



## Comparison of Freshman Baseline with First Year Seminar Assessment Results Academic Year 2018 – 2019

**Summer Assessment Team 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 Chris Sochor

### Executive Summary

#### *Background*

*Recommendations from the 2018 Assessment Workgroup (with current status in red)*

#### *Recommendations regarding baseline assessments*

1. The Assessment Team 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 during the first week of their UNI 100 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 100 course director. **The FYS Director worked with a group of faculty who developed new scenarios for baseline assessment in UNI 100. This team consulted with Mary Beth Reynolds, Chris Sochor, and the Associate**

Dean of Undergraduate Studies/Director of University College regarding the deployment of these scenarios as part of UNI 100 during the first week of the fall 2018 semester.

2. The Assessment Team 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. Beginning in fall 2018 the baseline and FYS exams mirrored each other in terms of content, format, and instructions. Each exam began by asking students first to evaluate each of seven documents for credibility and relevance, second to indicate information still needed to answer the question posed, and third to provide a detailed response to the question posed.
3. The Assessment Team recommended that both baseline and FYS assessments should make explicit to students the convention to be used for their recommendations. Some scenarios provided examples of the convention to be used in the response, but this was not done uniformly across all scenarios.
4. The Assessment Team also recommended that total page length of documents in document library should be similar across assessments. Page lengths continued to vary across documents for each scenario. However, we recognize the difficulty of creating authentic assignments that meet this criterion and discuss this issue in our recommendations for this year.
5. The Assessment Team recommended that we review this report, especially the section comparing student performance on each scenario, before beginning our analysis in spring 2019. Group members recommended more careful analysis of scenario differences. As has been done in the past, the report for 2019 will provide results regarding differences in student performance among scenarios.

### ***Procedures for 2019 Assessment***

#### ***General Procedures***

In August 2018, 1,324 incoming freshmen at Marshall University completed and 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 finished by incoming freshmen.

In May 2019 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 traits included *information needed* and *source acknowledgment* (Information Literacy), *evidence*, *viewpoints*, and *recommendation/position* (Inquiry-Based [Critical] Thinking), and *development*, *convention/format*, and *communication style* (Communication Fluency). This project was coordinated by the Office of Assessment and Quality Initiatives.

A random sample of 200 Marshall Freshman FYS exams was drawn from the pool of 1,234 (16%) of the total number of assessments available. Forty-two of the 200 students whose FYS exams were pulled for assessment had no corresponding baseline exam. This reduced the number of possible baseline/FYS pre-post comparisons to 158. Of these 158 pre/post comparisons we noted the following:

1. Sixteen students with scorable FYS exams uploaded baseline exams that were not able to be scored. This reduced the baseline/FYS pre-post comparisons to 142.
2. Five students with scorable baseline exams uploaded FYS exams that were not able to be scored. This further reduced the baseline/FYS pre-post comparisons to 137.
3. One student uploaded scorable baseline and FYS exams but included no information for the trait *information needed* in the baseline exam. One other student uploaded scorable baseline and FYS exams but included only information for the trait *information needed* in the FYS exam, but no information for the other seven traits. This reduced the baseline/FYS pre-post comparisons to 136.
4. Two students uploaded both baseline and FYS exams, but both exams for both students were not able to be scored. This reduced the baseline/FYS pre-post comparisons to their final number of 134 (67% of the original FYS sample of 200).

The reasons for the forty-two students who did not upload baseline exams are as follows:

Reason	Number of students
Completed UNI 100 in Fall 2017. Repeated FYS in Fall 2018 after having failed it during an earlier semester. Therefore, they did not have baselines from fall 2018.	5
Did not take UNI	14
Completed UNI in Fall 2017 but did not take FYS until academic year 2018-2019. Therefore, they had no baseline in fall 2018.	2
Took UNI 100 in Spring 2019, so were not included in the fall 2018 baseline sample.	1
Took UNI 100, but did not upload a baseline exam in Blackboard	17
Students were not able to be identified at the time of assessment because they submitted blank FYS exams	3
<b>Total</b>	<b>42</b>

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

Of the 200 exams in the FYS sample, 189 had usable scores for all traits and 191 also had usable scores for the Information Literacy: *information needed*. Mean scores (on a scale of 1 – 4) were 2.49 for Information Literacy: *information needed*, 2.18 for Information Literacy: *source acknowledgment*, 2.22 for Inquiry-Based (Critical) Thinking: *evidence*, 2.01 for Inquiry-Based (Critical) Thinking: *viewpoints*, 2.25 for Inquiry-Based (Critical) Thinking: *recommendations*, 2.24 for Communication Fluency: *development*, 2.25 for Communication Fluency: *convention/format*, and 2.47 for Communication Fluency: *communication style*. These means did not differ significantly from the mean FYS results for the 134 students whose FYS scores remained in the Baseline/FYS analysis.

One hundred fifty-eight of the students who uploaded FYS exams also uploaded baseline assessments. Of these, there were usable scores for 140 baseline assessments for the last seven traits, and 139 for Information Literacy: *information needed*. Mean scores (on a scale of 1 – 4) were 2.33 for Information Literacy: *information needed*, 1.78 for Information Literacy: *source acknowledgment*, 2.03 for Inquiry-Based (Critical) Thinking: *evidence*, 1.87 for Inquiry-Based (Critical) Thinking: *viewpoints*, 2.10 for Inquiry-Based (Critical) Thinking: *recommendations*, 2.03 for Communication Fluency: *development*, 1.96 for Communication Fluency: *convention/format*, and 2.38 for Communication Fluency: *communication style*. These means did not differ significantly from the mean baseline results for the 134 students whose baseline scores remained in the Baseline/FYS analysis.

The baseline and FYS means (and standard deviations) for the 134 students in the sample with scorable baseline and FYS exams are outlined 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 Information literacy), (.017 for Inquiry-Based [Critical] Thinking), and (.017 for Communication Fluency) showed significant mean differences between freshman baseline and FYS results for all traits of Information Literacy and Inquiry-Based (Critical) Thinking, with students performing better on FYS assessments than on baseline assessments. Likewise, performance improved significantly between baseline and FYS for two traits of Communication Fluency (*development* and *convention/format*). We note that Communication Fluency is not an explicit learning outcome addressed in FYS.

Outcome	Trait	Baseline Mean (SD)	FYS Mean (SD)	Statistical Significance
Information Literacy	Information Needed	2.34 (0.67)	2.52 (0.70)	$t(133) = 2.494, p = .014$
	Source Acknowledgment	1.77 (0.80)	2.27 (0.95)	$t(133) = 5.486, p < .001$
Inquiry-Based (Critical) Thinking	Evidence	2.02 (0.69)	2.25 (0.73)	$t(133) = 3.129, p = .002$
	Viewpoints	1.87 (0.50)	2.01 (0.47)	$t(133) = 3.075, p = .003$
	Recommendation/Position	2.09 (0.56)	2.25 (0.69)	$t(133) = 2.731, p = .007$
Communication Fluency	Development	2.02 (0.69)	2.24 (0.71)	$t(133) = 3.154, p = .002$
	Convention/Format	1.95 (0.76)	2.28 (0.80)	$t(133) = 3.890, p < .001$
	Communication Style	2.37 (0.57)	2.50 (0.60)	$t(133) = 2.405, p = .018$

A frequency analysis also showed that the following increases in students scoring between 2.5 and 4.0 on the rubric. 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	26%
Inquiry-Based (Critical) Thinking	Evidence	9%
	Viewpoints	8%
	Recommendation/Position	15%
Communication Fluency	Development	14%
	Convention/Format	23%
	Communication Style	13%

This year’s results showed some significant differences in performance based on scenario used for the FYS assessments. In general, students performed better on the College Costs scenario than on the High-Tech Policing scenario. Also, gain scores between students on our sample who completed FYS in fall 2018 ( $n = 70$ ) and those who completed FYS in spring 2019 ( $n = 64$ ) did not differ significantly on any outcome trait. Please refer to the supporting documentation for additional detail.

**Recommendations from the 2019 Assessment Team**

1. The Assessment Team recommends that a consistent format and rhetorical situation be adopted for all baseline and FYS scenarios. This year there were different formats, e.g., some scenarios asked students to write an essay, others a memo, and one asked that students develop a lesson plan. The lesson plan scenario resulted in depressed scores for all traits of Inquiry-Based (Critical) Thinking. The team felt that it was important that all scenarios allow students to grapple with a question, issue, or problem that has at least two possible answers or solutions. The format must allow the students to justify their position/recommendation by using evidence (which they have carefully evaluated for relevance and credibility), by considering multiple points of view and potential consequences of the recommendation they make.
2. The Assessment Team realizes the challenges of developing strong *authentic* scenarios that will engage students in significant critical thinking/problem solving. They further recognize the challenges of finding significant sources that are of uniform page length across multiple scenarios. For this reason, the team strongly suggests that the baseline and FYS exercises be divided into two parts. The first part should not be timed and should occur before the student takes the final part of the assessment. During Part I of the assessment, students should read each document thoroughly and evaluate each for credibility and relevance. They should also include a short summary of each document. For the second part of the assessment, which (for FYS) occurs during a two-hour time block during Marshall’s final exam week, students should bring their notes regarding each document, the documents themselves, and their summaries and evaluations of each document regarding its credibility and relevance. Then, during Part II of the assessment, students should complete the section of the assessment that asks them to outline additional information they would like to have to propose an answer or solution. Finally, with the

information they have, they should write their recommendation in the format required. Although we realize that, due to scheduling constraints, all students in each section of UNI 100 cannot complete baseline assessments at the same time, we recommend that, when students begin Part II of the assessment, Blackboard provide them with only a two-hour window to complete the exam to make it compatible with the timeframe for FYS exams. The Assessment Team hopes that this process will provide students with enough time to carefully read and evaluate each document and that students will have sufficient time to thoughtfully prepare their recommendations. The team feels that we should explore the possibility of a similar time sequence for the baseline assessment, which must occur in UNI 100 during the first week of the fall term.

3. The Assessment Team's final recommendation is that the timeframe to complete baseline assessments be extended through week two of the fall semester. This will enable students who enroll in UNI 100 after the first week to complete the assessment and give all students enough time to complete both parts of the task.



Supporting Documentation



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

Academic Year 2018 - 2019



# Review Procedures

- Two hundred (200) First Year Seminar final exam artifacts were used for this evaluation. The FYS final exam assessment sample represented approximately 16% of the 1,234 completed and submitted to Blackboard during fall 2018 and spring 2019.
- One hundred fifty-eight (158; 79%) of the 200 freshmen whose FYS assessments were sampled completed similar baseline assessments upon matriculation. Of these 158 students, we noted the following:
  1. Sixteen students with scorable FYS exams uploaded baseline exams that were not able to be scored. This reduced the baseline/FYS pre-post comparisons to 142.
  2. Five students with scorable baseline exams uploaded FYS exams that were not able to be scored. This reduced the baseline/FYS pre-post comparisons to 137.
  3. One student uploaded scorable baseline and FYS exams but included no information for the trait *information needed* in the baseline exam. One other student uploaded scorable baseline and FYS exams but included only information for the trait *information needed* in the FYS exam, but no information for the other seven traits. This reduced the baseline/FYS pre-post comparisons to 136.
  4. Two students uploaded both baseline and FYS exams, but both exams for both students were not able to be scored. This reduced the baseline/FYS pre-post comparisons to their final number of 134 (67% of the original FYS sample of 200).

# 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

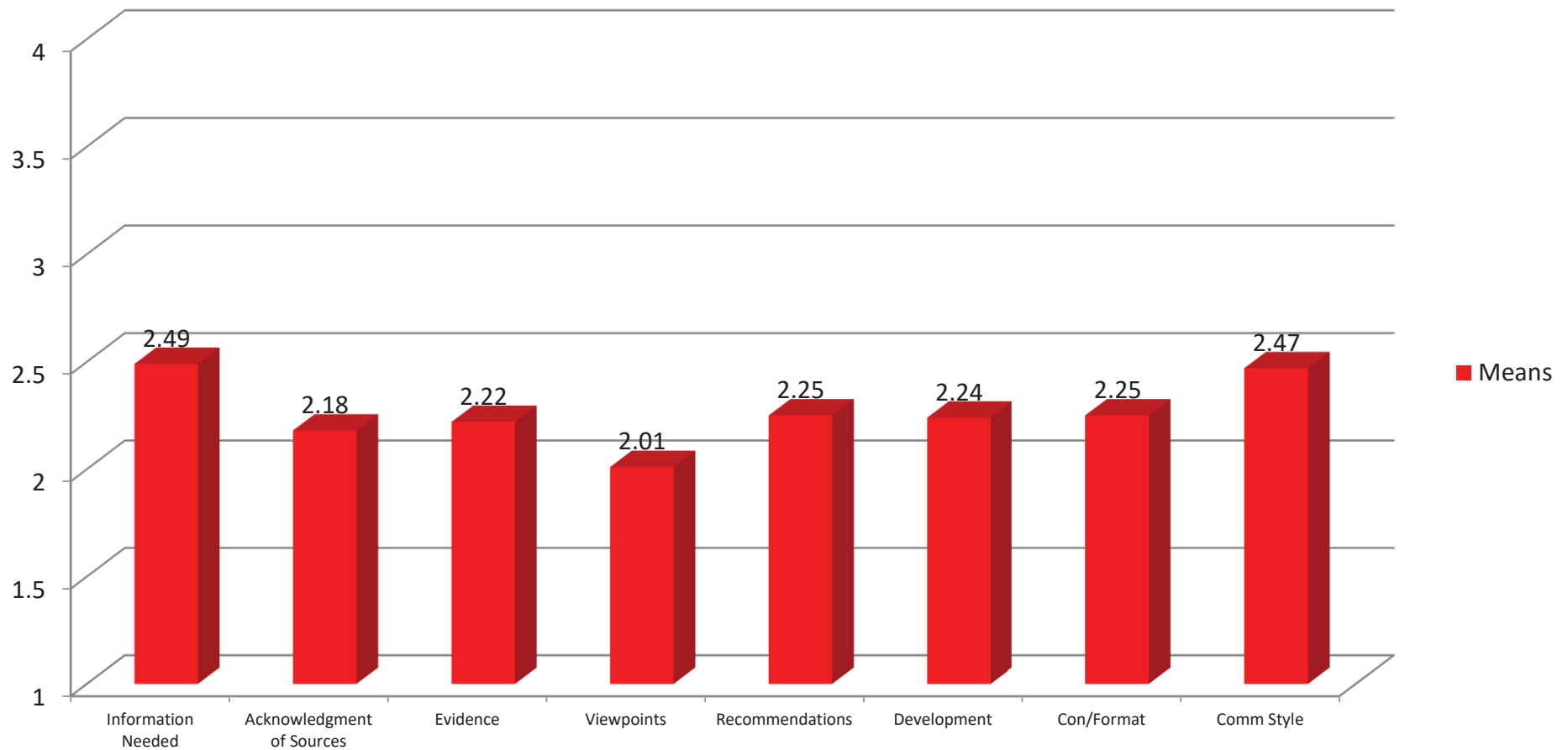
Baseline/FYS Assessment Rubric – Summer 2019 – updated 5-14-2018

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 <b>some</b> sources of information from the DL.	Clearly acknowledges <b>multiple</b> 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	<u>Either</u> does not make a recommendation <u>or</u> 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 <b>sophisticated</b> language that skillfully communicates meaning to readers with clarity and fluency, and is virtually error-free.

