Comparison of Freshman Baseline with First Year Seminar and Senior Exiting Assessment Results

Academic Year 2014 – 2015

Summer Assessment Workgroup Members: Marie Archambault, Harold Blanco, Kim DeTardo-Bora, Robert Ellison, Marty Laubach, Joan St. Germain, Gregg Twietmeyer, Anita Walz, Mary Welch, and Mary Beth Reynolds (Office of Assessment)

<u>**Executive Summary**</u>

Background

Recommendations from 2014 Assessment (with current status in red)

Assessments

Currently, students are given 90 minutes to complete the freshman baseline and senior assessments. The FYS final assessment, which consists of one additional section asking students to evaluate each piece of evidence that might help them to arrive at a recommendation for accuracy, bias, and relevance, is not included in the freshman baseline/senior assessments. Therefore, students are given 120 minutes to complete the FYS final assessment. The committee recommended standardization among all three assessments (freshman baseline, FYS, and senior) along the following parameters:

- Length of time period to complete – either 90 or 120 minutes. Due to time constraints during Week of Welcome, we were not able to accomplish this recommendation.
- All sections of assessment present for all. Due to the 90-minute limitation during Week of Welcome, we did not think it was feasible to have students explicitly evaluate <u>each</u> document for accuracy, bias, and relevance during the assessment. However, they are asked to keep these thoughts in mind as they develop a recommendation or position. The *CLA+* uses a similar approach. We also feel that the explicit teaching of document/evidence evaluation in FYS is important to the improvement of students' abilities to more carefully evaluate information and question viewpoints.
- All presented in electronic format (currently, all assessments are completed using paper and pen/pencil). Although we would like to do this, the lack of computer lab space and the need to proctor these assessments has precluded this from happening for baseline and senior assessments. However, all FYS assessments were completed electronically through BlackBoard during academic year 2014 – 2015.

The committee recommended implementing the problem-based senior assessment in

- Capstone classes or
- As part of a student's graduation requirements. Senior assessments were given to seniors from capstone classes this year. Not all capstone instructors participated, but those who did required their students to complete the assessments. We believe this helped our sample, not in size, but in representativeness and in the elimination of self-selection as a factor.

Rubrics

Last summer, the assessment workgroup made very specific recommendations to revise the rubric we used to evaluate baseline/senior/FYS assessments. We used all of last year's recommendations and added a few more following our norming session on the first day of the assessment session this summer (2015).

General Procedures for 2015 Assessment

In August 2014, 1,479 incoming freshmen at Marshall University completed baseline assessments (an additional 135 students completed the *Collegiate Learning Assessment* [*CLA+*]). Both assessments required students to analyze and evaluate information, solve problems, and write effectively. These skills are aligned to three of Marshall University's outcomes; Information Literacy, Inquiry-Based Thinking, and Communication Fluency. In the spring semester of 2015, 145 graduating seniors completed the same assessments (43 the Marshall assessment and 102 the *CLA+*). The 145 seniors who completed either the *CLA+* or Marshall's senior assessment did not differ significantly from the senior population in terms of gender or entering academic ability based on ACT or SAT performance. However, the sample had a slightly higher mean college GPA (3.2) than the senior population (3.1) Freshmen completing Marshall's mandatory First Year Seminar (FYS) completed assessments that were similar to those finished by incoming freshmen and graduating seniors.

In May 2015 a group of nine faculty representing several academic colleges from across the university evaluated a sample of Marshall's assessments using a rubric that allowed them to score each assessment across nine criteria (traits). These included *information needed* and *source acknowledgment* (Information Literacy), *evidence*, *viewpoints*, and *recommendation/position* (Inquiry-Based Thinking), and *development*, *convention/format*, and *communication style* (Communication Fluency). This project was coordinated by the Office of Assessment.

A random sample of 225 Marshall Freshman baseline assessments was drawn from the pool of 1,479 (15%) of the total number of assessments available. Since only 43 seniors completed the Marshall senior exiting assessment, we included all in our analysis, giving us a total of 268 assessments in our sample.

One hundred eighty-two of the 225 freshmen from our baseline sample (81%) completed FYS assessments. The reasons we had no FYS assessments from 43 of the students in the baseline sample were as follows: 16 were enrolled in, and received credit for, FYS, but did not

complete the final exam, 8 were enrolled in, but did not receive credit for FYS, 7 were not enrolled in FYS during academic year 2014-2015, and 12 students withdrew from Marshall University without completing FYS.

All assessments were de-identified and, for the freshman baseline/senior comparisons, raters did not know which were completed by freshmen and which by seniors. 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 Senior Exiting Results and to Results at the End of FYS

Mean scores (on a scale of 1 – 4) for seniors were significantly higher than freshman baseline measures on all criteria (traits). However, mean performance for seniors ranged from a low of 2.09 (*Inquiry-Based Thinking*: viewpoints) to a high of 3.1 (*Communication Fluency*: development), indicating, as was the case last year, that there is room for improvement among Marshall's graduating seniors. Mean differences between freshman baseline performance and senior exiting performance ranged from a low of 0.46 for *Communication Fluency*: communication style to a high of 0.96 for *Communication Fluency*: development.

The workgroup discussed the two-pronged approach that Marshall uses to compare student performance in *Information Literacy*, *Inquiry-Based Thinking* (aka *Critical Thinking*), and *Communication Fluency* between freshman baseline and senior exiting assessments, namely that some students take the nationally standardized *Collegiate Learning Assessment (CLA+)*, while the rest take a similar assessment developed by Marshall University faculty. This process works well for freshmen, but having representative senior samples that are large enough to draw meaningful conclusions remains problematic. We noted, however, that for the past several years the *CLA+* and Marshall Assessment results have mirrored each other. Results of the *CLA+* for the past two years (and of the *CLA* for several years prior to that) have shown that Marshall University's value-added in student growth in these outcomes between freshman and senior year has been at the statistically calculated "expected level." For the past two years, the average baseline *CLA+* score of our freshman has been at the *basic* level, while the average score of our seniors has been at the *proficient* level. Likewise, for the past three years our seniors have scored significantly higher than our freshmen on all outcomes/traits of the Marshall developed assessment. As noted in the preceding paragraph, despite these results there continues to be room for our seniors to improve in all outcomes addressed in these assessments.

For the 182 students who completed both baseline and FYS assessments, *paired-samples t-tests* using adjusted alpha levels to control for Type I error (.025 for information literacy), (.017 for Inquiry-Based Thinking), and (.017 for Communication Fluency) showed significant mean differences between freshman baseline and FYS results for the following *outcomes* (traits) *Information Literacy* (acknowledgment of sources), *Inquiry-Based Thinking* (evidence), and *Communication Fluency* (convention/format). Students showed the greatest improvement in performance in *Information Literacy* (acknowledgment of sources [.52]). These results are not as impressive as last year's results, where student

showed more improvement in using evidence to make recommendations. However, we note that, in this area, the baseline performance of our sample was higher than that of last year's sample, with the final scores at the end of FYS being similar between the two years.   As was the case last year, students did not demonstrate significant gains in questioning the viewpoints expressed in the pieces of evidence they examined, nor did they make significant gains in indicating the types of additional evidence they might need to make a recommendation. Therefore, as was the case last year, we recommend that the FYS Director and course instructors place additional emphasis on helping students to determine information need and critically examine various viewpoints surrounding real-world problems. Although we evaluated the FYS assessments for *Communication Fluency*, we note that this is not one of the outcomes of the FYS course.

Recommendations from the 2015 Assessment Workgroup

The workgroup noted that the revision of the FYS final assessment, which allows all students to complete the assessment online, was a positive step. However, members of the group expressed concern about the length of some of the documents the students must read and evaluate before making their recommendations for the problem they must solve. We noted that the FYS Advisory Board decided to begin using real documents in the faculty developed scenarios rather than documents created by faculty. The rationale for this was that the task would be more authentic because, in the real world, professionals are called upon to identify and evaluate such documents. However, members of the assessment workgroup pointed out that, in the real world, people typically have longer than two hours to do this. There was concern that the students had to spend so long reading the documents that they didn't have sufficient time to fully evaluate them and thoughtfully develop their recommendations. We note that two students' final assessments could not be evaluated because they had not included a recommendation, presumably running out of time before getting to that part of the assessment. The assessment workgroup recommended several options to try to remedy these issues:

- Release the documents before the final exam. Instructors would tell students they should have read the documents before arriving for the exam. Since the exam is administered in BlackBoard, one member suggested that it could be set up in two modules; first the documents, which would have to be read and evaluated for accuracy, relevance, and bias as a take-home part of the exam. Then, on the day of the exam itself, the second module allowing students to make a recommendation and indicate information still needed, would open.
- If the first option is not possible, the workgroup recommended that FYS faculty return to the previous method of using faculty created documents of a reasonable length.
- If students are instructed to give their recommendations in the form of a memorandum, the group recommended that one of the documents they read should be written in that format (or in whatever format they are asked to use to prepare their response).

Supporting Documentation

Comparison of Freshman Baseline and Senior Exiting Assessment Results

Academic Year 2014 – 2015

Review Procedures

- A total of 268 assessments (225 freshman and 43 senior) were used for assessment. Freshman assessments represented approximately 15% of the 1,479 completed during the University's Week of Welcome in August 2014. Only 43 seniors completed the Marshall Developed Senior Assessment in spring 2014 (an additional 102 seniors completed the Collegiate Learning Assessment [CLA+]), so all 43 were included in this sample. The 145 seniors who completed either the CLA+ or Marshall's Senior assessment did not differ significantly from the senior population in terms of gender or entering academic ability based on ACT or SAT performance. However, the sample had a slightly higher mean college GPA (3.2) than the senior population (3.1)

  - Assessments were de-identified and raters did not know which were completed by freshmen and which by seniors.
  - Each assessment was scored across nine criteria.

