Comparison of Freshman Baseline with First Year Seminar Assessment Results

Academic Year 2023 – 2024

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

Summer Assessment Support Staff: Mary Beth Reynolds, Adam Russell, Diana Adams, and Mary Welch

Executive Summary

Background

Recommendations from the 2023 Assessment Team (current status in red)

The Summer Assessment Team made the following recommendations:

1. That we reflect on the original purpose of the course we call "FYS," whose name is "First Year Seminar in **Critical Thinking.** We were concerned that this is the first year since we have been assessing change in outcomes related to *Information Literacy* and *Critical Thinking* that we saw no significant difference between student performance on their baseline assessments and assessments at the conclusion of their FYS experience. We recommend that additional support be provided to instructors to help them craft their pedagogy to focus on critical thinking during this course. This should be done by returning a faculty member to the position of FYS coordinator. Although there is not an official FYS coordinator, the Executive Director of the Center for Teaching and Learning oversees FYS instruction. We note that student performance in FYS has improved since last year.
2. 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. This recommendation is repeated from last year. Did they do this?

Procedures for the 2023 Assessment

General Procedures

In August 2023, 1,529 incoming freshmen at Marshall University appeared to have 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 1,362 FYS assessments uploaded into Blackboard. To obtain a sample of matched pairs of baseline and FYS assessments, we began by comparing lists of all FYS and baseline artifacts uploaded to Blackboard during academic year 2023-2024 to determine which students submitted both baseline and FYS artifacts. We identified 904 matches and, from there, chose a random sample of 175 matched pairs. Each pair was further examined to ensure that the artifacts were uploaded and complete. When this was not the case for either the baseline or FYS artifacts, that match was discarded, and another chosen until we had the desired 175 matched pairs.

In May 2024, a group of seven faculty representing three academic colleges (Liberal Arts, Science, and Business) 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*), 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. We note that, despite the time spent checking the artifacts before scoring began, reviewers were not able to access three of the baseline artifacts due to expired links. They scored the FYS pair for each of these artifacts, but because those scores did not change the FYS mean scores, the three FYS artifacts paired with the baselines with expired links were eliminated from our analysis. This left us with 172 scored matched pairs, for which we conducted *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*). Results showed significant differences between baseline and FYS results for all traits of each learning outcome. These results are shown in the table below. 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**	2.087 (0.5084)	2.247 (0.5632)	$t(171) = -2.994$, $p = .003$
	Source Acknowledgment	1.869 (0.7505)	2.361 (0.8330)	$t(171) = -6.374$, $p < .001$
Inquiry-Based (Critical) Thinking	**Evidence**	1.884 (0.6652)	2.323 (0.6660)	$t(171) = -6.986$, $p < .001$
	Viewpoints	1.765 (0.4727)	1.942 (0.5954)	$t(171) = -4.016$, $p < .001$
	Recommendation/ Position	2.105 (0.5954)	2.451 (0.6110)	$t(171) = -6.003$, $p < .001$
Communication Fluency	**Development**	2.076 (0.6353)	2.477 (0.5968)	$t(171) = -6.785$, $p < .001$

Outcome	Trait	Baseline Mean (*SD*)	FYS Mean (*SD*)	Statistical Significance
	Convention/Format	2.317 (0.8411)	2.637 (0.6668)	*t(171)* = -4.094, *p* < .001
	Communication Style	2.564 (0.5396)	2.799 (0.4034)	*t(171)* = -5.362, *p* < .001

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**	7%
	Source Acknowledgment	27%
Inquiry-Based (Critical) Thinking	**Evidence**	25%
	Viewpoints	1%
	Recommendation/Position	28%
Communication Fluency	**Development**	25%
	Convention/Format	23%
	Communication Style	17%

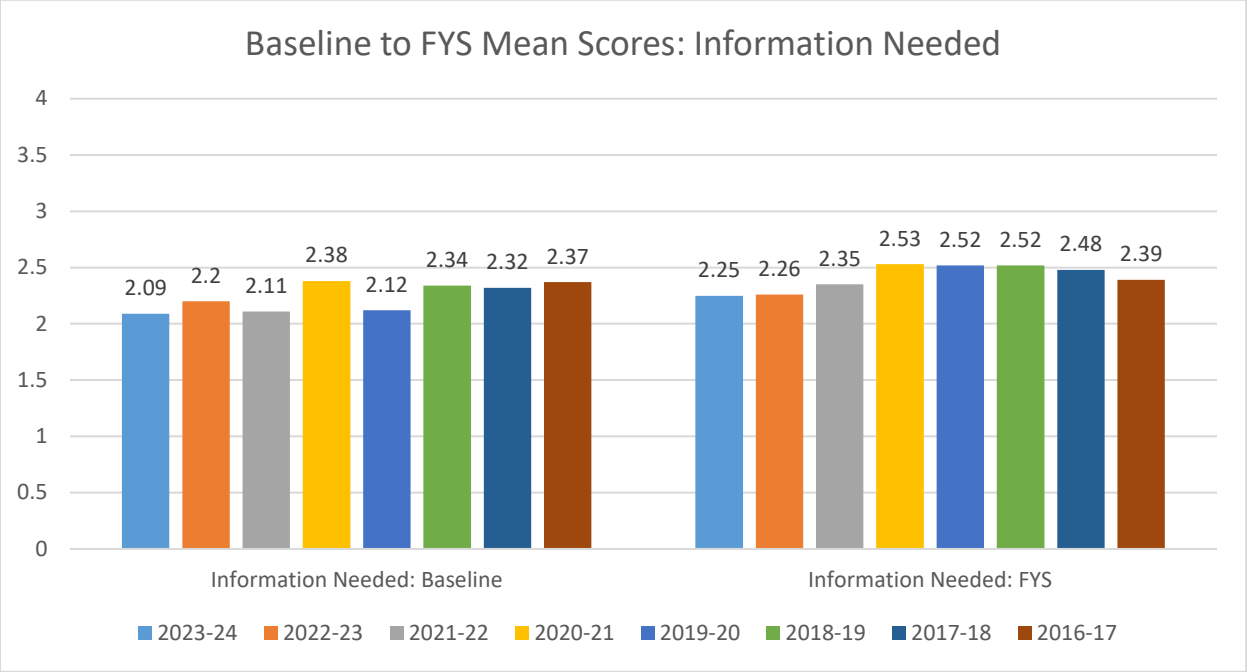
Since students enrolled in FYS completed their responses to one of four possible scenarios, we further analyzed results based on scenario. This year's results showed a significant difference in performance based on scenario used for the FYS assessments for one trait (source acknowledgment) of *Information Literacy*, and for one trait (convention/format) of *Communication Fluency*. On source acknowledgment, students scored significantly higher on GMO Foods than on Gaming. On convention/format, the opposite was the case, with students scoring significantly lower on GMO Foods than on Gaming.

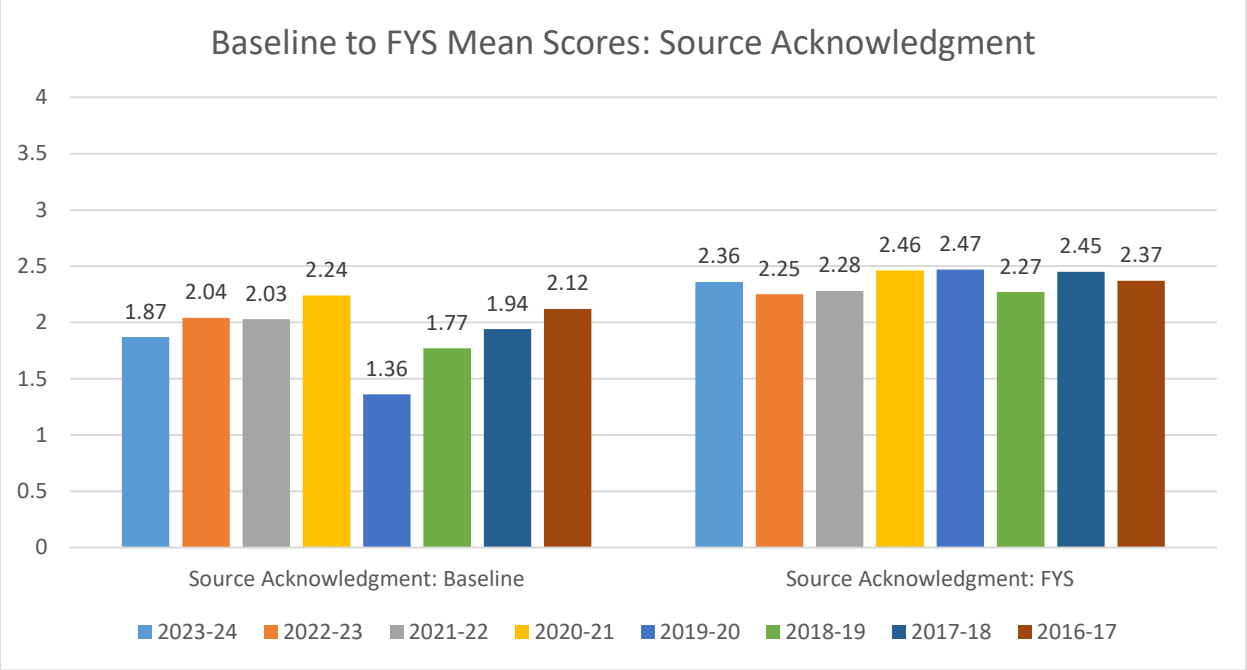
Gain scores between students in our sample who completed FYS in fall 2023 (*n* = 78) and those who completed FYS in spring 2024 (*n* = 94) did not differ significantly on any outcome. Please refer to the supporting documentation for additional detail.

