%)	2 (1%)	12 (7%)	6 (3%)	9 (5%)
Difference = 3 points	0	0	0	0	0
Total	**173 (100%)**	**173 (100%)**	**173 (100%)**	**173 (100%)**	**173 (100%)**

Freshman Baseline/FYS Comparisons

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

n = 156

Mean differences were statistically significant for *all traits*



Freshman Baseline/FYS Comparisons

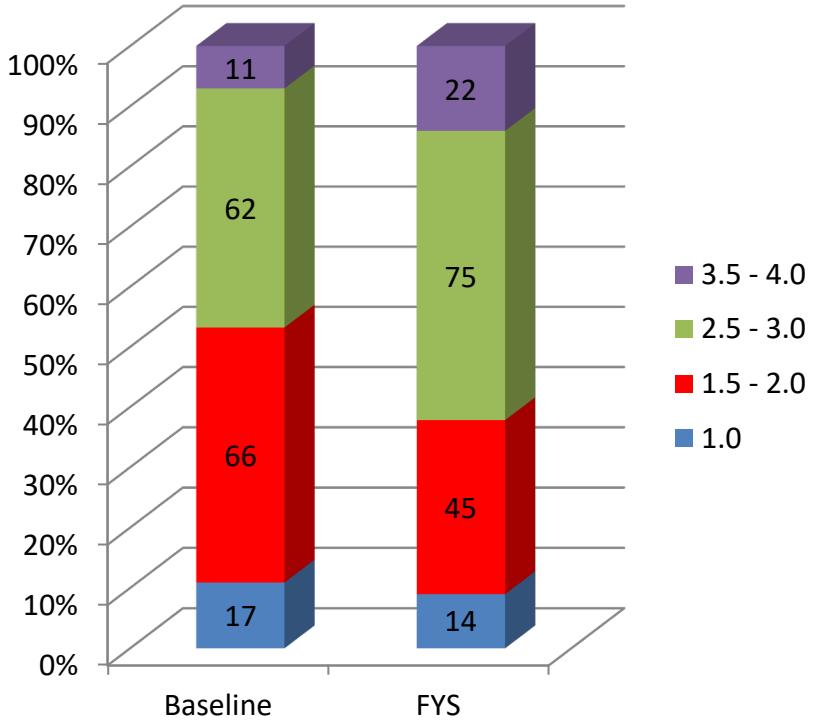
n = 156

Trait/ Performance Level	Development	Convention/Format	Communication Style
1.0 Baseline	17 (11%)	8 (5%)	2 (1%)
1.0 FYS	14 (9%)	15 (10%)	0
1.5 – 2.0 Baseline	66 (42%)	44 (28%)	36 (23%)
1.5 – 2.0 FYS	45 (29%)	24 (15%)	22 (14%)
2.5 – 3.0 Baseline	62 (40%)	91 (58%)	111 (71%)
2.5 – 3.0 FYS	75 (48%)	81 (52%)	123 (79%)
3.5 – 4.0 Baseline	11 (7%)	13 (8%)	7 (4%)
3.5 – 4.0 FYS	22 (14%)	36 (23%)	11 (7%)
Grand Total Baseline	**156 (100%)**	**156 (100%)**	**156 (100%)**
Grand Total FYS	**156 (100%)**	**156 (100%)**	**156 (100%)**