- Each assessment had two independent raters and scores were determined in the following manner:
  - If raters assigned the same score, that became the score for the artifact.
  - If raters' scores differed by one point or less, e.g. Rater 1 assigned a score of 1 and Rater 2 a score of 2, the final score was the mean, i.e. 1.5.
  - If raters' scores differed by more than one point, e.g. Rater 1 assigned a score of 1 and Rater 2 a score of 3, the raters met to discuss the rationale for their scores to see if they could agree on a score or, at minimum, scores that differed by no more than one point.
  - If raters' scores differed by more than one point and, after discussion, they were not able to resolve the differences, a third rater was assigned to review the assessment.

Rules for Arriving at Final Scores when there were Three Raters: These rules were followed for all assessments conducted.

1. If the third rater's score agreed with one of the first two, the score with the two agreements was used.

2. If the first two raters' scores were two points apart, e.g. 1 and 3 and the third rater's score was in the middle, e.g. 2, the third rater's score was used.

3. If the first two raters' scores were two points apart, e.g. 1 and 3, and the third rater's score was between them, but a decimal, e.g. 1.5 or 2.5, the third rater's score was used.

4. If the first two raters' scores were two points apart, e.g. 1 and 3, and the third rater's score was a "4", the two scores closer together were averaged, e.g. 3.5.

5. IF the first two raters' scores were three points apart, e.g. 1 and 4, the third rater's score was averaged with the closest other rater; e.g. if the third rater's score was 3, the final score was 3.5; if the third rater's score was 2, the final score was 1.5.

Rubric Used for Scoring

Baseline/Senior Assessment Rubric – Summer 2015 – updated 5-18-2015

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 in the DL or evidence is used without appropriate interpretation/evaluation (i.e. poor job).	Evidence is taken from relevant and valid sources in the DL with some interpretation/evaluation, but not enough to develop a coherent analysis or synthesis (i.e. adequate job).	Evidence is taken from relevant and valid sources in the DL with enough interpretation/evaluation to develop a coherent analysis or synthesis (i.e. good/excellent job).
	Viewpoints	Ignores viewpoints expressed in the DL.	Viewpoints expressed in the DL are taken as mostly fact, with little or no question.	Questions some viewpoints expressed in the DL.	Thoroughly questions and evaluates viewpoints expressed in the DL.
	Recommendation/Position	Either does not make a recommendation or makes a recommendation, but does not justify it in any way.	Recommendation is justified, but does not acknowledge different sides of the issue.	Recommendation is justified and takes into account different sides/complexities of the issue.	Recommendation takes into account the complexities of the issue. Any limits to the recommendation are acknowledged.
Communication Fluency	Development	Shows little or no evidence of developing his/her ideas.	Shows some development of ideas.	Shows a strong, but perhaps somewhat incomplete, development of ideas.	Produces a document in which the ideas have been fully developed.
	Convention/Format	Demonstrates minimal attention to basic organization and presentation and stylistic conventions.	Demonstrates some awareness of basic organization, content, and presentation and stylistic conventions.	Demonstrates consistent use of important conventions particular to a specific writing task, including organization, content, presentation, and stylistic choices.	Demonstrates detailed attention to and successful execution of a wide range of conventions particular to a specific writing task including organization, content, presentation, formatting, and stylistic choices.
	Communication Style	Uses language that impedes meaning because of errors in usage/mechanics.	Uses language that generally conveys meaning to readers with clarity, although writing may include some errors.	Uses straightforward language that generally conveys meaning to readers. The language in the document has few errors.	Uses graceful language that skillfully communicates meaning to readers with clarity and fluency, and is virtually error-free.

Freshman Baseline/Senior Exiting Comparisons
Mean Scores on a scale of 1 – 4, with 4 being the highest possible score
Freshman *n* = 225; Senior *n* = 43
All differences statistically significant
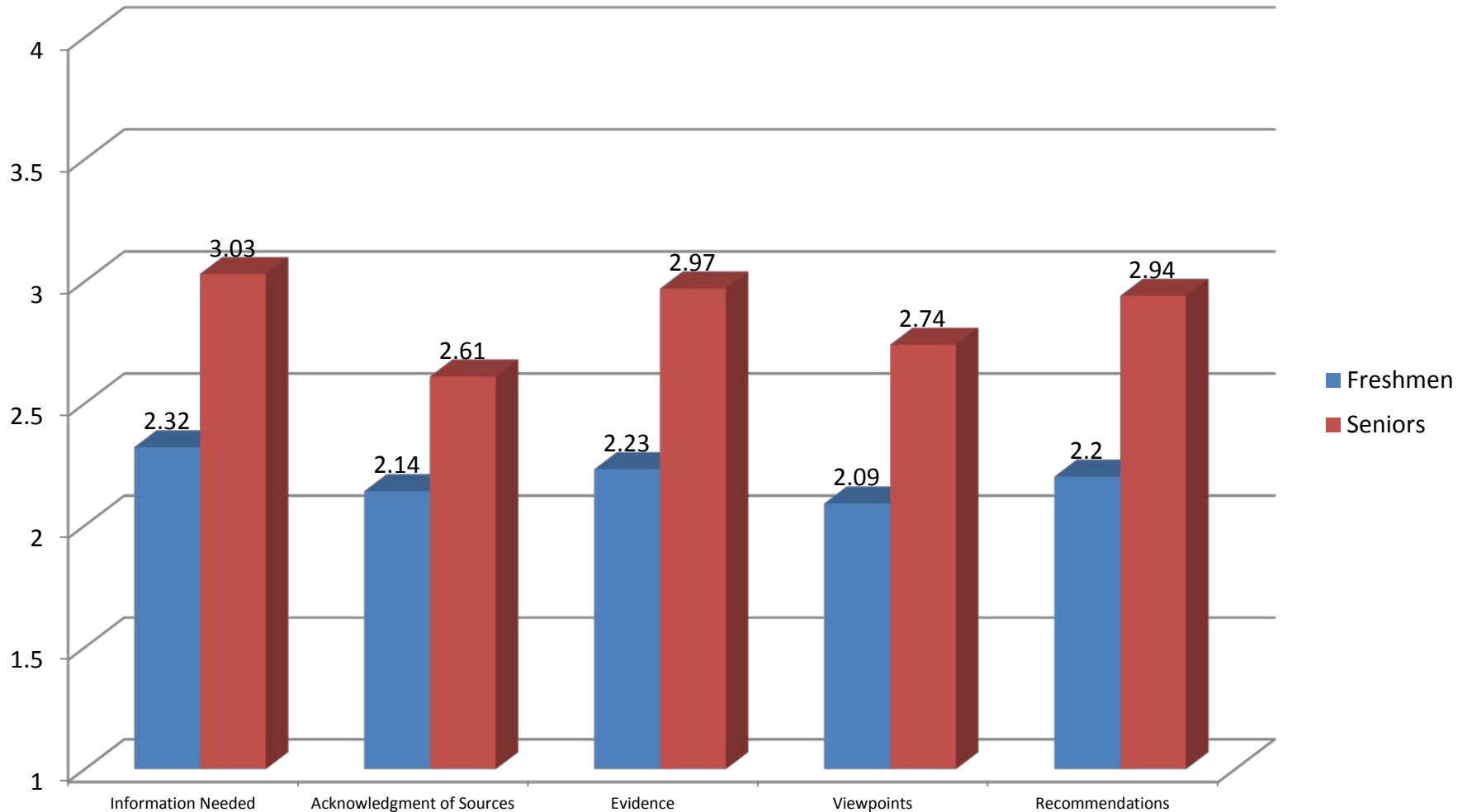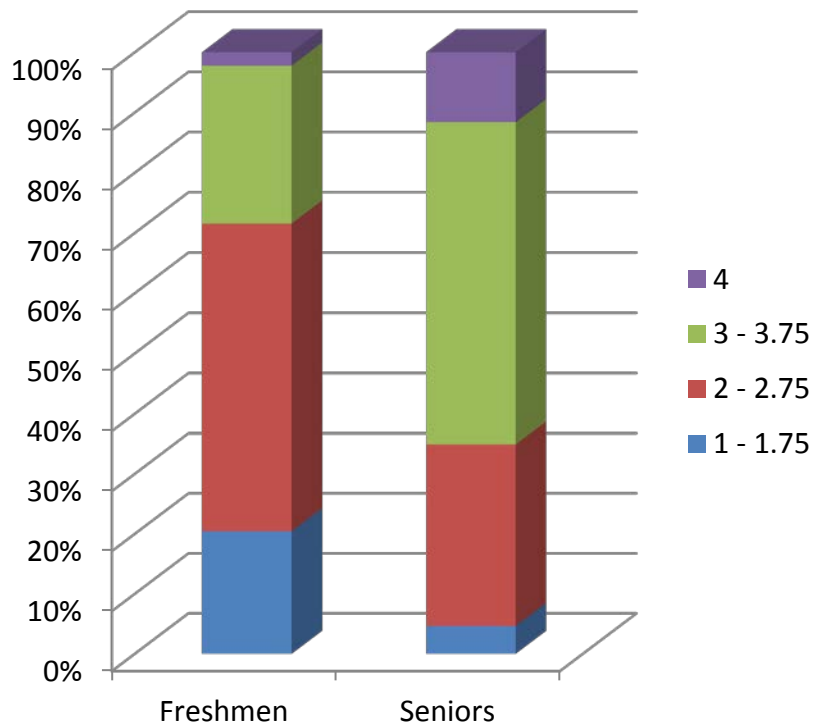
Freshman Baseline/Senior Exiting Comparisons
Freshman *n* = 225; Senior *n* = 43