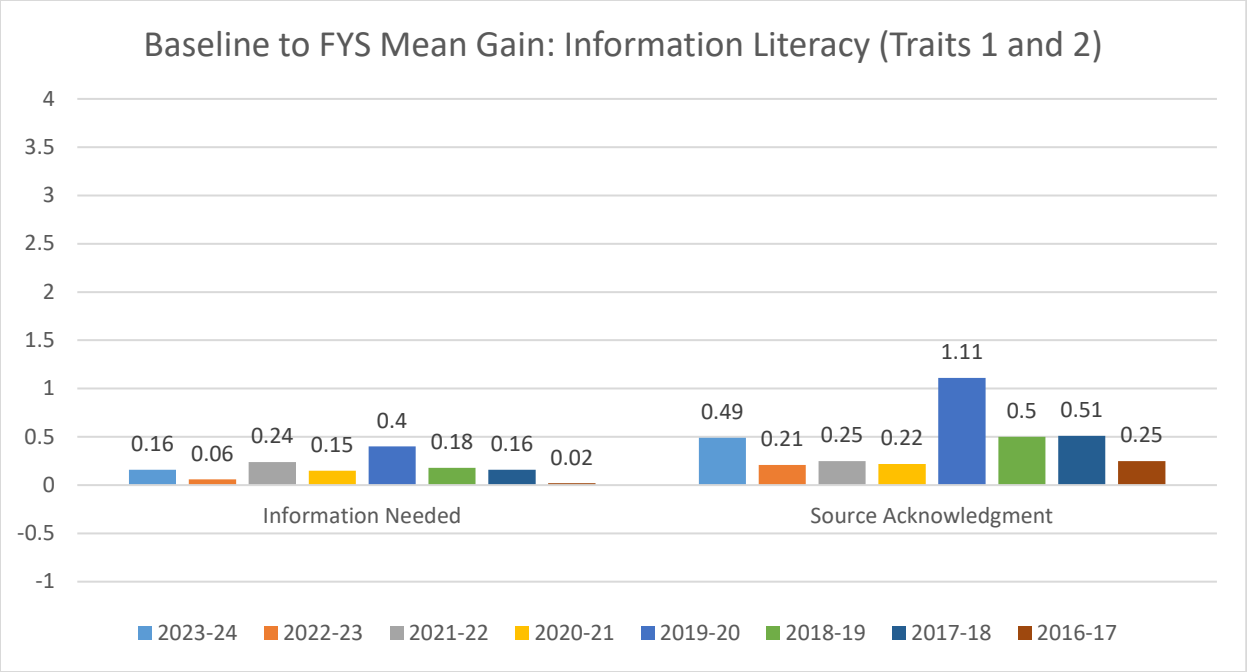
Conclusions

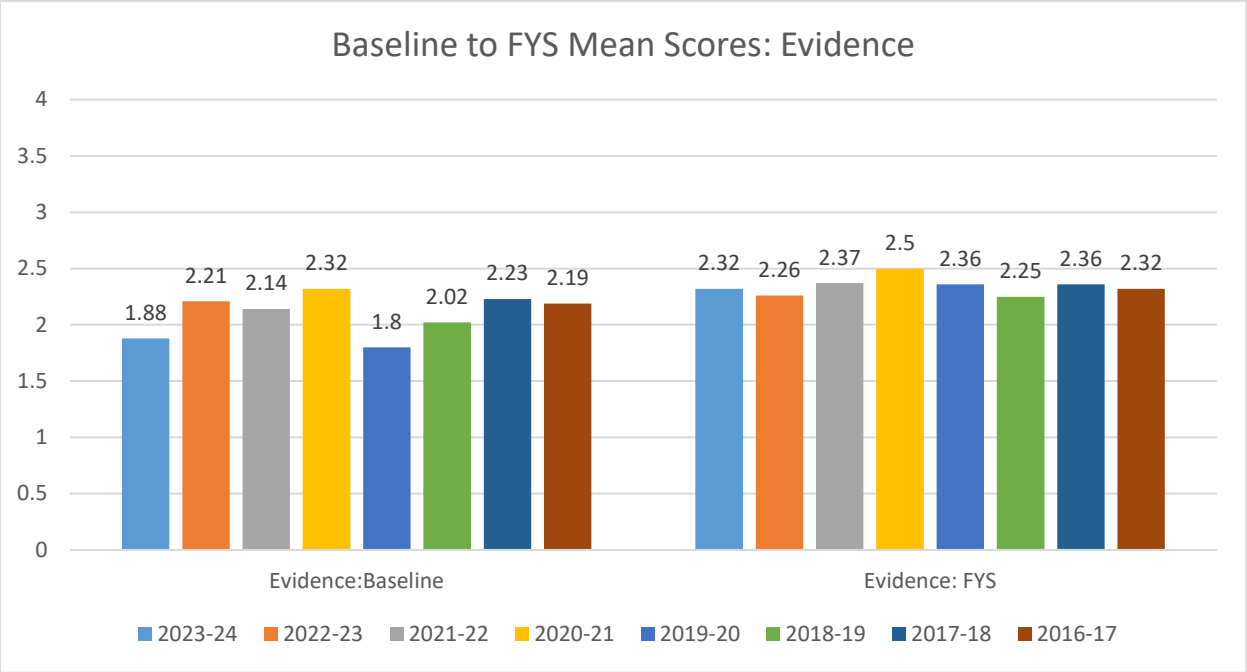
Although we have not performed statistical analyses to compare the results across years, we were concerned about our results in 2022-2023 because that was the only year we had not seen statistically significant improvement between baseline and FYS in at least some traits of *Critical Thinking* since we began analyzing student performance in 2013. After comparing trends for baseline and FYS means from 2016-2017 through 2022-2023, we concluded that last year's results were not due to higher than usual baseline scores, but rather to lower FYS scores than in past years.

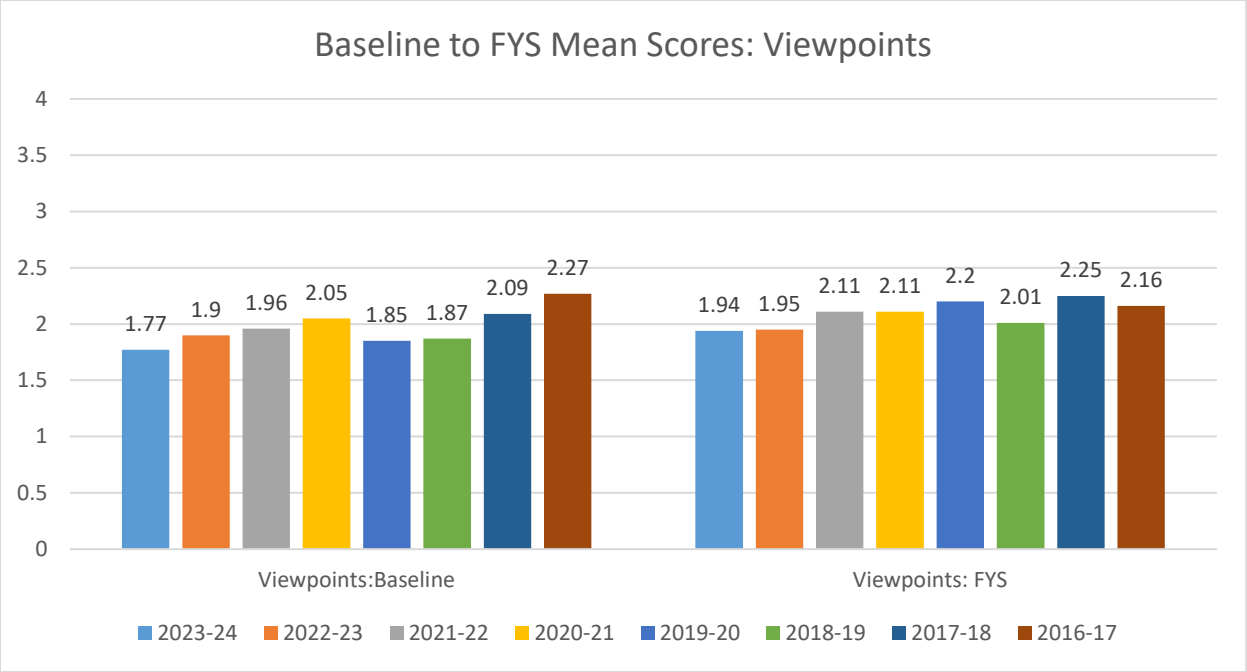
As noted, this year's results did show significant differences for all traits of the three outcomes (*Information Literacy, Critical Thinking*, and *Communication Fluency*) assessed. In reviewing these results, we noted that this year's students scored lower on baseline and higher on FYS than students from 2022-2023. This led us to examine two metrics – 1) gain score for *Information Literacy* and *Critical Thinking* for our samples from 2016 to the present, and 2) Baseline and FYS means for the same period. This information is shown below.

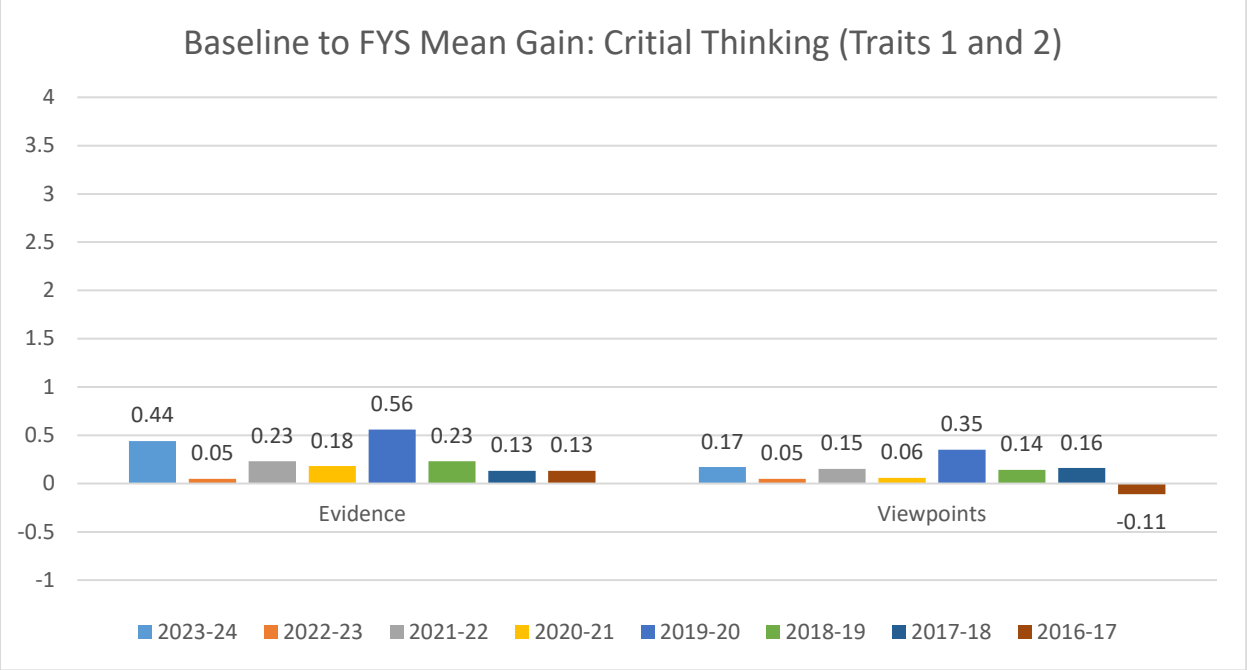

Baseline to FYS Mean Scores: Information Needed
4
3.5
3
2.5
2
1.5
1
0.5
0
2.09
2.2
2.11
2.38
2.12
2.34
2.32
2.37
2.25
2.26
2.35
2.53
2.52
2.52
2.48
2.39
Information Needed: Baseline
Information Needed: FYS
2023-24
2022-23
2021-22
2020-21
2019-20
2018-19
2017-18
2016-17

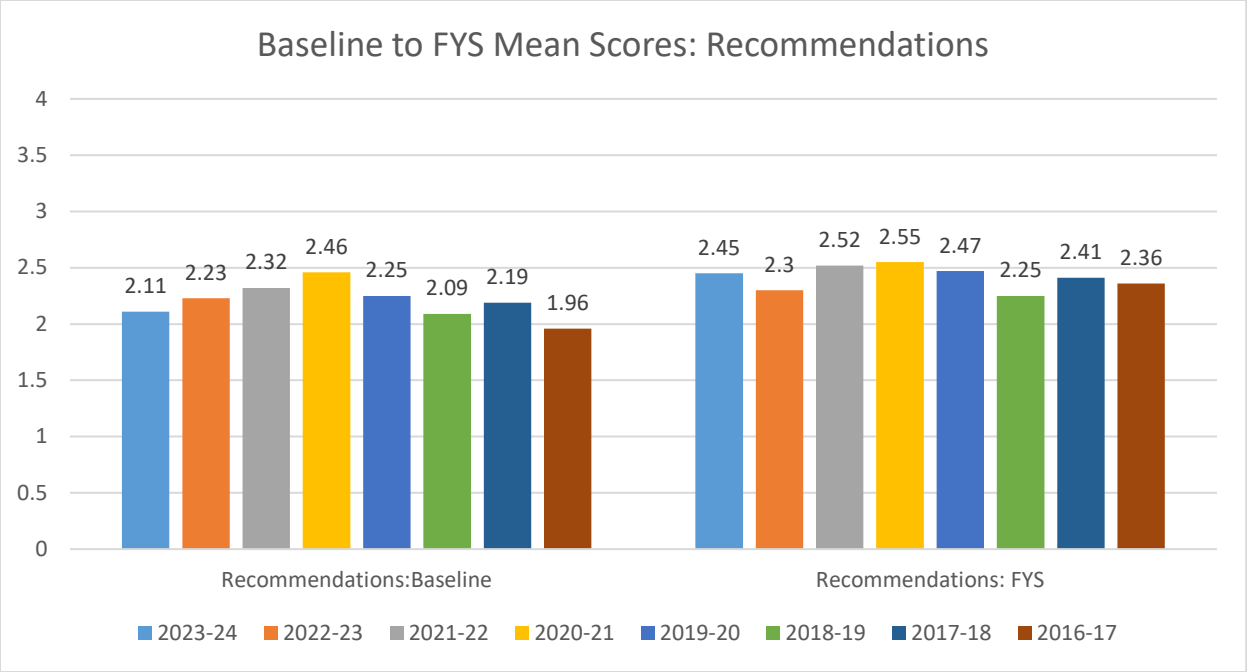

Baseline to FYS Mean Scores: Source Acknowledgment
4
3.5
3
2.5
2
1.5
1
0.5
0
1.87
2.04
2.03
2.24
1.36
1.77
1.94
2.12
2.36
2.25
2.28
2.46
2.47
2.27
2.45
2.37
Source Acknowledgment: Baseline
Source Acknowledgment: FYS
2023-24
2022-23
2021-22
2020-21
2019-20
2018-19
2017-18
2016-17