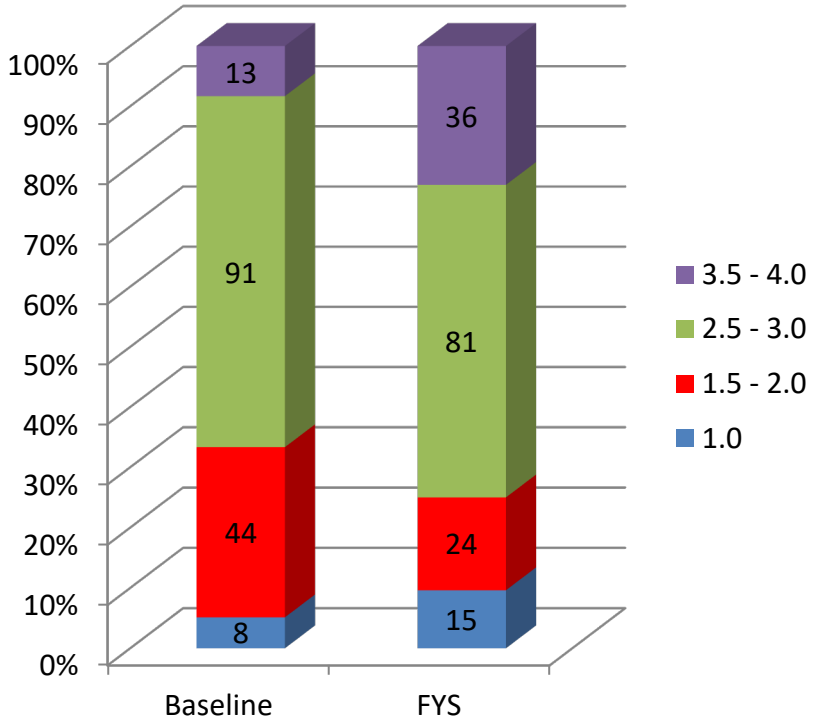
Freshman Baseline/FYS Comparisons

n = 156

Development



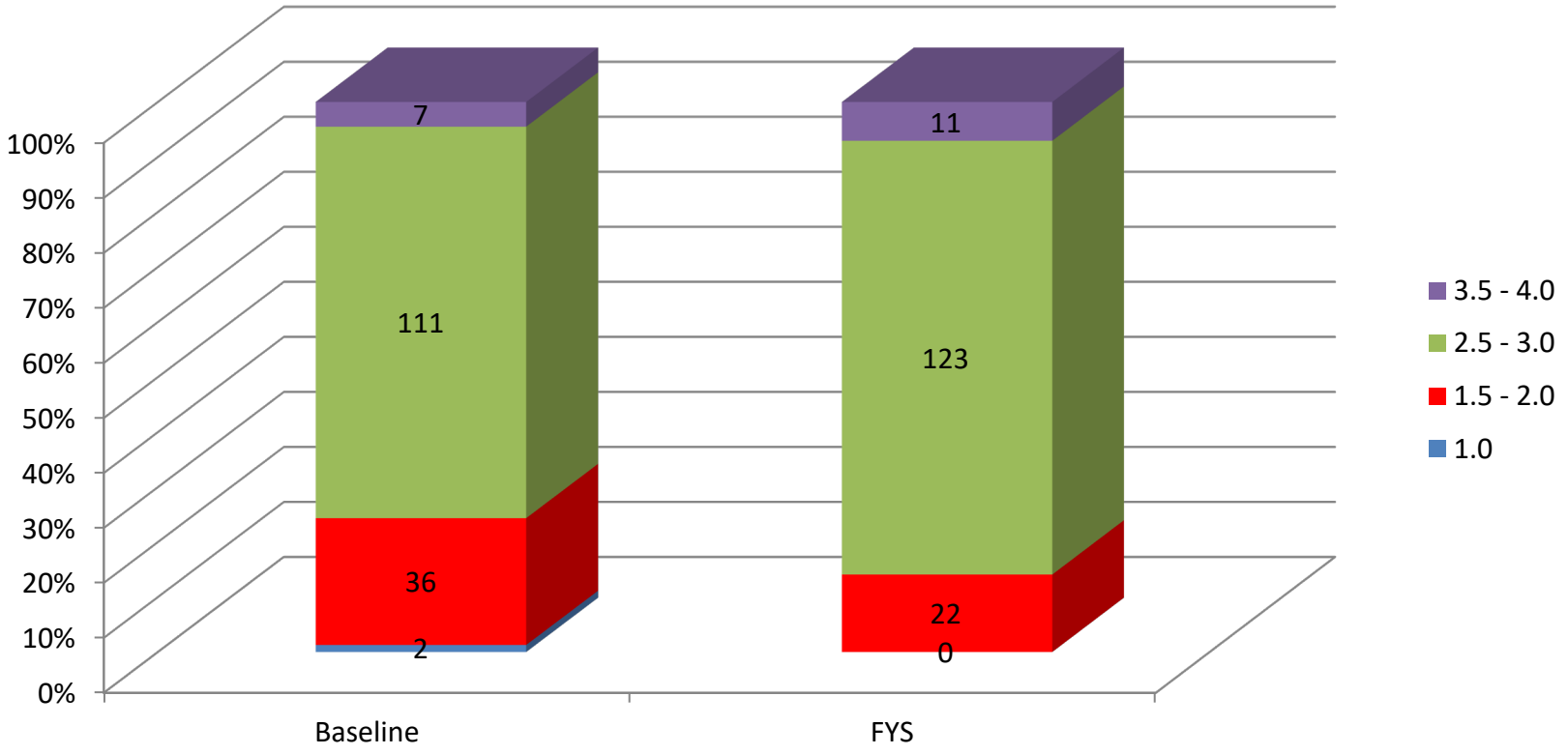
Convention/Format



Freshman Baseline/FYS Comparisons

n = 156

Communication Style



Baseline Inter-Rater Agreement Results

Trait/ Agreement	Development: Cohen's Kappa (Liberal) = .977	Convention/Format: Cohen's Kappa (Liberal) = .904	Communication Style: Cohen's Kappa (Liberal) = .876
Agree on score	86 (54%)	63 (40%)	84 (53%)
Difference = 1 point	69 (44%)	83 (53%)	60 (38%)
Difference = 2 points	3 (2%)	12 (8%)	14 (9%)
Difference = 3 points	0	0	0
Total	**158 (100%)**	**158 (100%)**	**158 (100%)**

FYS Inter-Rater Agreement Results

Trait/ Agreement	Development: Cohen's Kappa (Liberal) = .979	Convention/Format: Cohen's Kappa (Liberal) = .929	Communication Style: Cohen's Kappa (Liberal) = .947
Agree on score	96 (55%)	99 (57%)	100 (58%)
Difference = 1 point	74 (43%)	64 (37%)	67 (39%)
Difference = 2 points	3 (2%)	9 (5%)	6 (3%)
Difference = 3 points	0	1 (1%)	0
Total	**173 (100%)**	**173 (100%)**	**173 (100%)**