Trait/ Performance Level	Info Needed	Acknowledgment of Sources	Evidence	Viewpoints	Recommendations
1 – 1.75 Freshmen	46 (20%)	57 (25%)	56 (25%)	62 (28%)	64 (28%)
1 – 1.75 Seniors	2 (5%)	4 (9%)	2 (5%)	4 (9%)	4 (9%)
2 – 2.75 Freshmen	115 (51%)	131 (58%)	119 (53%)	135 (60%)	118 (52%)
2 – 2.75 Seniors	13 (30%)	24 (56%)	15 (35%)	17 (40%)	13 (30%)
3 – 3.75 Freshmen	59 (26%)	32 (14%)	47 (21%)	26 (12%)	41 (18%)
3 – 3.75 Seniors	23 (53%)	10 (23%)	22 (51%)	19 (44%)	22 (51%)
4 Freshmen	5 (2%)	5 (2%)	3 (1%)	2 (1%)	2 (1%)
4 Seniors	5 (12%)	5 (12%)	4 (9%)	3 (7%)	4 (9%)
Grand Total Freshmen	225 (100%)	225 (100%)	225 (100%)	225 (100%)	225 (100%)
Grand Total Seniors	43 (100%)	43 (100%)	43 (100%)	43 (100%)	43 (100%)

Freshman Baseline/Senior Exiting Comparisons
Freshman *n* = 225; Senior *n* = 43

Information Needed
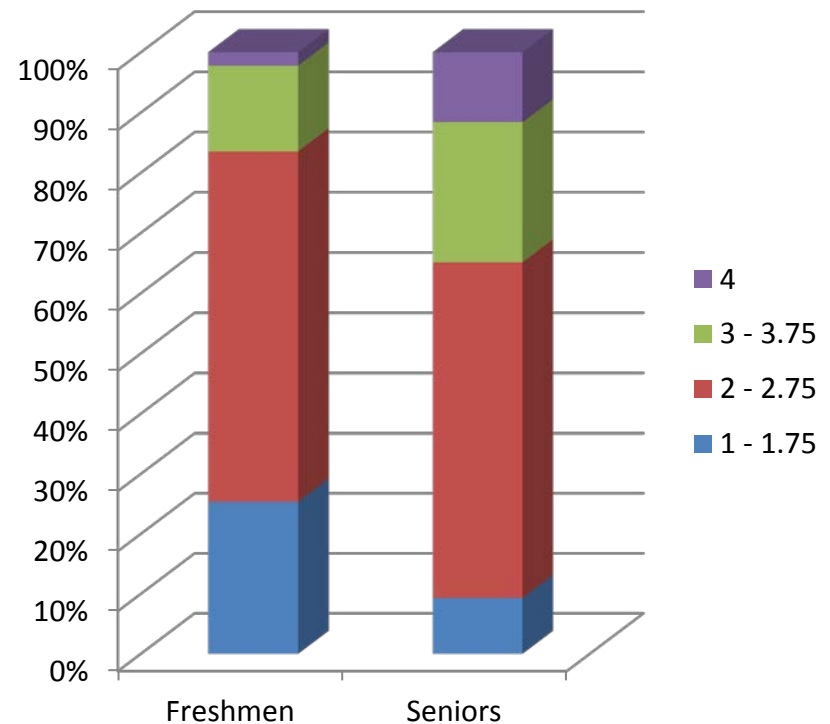
Acknowledgment of Sources

Freshman Baseline/Senior Exiting Comparisons
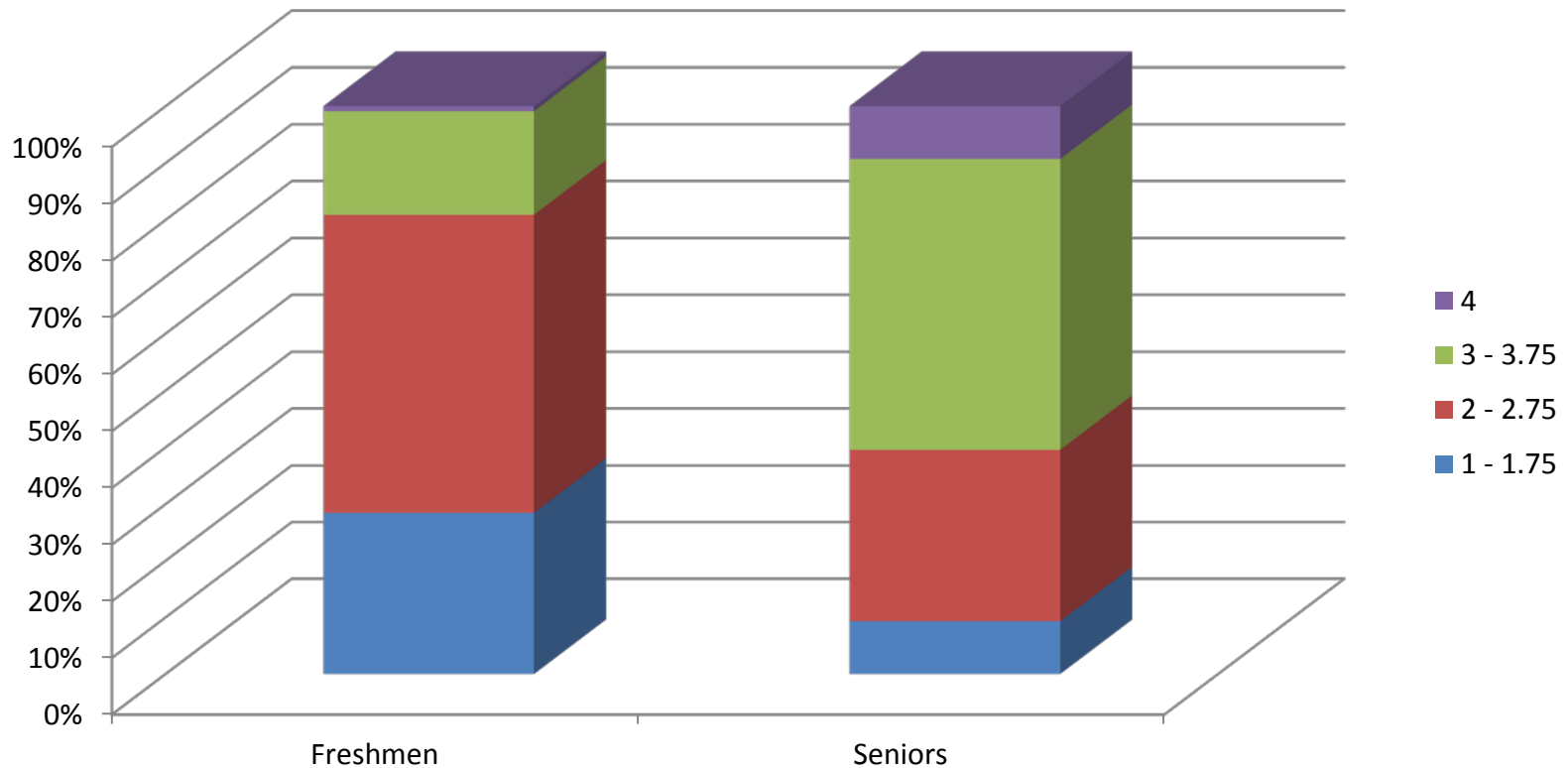Freshman *n* = 225; Senior *n* = 43
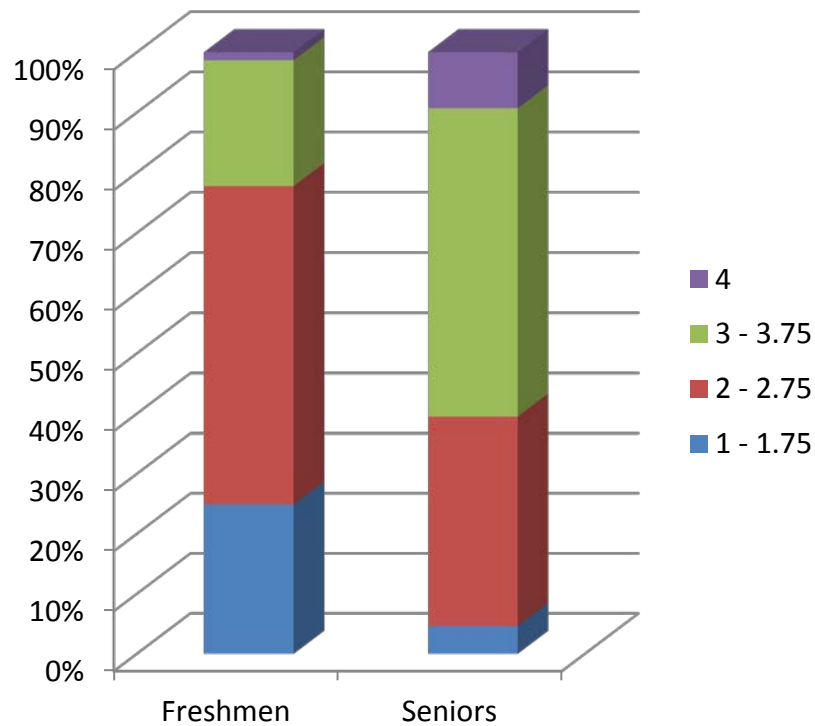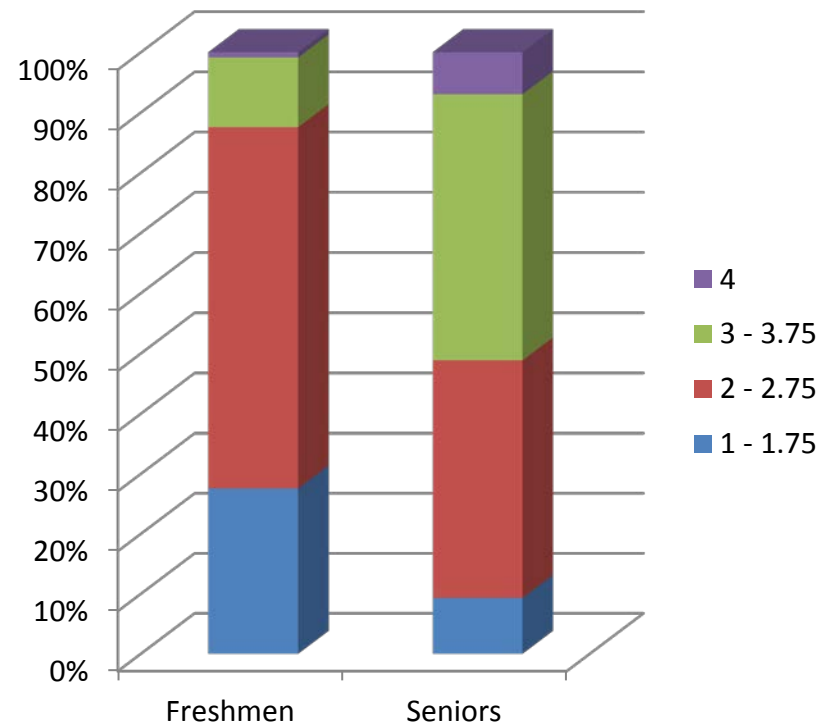
Freshman Baseline/Senior Exiting Comparisons
Freshman *n* = 225; Senior *n* = 43