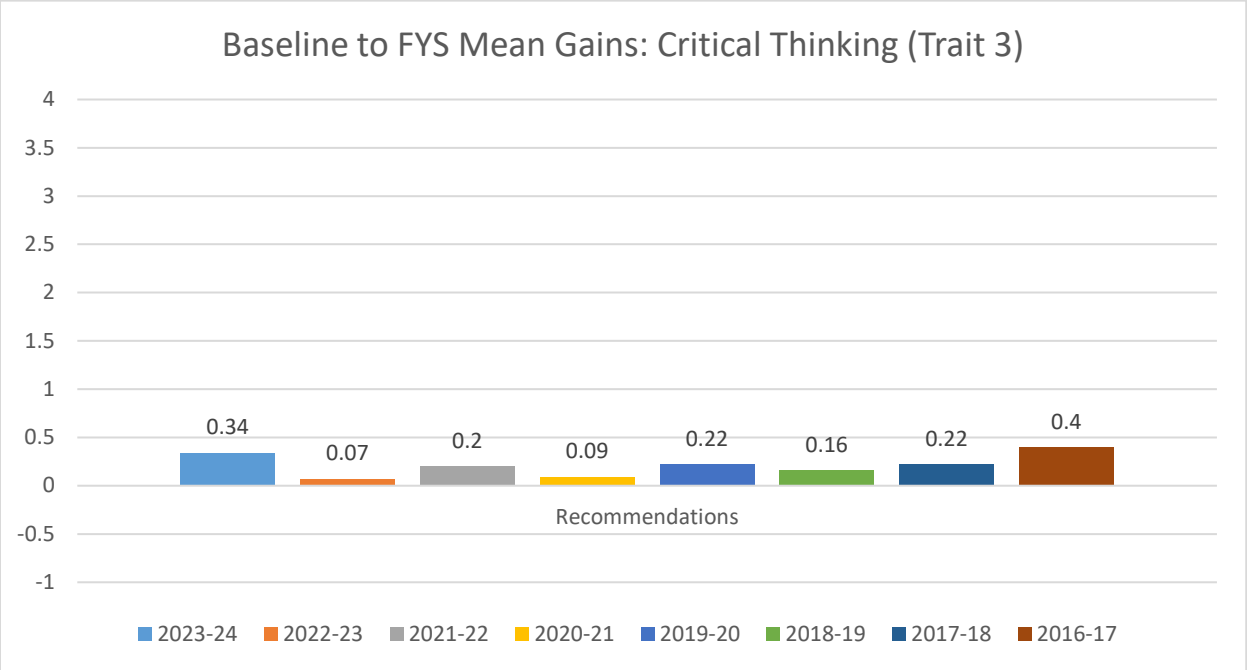

Baseline to FYS Mean Gain: Information Literacy (Traits 1 and 2)
4
3.5
3
2.5
2
1.5
1
0.5
0
-0.5
-1
0.16
0.06
0.24
0.15
0.4
0.18
0.16
0.02
0.49
0.21
0.25
0.22
1.11
0.5
0.51
0.25
Information Needed
Source Acknowledgment
2023-24
2022-23
2021-22
2020-21
2019-20
2018-19
2017-18
2016-17


Baseline to FYS Mean Scores: Evidence
4
3.5
3
2.5
2
1.5
1
0.5
0
1.88
2.21
2.14
2.32
1.8
2.02
2.23
2.19
2.32
2.26
2.37
2.5
2.36
2.25
2.36
2.32
Evidence:Baseline
Evidence: FYS
2023-24
2022-23
2021-22
2020-21
2019-20
2018-19
2017-18
2016-17


Baseline to FYS Mean Scores: Viewpoints
4
3.5
3
2.5
2
1.5
1
0.5
0
1.77
1.9
1.96
2.05
1.85
1.87
2.09
2.27
1.94
1.95
2.11
2.11
2.2
2.01
2.25
2.16
Viewpoints:Baseline
Viewpoints: FYS
2023-24
2022-23
2021-22
2020-21
2019-20
2018-19
2017-18
2016-17


Baseline to FYS Mean Gain: Critial Thinking (Traits 1 and 2)
4
3.5
3
2.5
2
1.5
1
0.5
0
-0.5
-1
0.44
0.05
0.23
0.18
0.56
0.23
0.13
0.13
0.17
0.05
0.15
0.06
0.35
0.14
0.16
-0.11
Evidence
Viewpoints
2023-24
2022-23
2021-22
2020-21
2019-20
2018-19
2017-18
2016-17





Examination of the charts above show that, for the most part, mean scores at the end of FYS reach between 2.0 and 2.4 on a 4-point rubric scale, with students making gains between baseline and FYS for all rubric traits except *Critical Thinking* (viewpoints) in one out of the eight years examined.

Recommendations from the 2024 Assessment Team

The Summer Assessment Team made the following recommendations:

To be determined.

Supporting Documentation

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

Academic Year 2023 - 2024

Review Procedures

- One hundred seventy-five (175) FYS critical thinking artifacts were matched with 175 baseline critical thinking artifacts. This number represented 13% of the 1,362 FYS artifacts and 11% of the 1,529 baseline artifacts uploaded to Blackboard.

- During the evaluation we discovered that the links for three baseline artifacts had expired, so were not able to be assessed. All 175 of the FYS artifacts were accessible and reviewers scored each. However, since we use matched pairs to evaluate the change in student performance between baseline and FYS and we determined that elimination of the unmatched FYS artifacts did not change the mean across the remaining 172, the three unpaired FYS artifacts were not included in the analysis.

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 2023 – updated 5-8-2023

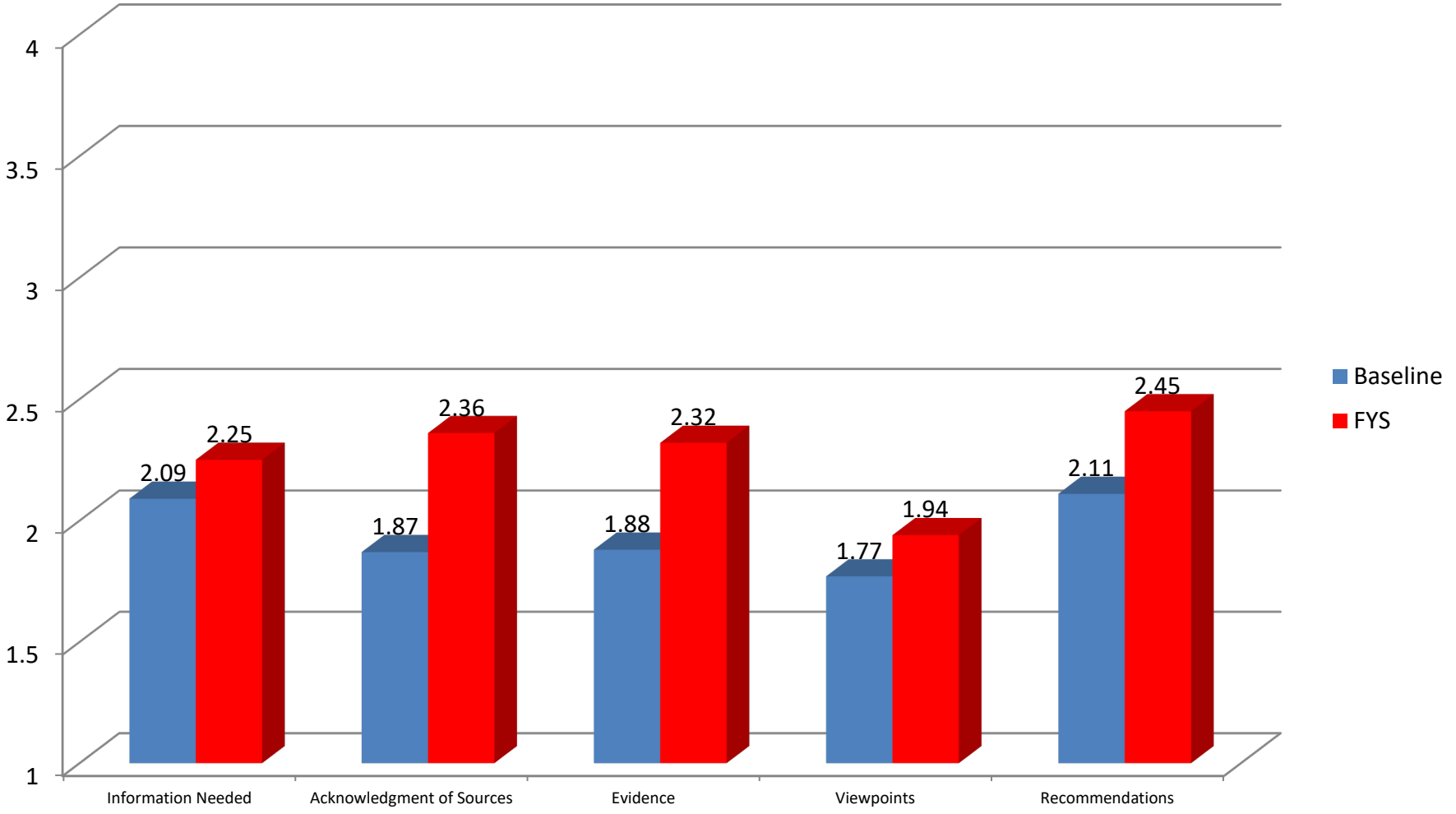
Outcomes	Traits	Performance Levels			
		1	2	3	4
Information Literacy	**Information Needed**	Does not acknowledge or assess the need for more information.	Acknowledges the need for more information but does not identify research methods/sources (or those identified are not feasible) that would address 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 (e.g., only one or two pieces of evidence) 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 their 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 errors in usage/mechanics impedes smooth reading of the document.	Uses straightforward language that conveys meaning to readers. The language in the document has few errors.	Uses 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 = 172 matched pairs

Mean differences between baseline and FYS were statistically significant for all traits.



Freshman Baseline/FYS Comparisons

n = 172 matched pairs

Trait/ Performance Level	Info Needed (Baseline)	Info Needed (FYS)	Acknowledgment of Sources (Baseline)	Acknowledgment of Sources (FYS)
1.0	13 (8%)	6 (3%)	47 (27%)	27 (16%)
1.5 – 2.0	102 (59%)	96 (56%)	74 (43%)	47 (27%)
2.5 – 3.0	57 (33%)	66 (38%)	43 (25%)	83 (48%)
3.5 – 4.0	0	4 (2%)	8 (5%)	15 (9%)
Totals	**172**	**172**	**172**	**172**

Freshman Baseline/FYS Comparisons

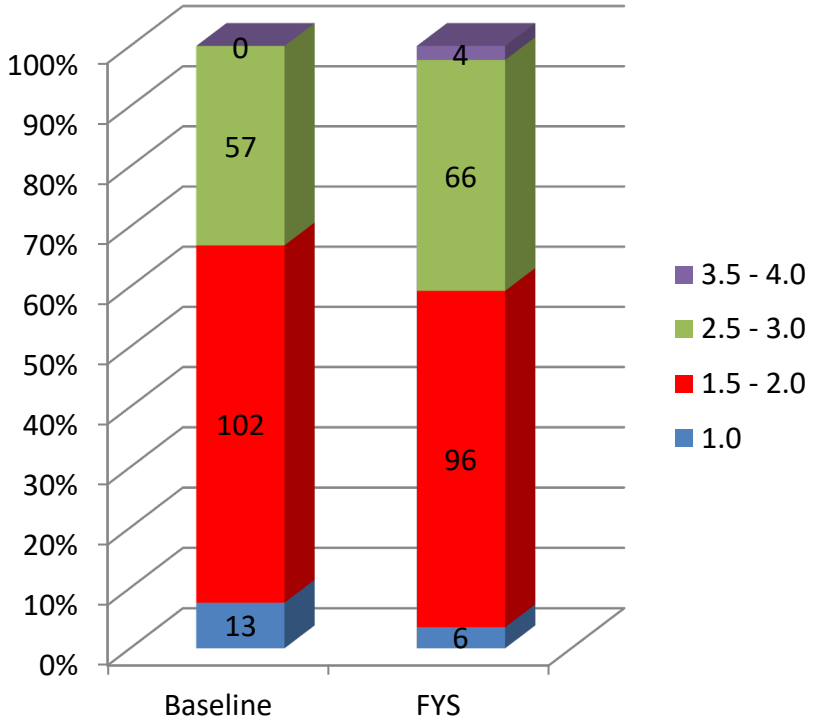
n = 172 matched pairs

Trait/ Performance Level	Evidence Baseline	Evidence FYS	Viewpoints Baseline	Viewpoints FYS	Recommendations Baseline	Recommendations FYS
1.0	36 (21%)	14 (8%)	31 (18%)	14 (8%)	17 (10%)	6 (3%)
1.5 – 2.0	86 (50%)	64 (37%)	117 (68%)	133 (77%)	87 (51%)	51 (30%)
2.5 – 3.0	46 (27%)	85 (49%)	24 (14%)	25 (15%)	66 (38%)	103 (60%)
3.5 – 4.0	4 (2%)	9 (5%)	0	0	2 (1%)	12 (7%)
Totals	**172**	**172**	**172**	**172**	**172**	**172**