MARSHALL UNIVERSITY®

Comparison of FYS Results for Each Trait by Scenario

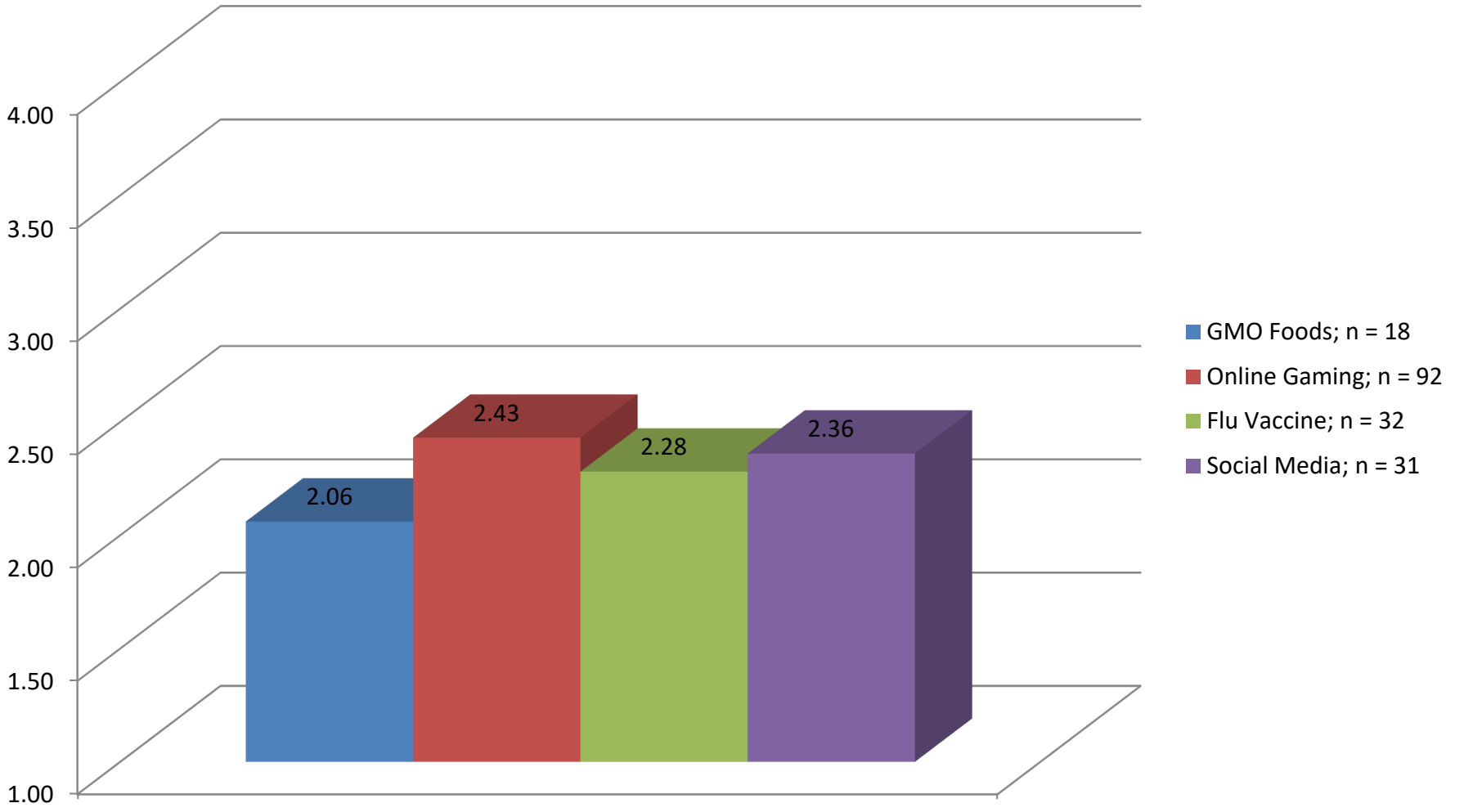
Academic Year 2021 - 2022

FYS Comparisons by Scenario for IL: Information Needed

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

Total *n* for FYS = 173

A One-Way ANOVA revealed no statistically significant differences in means across the scenarios.

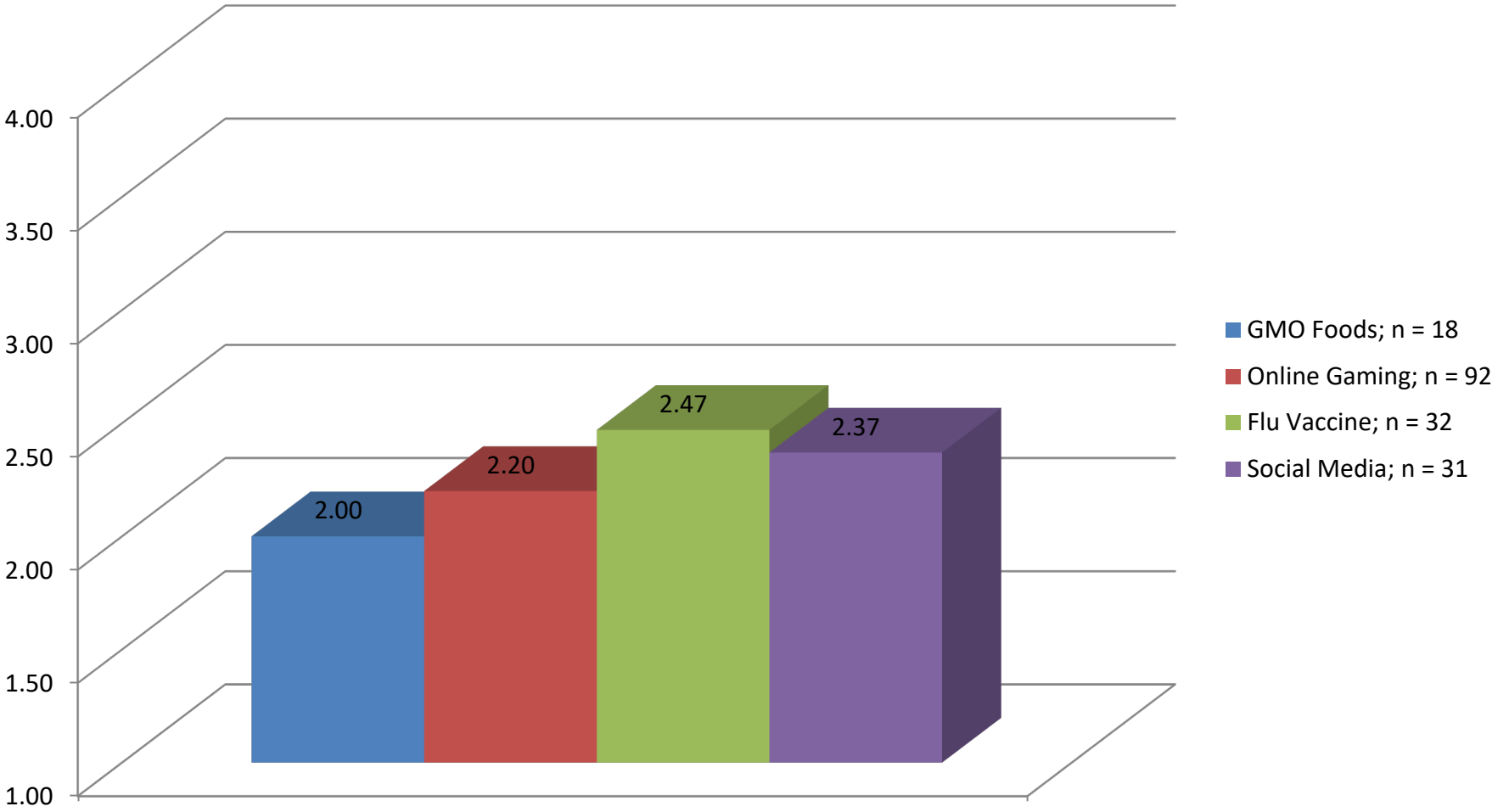


FYS Comparisons by Scenario for IL: Source Acknowledgment

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

Total *n* for FYS = 173

A One-Way ANOVA revealed no statistically significant differences in means across the scenarios