Freshman Baseline/Senior Exiting Assessment Inter-Rater Agreement Results

| Trait/ Agreement	Info Needed (Conservative Kappa = .308; Liberal Kappa = .948)	Acknowledgment of Sources (Conservative Kappa = .278; Liberal Kappa = .941)	Evidence (Conservative Kappa = .212; Liberal Kappa = .929)	Viewpoints (Conservative Kappa = .214; Liberal Kappa = .925)	Recommendations (Conservative Kappa = .154; Liberal Kappa = .885)
Agree	135 (50%)	151 (56%)	122 (46%)	134 (50%)	103 (38%)
Difference = 1 point or less	121 (45%)	105 (39%)	128 (48%)	118 (44%)	139 (52%)
Difference = 1.5 to 2 points	10 (4%)	11 (4%)	18 (7%)	16 (6%)	19 (7%)
Difference = 2.5 to 3 points	2 (1%)	1 (0%)	0 (0%)	10(0%)	7 (3%)
Total	268 (100%)	268 (100%)	268 (100%)	268 (100%)	268 (100%)

Freshman Baseline/Senior Exiting Comparisons
Mean Scores on a scale of 1 – 4, with 4 being the highest possible score
Freshman *n* = 225; Senior *n* = 43
All differences statistically significant
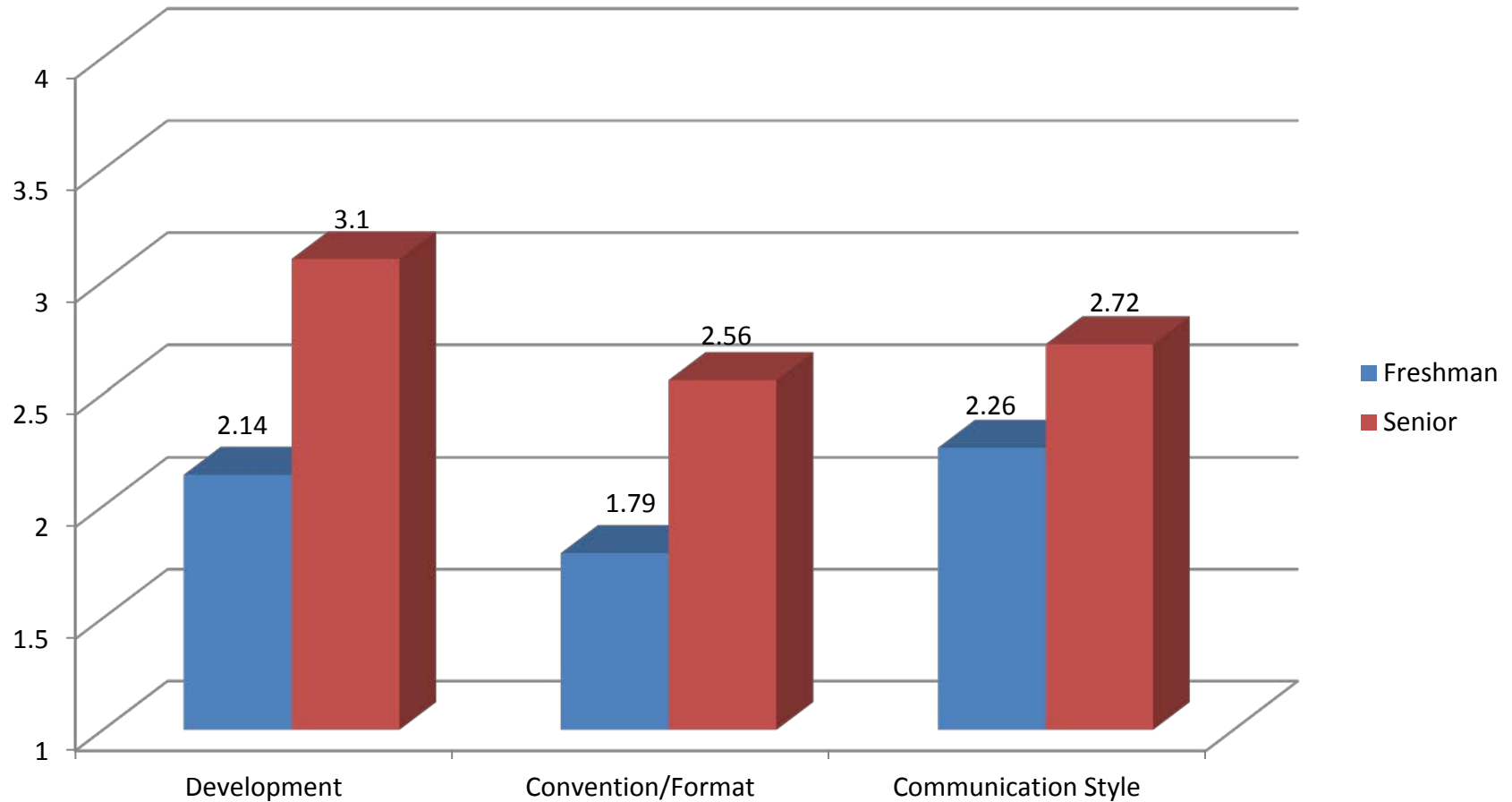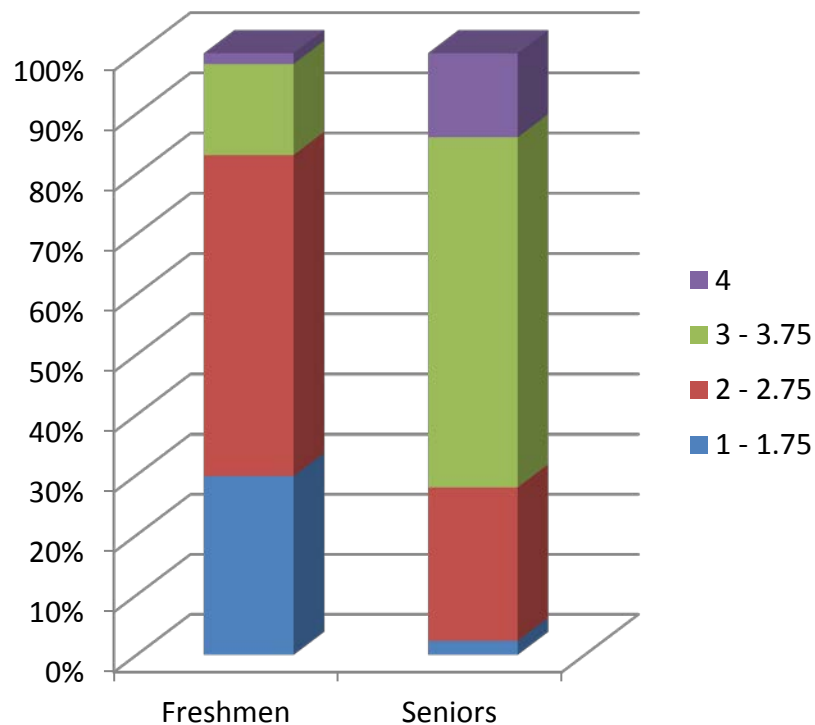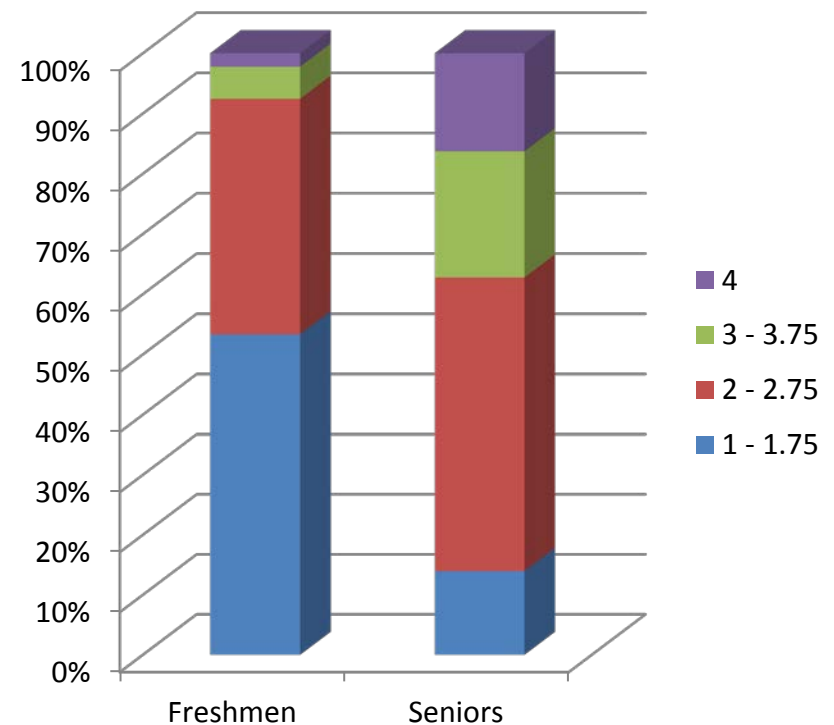
Freshman Baseline/Senior Exiting Comparisons
Freshman *n* = 225; Senior *n* = 43