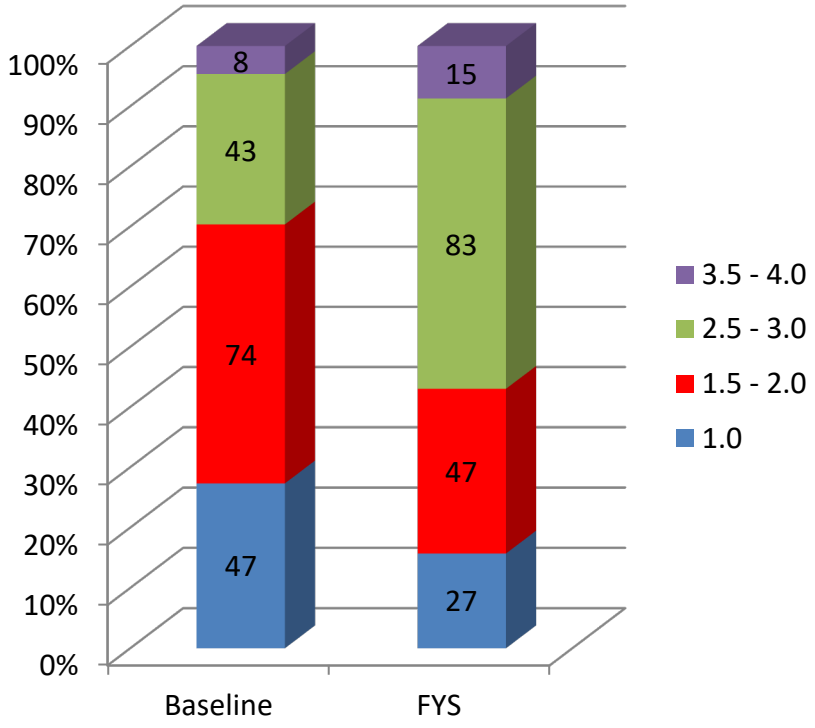
Freshman Baseline/FYS Comparisons

n = 172 matched pairs

Information Needed



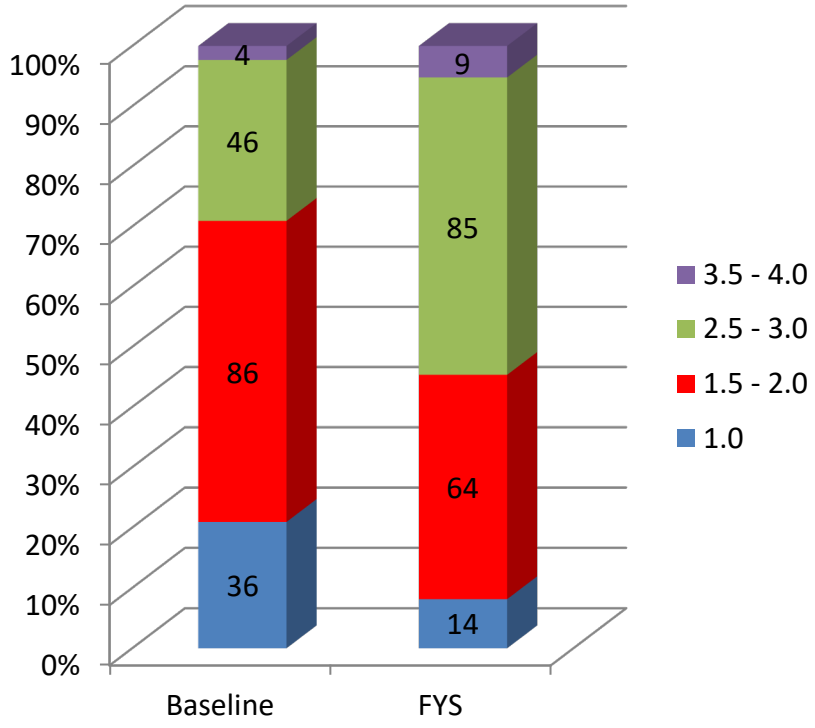
Acknowledgment of Sources



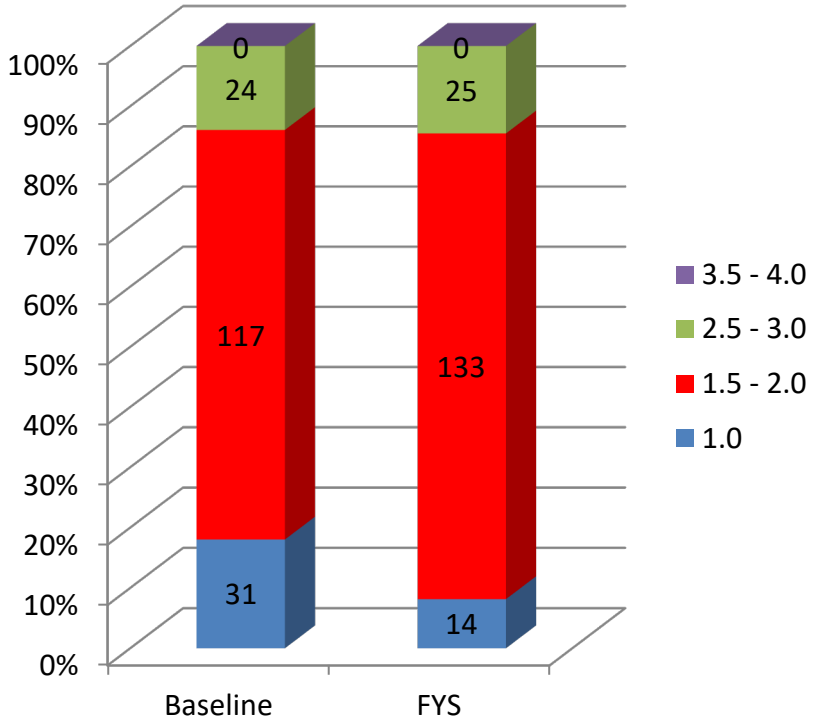
Freshman Baseline/FYS Comparisons

n = 172 matched pairs

Evidence



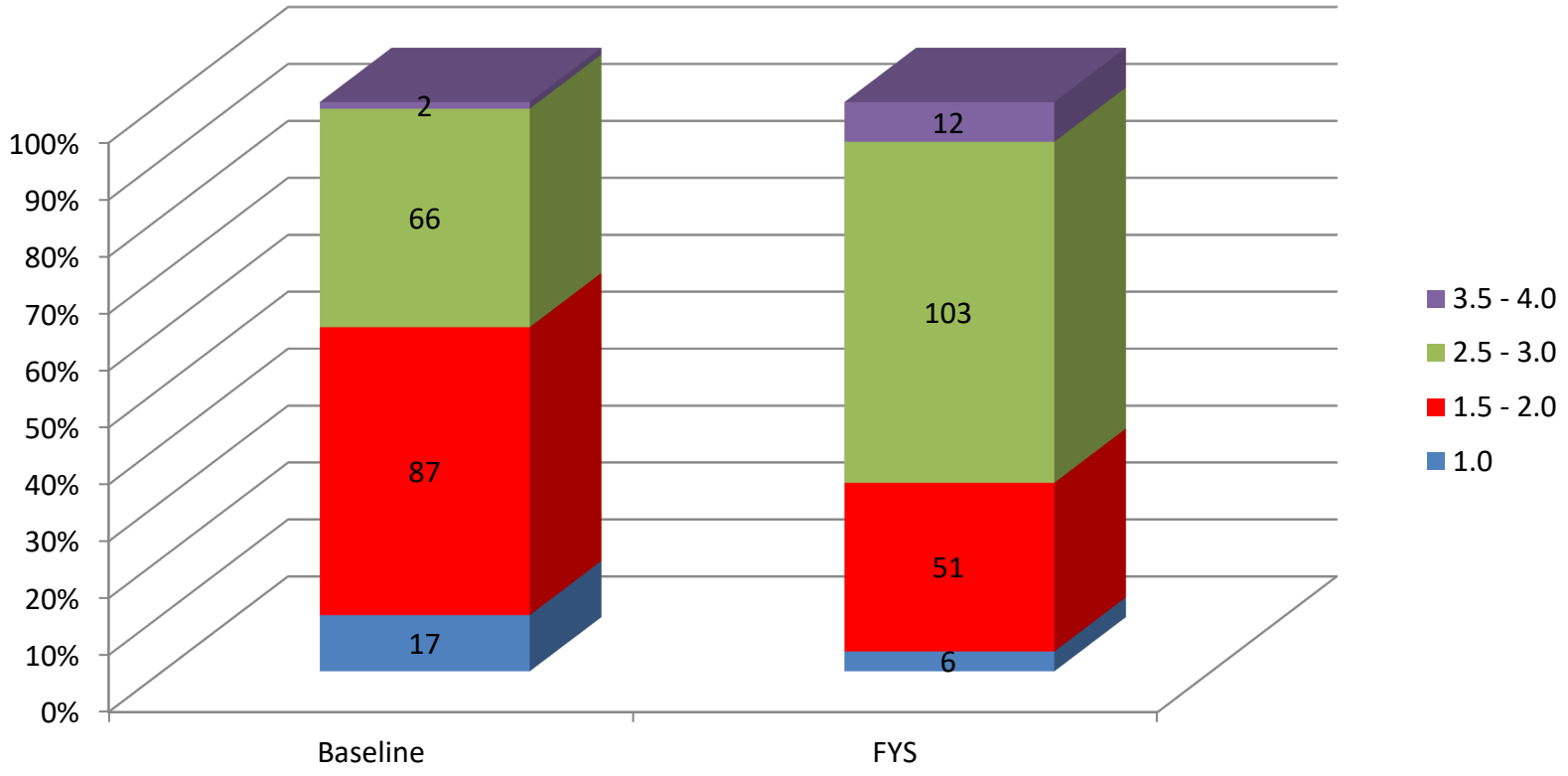
Viewpoints



Freshman Baseline/FYS Comparisons

n = 172 matched pairs

Recommendations



Baseline Inter-Rater Agreement Results

Includes 172 baseline assessments scored

Trait/ Agreement	Info Needed : Cohen's Kappa (Liberal) = 1.000	Acknowledgment of Sources: Cohen's Kappa (Liberal) = .993	Evidence: Cohen's Kappa (Liberal) = .949	Viewpoints: Cohen's Kappa (Liberal) = .959	Recommendations: Cohen's Kappa (Liberal) = .977
Agree on score	114 (66%)	102 (59%)	92 (54%)	103 (60%)	99 (58%)
Difference = 1 point	58 (34%)	69 (40%)	73 (42%)	64 (37%)	70 (41%)
Difference = 2 points	0	1 (1%)	7 (4%)	5 (3%)	3 (2%)
Difference = 3 points	0	0	0	0	0
Total	**172**	**172**	**172**	**172**	**172**

FYS Inter-Rater Agreement Results

Includes all 175 FYS assessments scored

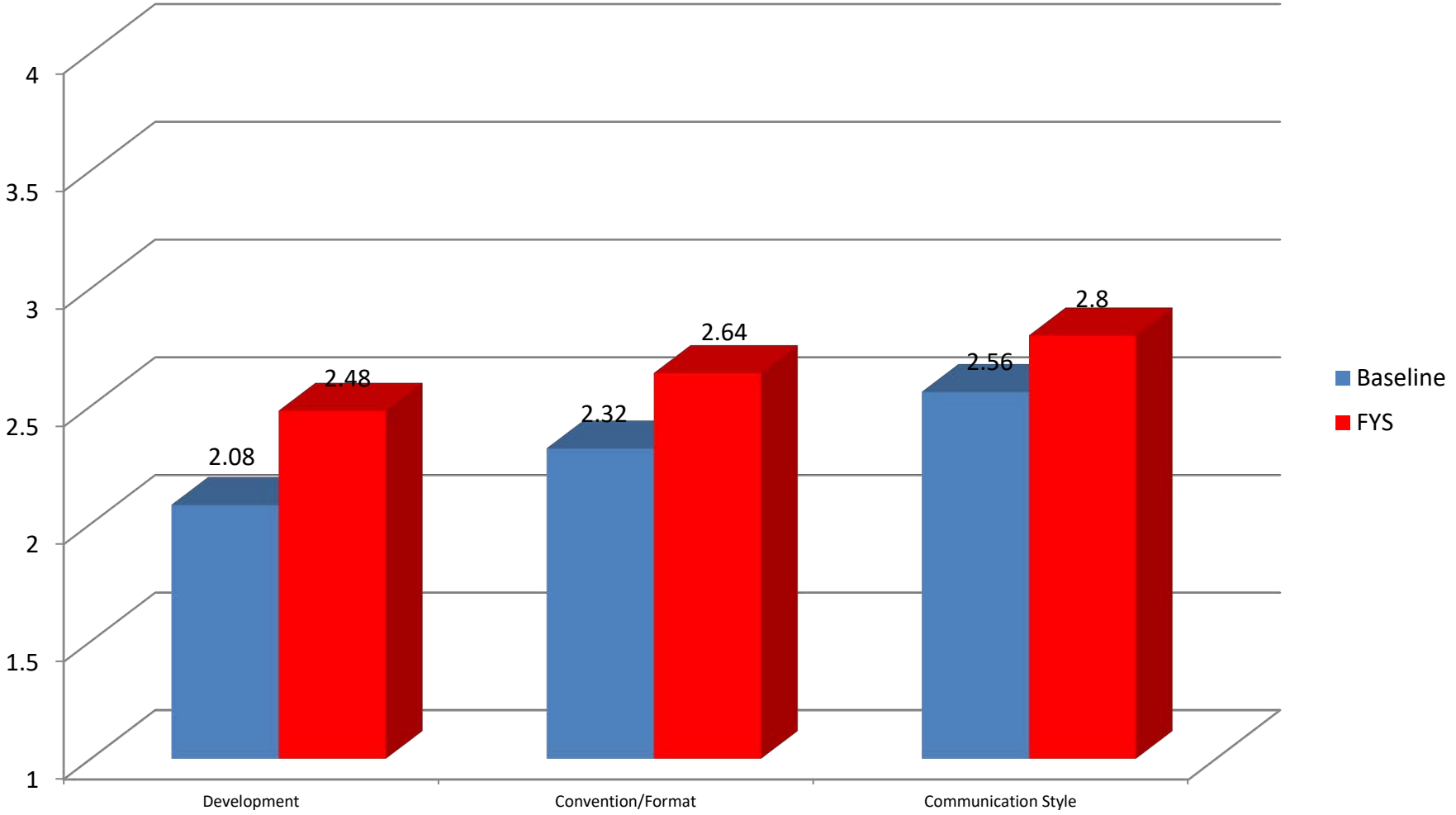
Trait/ Agreement	Info Needed : Cohen's Kappa (Liberal) = .960	Acknowledgment of Sources: Cohen's Kappa (Liberal) = .978	Evidence: Cohen's Kappa (Liberal) = .978	Viewpoints: Cohen's Kappa (Liberal) = 1.000	Recommendations: Cohen's Kappa (Liberal) = .911
Agree on score	120 (69%)	123 (70%)	104 (59%)	132 (75%)	83 (47%)
Difference = 1 point	50 (29%)	49 (28%)	68 (39%)	43 (25%)	80 (46%)
Difference = 2 points	5 (3%)	3 (2%)	3 (2%)	0	12 (7%)
Difference = 3 points	0	0	0	0	0
Total	**175**	**175**	**175**	**175**	**175**