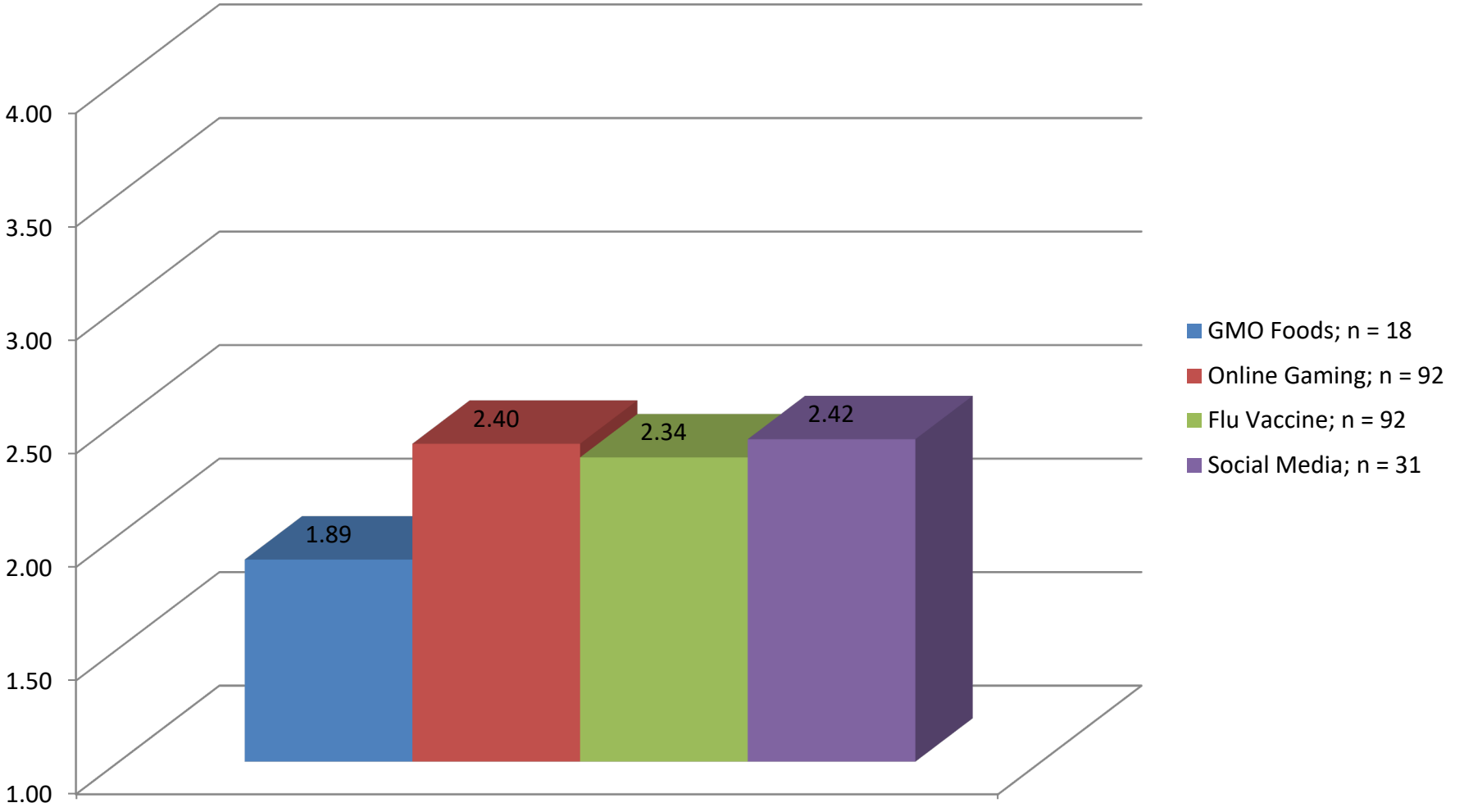


FYS Comparisons by Scenario for BT: Evidence

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

Total *n* for FYS = 173

A One-Way ANOVA revealed statistically significant differences in means across the scenarios; *the mean for GMO Foods was significantly lower than means for Online Gaming and Social Media.*