# Freshman FYS Exam Means

Mean Scores on a scale of 1 – 4, with 4 being the highest possible score  
 $n = 191$  for Information Needed and 189 for All Other Traits

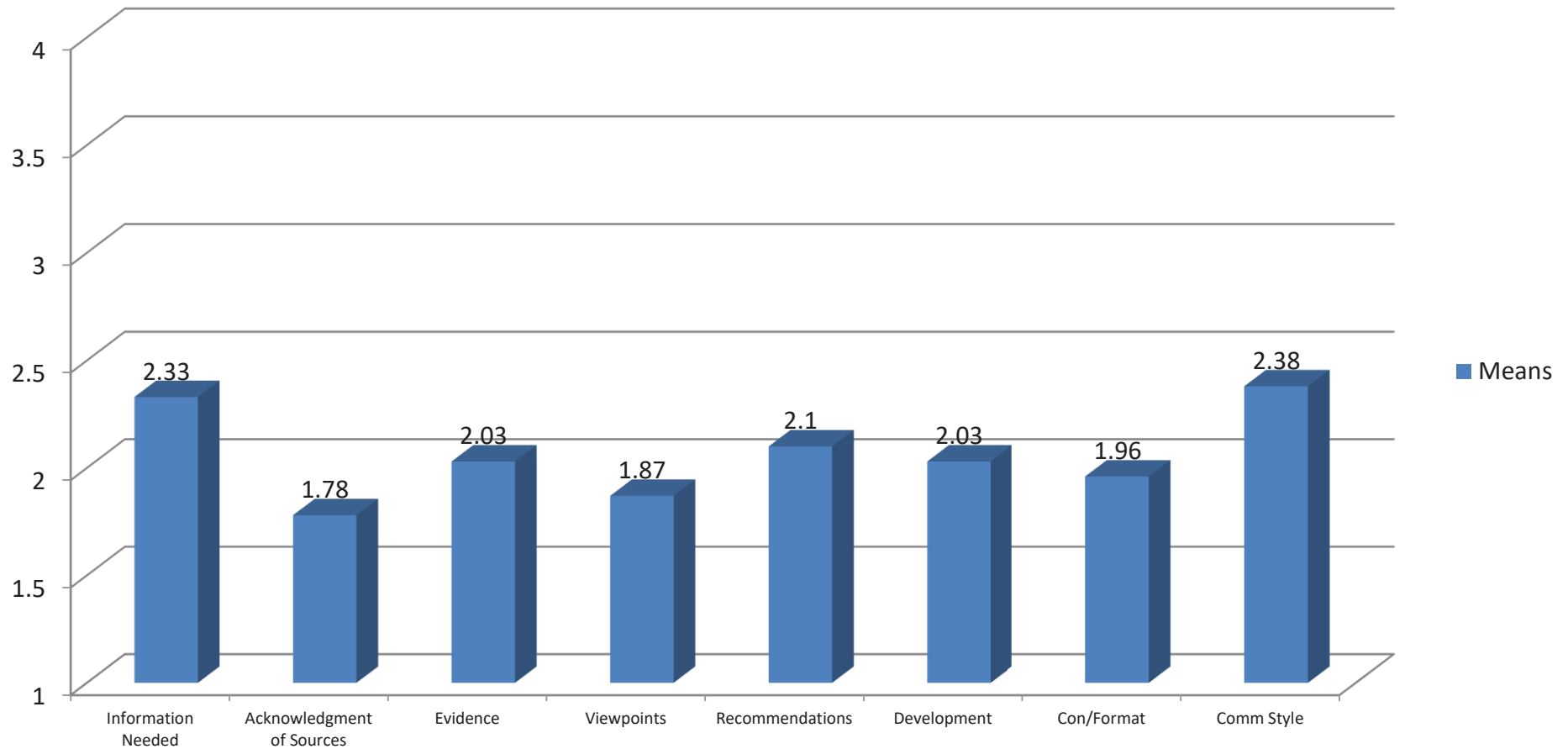
## FYS Final Exam Sample



# Freshman Baseline Exam Means

Mean Scores on a scale of 1 – 4, with 4 being the highest possible score  
 $n = 139$  for Information Needed and 140 for All Other Traits

## Baseline Exam Sample

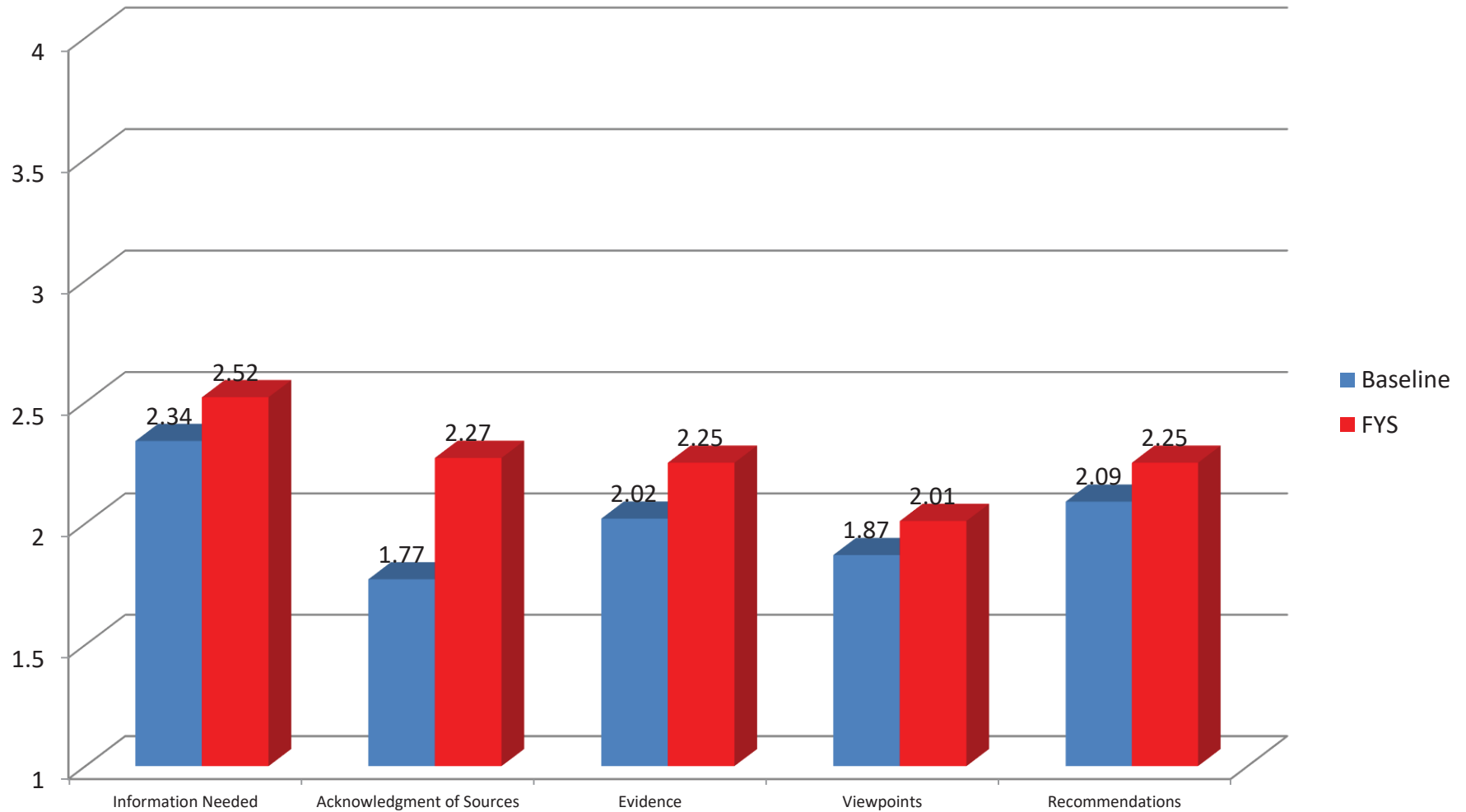


# Freshman Baseline/FYS Comparisons

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

$n = 134$  for all traits

Mean differences are statistically significant for all traits.



# Freshman Baseline/FYS Comparisons

*n* = 134

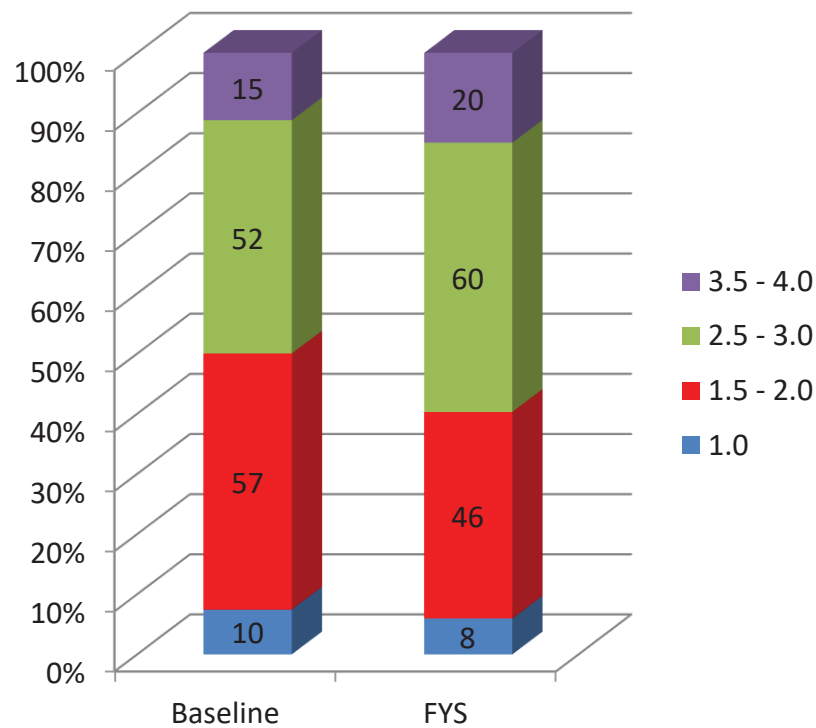
Trait/ Performance Level	Info Needed	Acknowledgment of Sources	Evidence	Viewpoints	Recommendations
1.0 Baseline	10 (7%)	46 (34%)	21 (16%)	19 (14%)	11 (8%)
1.0 FYS	8 (6%)	28 (21%)	15 (11%)	10 (7%)	14 (10%)
1.5 – 2.0 Baseline	57 (43%)	57 (43%)	66 (49%)	89 (66%)	75 (56%)
1.5 – 2.0 FYS	46 (34%)	41 (31%)	60 (45%)	86 (64%)	51 (38%)
2.5 – 3.0 Baseline	52 (39%)	21 (16%)	41 (31%)	25 (19%)	47 (35%)
2.5 – 3.0 FYS	60 (45%)	44 (33%)	49 (37%)	38 (28%)	63 (47%)
3.5 – 4.0 Baseline	15 (11%)	10 (7%)	6 (4%)	1 (1%)	1 (1%)
3.5 – 4.0 FYS	20 (15%)	21 (16%)	10 (7%)	0	6 (4%)
<b>Grand Total Baseline</b>	<b>134 (100%)</b>	<b>134 (100%)</b>	<b>134 (100%)</b>	<b>134 (100%)</b>	<b>134 (100%)</b>
<b>Grand Total FYS</b>	<b>134 (100%)</b>	<b>134 (100%)</b>	<b>134 (100%)</b>	<b>134 (100%)</b>	<b>134 (100%)</b>