Freshman Baseline/FYS Comparisons

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

n = 172 matched pairs

Mean differences between baseline and FYS were statistically significant for *all traits.*



Freshman Baseline/FYS Comparisons

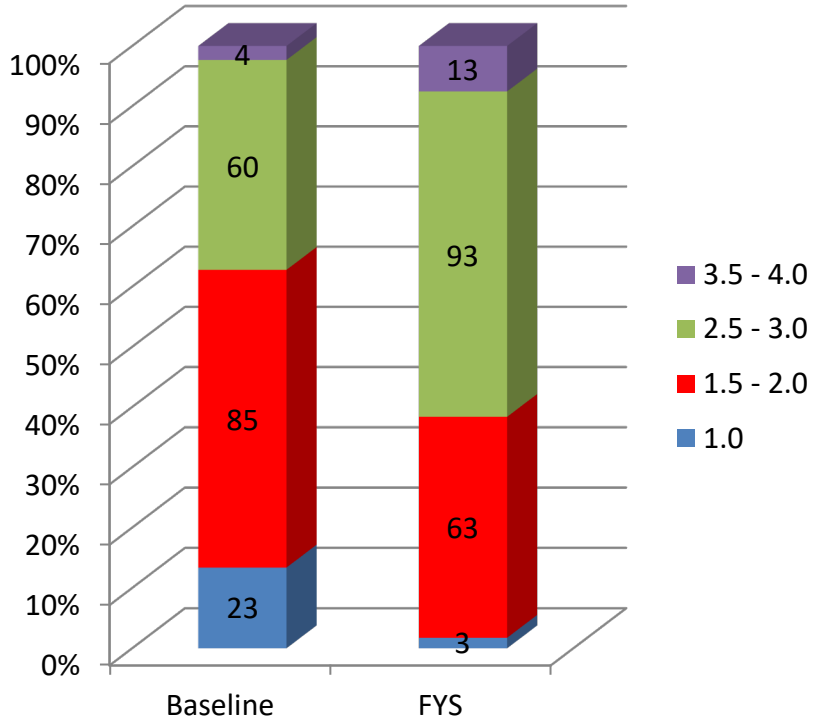
n = 172 matched pairs

Trait/ Performance Level	Development Baseline	Development FYS	Convention/ Format Baseline	Convention/ Format FYS	Communication Style Baseline	Communication Style FYS
1.0	23 (13%)	3 (2%)	25 (15%)	10 (6%)	2 (1%)	1 (1%)
1.5 – 2.0	85 (49%)	63 (37%)	51 (30%)	27 (16%)	43 (25%)	15 (9%)
2.5 – 3.0	60 (35%)	93 (54%)	74 (43%)	115 (67%)	118 (69%)	148 (86%)
3.5 – 4.0	4 (2%)	13 (8%)	22 (13%)	20 (12%)	9 (5%)	8 (5%)
Totals	**172**	**172**	**172**	**172**	**172**	**172**

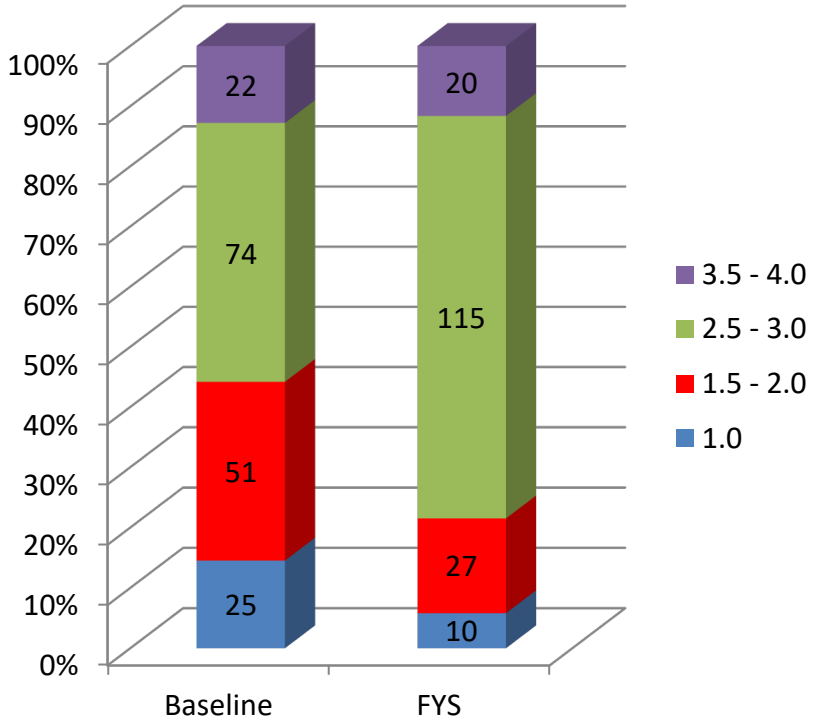
Freshman Baseline/FYS Comparisons

n = 172 matched pairs

Development



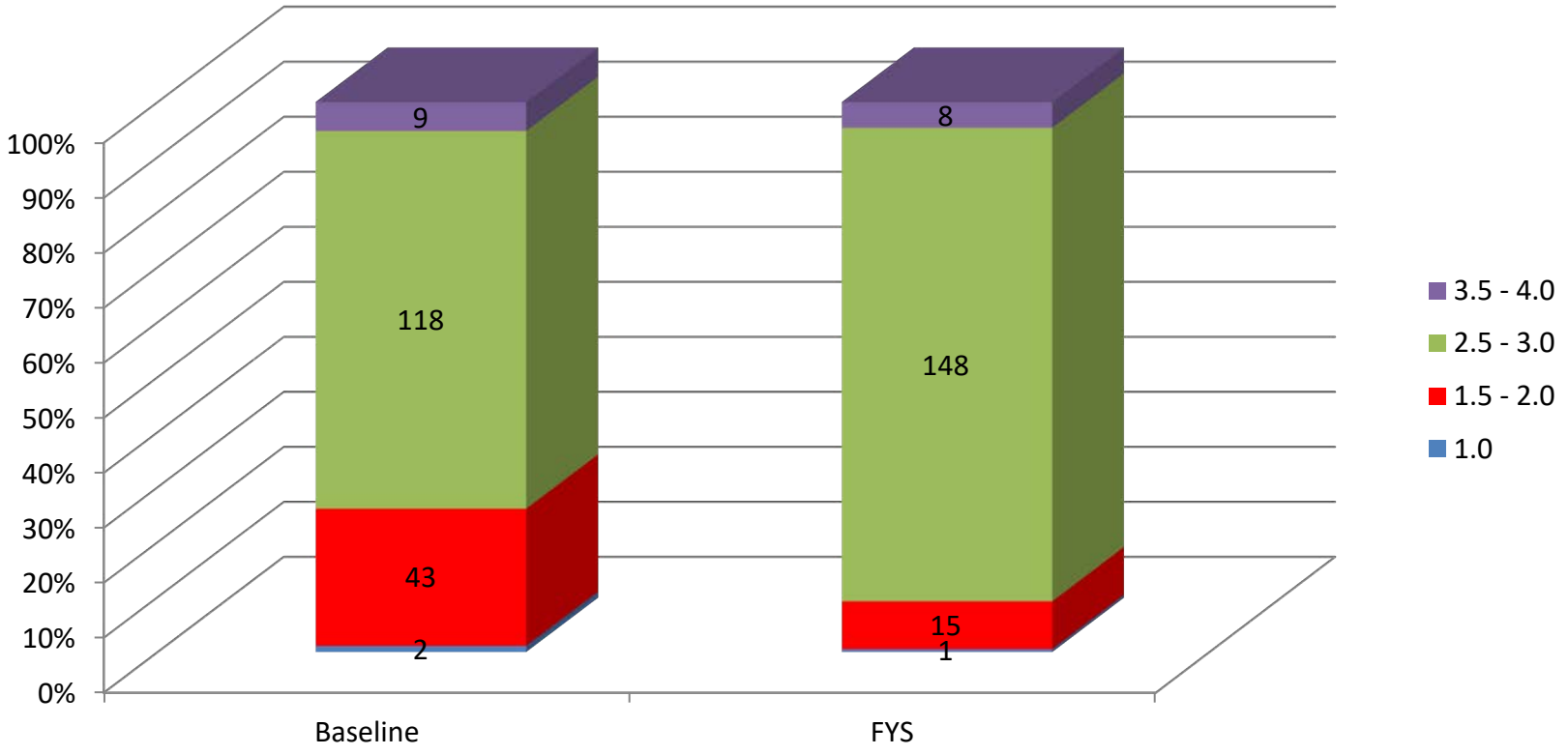
Convention/Format



Freshman Baseline/FYS Comparisons

n = 172 matched pairs

Communication Style



Baseline Inter-Rater Agreement Results

Includes 172 baseline assessments scored

Trait/ Agreement	Development: Cohen's Kappa (Liberal) = .985	Convention/Format: Cohen's Kappa (Liberal) = .908	Communication Style: Cohen's Kappa (Liberal) = .976
Agree on score	106 (62%)	89 (52%)	97 (56%)
Difference = 1 point	64 (37%)	70 (41%)	72 (42%)
Difference = 2 points	2 (1%)	13 (8%)	3 (2%)
Difference = 3 points	0	0	0
Total	**172**	**172**	**172**

FYS Inter-Rater Agreement Results

Includes all 175 baseline assessments scored

Trait/ Agreement	Development: Cohen's Kappa (Liberal) = .963	Convention/Format: Cohen's Kappa (Liberal) = .930	Communication Style: Cohen's Kappa (Liberal) = 1.000
Agree on score	104 (59%)	97 (55%)	123 (70%)
Difference = 1 point	66 (38%)	69 (39%)	52 (30%)
Difference = 2 points	4 (2%)	9 (5%)	0
Difference = 3 points	1 (1%)	0	0
Total	**175**	**175**	**1745**