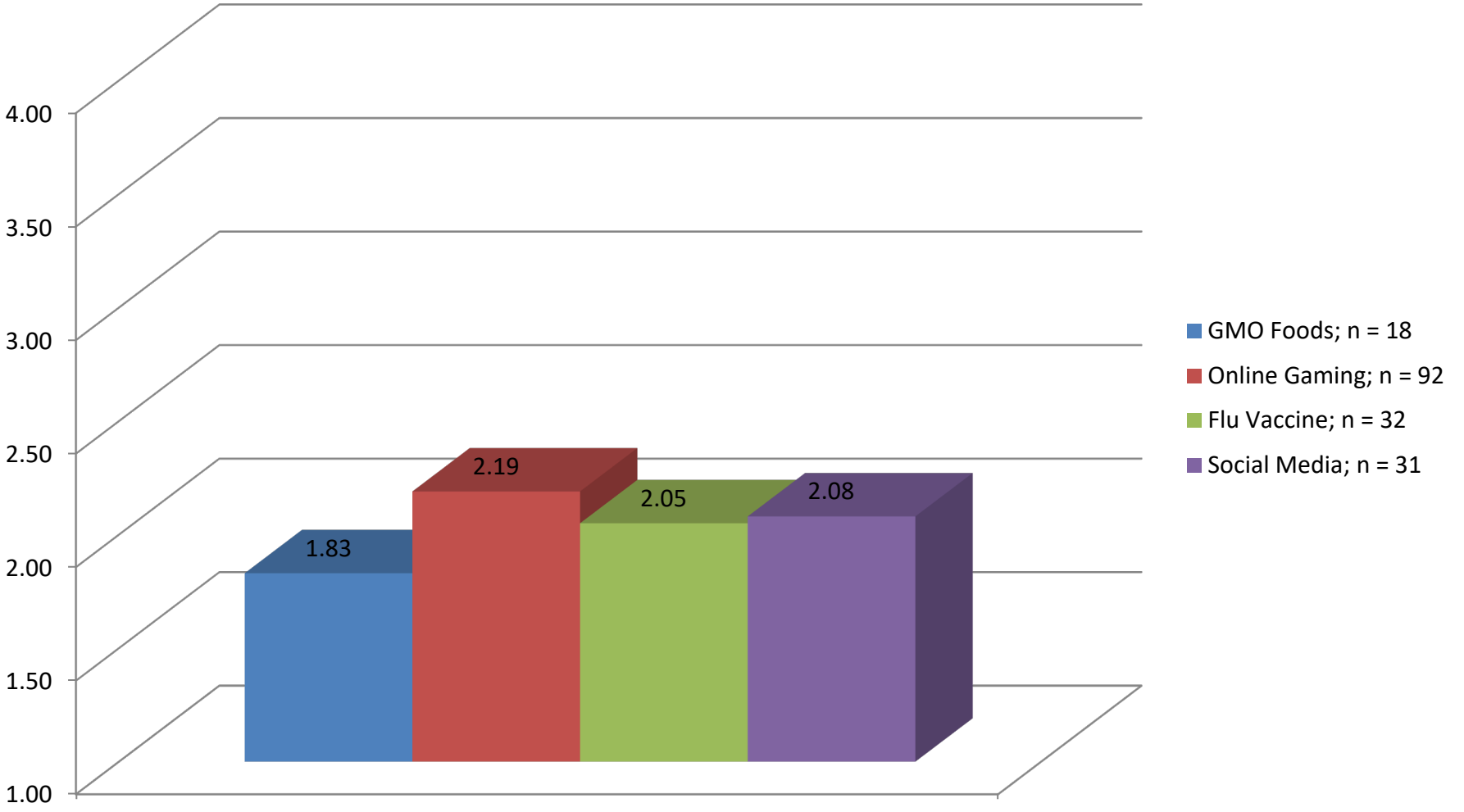


FYS Comparisons by Scenario for IBT: Viewpoints

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

Total *n* for FYS = 173

A One-Way ANOVA revealed no statistically significant differences in means across the scenarios.

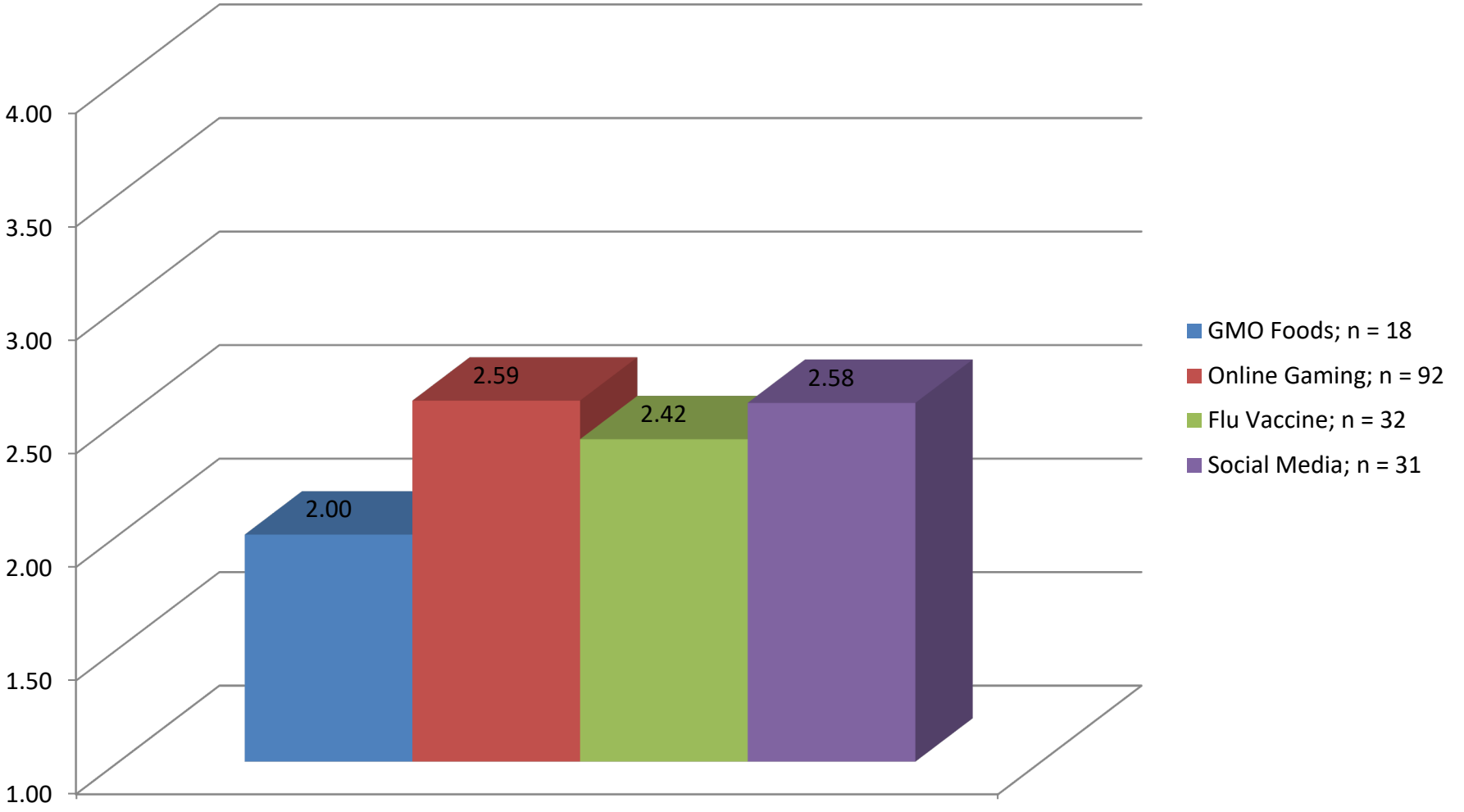


FYS Comparisons by Scenario for IBT: Recommendation/Position

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

Total *n* for FYS = 173

A One-Way ANOVA revealed statistically significant differences in means across the scenarios; *the mean for GMO Foods was significantly lower than means for Online Gaming and Social Media.*