Trait/ Performance Level	Development	Convention/Format	Communication Style
1 – 1.75 Freshmen	67 (30%)	120 (53%)	53 (24%)
1 – 1.75 Seniors	1 (2%)	6 (14%)	5 (12%)
2 – 2.75 Freshmen	120 (53%)	88 (39%)	119 (53%)
2 – 2.75 Seniors	11 (26%)	21 (49%)	15 (35%)
3 – 3.75 Freshmen	34 (15%)	12 (5%)	53 (24%)
3 – 3.75 Seniors	25 (58%)	9 (21%)	21 (49%)
4 Freshmen	4 (2%)	5 (2%)	0 (0%)
4 Seniors	6 (14%)	7 (16%)	2 (5%)
Grand Total Freshmen	225 (100%)	225 (100%)	225 (100%)
Grand Total Seniors	43 (100%)	43 (100%)	43 (100%)

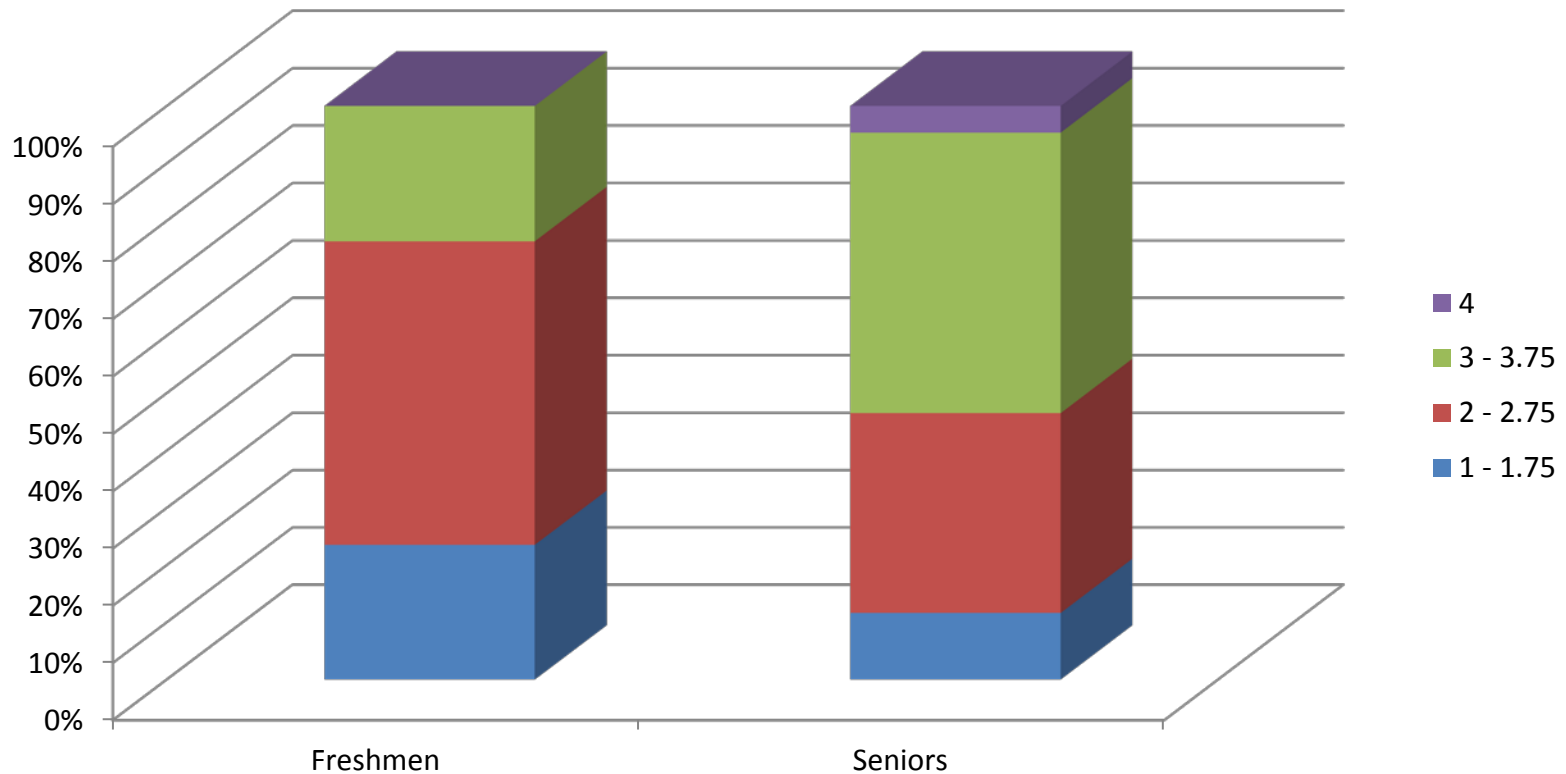# Freshman Baseline/Senior Exiting Comparisons
Freshman *n* = 225; Senior *n* = 43

Freshman Baseline/Senior Exiting Comparisons
Freshman *n* = 225; Senior *n* = 43

Communication Style

Freshman Baseline/Senior Exiting Assessment Inter-Rater Agreement Results

Trait/ Agreement	Development (Conservative Kappa = .230; Liberal Kappa = .915)	Convention/Format (Conservative Kappa = .330; Liberal Kappa = .967)	Communication Style (Conservative Kappa = .163; Liberal Kappa = .912)
Agree	123 (46%)	151 (56%)	118 (44%)
Difference = 1 point or less	127 (47%)	110 (41%)	131 (49%)
Difference = 1.5 to 2 points	18 (7%)	7 (3%)	18 (7%)
Difference = 2.5 to 3 points	0 (0%)	0 (0%)	1 (0%)
Total	268 (100%)	268 (100%)	268 (100%)

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

Academic Year 2014 - 2015

Review Procedures

- One hundred eighty-two (182) of the 225 freshmen who had completed baseline assessments during Week of Welcome completed similar assessments at the end of First Year Seminar (FYS). FYS assessments were evaluated across the same eight criteria (traits) used to score freshman baseline assessments. Scoring methodology also was the same.

Freshman Baseline/FYS Comparisons

Mean Scores on a scale of 1 – 4, with 4 being the highest possible score
n = 182 (Information Needed, Viewpoint, and Recommendations ns)
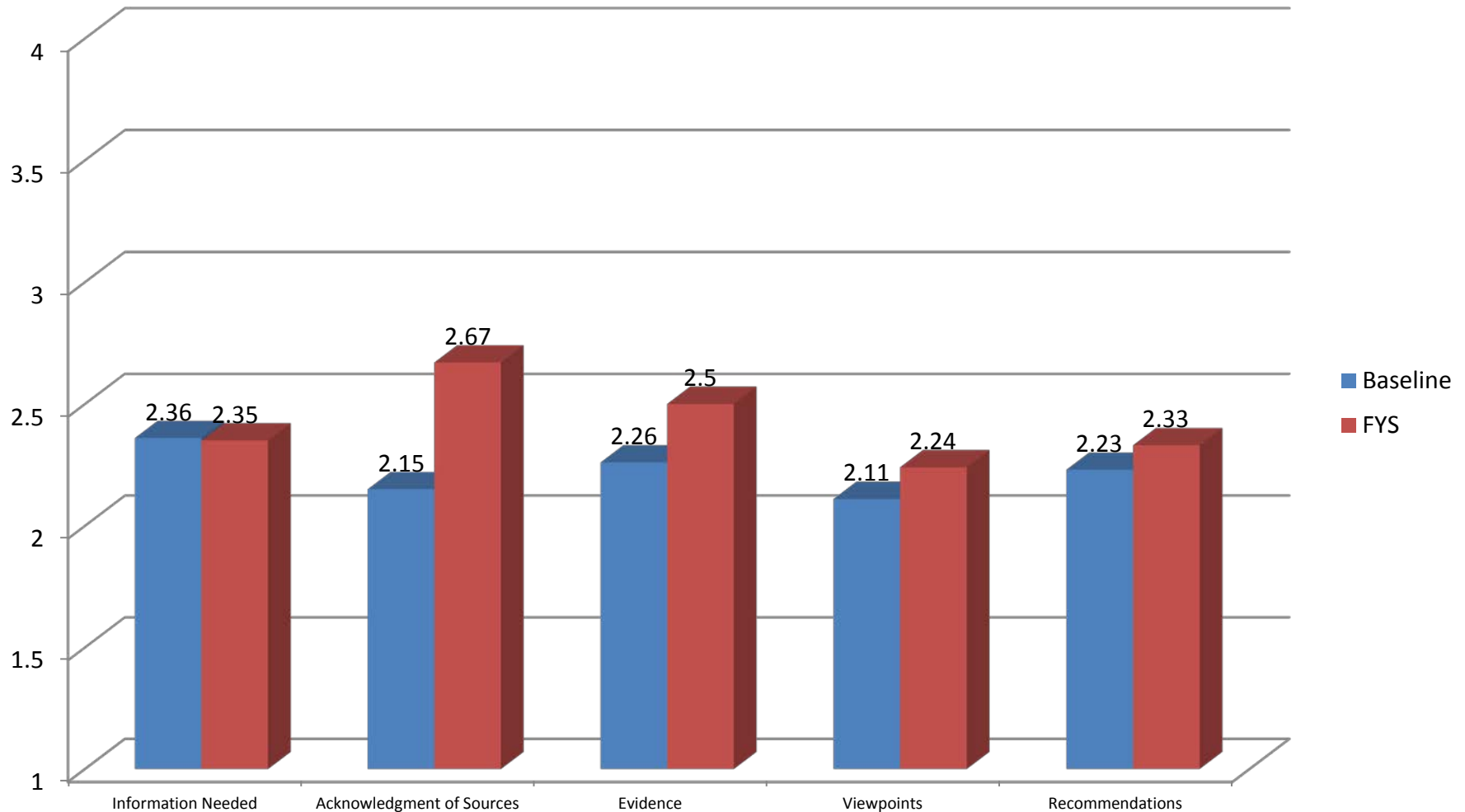
Freshman Baseline/FYS Comparisons
n = 182

