Analysis of Artifacts from Marshall's General Education Assessment Repository

Spring Semester 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)

Executive Summary

Background

Recommendations from 2014 Assessment (with current status in red)

GEAR Upload Process

1. Design GEAR so that instructors must upload assignment instructions before students can upload artifacts. Although not statistically significant in most cases, we noted a trend for a greater number of scores of 100 (assignments misaligned to outcomes) when the instructor had failed to upload the assignment instructions. Beginning with the spring 2015 GEAR assignments, instructors could not create assignments without uploading an assignment instruction file.
2. Redesign GEAR so that instructors (or students) must tag the assignment's outcome(s)/trait(s) and the outcome/trait performance levels to which the assignment is written. The Workgroup felt that this step would cause instructors and students to think more carefully about exactly what knowledge/skills are demonstrated in the artifact, as there are different outcome statements for each trait at each performance level. Beginning with the spring 2015 GEAR assignments, instructors were asked to indicate the performance level (introductory, milestone, capstone, advanced) of each trait to which the assignment was aligned.
3. Redesign GEAR so that, if instructors or students align an assignment/artifact to more than one outcome or to more than two outcome traits, they will be required to indicate a rank-order for the outcomes/traits tagged. In other words, reviewers would like to know if the outcome/traits they are assessing were the primary focus of the assignment, or a secondary focus. Beginning in spring 2015, instructors were required to indicate the primary outcome to which their assignment aligned.
4. Concern was expressed about the small percentage of outcomes assessed this year. To increase the number of artifacts reviewed from each outcome, the workgroup recommended that we rotate outcomes on a two-three year basis. For example, we might review artifacts tagged to only three-four outcomes in year 1, the next three-four in year 2, etc. For the summer 2015 assessment, we assessed artifacts that aligned with the following outcomes: *Intercultural Thinking, Ethical and Civic Thinking*, and *Communication Fluency.*

5. The workgroup strongly recommended that uploaded artifacts be summative in nature. The nature of the artifacts (summative or formative) continues to vary by course.
6. The workgroup recommended that we continue to assess artifacts for one outcome (can have multiple traits tagged for outcome) at a time. We continued this process. Reviewers assessed artifacts aligned to each of these outcomes, spending two days on artifacts from each: *Intercultural Thinking*, *Ethical and Civic Thinking*, and *Communication Fluency*.
7. The workgroup recommended that uploaded artifacts include process papers when tagged to an outcome/trait/performance level that addresses process rather than product. This recommendation has not yet been accomplished.
8. The workgroup recommended that instructors be provided with clearer definitions of rubric traits, especially for those of Inquiry-Based Thinking. This recommendation has not yet been accomplished.
9. The workgroup did not find the GEAR free text box asking students why they (or their instructors) had aligned artifacts with specific outcome(s)/trait(s). They recommended that we rely instead on formal process papers for the process-based outcome(s)/trait(s). This recommendation has not yet been accomplished.

General Procedures for 2015 Assessment

Recommended changes outlined above in red were made to GEAR before the spring semester of 2015. All students enrolled in FYS as well as in courses carrying multicultural, international, writing intensive, service learning, and critical thinking (CT) designations were asked to upload artifacts to GEAR. Instructors were asked to create assignments aligned to *Communication Fluency* (writing intensive courses), *Ethical and Civic Thinking* (service learning courses), and *Intercultural Thinking* (multicultural and international courses). Instructors were told that it was not necessary to align the assignments to <u>all</u> traits for the specified learning outcome; that they should align them only to those traits the assignment specifically addressed. Instructors also were asked to indicate the *performance level* they expected students to achieve. Since FYS addresses five of the University's outcomes (*Information Literacy* and *Inquiry-Based*, *Integrative*, *Intercultural*, and *Metacognitive Thinking*) and CT courses address *Integrative Thinking* and four additional university outcomes of their choice, it was left to instructors and/or students to decide to which of the course's outcome(s) their assignments aligned. It was possible for a single assignment to align to any number of outcomes and traits. However, as noted above, instructors were required to specify the <u>primary </u>outcome to which the assignment aligned.