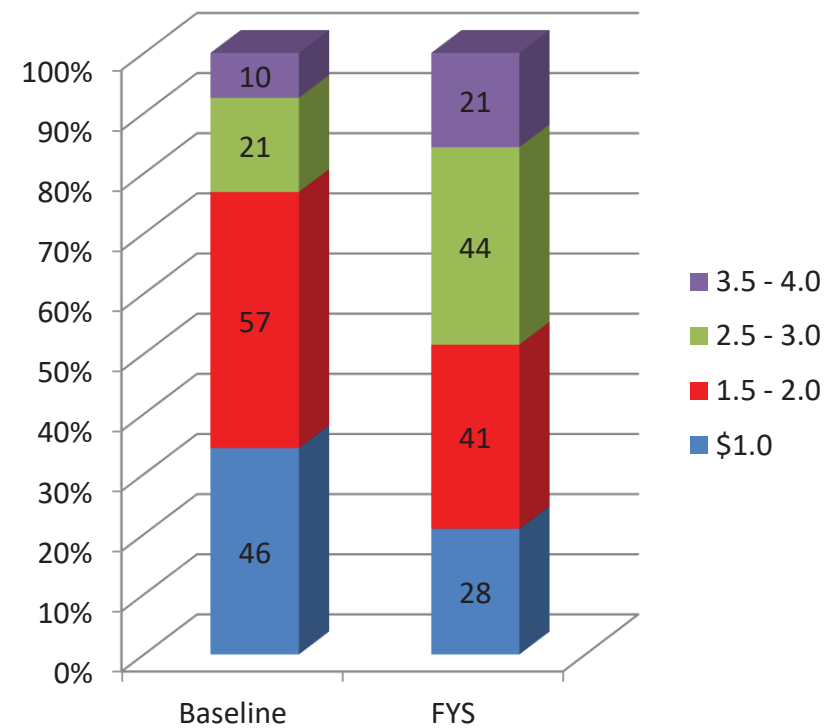
# Freshman Baseline/FYS Comparisons

$n = 134$

## Information Needed



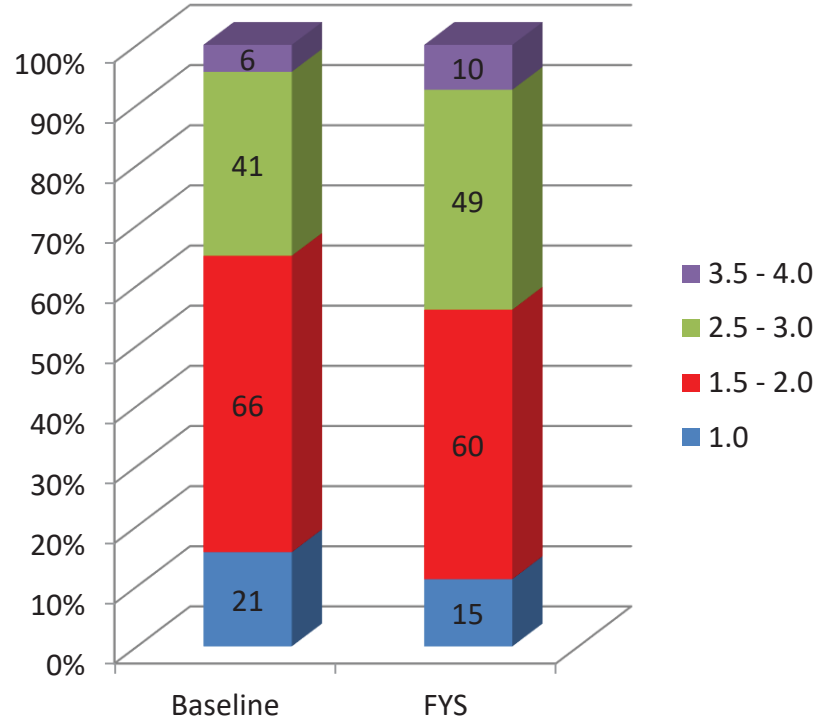
## Acknowledgment of Sources



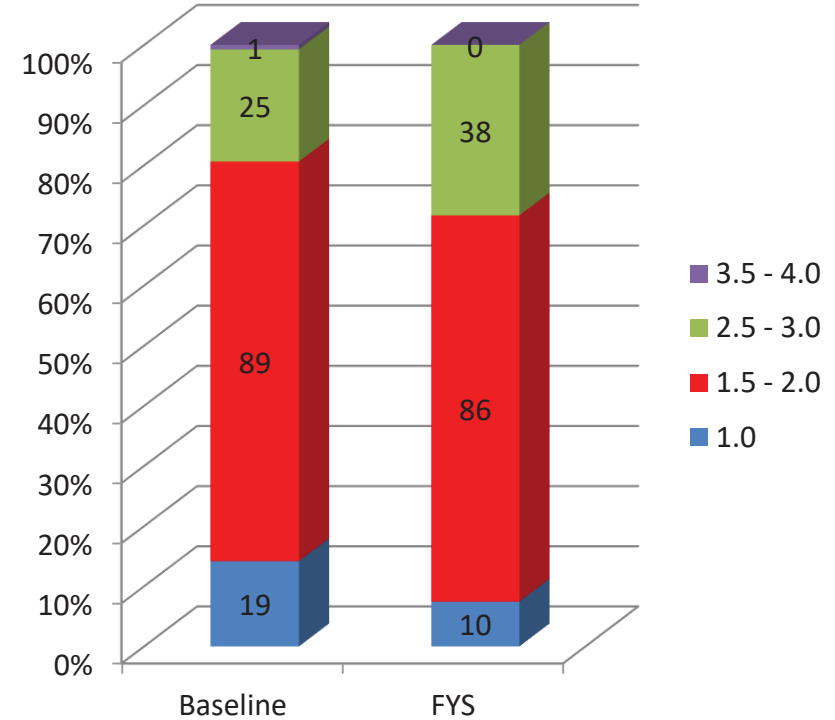
# Freshman Baseline/FYS Comparisons

$n = 134$

## Evidence



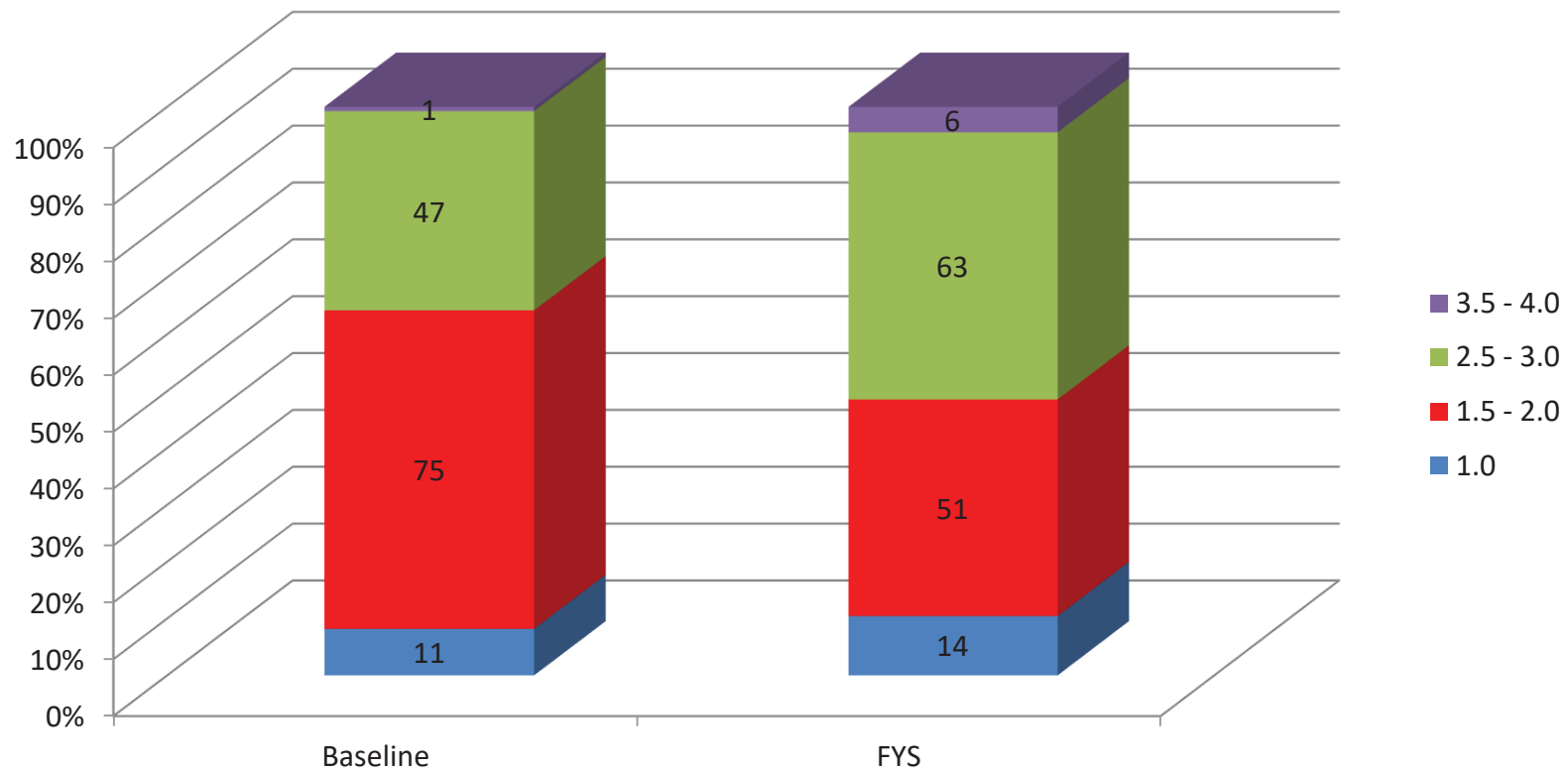
## Viewpoints



# Freshman Baseline/FYS Comparisons

$n = 134$

## Recommendations



# FYS Inter-Rater Agreement Results

Trait/ Agreement	Info Needed : Cohen's Liberal Kappa = .938	Acknowledgment of Sources: Cohen's Liberal Kappa = .968	Evidence: Cohen's Liberal Kappa = .914	Viewpoints: Cohen's Liberal Kappa = .969	Recommendations: Cohen's Liberal Kappa = .947
Agree on score	115 (60%)	120 (63%)	113 (60%)	114 (60%)	102 (54%)
Difference = 1 point	67 (35%)	64 (34%)	63 (33%)	71 (38%)	79 (42%)
Difference = 2 points	8 (4%)	4 (2%)	12 (6%)	4 (2%)	8 (4%)
Difference = 3 points	1 (1%)	1 (1%)	1 (1%)	0	0
<b>Total</b>	<b>191 (100%)</b>	<b>189 (100%)</b>	<b>189 (100%)</b>	<b>189 (100%)</b>	<b>189 (100%)</b>

# Baseline Inter-Rater Agreement Results

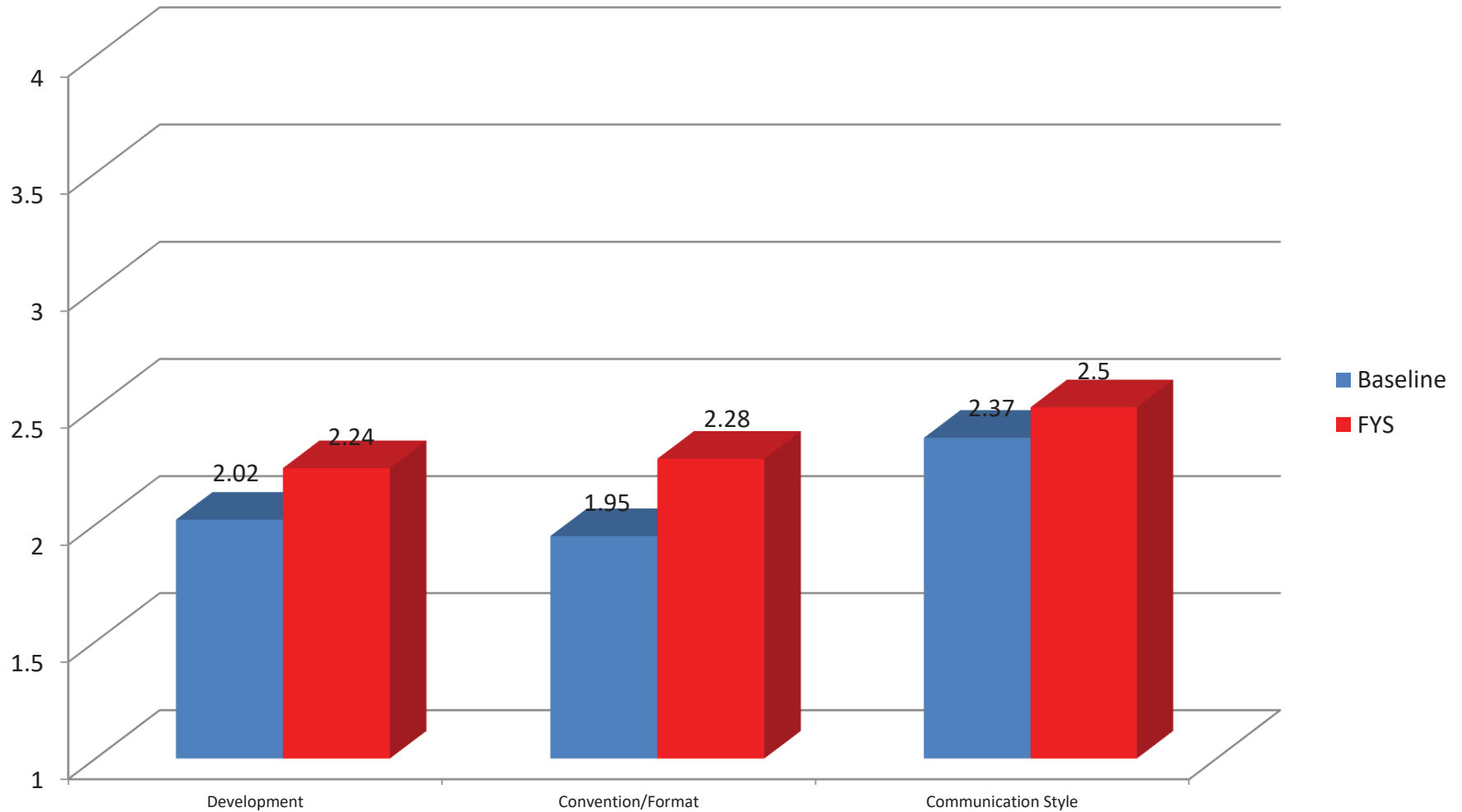
Trait/ Agreement	Info Needed : Cohen's Liberal Kappa = .963	Acknowledgment of Sources: Cohen's Liberal Kappa = .982	Evidence: Cohen's Liberal Kappa = .955	Viewpoints: Cohen's Liberal Kappa = .960	Recommendations: Cohen's Liberal Kappa = .953
Agree on score	83 (60%)	86 (61%)	80 (57%)	84 (60%)	82 (59%)
Difference = 1 point	52 (37%)	52 (37%)	55 (39%)	52 (37%)	53 (38%)
Difference = 2 points	4 (3%)	2 (1%)	5 (4%)	4 (3%)	3 (2%)
Difference = 3 points	0	0	0	0	2 (1%)
<b>Total</b>	<b>139 (100%)</b>	<b>140 (100%)</b>	<b>140 (100%)</b>	<b>140 (100%)</b>	<b>140 (100%)</b>

# Freshman Baseline/FYS Comparisons

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

$n = 134$

Mean differences are statistically significant except for *Communication Style*



# Freshman Baseline/FYS Comparisons

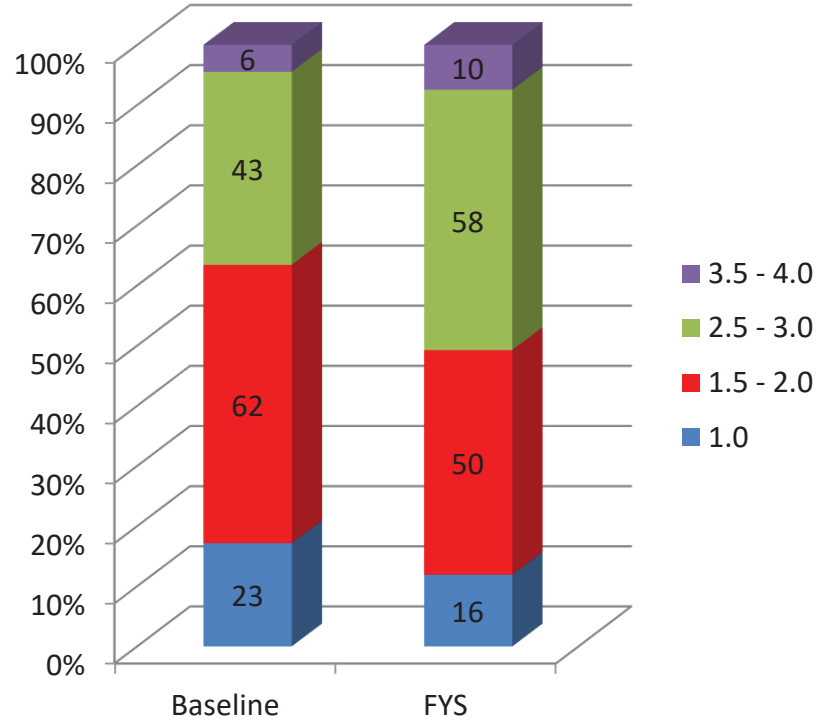
$n = 134$

Trait/ Performance Level	Development	Convention/Format	Communication Style
1.0 Baseline	23 (17%)	27 (20%)	4 (3%)
1.0 FYS	16 (12%)	17 (13%)	5 (4%)
1.5 – 2.0 Baseline	62 (46%)	65 (49%)	51 (38%)
1.5 – 2.0 FYS	50 (37%)	44 (33%)	33 (25%)
2.5 – 3.0 Baseline	43 (32%)	34 (25%)	74 (55%)
2.5 – 3.0 FYS	58 (43%)	58 (43%)	88 (66%)
3.5 – 4.0 Baseline	6 (4%)	8 (6%)	5 (4%)
3.5 – 4.0 FYS	10 (7%)	15 (11%)	8 (6%)
<b>Grand Total Baseline</b>	<b>134 (100%)</b>	<b>134 (100%)</b>	<b>134 (100%)</b>
<b>Grand Total FYS</b>	<b>134 (100%)</b>	<b>134 (100%)</b>	<b>134 (100%)</b>