| Trait/ Performance Level	Info Needed	Acknowledgment of Sources	Evidence	Viewpoints	Recommendations
1 – 1.75 Baseline	35 (19%)	43 (24%)	42 (23%)	46 (25%)	49 (27%)
1 – 1.75 FYS	23 (13%)	26 (14%)	20 (11%)	27 (15%)	34 (19%)
2 – 2.75 Baseline	92 (51%)	109 (60%)	99 (54%)	113 (62%)	95 (52%)
2 – 2.75 FYS	116 (64%)	64 (35%)	98 (54%)	122 (67%)	103 (57%)
3 – 3.75 Baseline	51 (28%)	25 (14%)	39 (21%)	21 (12%)	36 (20%)
3 – 3.75 FYS	41 (23%)	77 (42%)	61 (34%)	33 (18%)	42 (23%)
4 Baseline	4 (2%)	5 (3%)	2 (1%)	2 (1%)	2 (1%)
4 FYS	2 (1%)	15 (8%)	3 (2%)	0 (0%)	3 (2%)
Grand Total Baseline	182 (100%)	182 (100%)	182 (100%)	182 (100%)	182 (100%)
Grand Total FYS	182 (100%)	182 (100%)	182 (100%)	182 (100%)	182 (100%)

Freshman Baseline/FYS Comparisons
n = 182

Information Needed
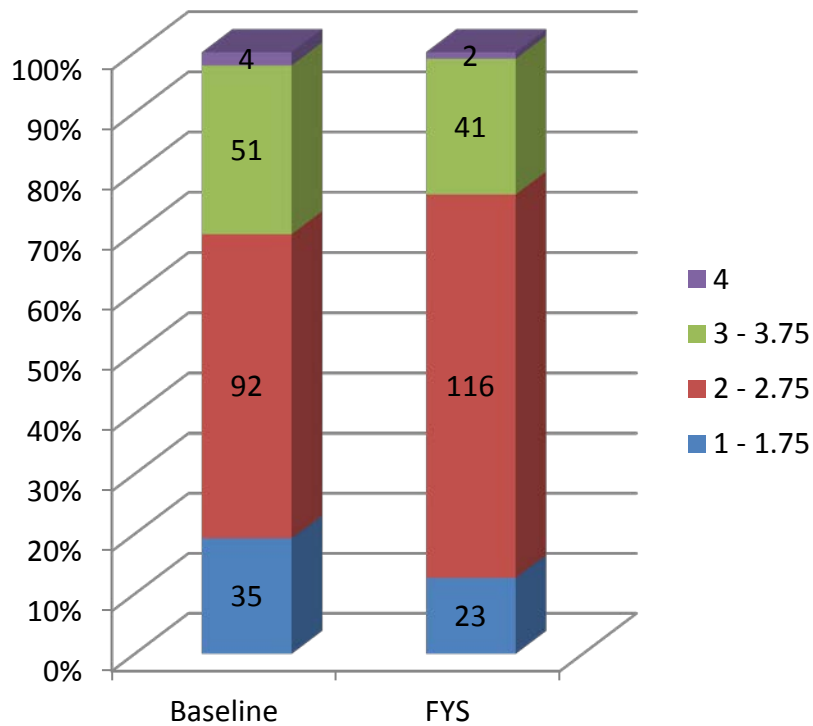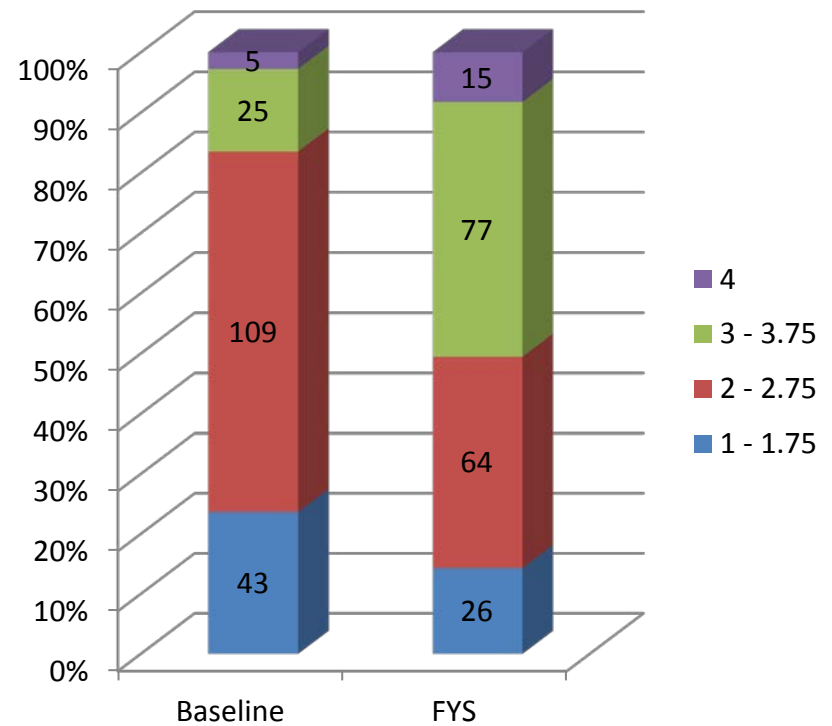
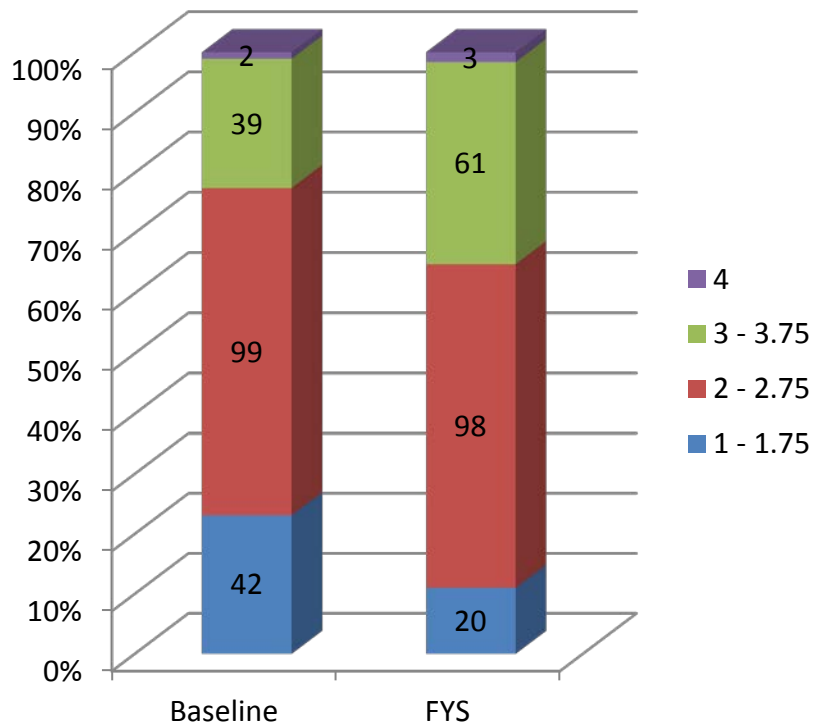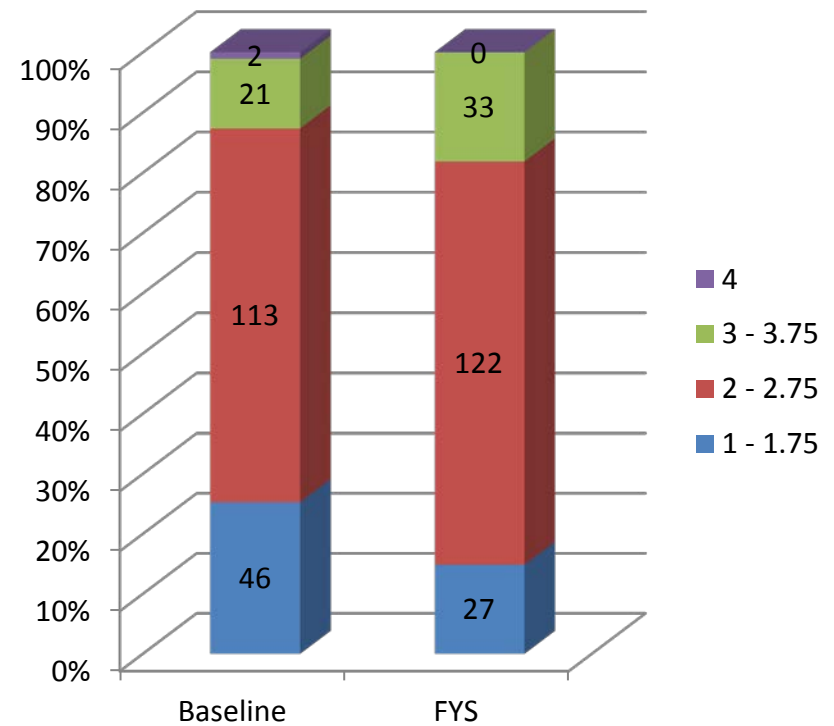**Acknowledgment of Sources**