In May 2015 a group of nine faculty representing several academic colleges from across the university evaluated a sample of these artifacts using outcome specific rubrics. These rubrics, which can be accessed by clicking on the hyperlink for each Domain of Critical Thinking at www.marshall.edu/assessment/LearningOutcomes.aspx, were developed as a series of outcome statements for each trait, specifying what students should be able to do at four levels of increasing challenge (introductory, milestone, capstone, and advanced). For purposes of Marshall's Baccalaureate Degree Profile, we expect students to perform at Level 3 (capstone) by the time of graduation. Based on last spring's recommendations, we focused our assessment efforts on three of the university's outcomes; *Communication Fluency*, *Ethical and Civic Thinking*, and *Intercultural Thinking*. This also allowed us to assess five course types (Writing Intensive [*Communication Fluency*], Multicultural [*Intercultural Thinking*], International [*Intercultural Thinking*], Service Learning [*Ethical and Civic Thinking*], and CT courses [potentially all three

outcomes]. In an effort to obtain a sample that would be as free as possible from assignments that were misaligned (i.e. not aligned to the correct outcomes), we decided to restrict our sample for each of the outcomes to those that specified these outcomes as the *primary* focus of the course assignment. However, due to an initial error in sampling (which was quickly corrected), five artifacts included in the sample for *Intercultural Thinking* were aligned to that outcome as a secondary outcome.   Our final sample consisted of 324 artifacts, 108 per outcome. Each artifact was read by two independent reviewers. This project was coordinated by the Office of Assessment and Quality Initiatives.

Scoring Procedures

Evaluators assessed each artifact using the following scale:

Special Scoring Codes	
Score	**Explanation**
100	In the opinion of the evaluator, the artifact was misaligned with the outcome/trait to which the instructor or student had tagged it.
99	The student did not upload the correct assignment or there was a technical problem with the upload that prevented the artifact from being assessed.
Regular Scoring Codes	
These codes were given to artifacts that, in the opinion of the evaluator, were aligned with appropriate outcomes/traits and contained enough information to allow assessment.	
0	The artifact did not demonstrate the minimum level of performance expected at the introductory level.
1	The artifact demonstrated introductory level performance.
2	The artifact demonstrated milestone level performance.
3	The artifact demonstrated capstone level performance.
4	The artifact demonstrated advanced level performance. We should note that this is the performance level expected of graduate students, so we would expect it to be rarely achieved at the undergraduate level.

Please see the supporting information that follows this summary for a detailed explanation of scoring procedures.

General Information about the Sample

Approximately 42% (137) of the artifacts in our sample were drawn from courses at the 100/200 level, with the remaining 58% (187) drawn from courses at the 300/400 level. The reason why a greater proportion of artifacts were pulled from upper level courses was because we wished to assess the University's Service Learning Courses, which address *Ethical and Civic Thinking*, and Marshall offers more Service Learning courses at the 300/400 level than at the 100/200 level. This differentiates our sample from that used in summer 2014, which contained twice as many artifacts from 100/200 as from 300/400 level courses.   Approximately 40% of the students in the sample were seniors, which also differed from

last year's sample, which was weighted toward freshmen. Unlike last year's sample, the sample this year had equal numbers of artifacts (108) aligned to each of the three outcomes assessed.

Results and Analysis

One challenge in reporting results of GEAR assessment is that, although we assessed 324 artifacts, results were analyzed by each outcome trait. As previously noted, instructors or students were free to align assignments/artifacts to as many (or as few) outcomes and traits as they deemed appropriate. Although we assessed each artifact for only one outcome, most of these aligned to more than one of the outcome's traits. For purposes of this assessment, we also added a trait (global contexts) to the *Intercultural Thinking* outcome and deleted a trait (context/audience) to the *Communication Fluency* outcome, bringing the total number of traits across the three outcomes to 13 (3 for *Communication Fluency*, 4 for *Ethical and Civic Thinking*, and 6 for *Intercultural Thinking*). A perusal of our supporting documentation shows that the artifacts evaluated by the Assessment Workgroup tagged to a total of 799 traits. However, scores for only 661 (83%) of those traits were usable for calculating means. One hundred thirty-eight were discarded either because they were judged not to align with the traits (91; 11%) or were not able to be assessed because of student upload error (47; 6%). The chart below shows the number of artifacts aligned to each trait, the number excluded from the analysis due to receiving scores of 100 (misalignment) or 99 (student upload error), and the resulting number of scores able to be used for the analysis of means. Focusing on assessing three outcomes this year helped us to significantly increase the number of scores able to be used over last year and assessing only artifacts that had been tagged to primary outcomes reduced the number of misalignments.