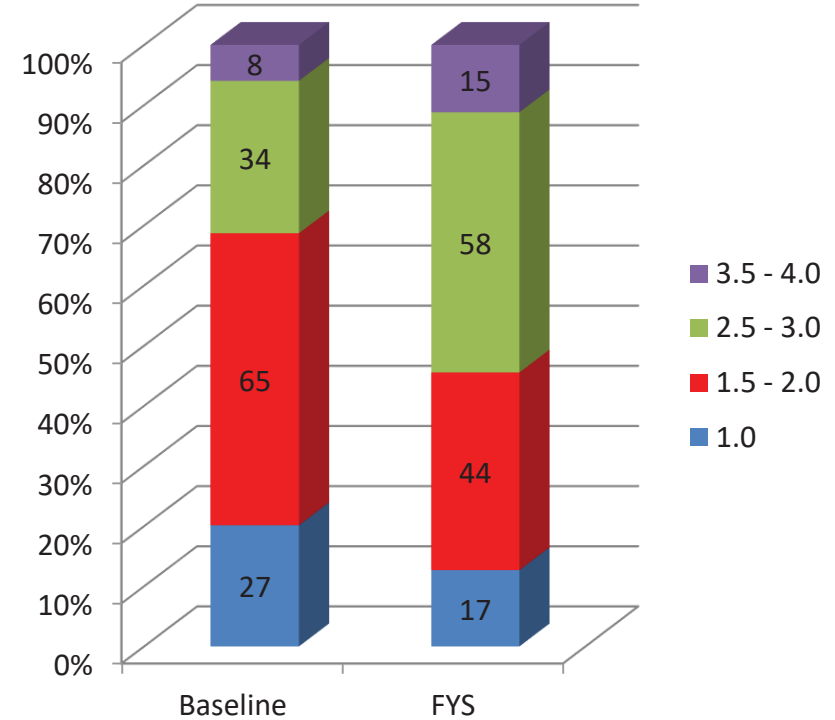
# Freshman Baseline/FYS Comparisons

$n = 134$

## Development



## Convention/Format

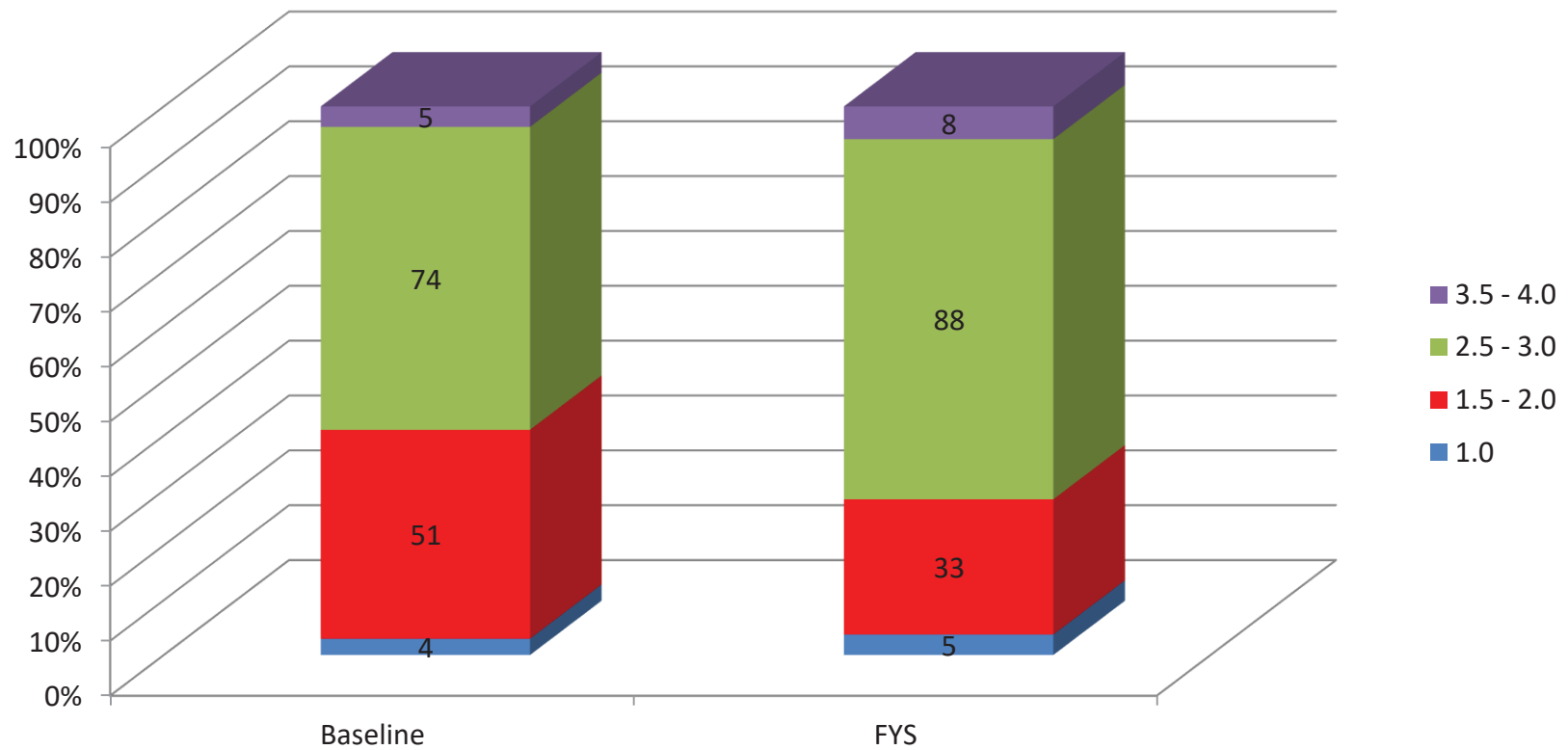




# Freshman Baseline/FYS Comparisons

$n = 134$

## Communication Style



# FYS Inter-Rater Agreement Results

Trait/ Agreement	Development: Cohen's Liberal Kappa = .961	Convention/Format: Cohen's Liberal Kappa = .898	Communication Style: Cohen's Liberal Kappa = .972
Agree on score	101 (53%)	91 (48%)	96 (51%)
Difference = 1 point	82 (43%)	82 (43%)	89 (47%)
Difference = 2 points	6 (3%)	15 (8%)	4 (2%)
Difference = 3 points	0	1 (1%)	0
<b>Total</b>	<b>189 (100%)</b>	<b>189 (100%)</b>	<b>189 (100%)</b>

# Baseline Inter-Rater Agreement Results

Trait/ Agreement	Development: Cohen's Liberal Kappa = .982	Convention/Format: Cohen's Liberal Kappa = .938	Communication Style: Cohen's Liberal Kappa = .981
Agree on score	78 (56%)	78 (56%)	76 (54%)
Difference = 1 point	60 (43%)	55 (39%)	62 (44%)
Difference = 2 points	2 (1%)	6 (4%)	2 (1%)
Difference = 3 points	0	1 (1%)	0
<b>Total</b>	<b>140 (100%)</b>	<b>140 (100%)</b>	<b>140 (100%)</b>



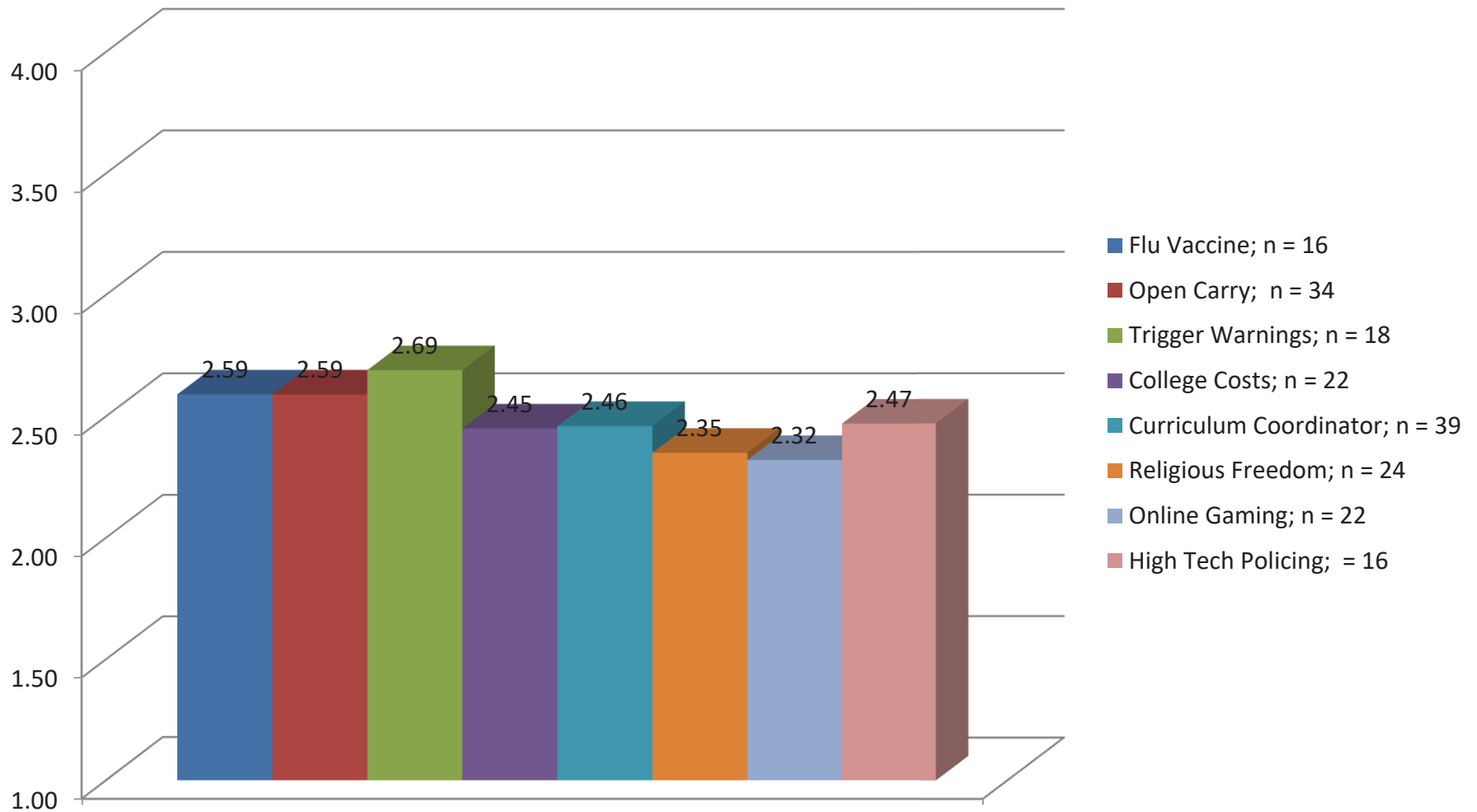
# Comparison of FYS Results for Each Trait by Scenario

Academic Year 2018 - 2019

# FYS Comparisons by Scenario for IL: Information Needed

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

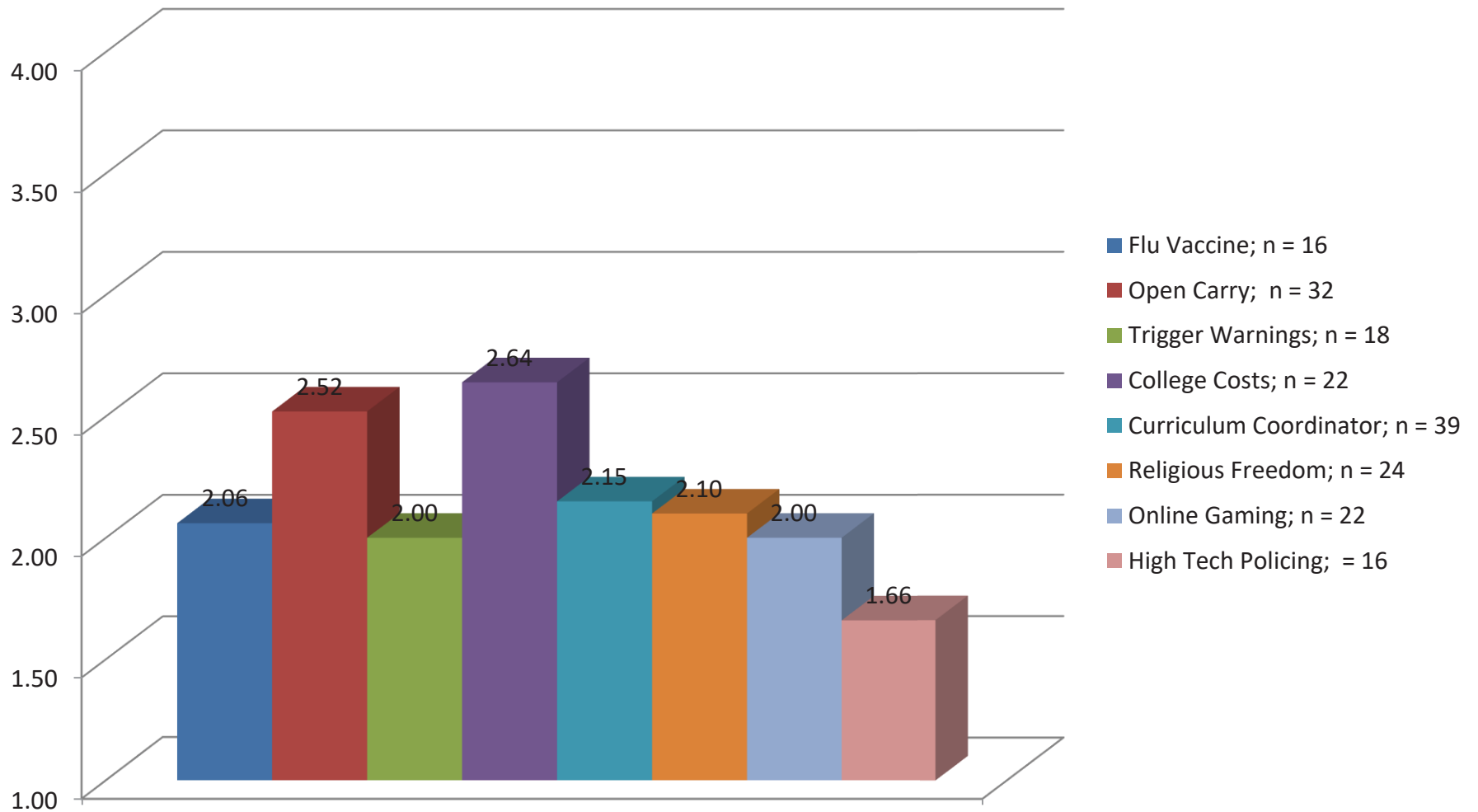
A One-Way ANOVA did not reveal any statistically significant differences in means across the scenarios.