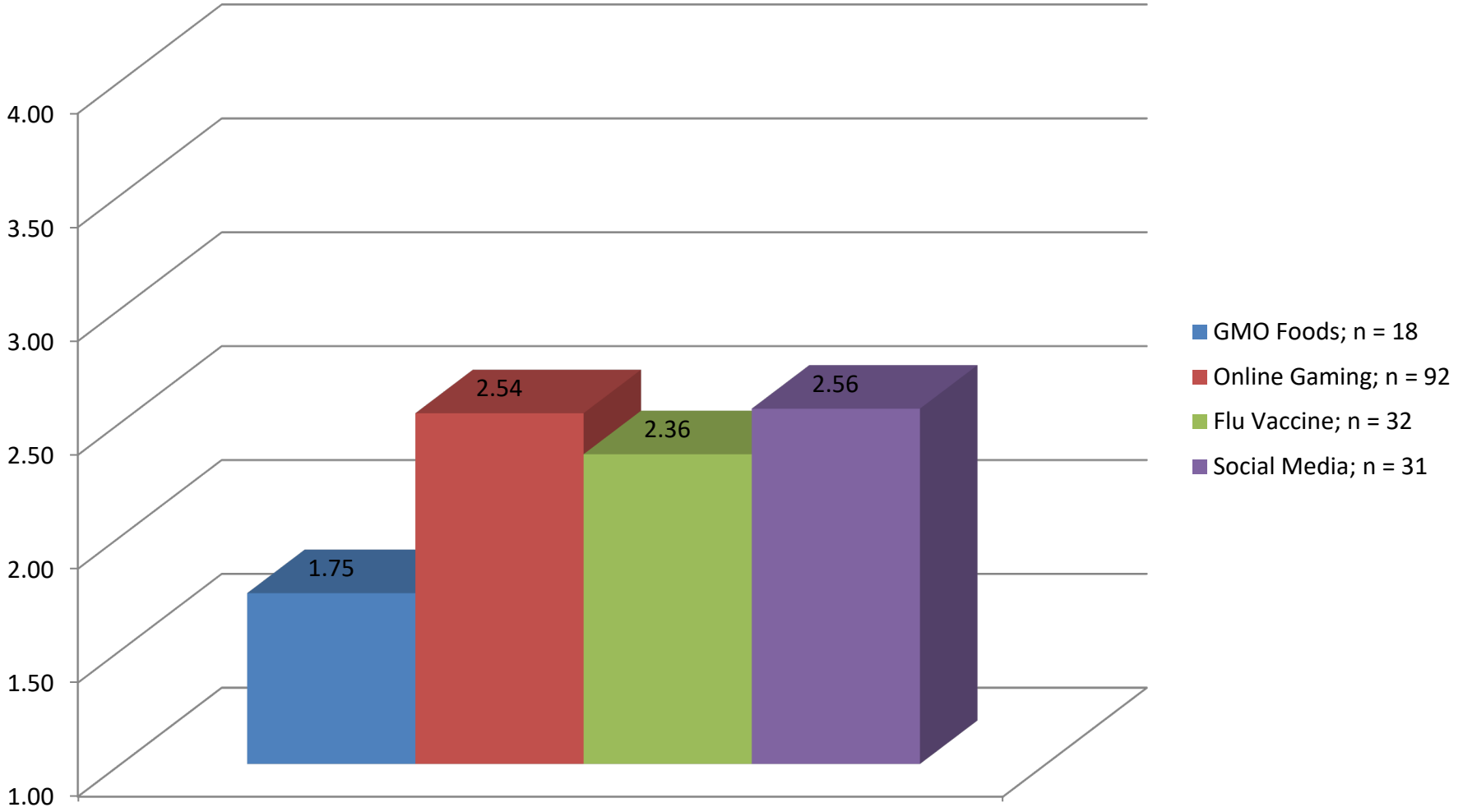


FYS Comparisons by Scenario for CF: Development

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

Total *n* for FYS = 173

A One-Way ANOVA revealed statistically significant differences in means across the scenarios; *the mean for GMO Foods was significantly lower than means for all other scenarios.*

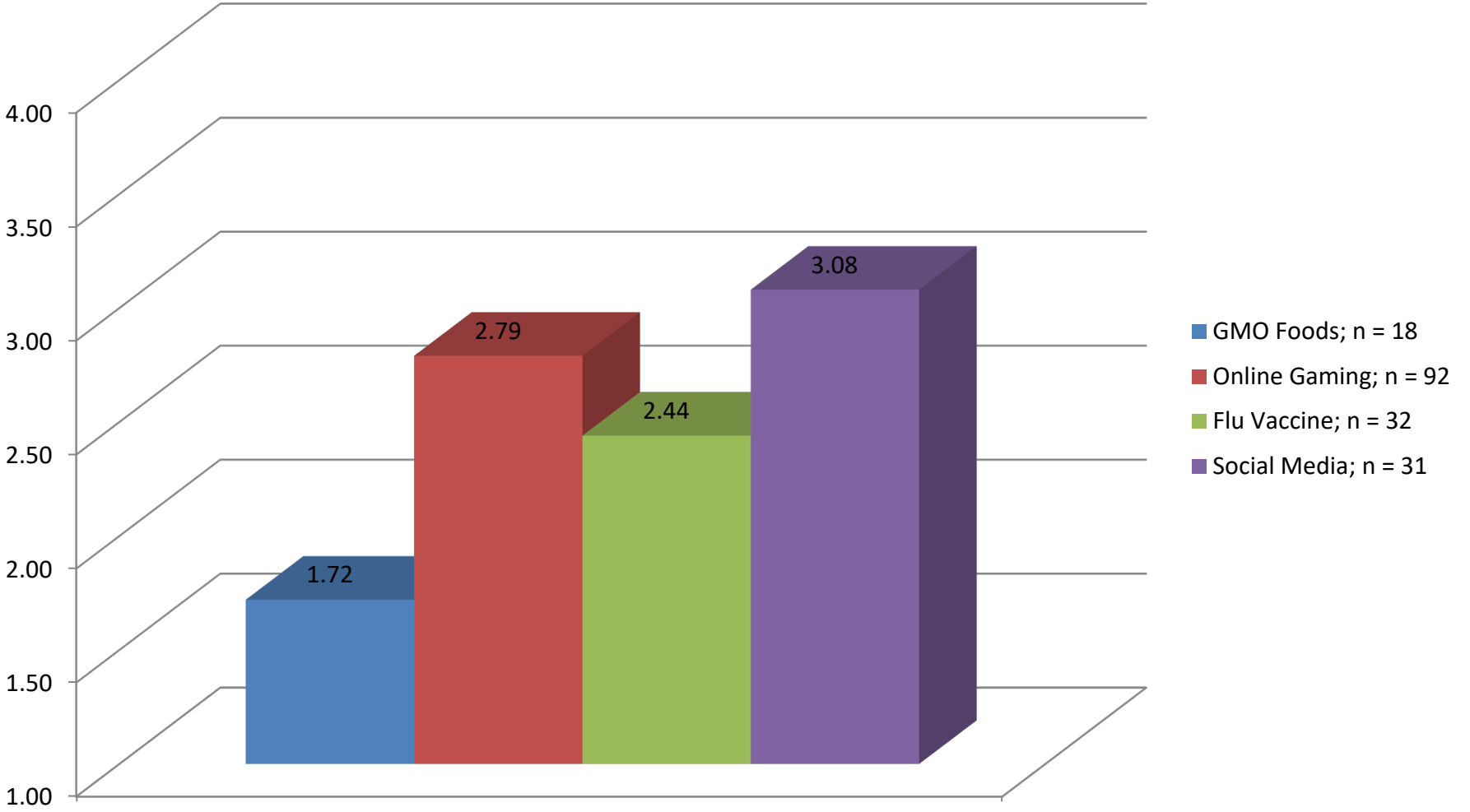


FYS Comparisons by Scenario for CF: Convention/Format

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

Total *n* for FYS = 173

A One-Way ANOVA revealed statistically significant differences in means across the scenarios; *the mean for GMO Foods was significantly lower than means for all other scenarios.*