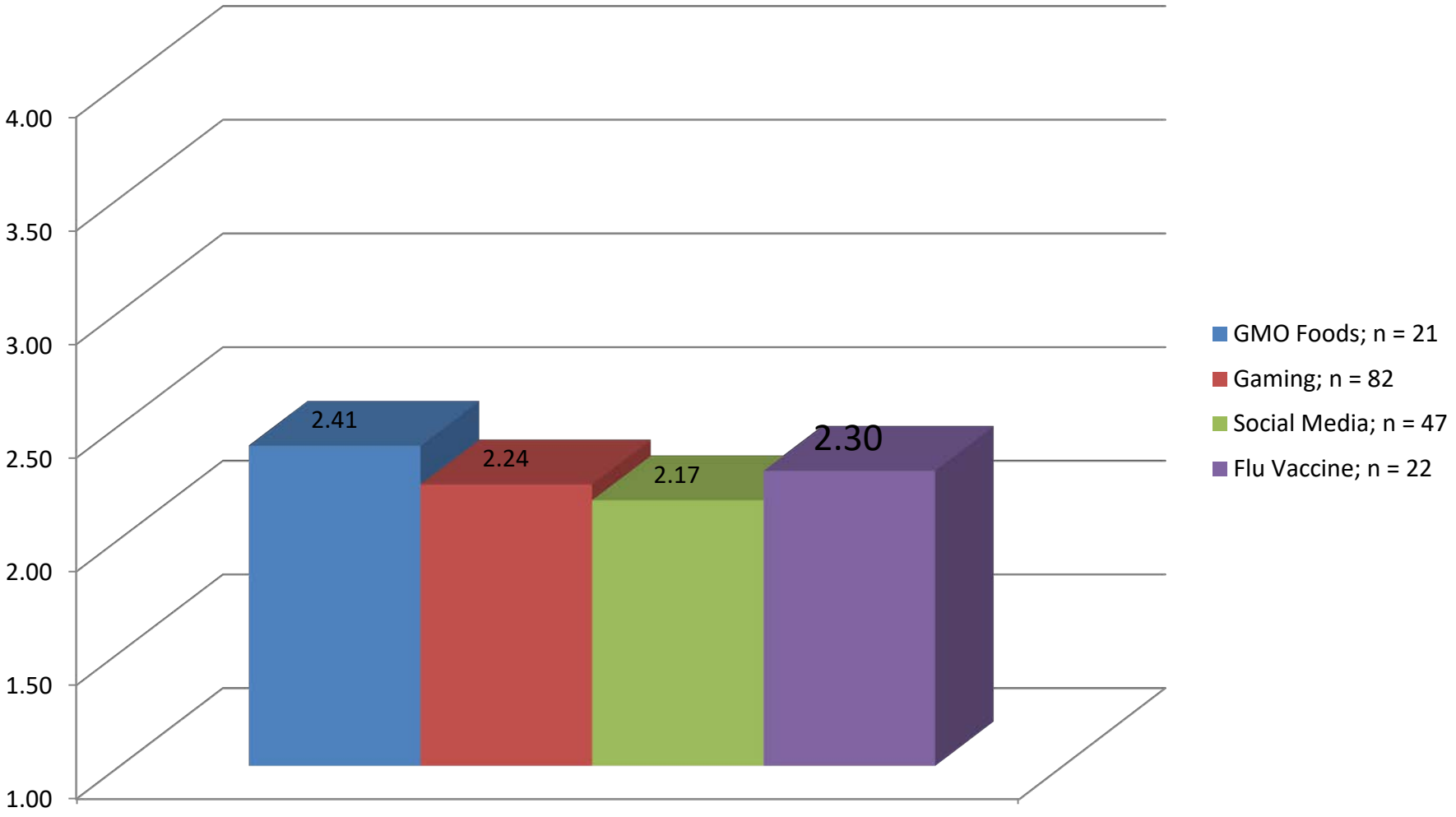
Comparison of FYS Results for Each Trait by Scenario

Academic Year 2023 - 2024

FYS Comparisons by Scenario for IL: Information Needed

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

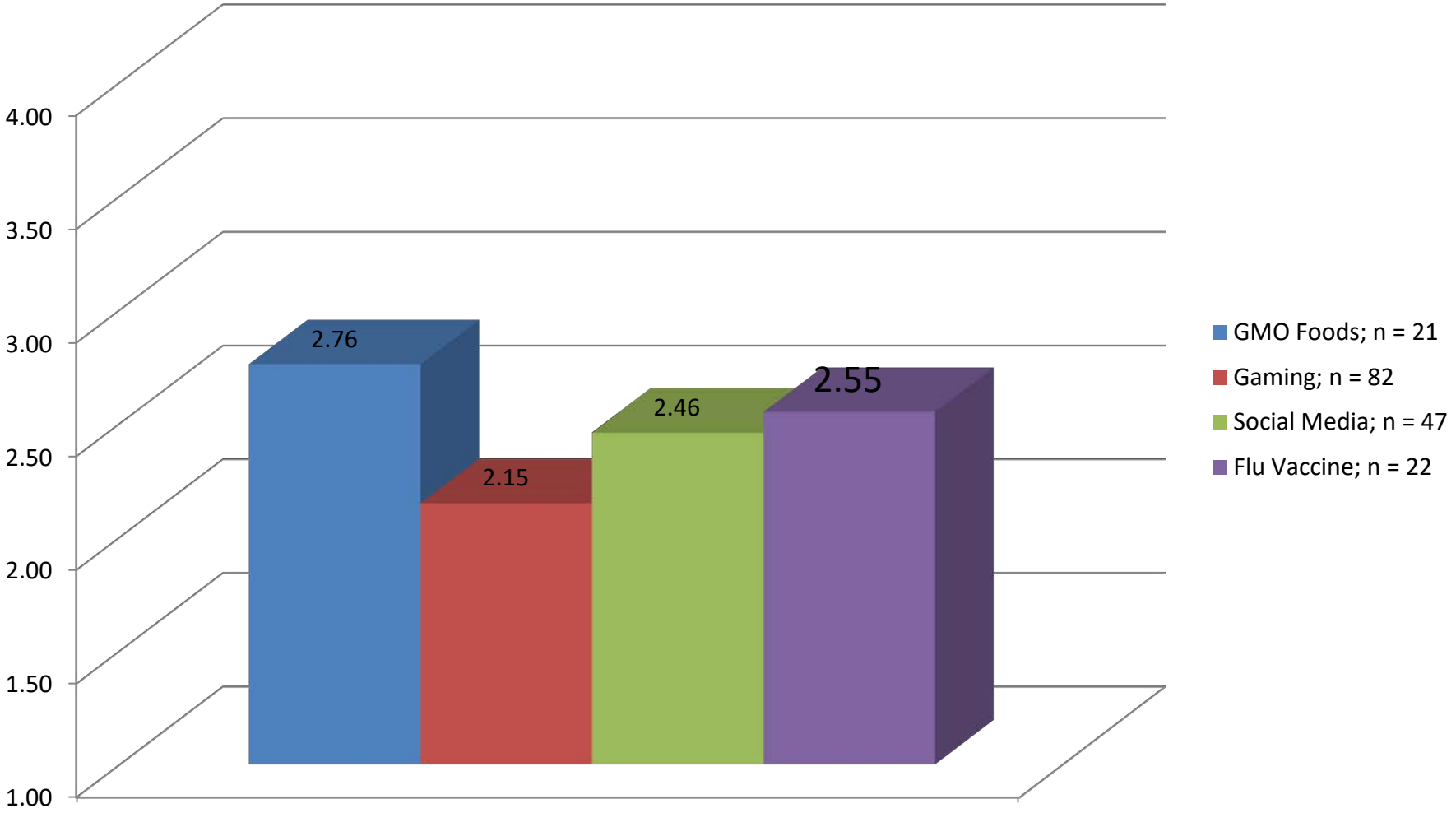
A One-Way ANOVA revealed no statistically significant differences in means across the scenarios.



FYS Comparisons by Scenario for IL: Source Acknowledgment

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

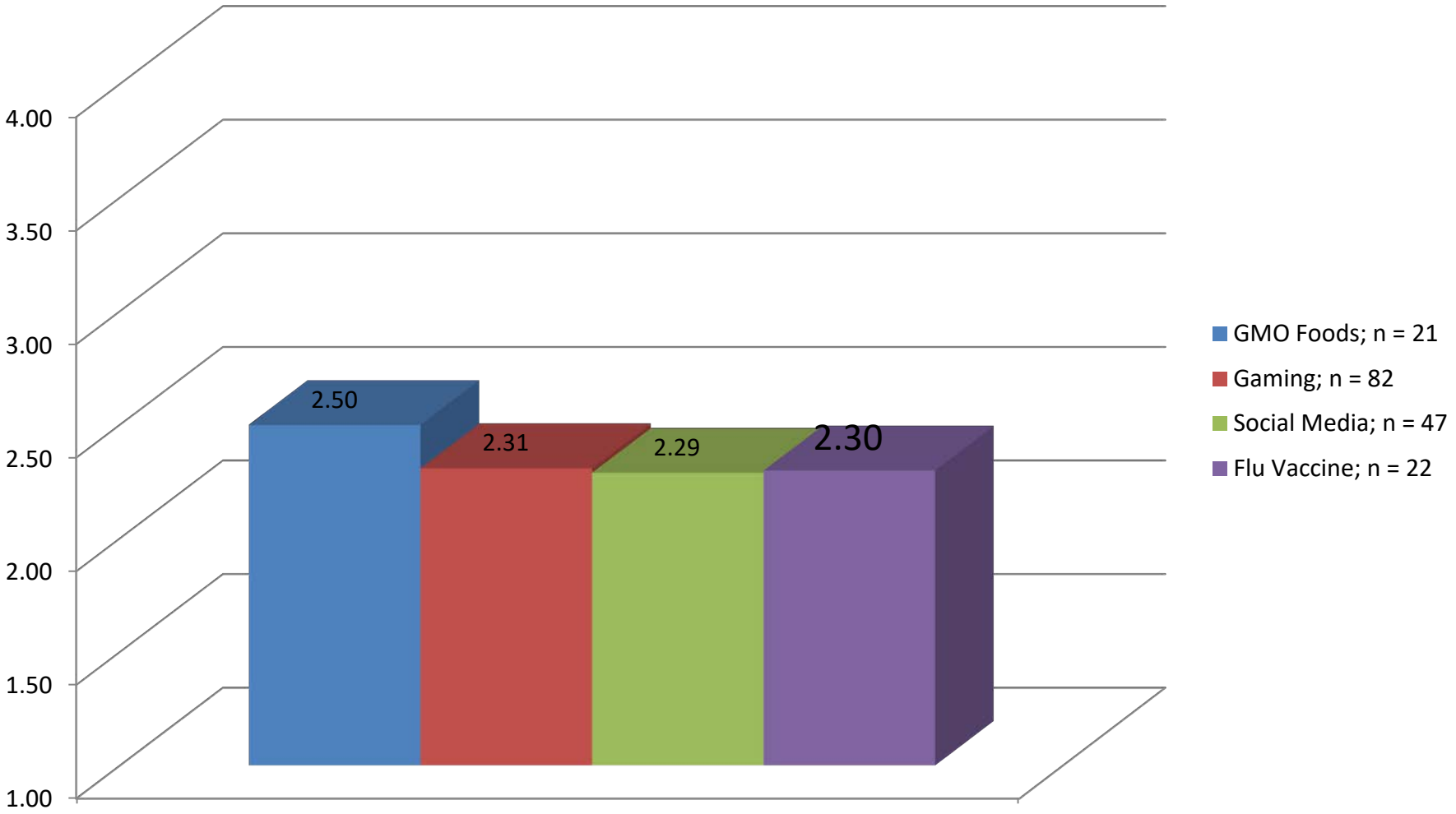
A One-Way ANOVA showed statistical significance across the scenarios. Bonferroni post-hoc analysis revealed that the mean for GMO Foods was significantly higher than the mean for Gaming.



FYS Comparisons by Scenario for CT: Evidence

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

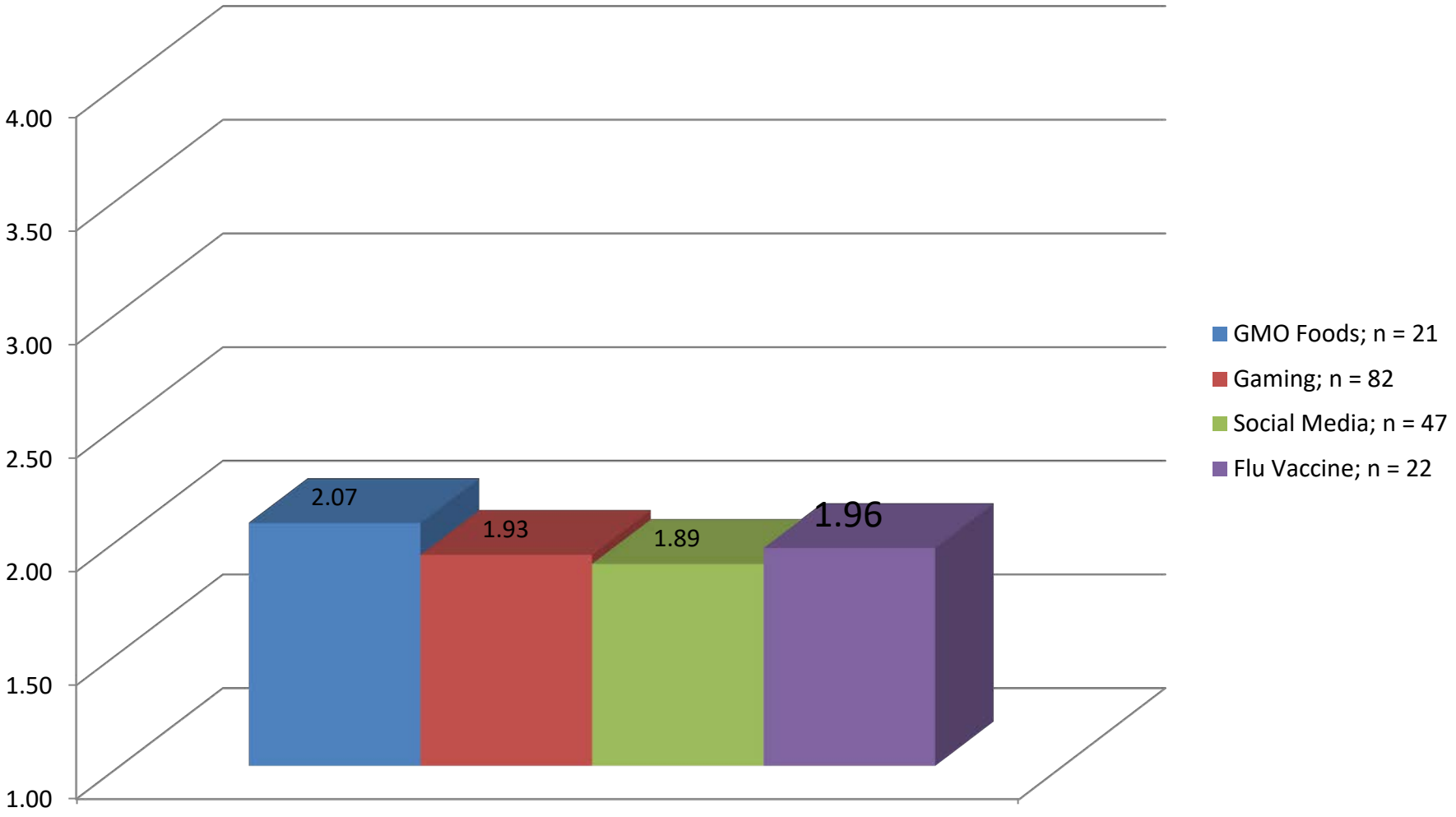
A One-Way ANOVA revealed no statistically significant differences in means across the scenarios.