# FYS Comparisons by Scenario for IL: Source Acknowledgment

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

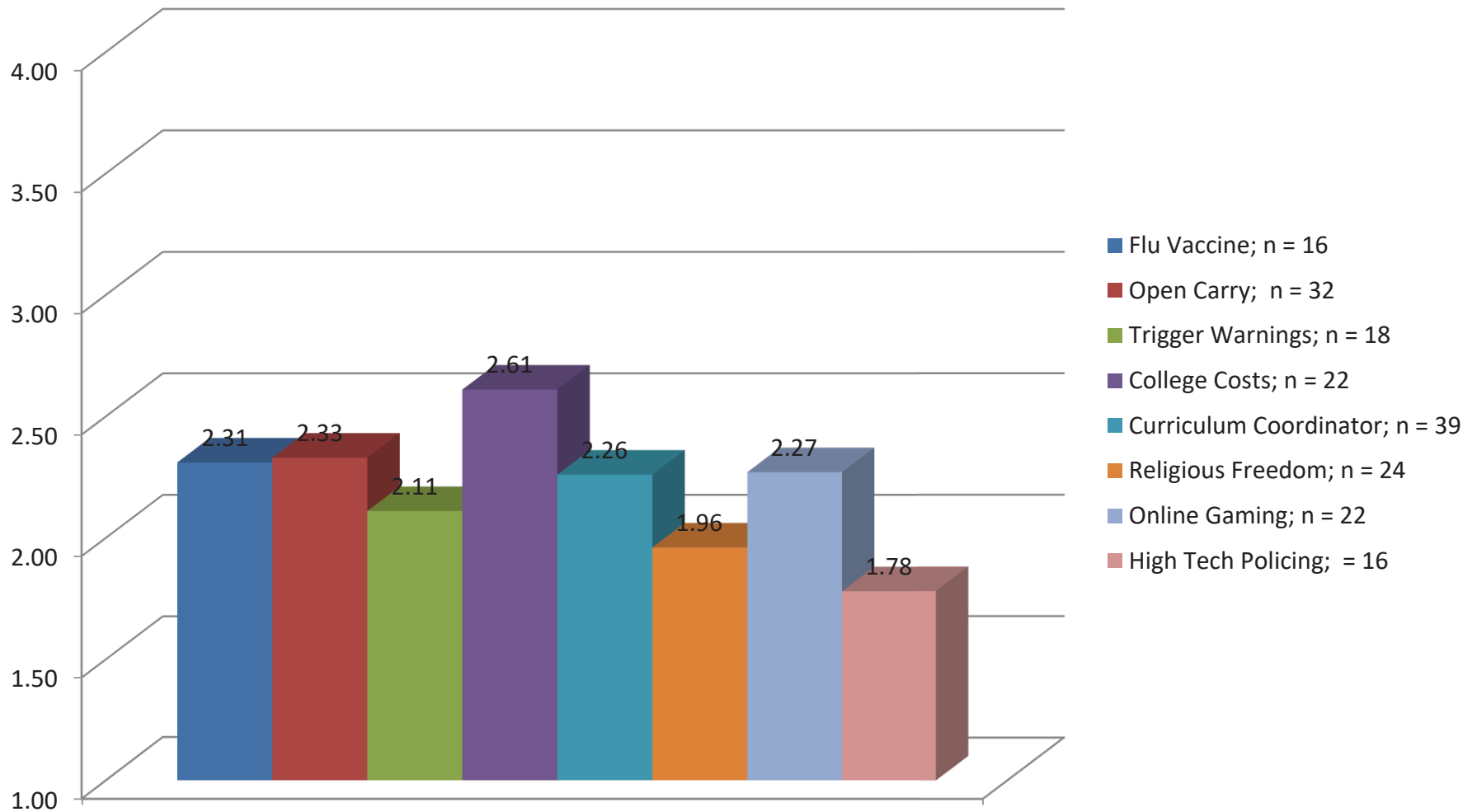
A One-Way ANOVA revealed statistical significance; a Tukey Post-Hoc Analysis revealed that student performance on College Costs was significantly higher than performance on High Tech Policing.



# FYS Comparisons by Scenario for IBT: Evidence

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

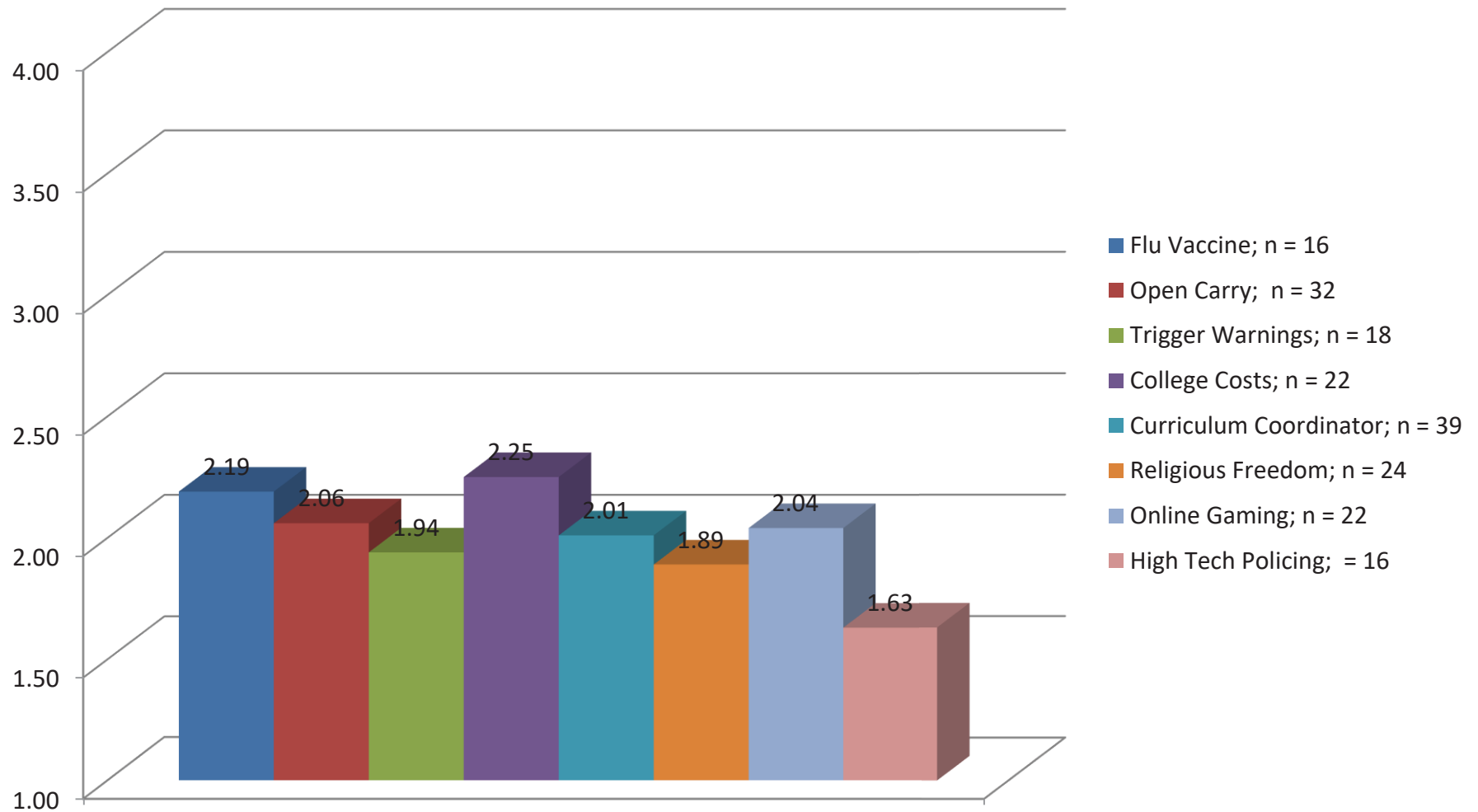
A One-Way ANOVA revealed statistical significance; a Tukey Post-Hoc Analysis revealed that student performance on College Costs was significantly higher than performance on High Tech Policing or on Religious Freedom.



# FYS Comparisons by Scenario for IBT: Viewpoints

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

A One-Way ANOVA revealed statistical significance; a Tukey Post-Hoc Analysis revealed that student performance on College Costs, Flu Vaccine, and Open Carry were significantly higher than performance on High Tech Policing.

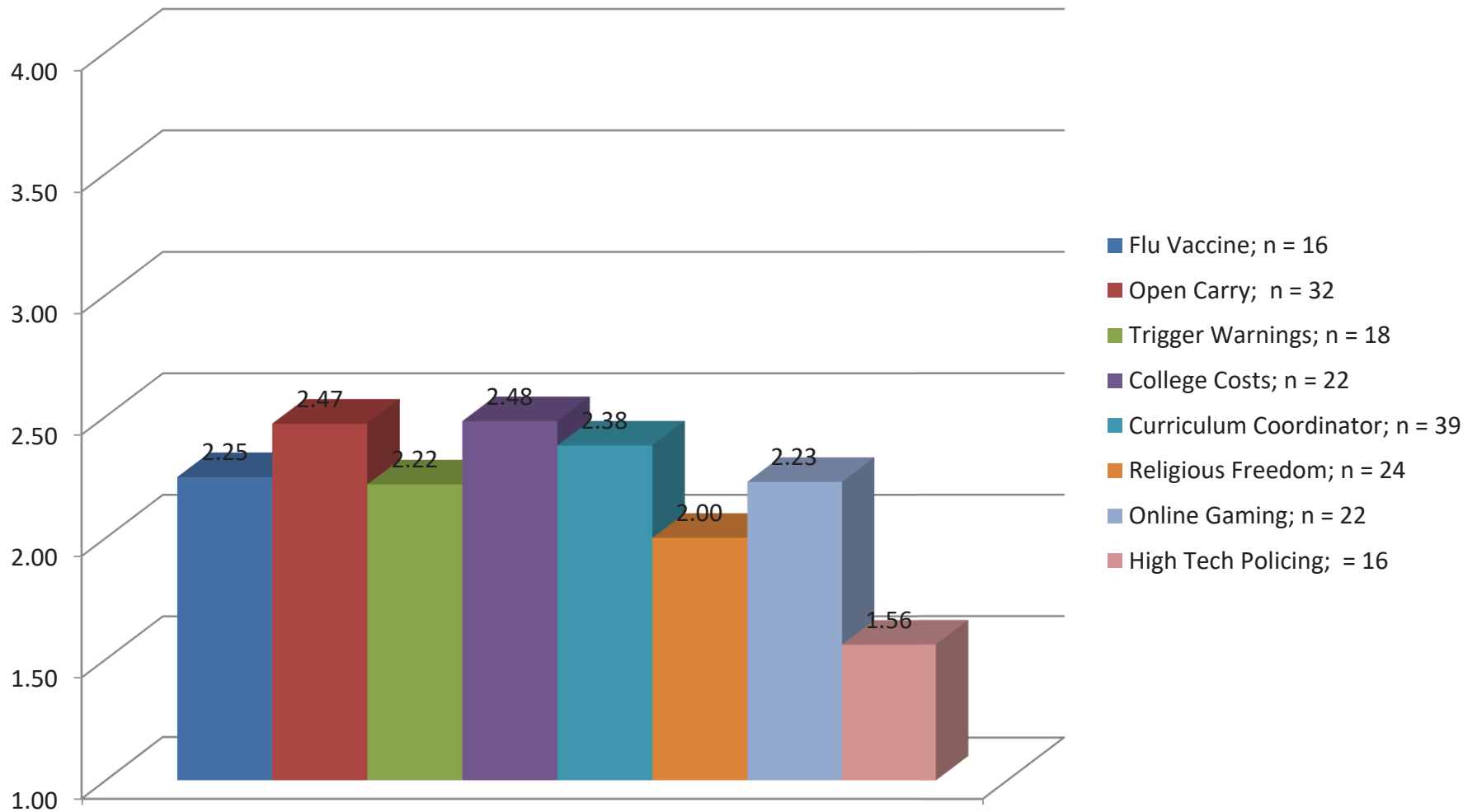




# FYS Comparisons by Scenario for IBT: Recommendations

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

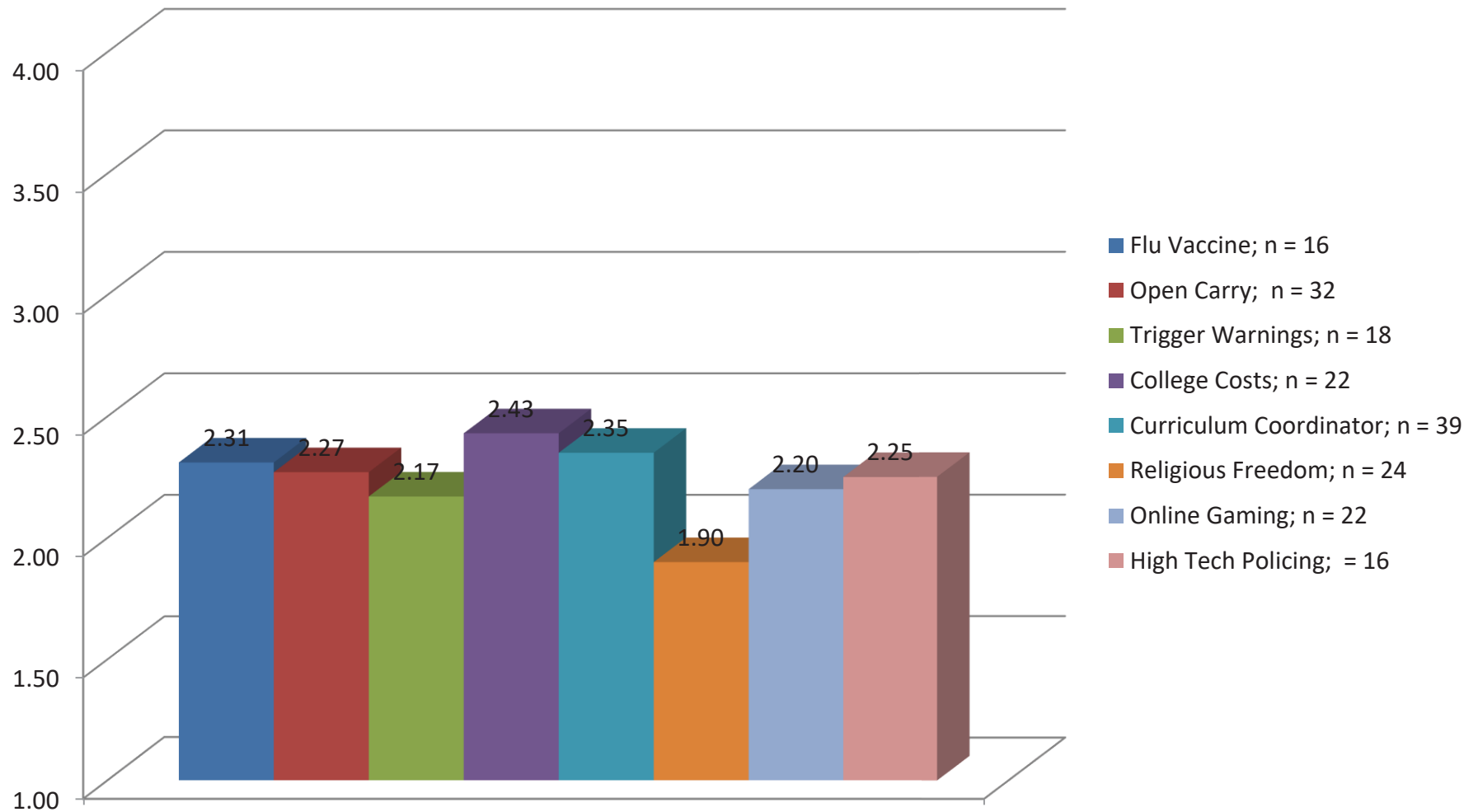
A One-Way ANOVA revealed statistical significance; a Tukey Post-Hoc Analysis revealed that student performance on College Costs, Curriculum Coordinator, Online Gaming, and Open Carry were significantly higher than performance on High Tech Policing.