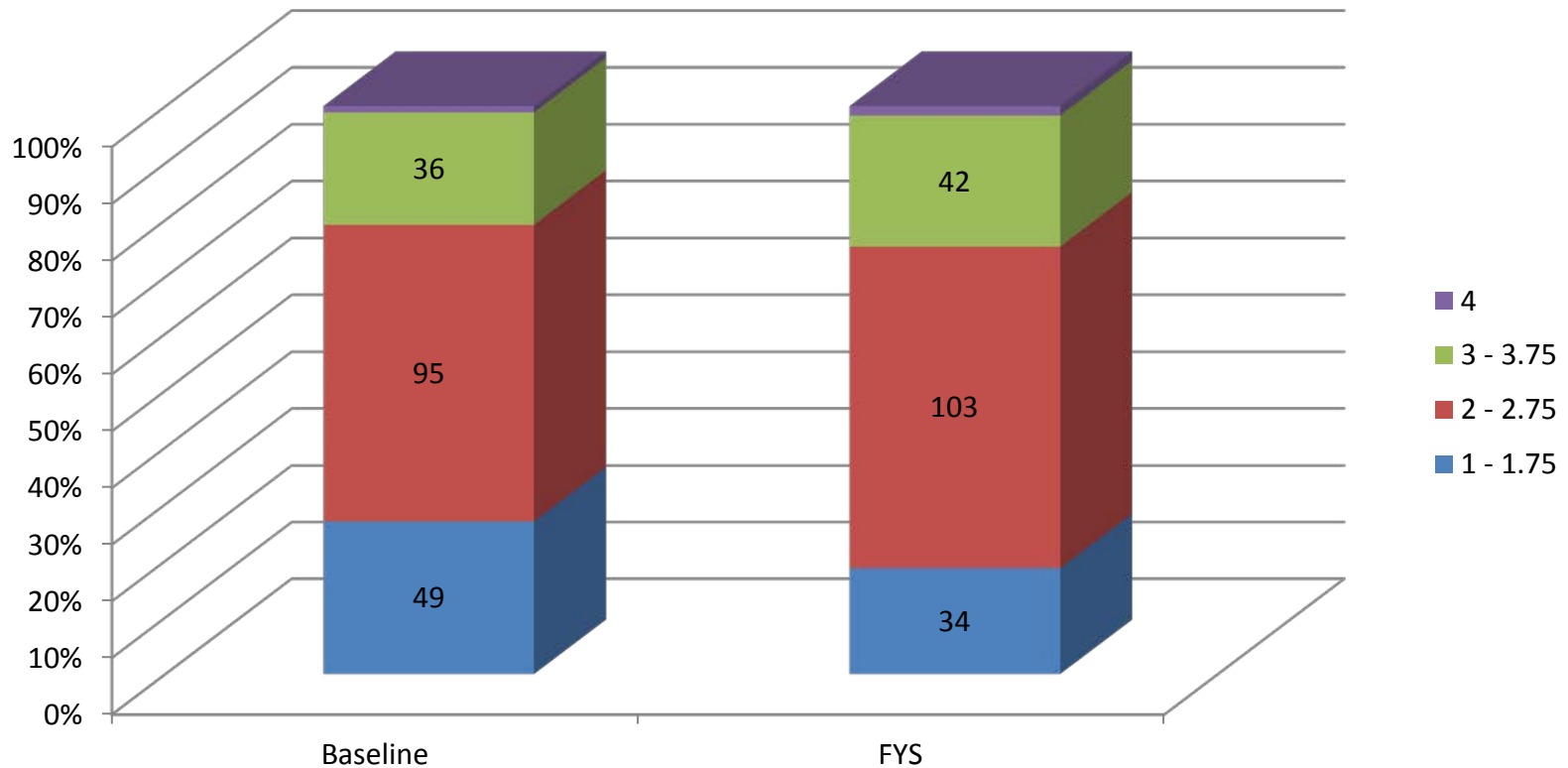# Freshman Baseline/FYS Comparisons
n = 182

Freshman Baseline/FYS Comparisons
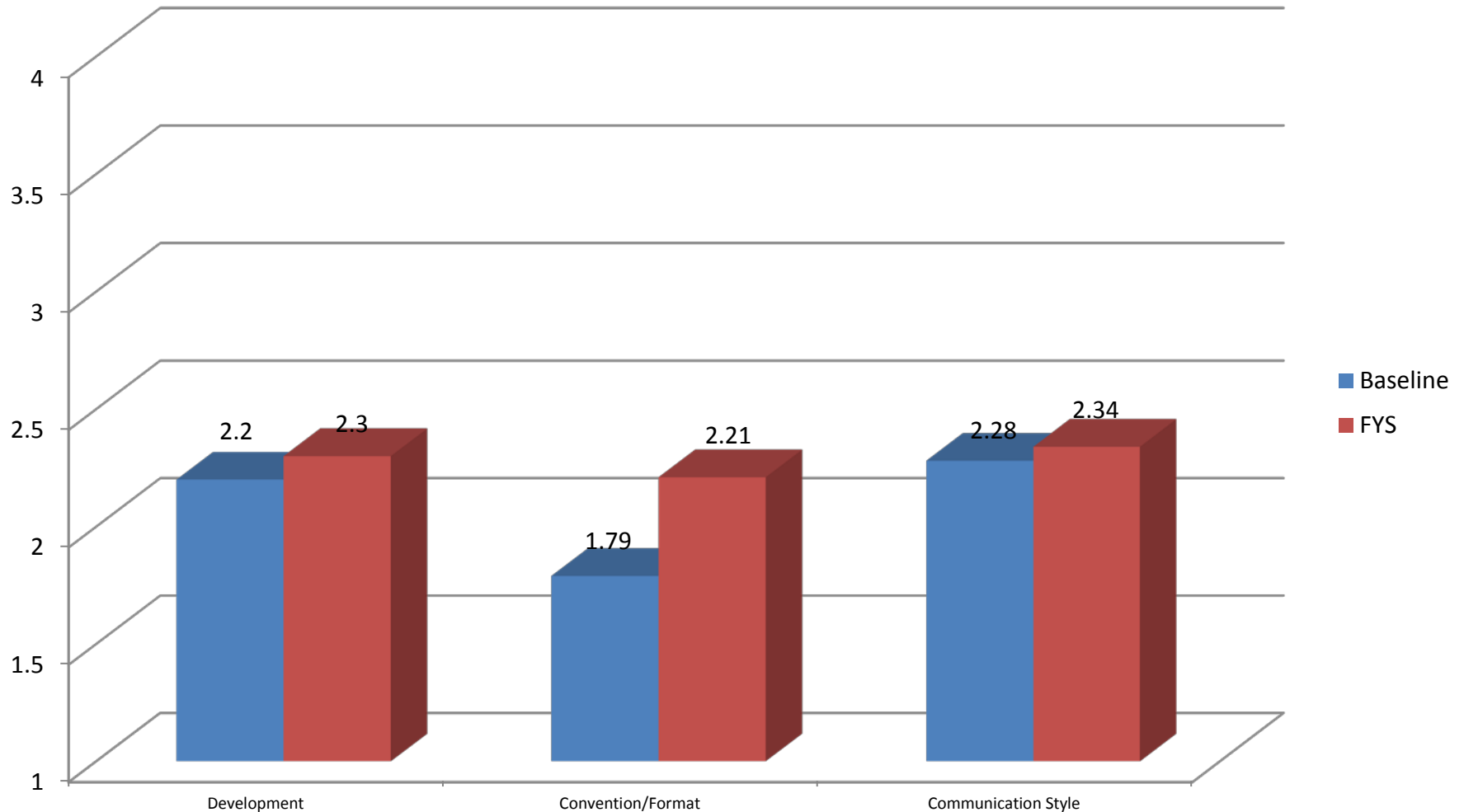n = 182

Recommendations

FYS Inter-Rater Agreement Results

Trait/ Agreement	Info Needed (Conservative Kappa = .299; Liberal Kappa = .959)	Acknowledgment of Sources (Conservative Kappa = .379; Liberal Kappa = .954)	Evidence (Conservative Kappa = .240; Liberal Kappa = .952)	Viewpoints (Conservative Kappa = .244; Liberal Kappa = .926)	Recommendations (Conservative Kappa = .192; Liberal Kappa = .940)
Agree	98 (54%)	100 (55%)	92 (51%)	103 (57%)	83 (46%)
Difference = 1 point or less	78 (43%)	75 (41%)	84 (46%)	69 (38%)	90 (49%)
Difference = 1.5 to 2 points	6 (3%)	6 (3%)	6 (3%)	69(5%)	9 (5%)
Difference = 2.5 to 3 points	0 (0%)	1 (1%)	0 (0%)	1 (1%)	0 (0%)
Total	182 (100%)	182 (100%)	182 (100%)	182 (100%)	182 (100%)