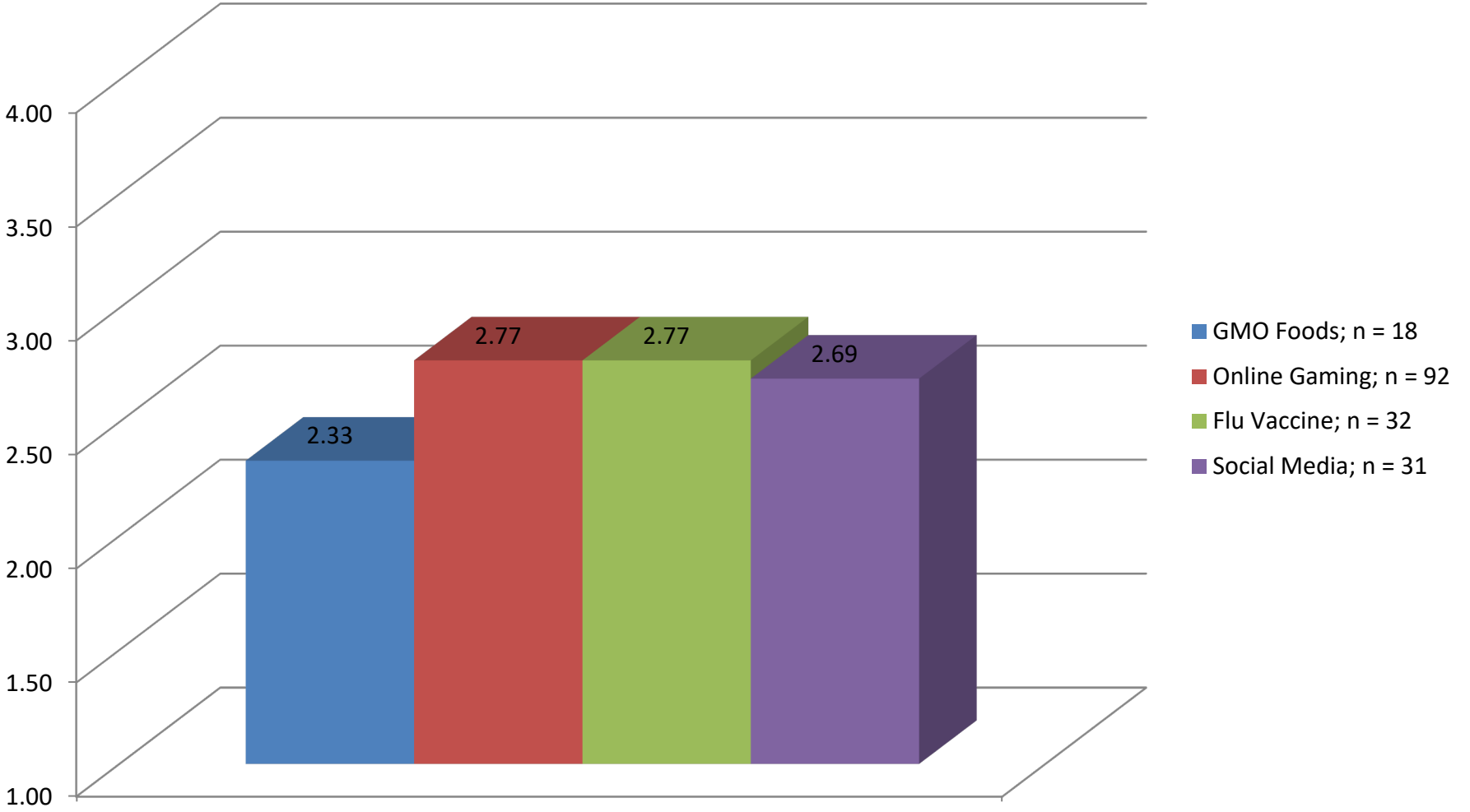


FYS Comparisons by Scenario for CF: Communication Style

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

Total *n* for FYS = 173

A One-Way ANOVA revealed statistically significant differences in means across the scenarios; *the mean for GMO Foods was significantly lower than means for Online Gaming and Flu Vaccine.*