# FYS Comparisons by Scenario for CF: Development

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

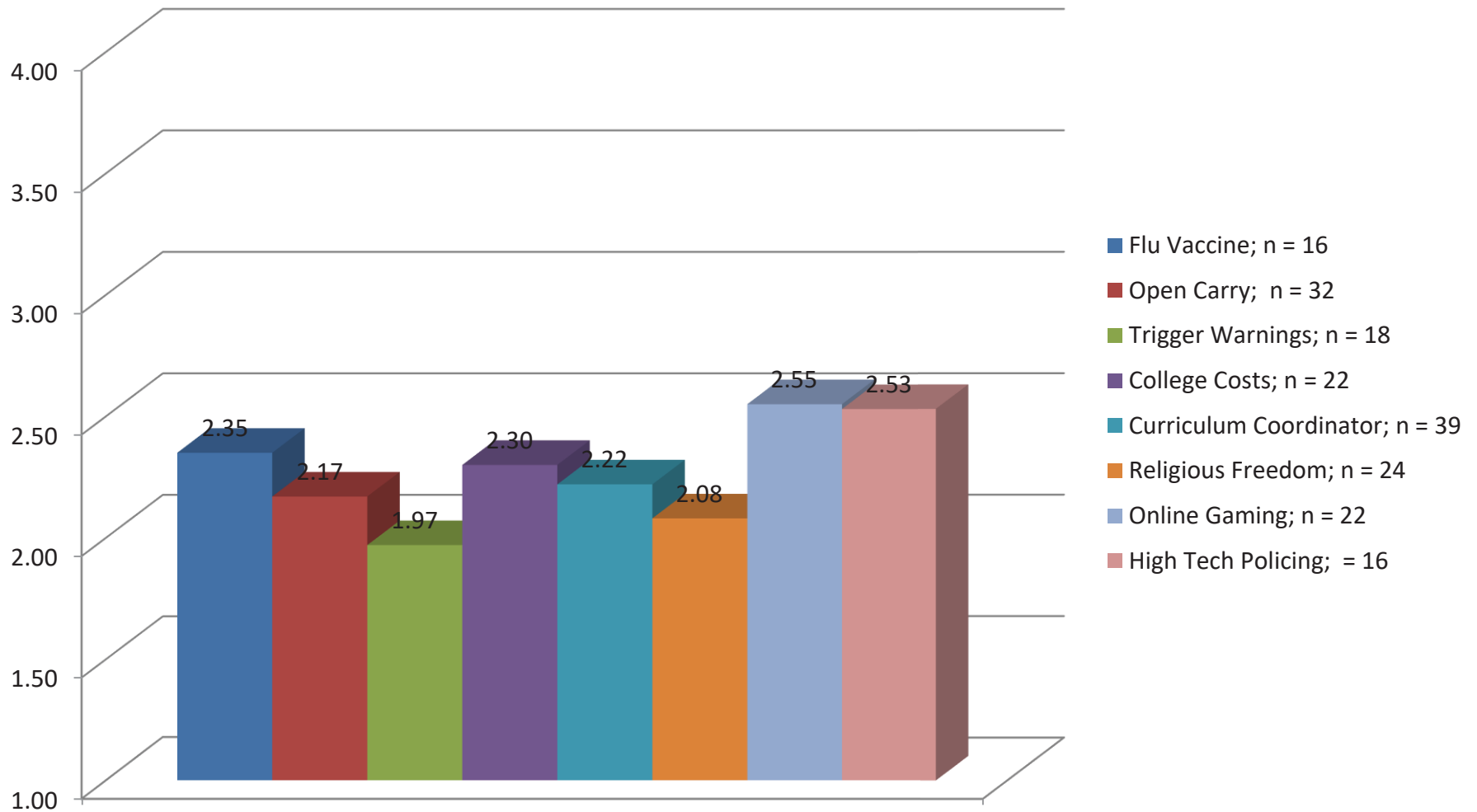
A One-Way ANOVA did not reveal any statistically significant differences in means across the scenarios.



# FYS Comparisons by Scenario for CF: Convention/Format

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

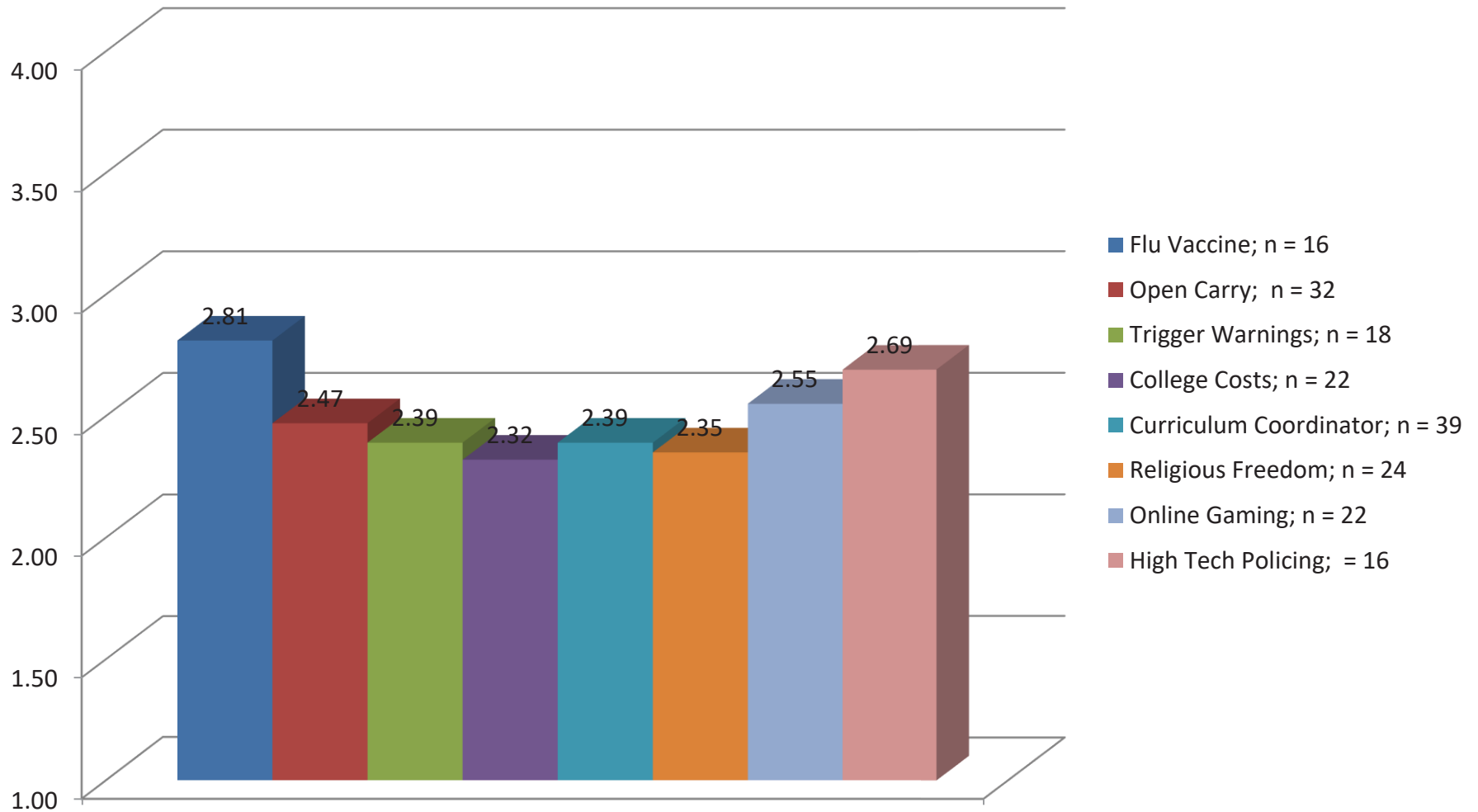
A One-Way ANOVA did not reveal any statistically significant differences in means across the scenarios.



# 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.





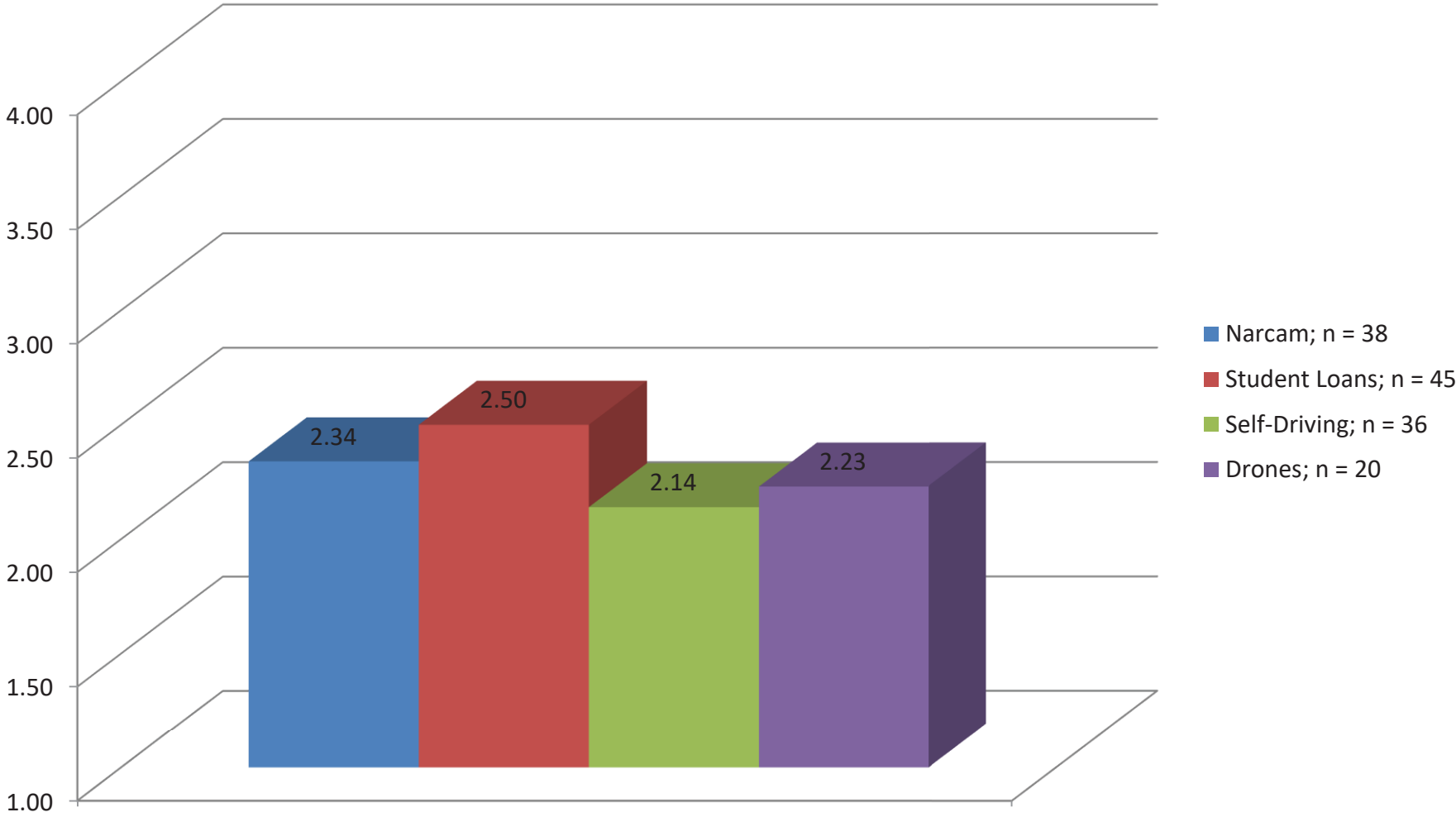
# Comparison of Baseline Results for Each Trait by Scenario

Academic Year 2018 - 2019

# Baseline Comparisons by Scenario for IL: Information Needed

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

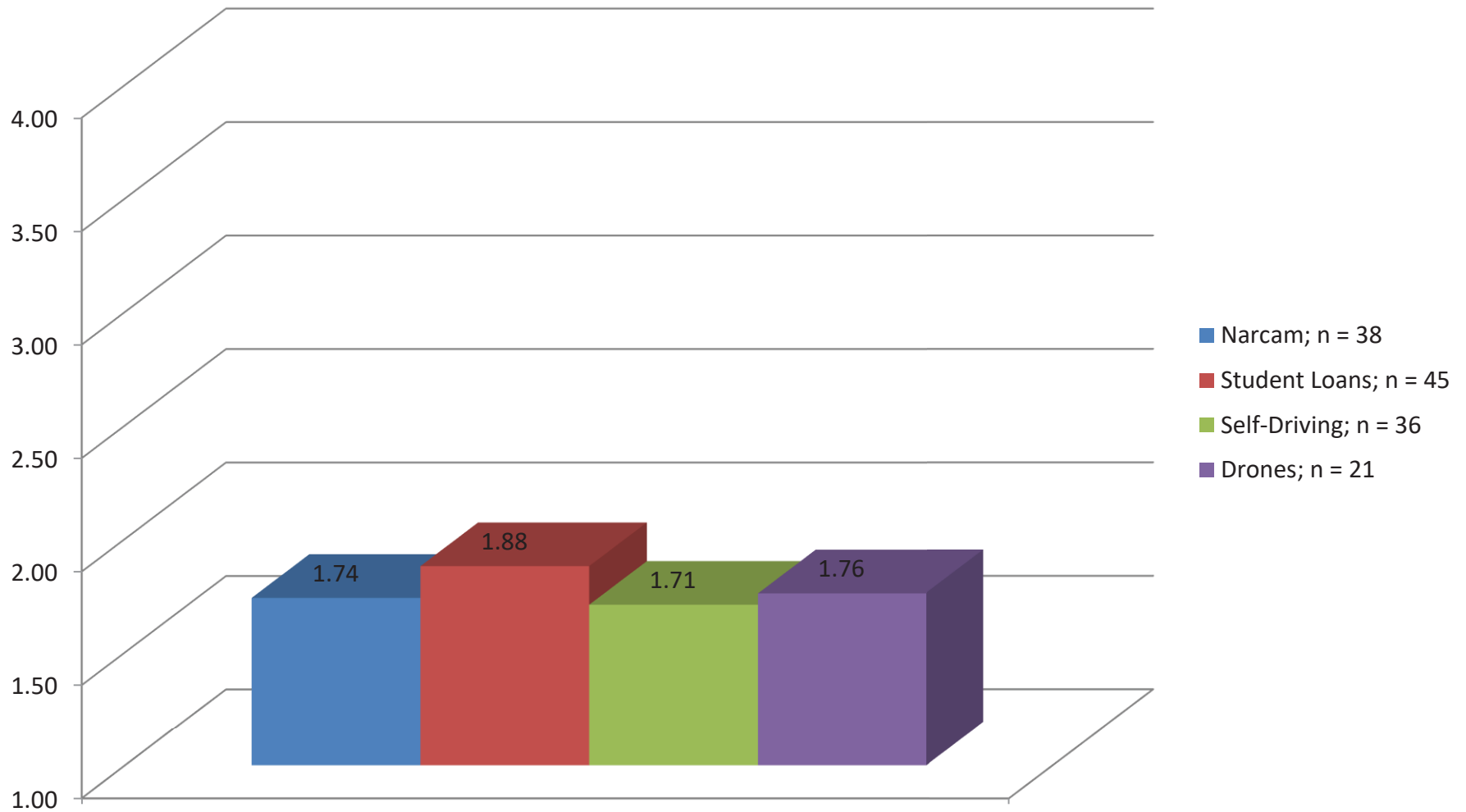
A One-Way ANOVA did not reveal any statistically significant differences in means across the scenarios.