Outcome	Trait	Total Artifacts Aligned	# Misaligned (Scores of 100)	# Not Able to be Assessed (Score of 99)	Total # Excluded from Analysis of Means	Total Usable Artifacts
Communication Fluency	Design/Organization	88	0	8	8	80
	Diction	58	0	5	5	53
	Communication Style	77	0	7	7	70
Ethical and Civic Thinking	Ethical Self-Awareness	87	7	1	8	79
	Professional Rules and Standards of Conduct	59	9	1	10	49
	Civic Well-Being	80	2	1	3	77
	Complex Ethical Issues	44	8	1	9	35
Intercultural Thinking	Own Culture	68	6	6	12	56
	Other Cultures	81	5	5	10	71
	Communication with	15	4	2	6	9

	Others from Different Cultures					
	Global Awareness	58	21	3	**24**	**34**
	Cultural Conflict	50	7	4	**11**	**39**
	Global Contexts	34	22	3	**25**	**9**

Results for *Communication Fluency* (diction and communication style) showed that mean scores of students in 300/400 level courses were significantly higher than those for students in 100/200 level courses. Results did not differ by course level for any trait of *Intercultural Thinking* and the small number of artifacts from 100/200 level courses for *Ethical and Civic Thinking* made course level comparison difficult. Juniors and seniors outperformed freshmen and sophomores on *Communication Fluency* (diction and communication style), but mean differences based on class rank were not significant for the other two outcomes.

Overall results showed mean performance for traits to range from 1.01 (*Ethical and Civic Thinking*: complex ethical issues) to 2.43 (*Communication Fluency:* design/organization). Mean performance for artifacts uploaded from 100/200 level courses ranged from 1.17 (*Intercultural Thinking:* communication with other cultures) to 2.11 (*Communication* Fluency: design/organization) and from 300/400 level courses from 1.01 (*Ethical and Civic Thinking*: complex ethical issues) to 2.54 (*Communication Fluency:* design/organization). Consistent with last year's results, *Communication Fluency* appears to be a relative strength for our students.

Results for Course Type

Writing Intensive Courses

The primary outcome to which artifacts from writing intensive courses aligned was *Communication Fluency.* Usable scores were obtained by trait as follows:

Trait	Course Level	Number	Mean Score
Design/Organization	100/200	25	2.09
	300/400	54	2.58
Diction	100/200	18	1.78
	300/400	34	2.53
Communication Style	100/200	24	1.53
	300/400	45	2.17

Mean scores for diction and communication style were significantly higher for 300/400 level courses than for 100/200 level courses.

Multicultural Courses

The primary outcome to which artifacts from multicultural courses aligned was *Intercultural Thinking*. Multicultural courses were most likely to align to the first two traits of the *Intercultural outcome* (own culture and other cultures).   Usable scores were obtained by trait as follows:

Trait	Course Level	Number	Mean Score
Own Culture	100/200	35	1.22
	300/400	10	1.55
Other Cultures	100/200	41	1.7
	300/400	13	1.73
Communication with Others from Different Cultures	100/200	2	1.0
	300/400	4	1.25
Global Awareness	100/200	5	1.4
	300/400	9	1.28
Cultural Conflict	100/200	7	1.43
	300/400	7	1.57
Global Contexts	100/200	0	---
	300/400	4	1.06

Although there were no significant differences between these means based on course level, we note the small number of alignments in each cell for the last four traits.