FYS Comparisons by Scenario for CT: Viewpoints

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

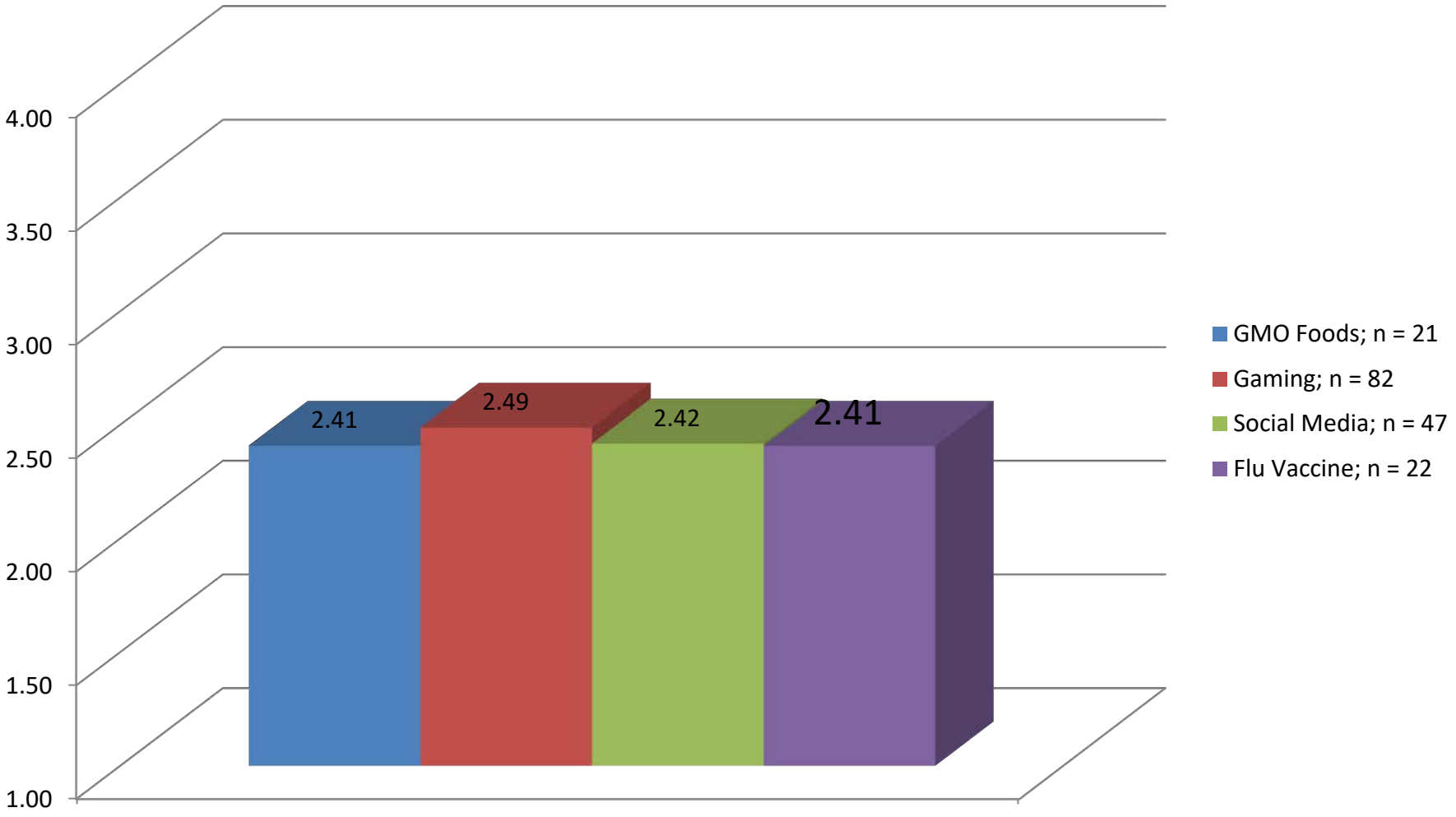
A One-Way ANOVA revealed no statistically significant differences in means across the scenarios.



FYS Comparisons by Scenario for CT: Recommendation

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

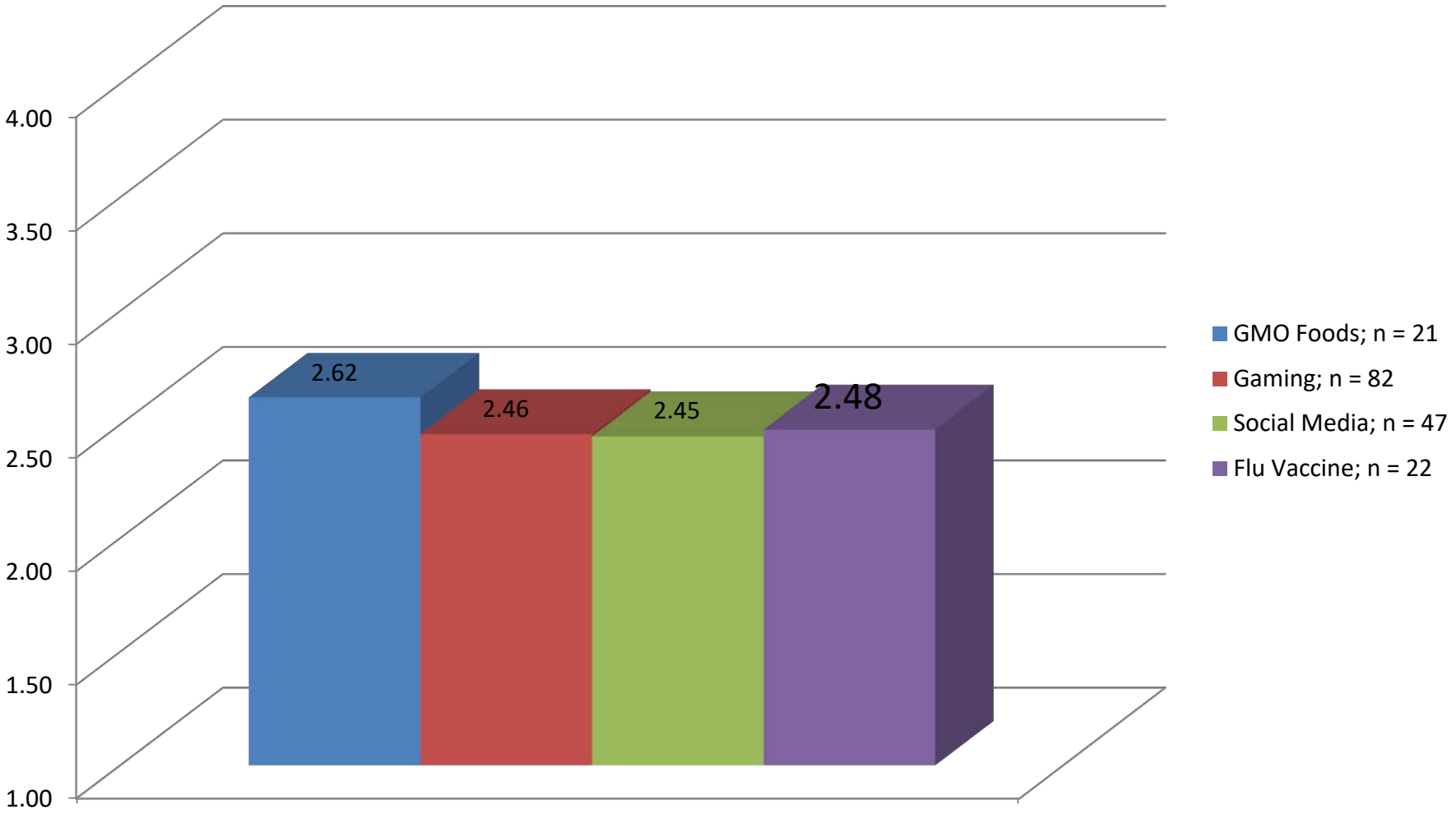
A One-Way ANOVA revealed no statistically significant differences in means across the scenarios.



FYS Comparisons by Scenario for CF: Development

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

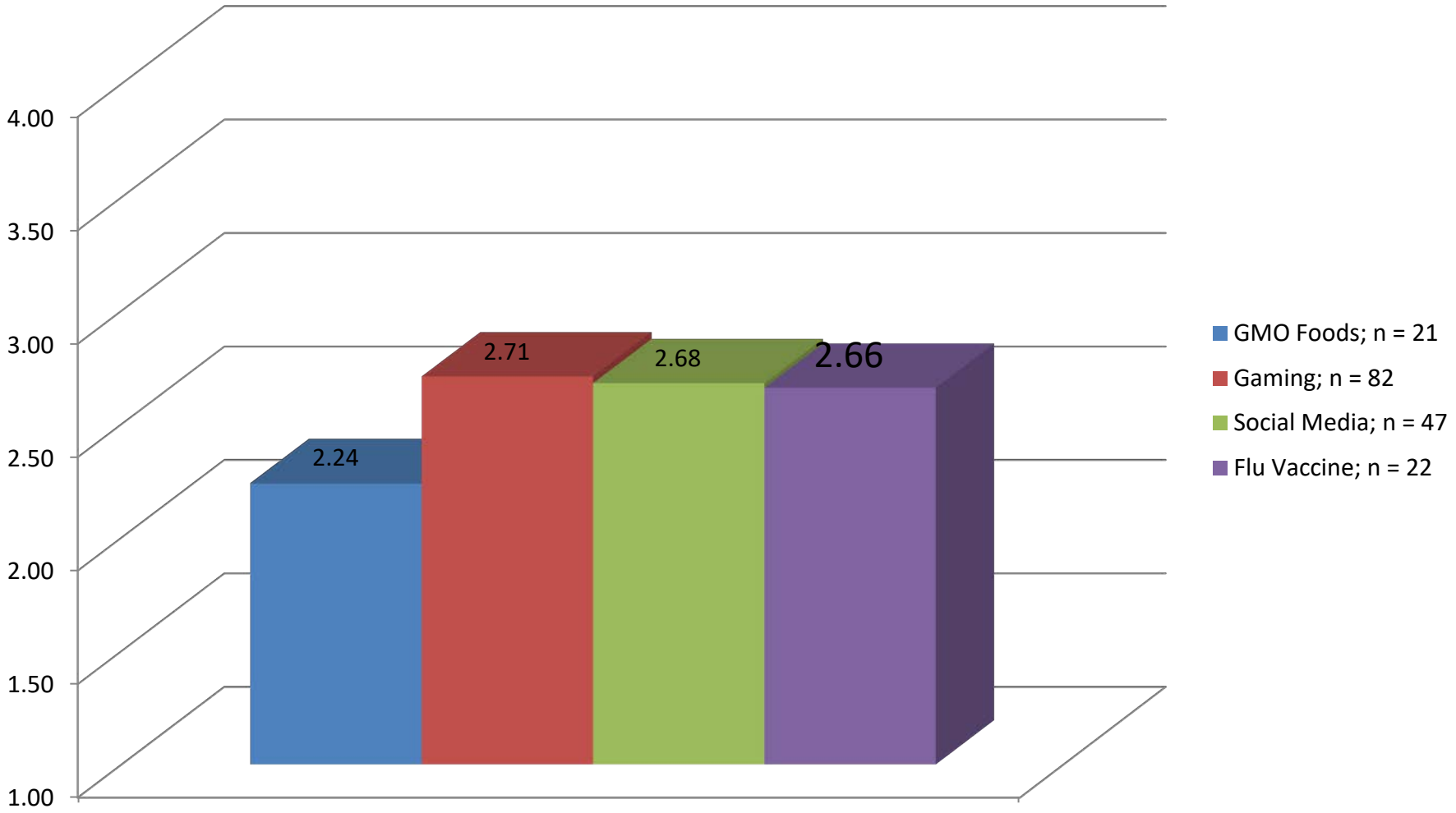
A One-Way ANOVA revealed no statistically significant differences in means across the scenarios.