## Baseline Comparisons by Scenario for IL: Source Acknowledgment

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

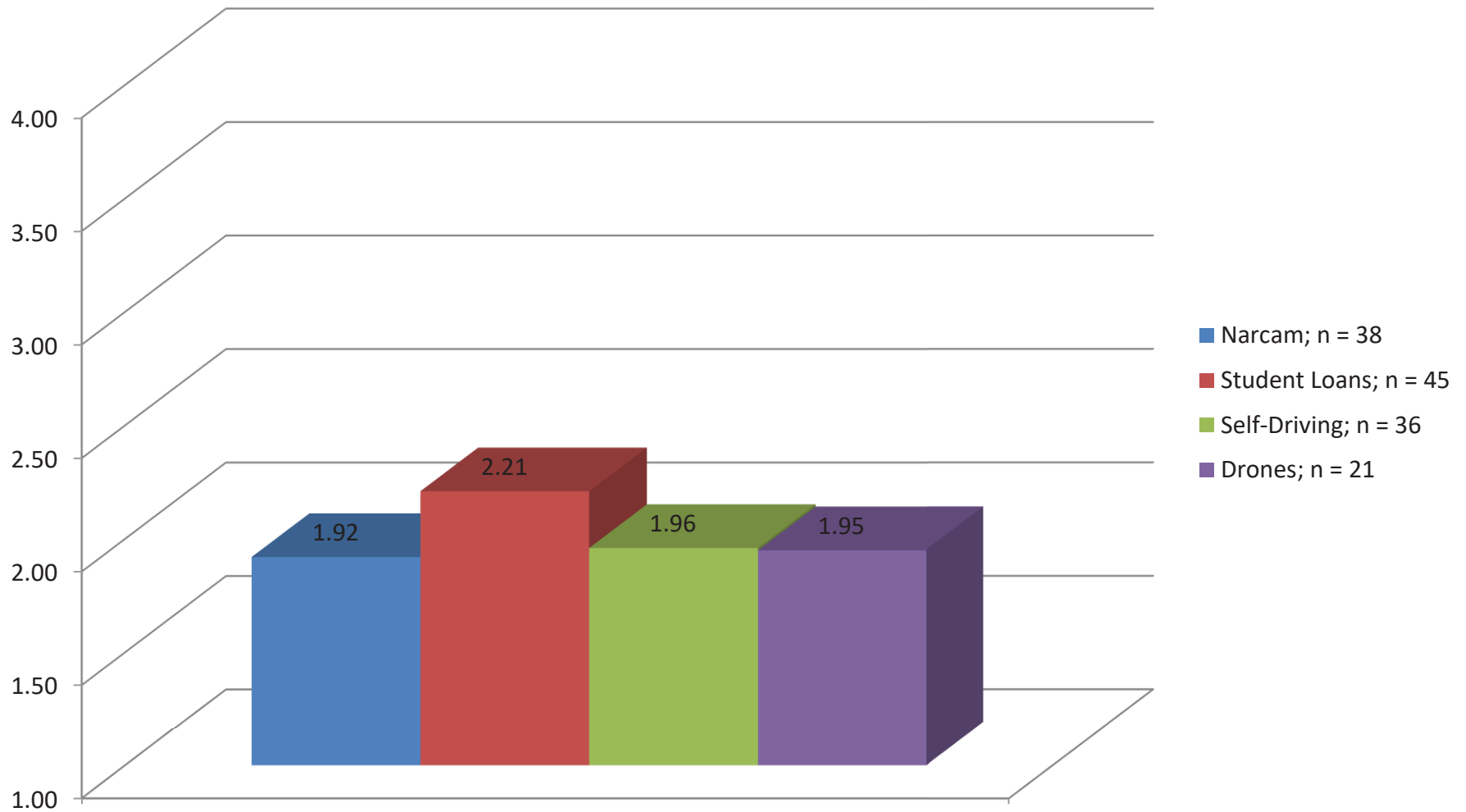
A One-Way ANOVA did not reveal any statistically significant differences in means across the scenarios.



## Baseline 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.

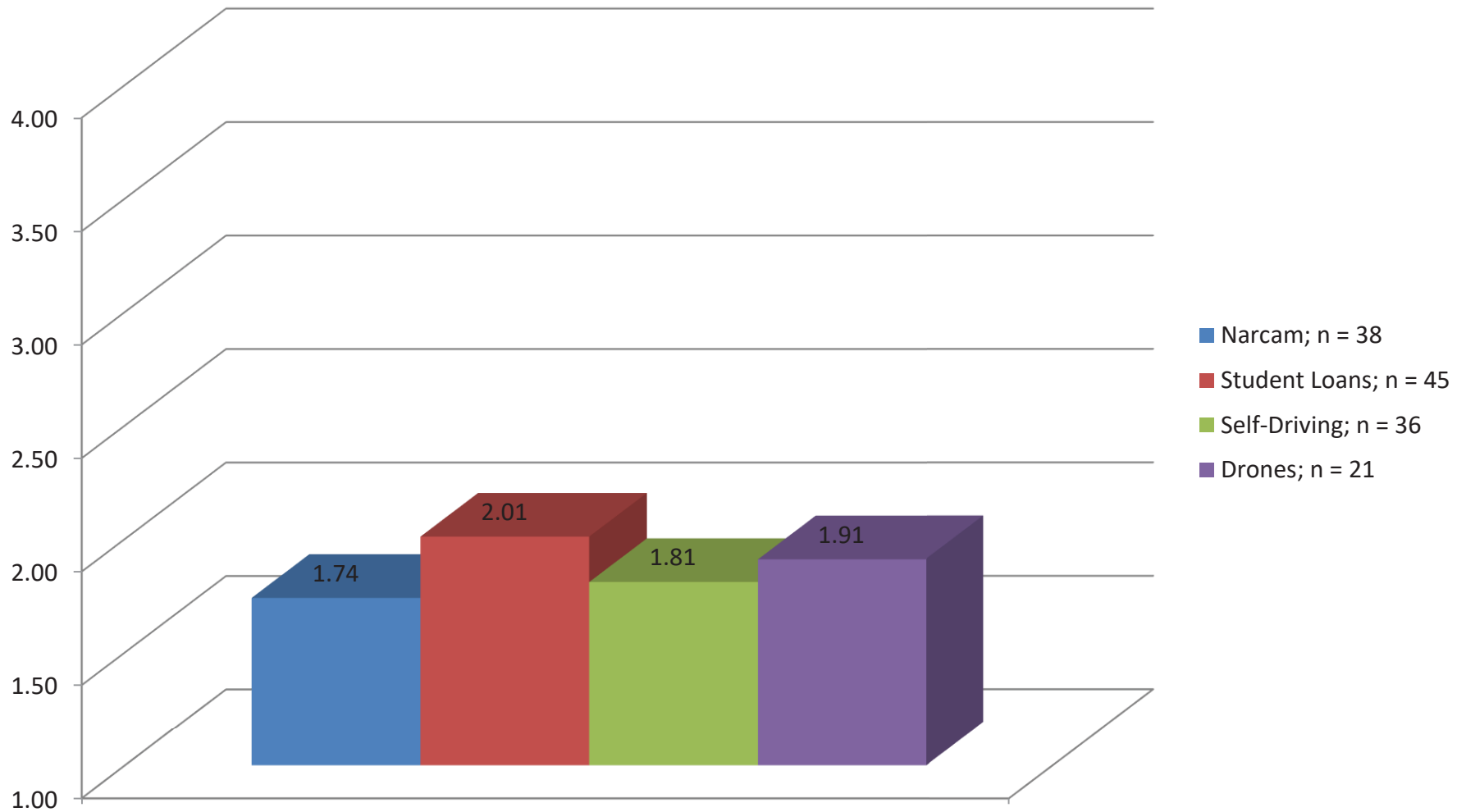




## Baseline Comparisons by Scenario for IBT: Viewpoints

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

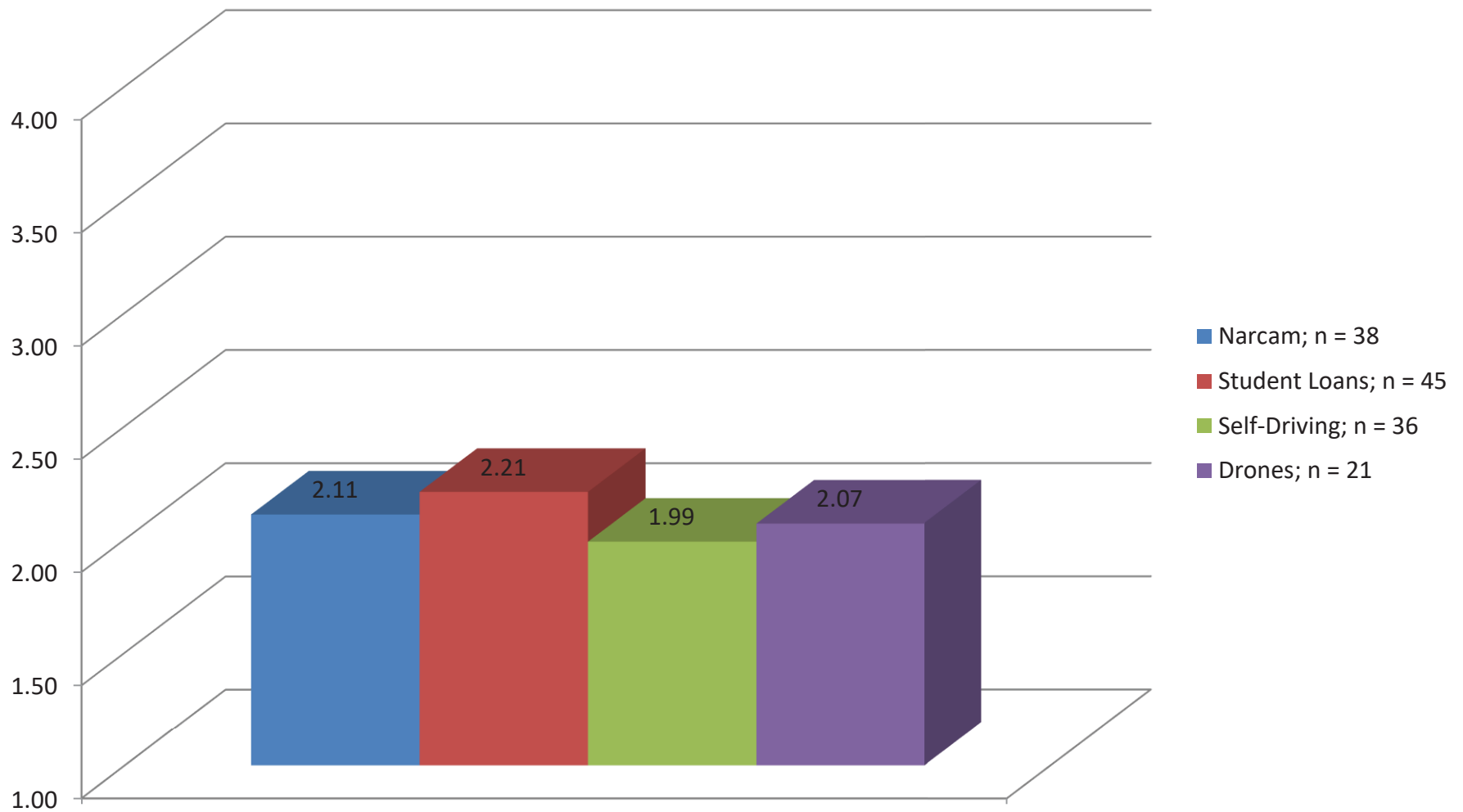
A One-Way ANOVA did not reveal any statistically significant differences in means across the scenarios.



## Baseline Comparisons by Scenario for IBT: Recommendations

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

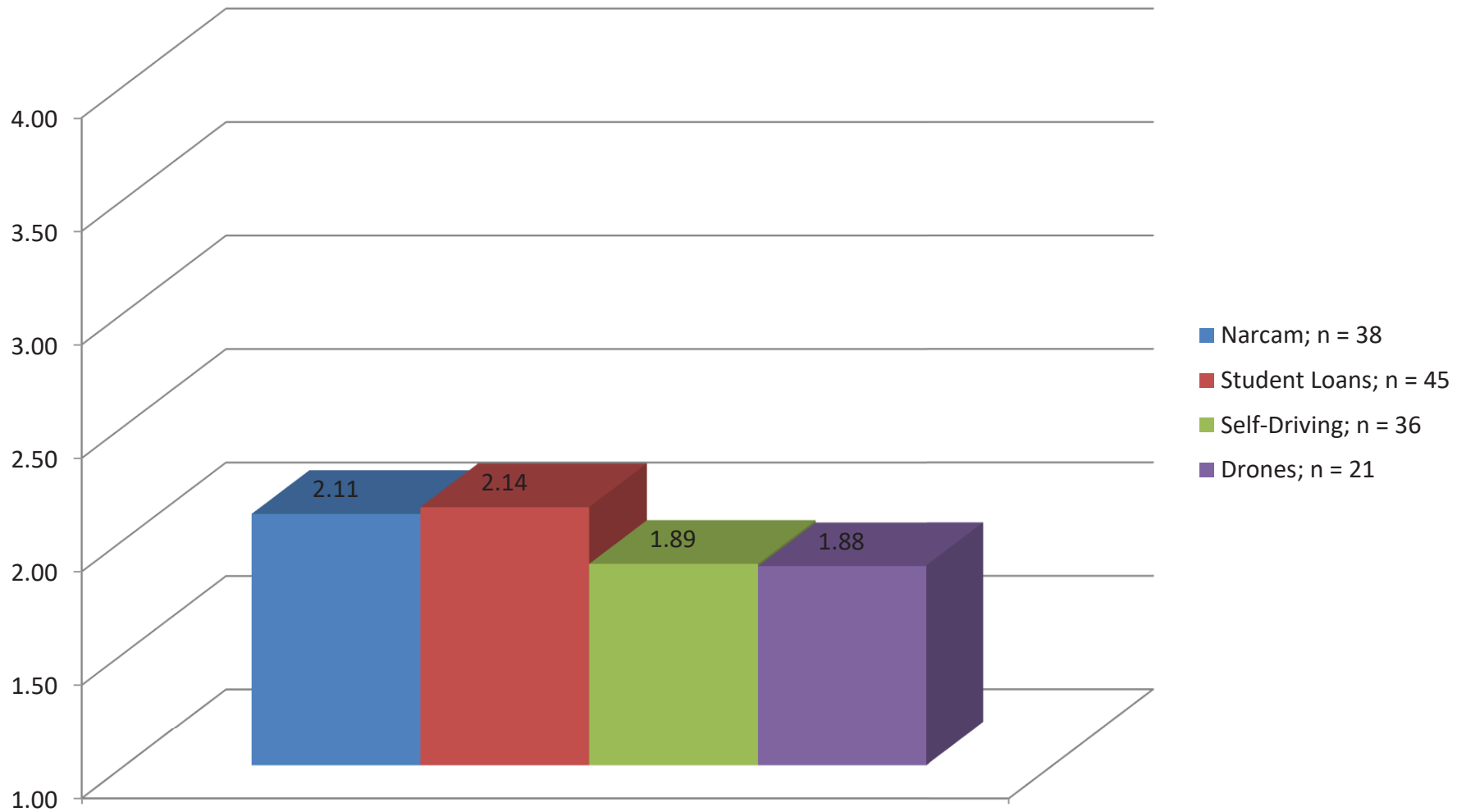
A One-Way ANOVA did not reveal any statistically significant differences in means across the scenarios.