Freshman Baseline/FYS Comparisons

Mean Scores on a scale of 1 – 4, with 4 being the highest possible score
n = 182 (Development and Communication Style *ns*)
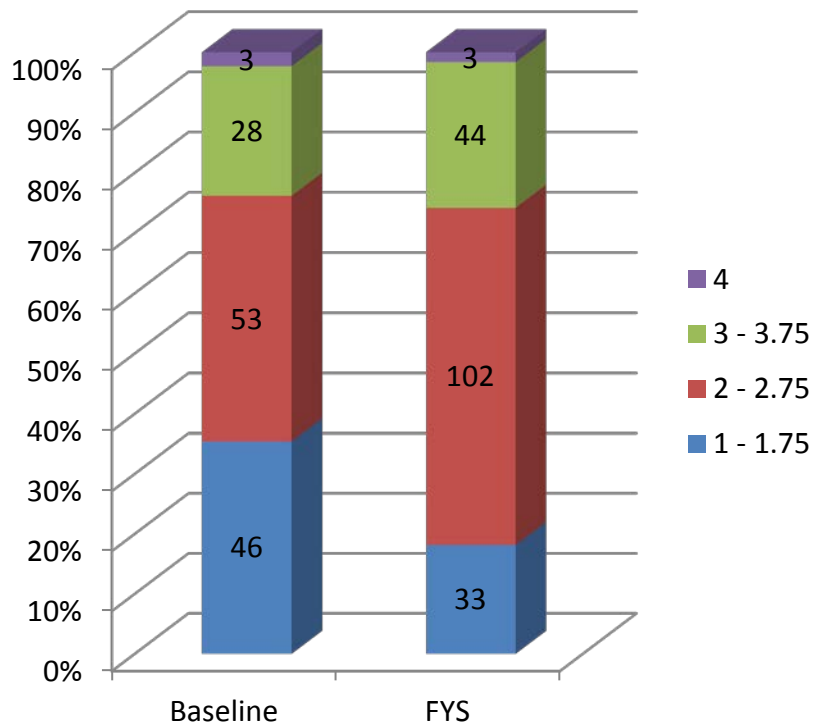
Freshman Baseline/FYS Comparisons
n = 182

| Trait/ Performance Level	Development	Convention/Format	Communication Style
1 – 1.75 Baseline	46 (35%)	96 (53%)	40 (22%)
1 – 1.75 FYS	33 (18%)	50 (27%)	31 (17%)
2 – 2.75 Baseline	53 (41%)	73 (40%)	97 (53%)
2 – 2.75 FYS	102 (56%)	100 (55%)	107 (59%)
3 – 3.75 Baseline	28 (22%)	8 (4%)	45 (25%)
3 – 3.75 FYS	44 (24%)	30 (16%)	44 (24%)
4 Baseline	3 (2%)	5 (3%)	0 (0%)
4 FYS	3 (3%)	2 (1%)	0 (0%)
Grand Total Baseline	182 (100%)	182 (100%)	182 (100%)
Grand Total FYS	182 (100%)	182 (100%)	182 (100%)

Freshman Baseline/FYS Comparisons
n = 182
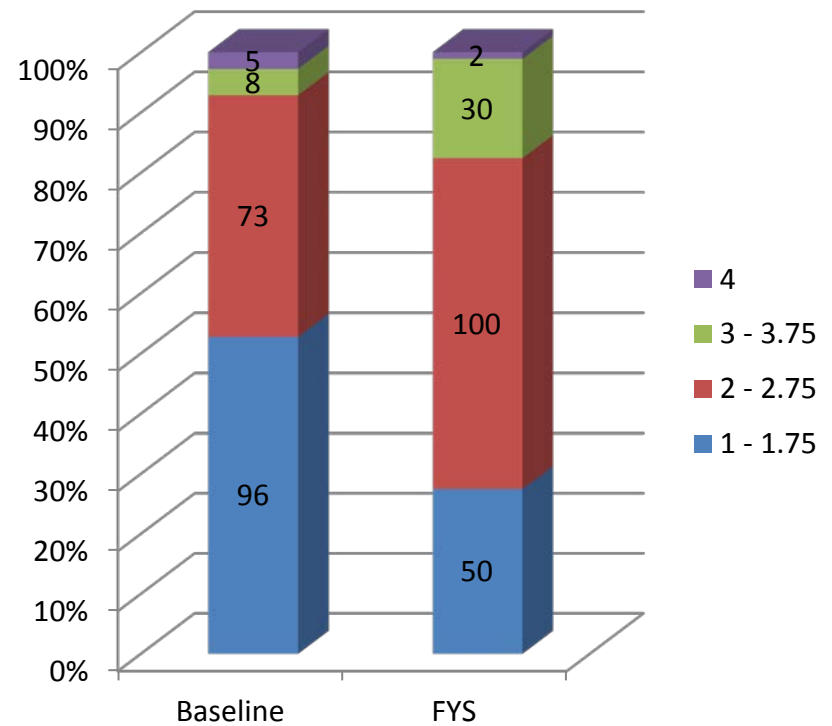
Freshman Baseline/FYS Comparisons
n = 182

Communication Style

Legend:
- 4
- 3 - 3.75
- 2 - 2.75
- 1 - 1.75

Baseline: 0, 45, 97, 40
FYS: 0, 44, 107, 31
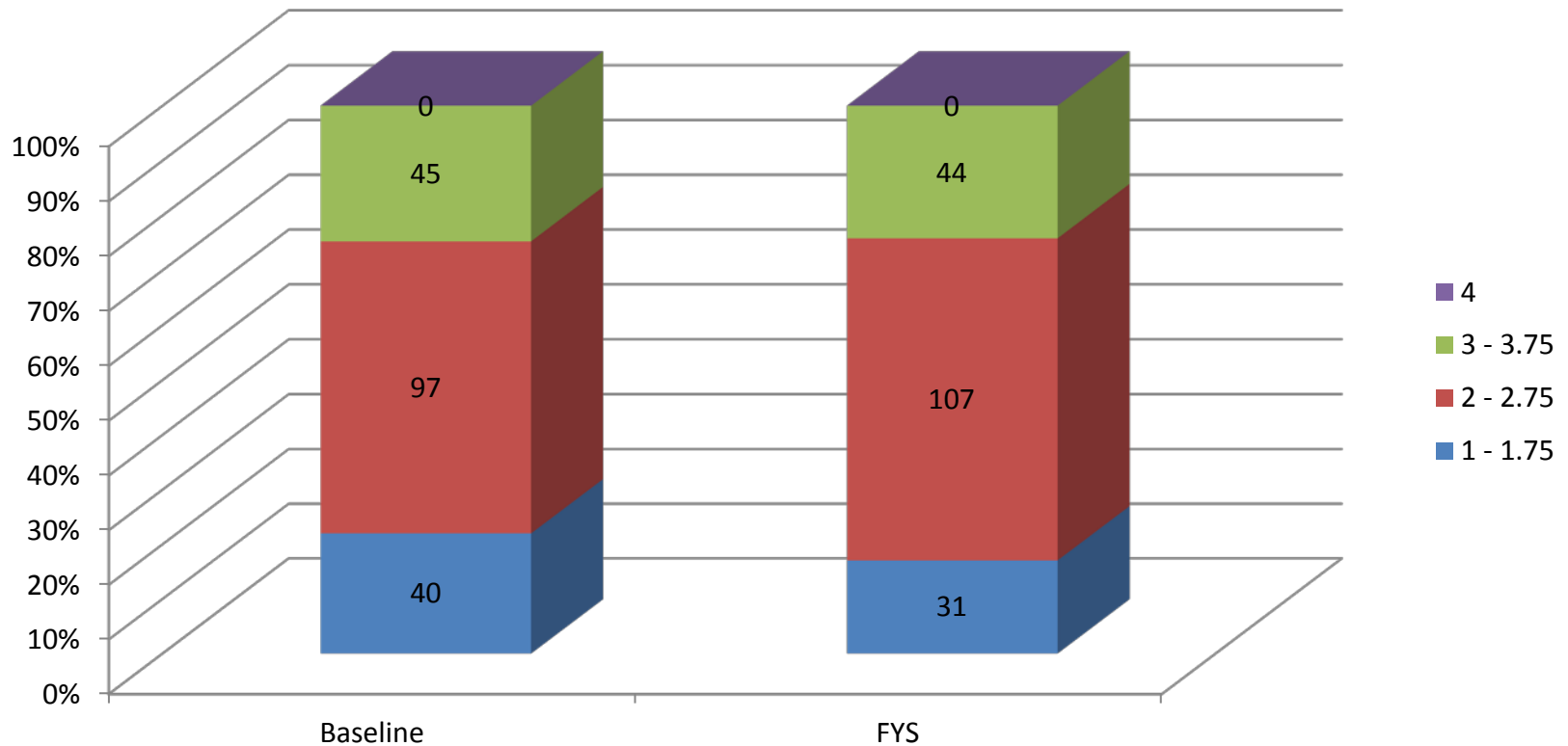
FYS Inter-Rater Agreement Results

Trait/ Agreement	Development (Conservative Kappa = .388; Liberal Kappa = .938)	Convention/Format (Conservative Kappa = .044; Liberal Kappa = .872)	Communication Style (Conservative Kappa = .092; Liberal Kappa = .903)
Agree	109 (60%)	64 (35%)	77 (42%)
Difference = 1 point or less	66 (36%)	99 (54%)	91 (50%)
Difference = 1.5 to 2 points	6 (3%)	19 (10%)	14 (8%)
Difference = 2.5 to 3 points	1 (1%)	0 (0%)	0 (0%)
Total	182 (100%)	182 (100%)	182 (100%)