FYS Comparisons by Scenario for CF: Convention/Format

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

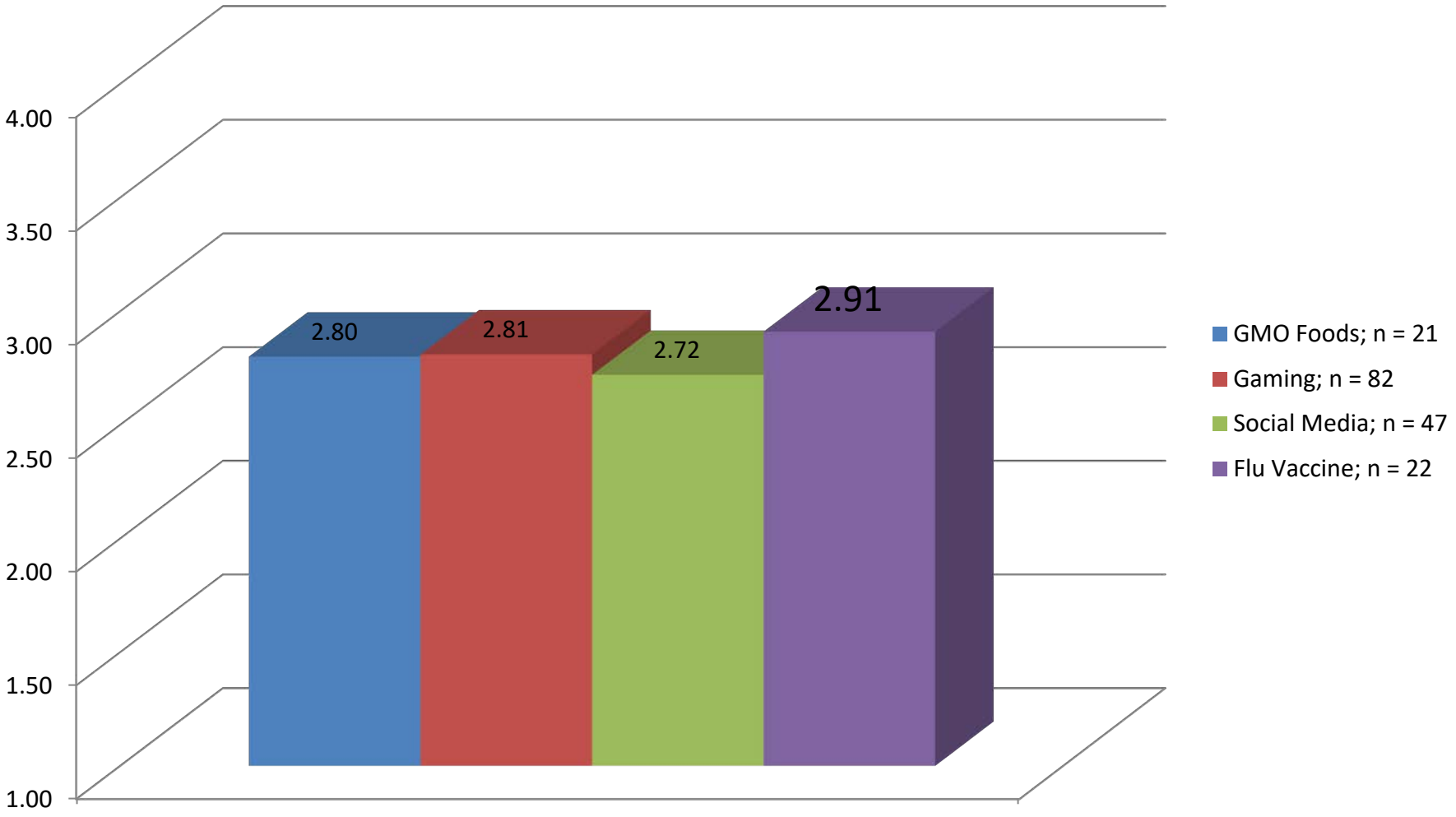
A One-Way ANOVA showed statistical significance across the scenarios. Bonferroni post-hoc analysis revealed that the mean for GMO Foods was significantly lower than the mean for Gaming.



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 no statistically significant differences in means across the scenarios.



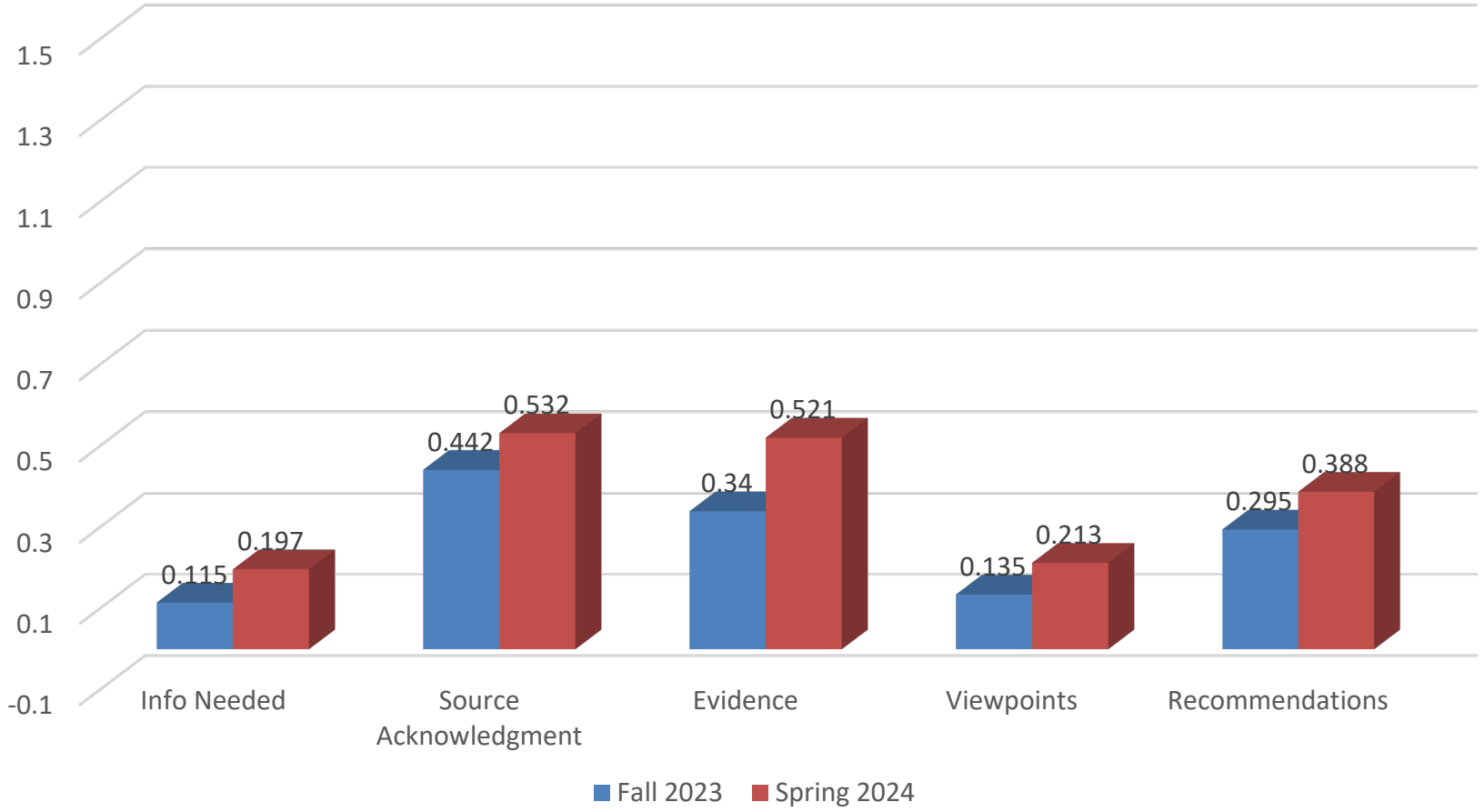
Comparison of Baseline to FYS Mean Gain Score for Each Trait by Semester of FYS

Academic Year 2023 - 2024

Baseline to FYS Mean Gain Scores for Each Trait

n = 78 in fall and 94 in spring

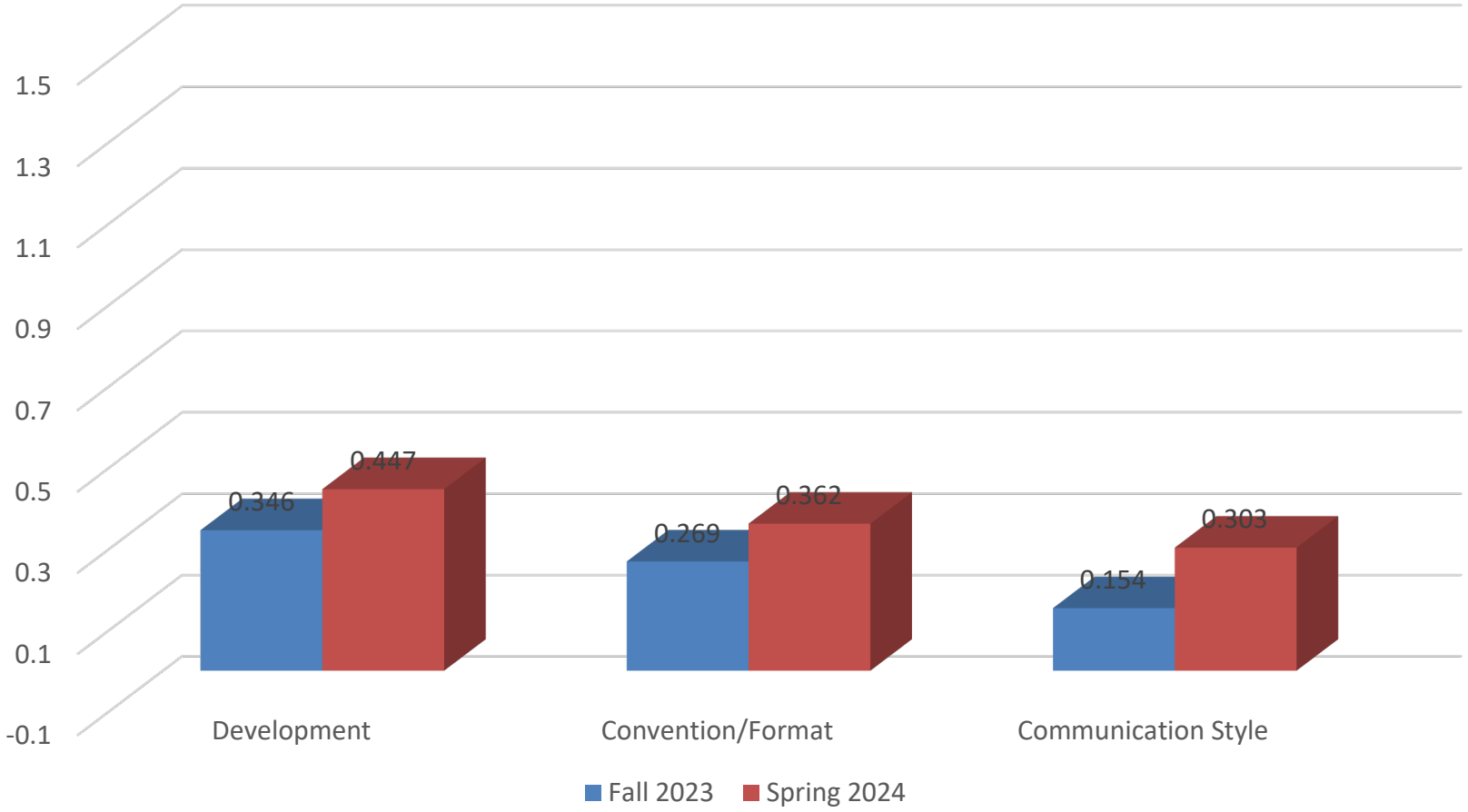
(Mean differences between fall and spring were not statistically significant)



Baseline to FYS Mean Gain Scores for Each Trait

n = 78 in fall and 94 in spring

(Mean 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.