## Baseline Comparisons by Scenario for CF: Development

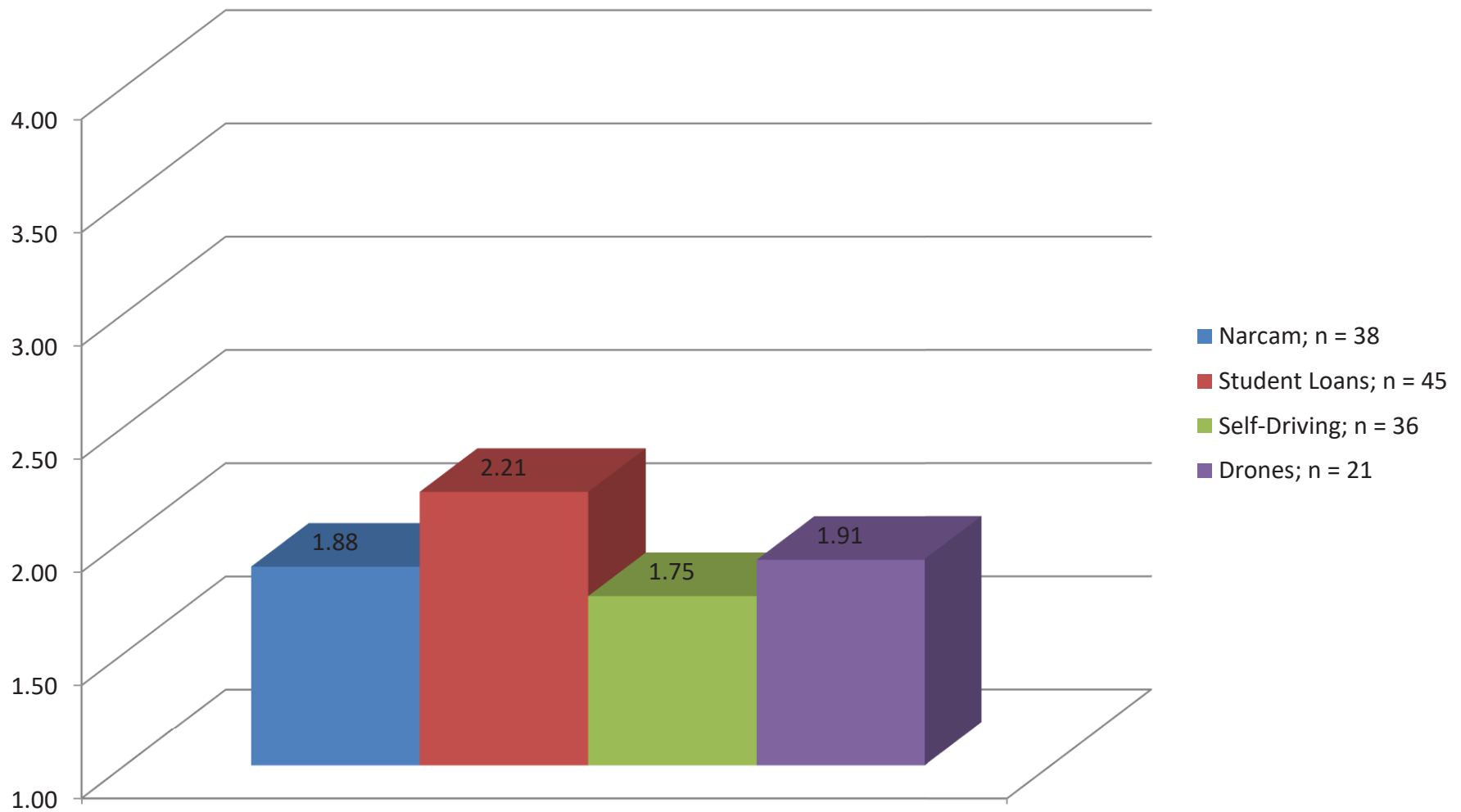
Mean Scores on a scale of 1 – 4, with 4 being the highest possible score

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



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

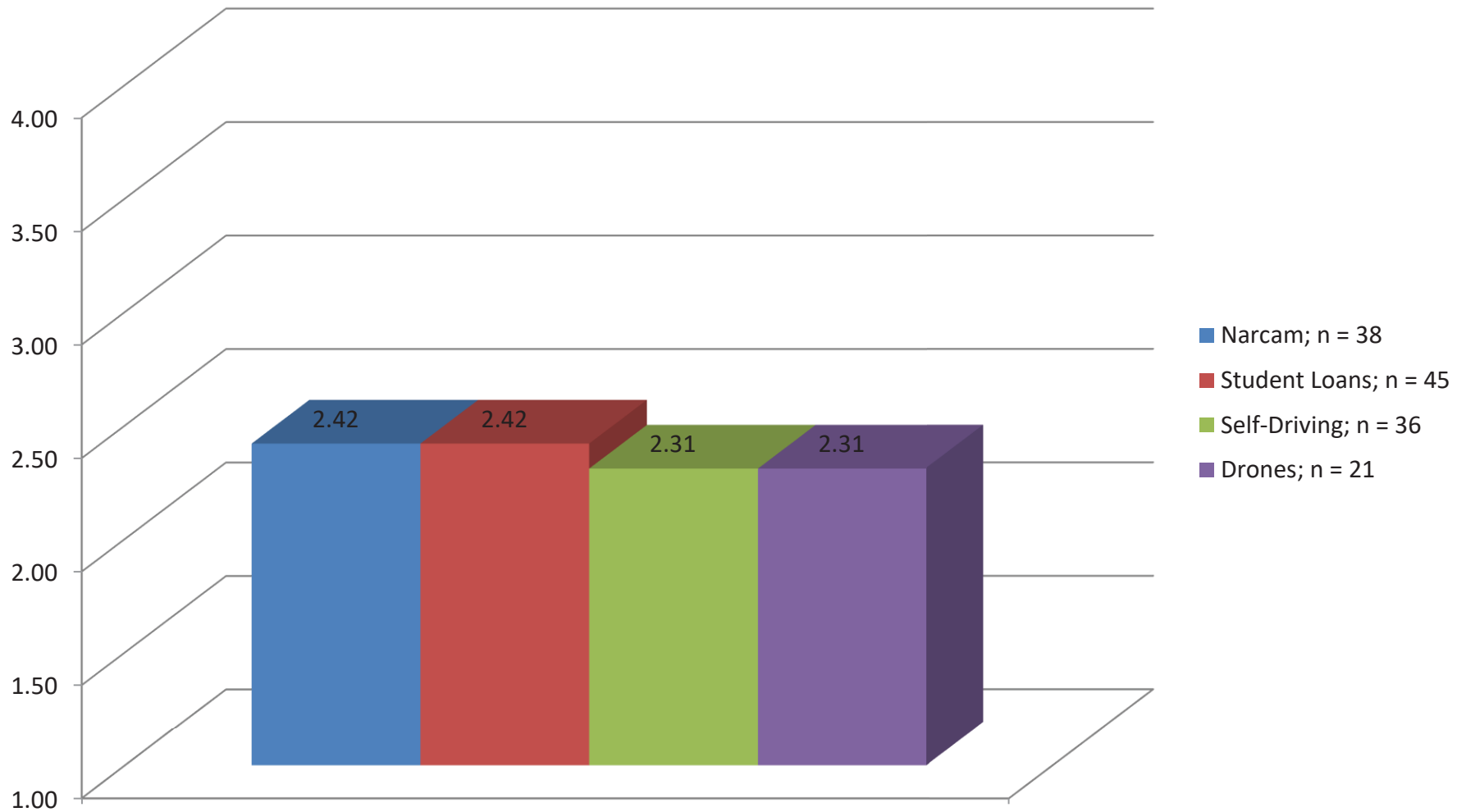
A One-Way ANOVA revealed statistical significance; a Tukey Post-Hoc Analysis revealed that student performance on Student Loans was significantly higher than performance on Self-Driving.



## Baseline 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.





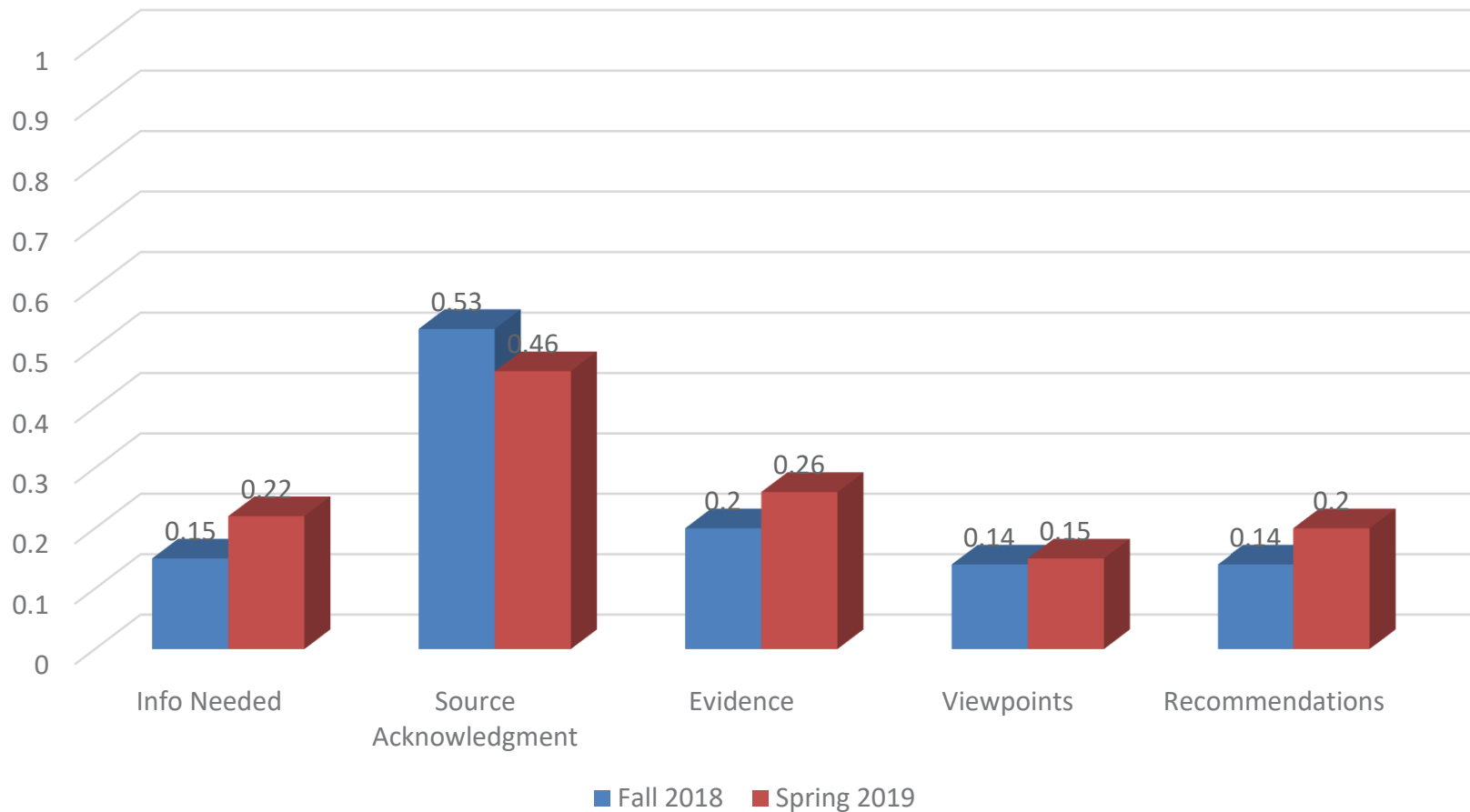
Comparison of Baseline to FYS Mean  
Gain Score for Each Trait by Semester of  
FYS

Academic Year 2018 - 2019

# Baseline to FYS Mean Gain Scores for Each Trait

$n = 70$  in fall and  $64$  in spring

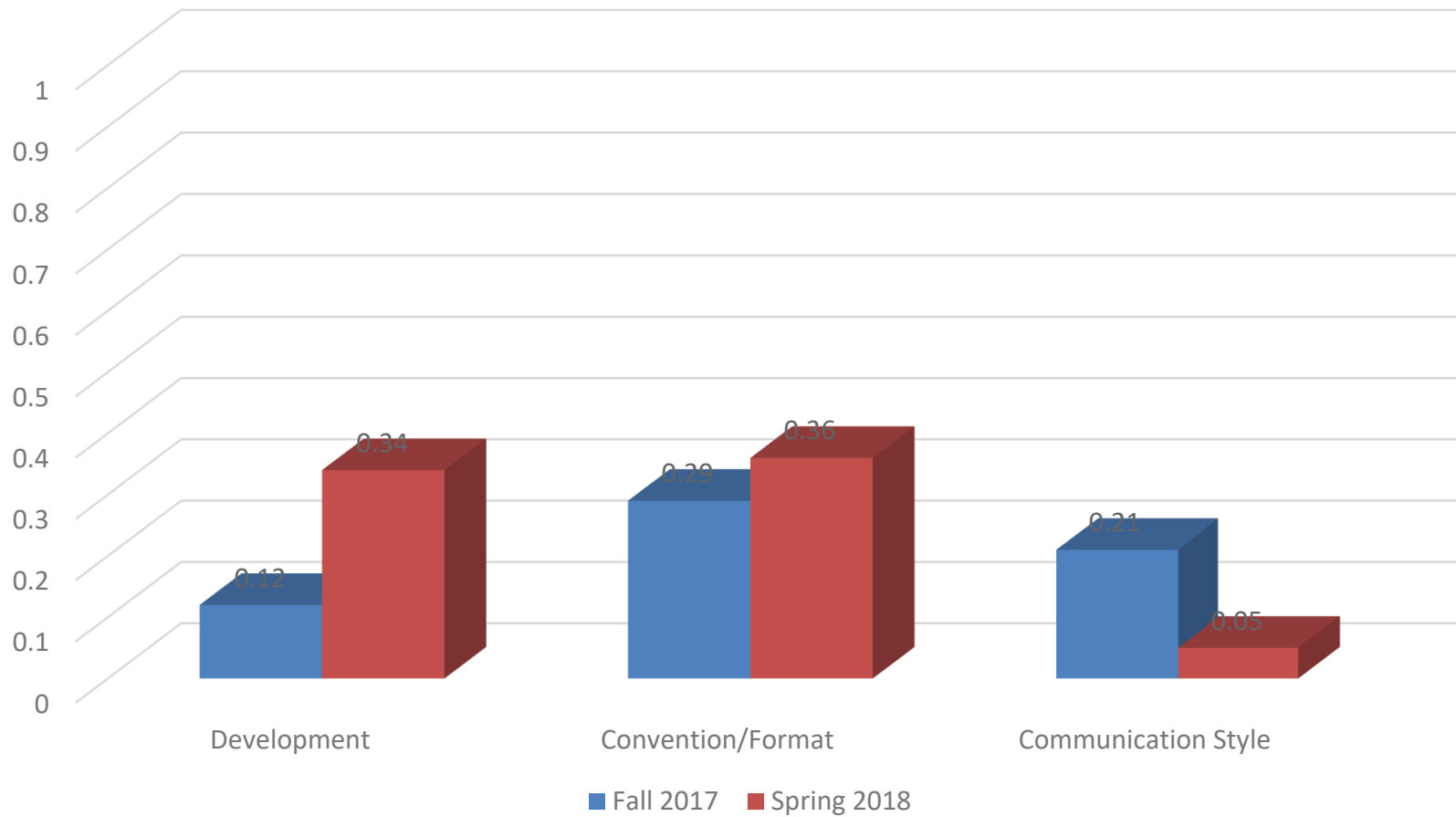
(Differences between fall and spring were not statistically significant)



# Baseline to FYS Mean Gain Scores for Each Trait

$n = 70$  in fall and  $64$  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.