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

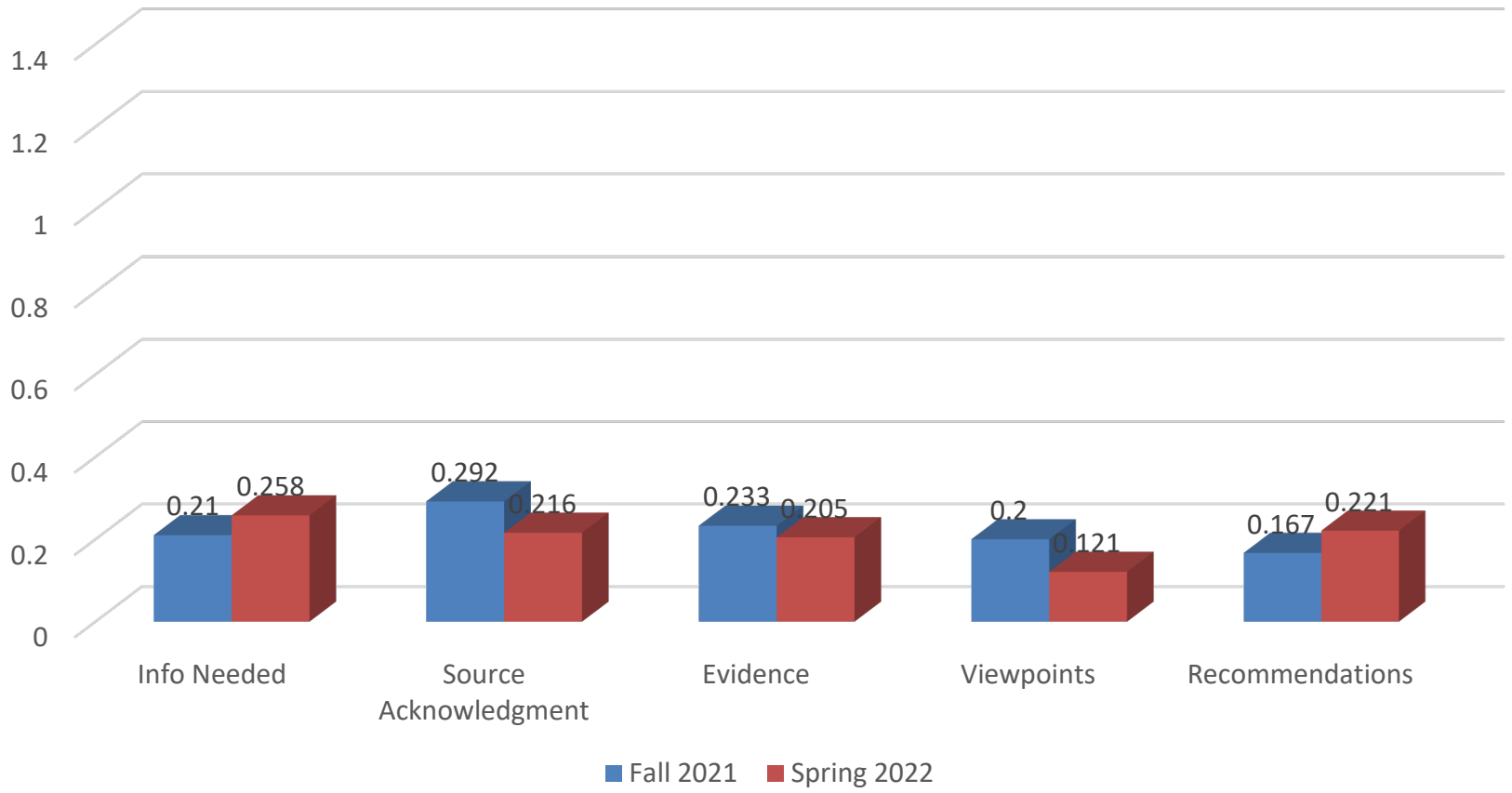
Academic Year 2021 - 2022

Baseline to FYS Mean Gain Scores for Each Trait

n = 62 in fall and 95 in spring (Information Needed)

n = 60 in fall and 95 in spring (All Other Traits)

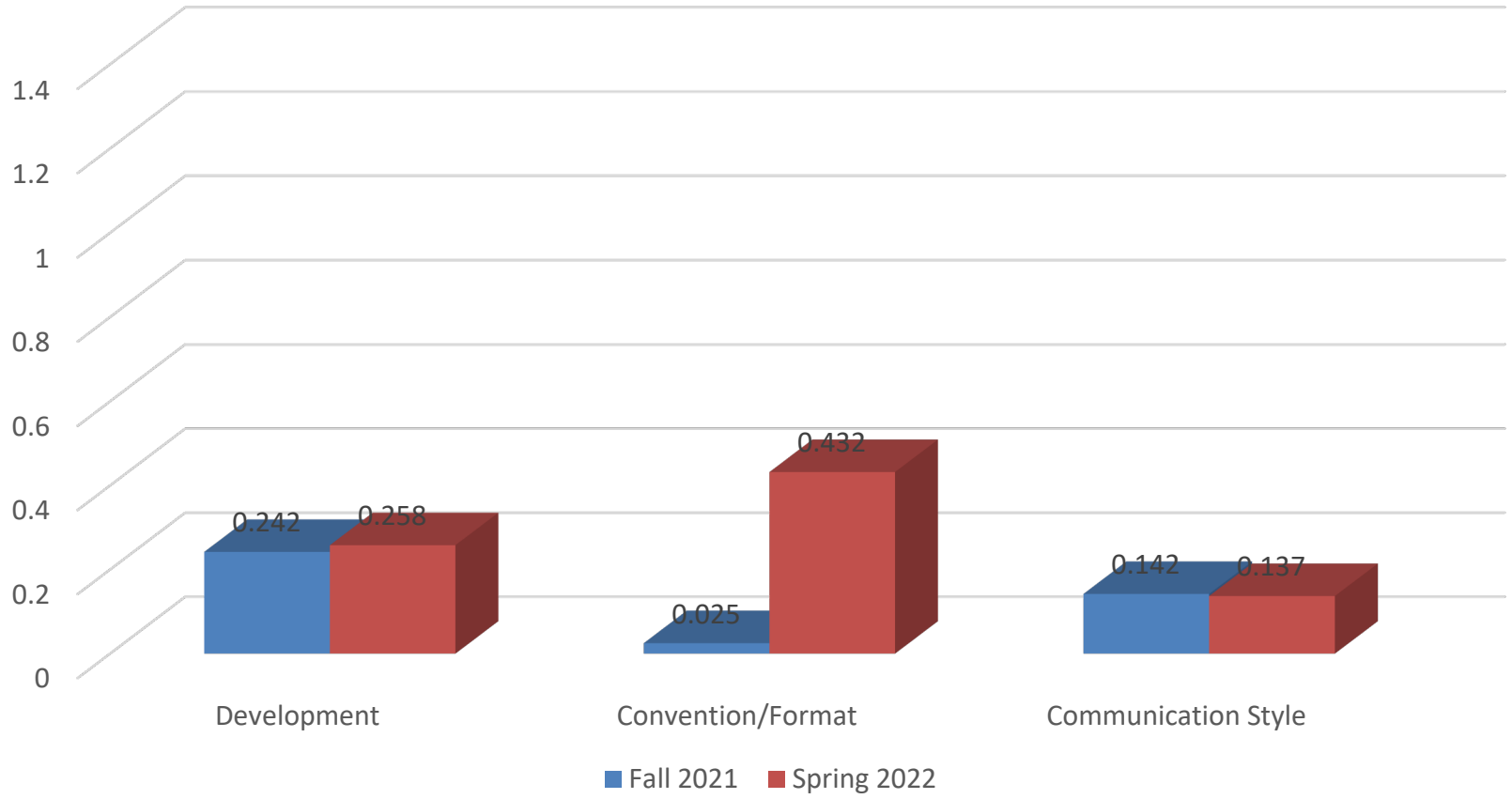
(Mean differences between fall and spring were not statistically significant)



Baseline to FYS Mean Gain Scores for Each Trait

n = 60 in fall and 95 in spring

(Mean difference between fall and spring for *Convention/Format* were significant, *t* (110.183) = -2.806; *p* = .006. *Mean differences for the other traits were not 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.