International Courses

The primary outcome to which artifacts from international courses aligned was *Intercultural Thinking*. Although the overall number of International courses in the sample was smaller than the number of Multicultural courses, we see that these courses were more likely to align assignments to the fourth and fifth outcomes of the *Intercultural outcome* (global awareness and cultural conflict).   Usable scores were obtained by trait as follows:

Trait	Course Level	Number	Mean Score
Own Culture	100/200	9	1.56
	300/400	0	----
Other Cultures	100/200	10	2.08
	300/400	6	1.88
Communication with Others from Different Cultures	100/200	0	----
	300/400	2	2.0
Global Awareness	100/200	15	1.87

	300/400	5	1.6
Cultural Conflict	100/200	18	1.88
	300/400	2	1.13
Global Contexts	100/200	0	---
	300/400	5	1.6

There were no significant differences between these means based on course level; however we note that only the traits *other cultures, global awareness, and cultural conflict* had /n/s larger than 100 and these were in 100/200 level courses.

Service Learning Courses

The primary outcome to which artifacts from service learning courses aligned was *Ethical and Civic Thinking.* Usable scores were obtained by trait as follows:

Trait	Course Level	Number	Mean Score
Ethical Self-Awareness	100/200	18	1.99
	300/400	61	1.4
Professional Rules and Standards of Conduct	100/200	4	1.88
	300/400	45	1.27
Civic Well-Being	100/200	0	----
	300/400	77	1.56
Complex Ethical Issues	100/200	0	----
	300/400	35	1.01

The only statistically significant difference based on course level showed that students in 100/200 level courses scored significantly higher than students in 300/400 level courses in *ethical self-awareness*. We note, however, that the 300/400 course level sample for this trait has almost 3.5 times more students than the sample for 100/200 level courses.

Critical Thinking (CT) Courses

CT courses included in the assessment sample aligned to either *Communication Fluency or to Intercultural Thinking.* All CT courses are at the 100/200 level. Results are below:

Communication Fluency			Intercultural Thinking		
Trait	Number	Mean Score	**Trait**	Number	Mean Score
Design/Organization	25	2.09	**Own Culture**	19	1.71
Diction	18	1.78	**Other Cultures**	20	1.88
Communication Style	24	1.53	**Communication with Others from Different Cultures**	2	1.0

			Global Awareness	18	1.83
			Cultural Conflict	26	1.78
			Global contexts	0	-----

Recommendations from the 2015 Assessment Workgroup

Recommendations Specific to the Outcomes and Assessment Rubrics

1. Redesign all university rubrics so that they are continuous in nature. This should be done by stating the Baccalaureate Degree Profile outcome statements for <u>each</u> trait and then describing four levels of increasingly sophisticated levels of performance. A revised rubric for *Intercultural Thinking* might look like this:

Trait Outcome Statements	Performance Levels			
	Level 1	Level 2	Level 3	Level 4
Evaluates generalizations about one's own cultural group(s).				
Critiques generalizations and expressions of bias about a specific cultural group.				
Analyzes how specific cultural beliefs, values and sensibilities might affect the way people in different cultural groups communicate with each other.				
Evaluates how specific approaches to global issues will affect multiple cultural communities or political institutions.				
Analyzes and **untangles** competing economic, religious, social, political, institutional, or geographical				

interests of cultural groups in conflict.				
Evaluates practical solutions to global challenges that are appropriate to their contexts.				

Reasons for this recommendation include:

- We believe that all assignments should be written to the outcome specified in the Baccalaureate Degree Profile. This will provide students with the maximum amount of practice in achieving the goals Marshall University has set for them by the time of graduation. It will have the added advantage of students seeing these outcomes occurring across courses within the Core Curriculum, thus promoting integration of outcomes across courses.
- This will reduce confusion among instructors as to what their assignments need to address. At present, most rubrics consist of outcome statements for each performance level, allowing assignments that vary across courses in terms of what students are expected to do.
- Interrater reliability continues to be problematic when using these rubrics, with the greatest problem occurring with misalignments. And, a quick perusal of the interrater reliability data show that often one rater feels that the assignment has been misaligned with the rubric, but the other does not. This was especially true for several trait of the *Intercultural Thinking* rubric.

2. Form committees consisting of key stakeholders for each university outcome to revise the university outcomes (if needed) and to revise the rubrics. For example, the committee that reviews the *Intercultural Thinking* outcome and rubric should consist of faculty who teach *International* and *Multicultural* courses, a representative from the Office of Intercultural Affairs, a representative from INTO-Marshall, and other key stakeholders as deemed appropriate. The committee that reviews the rubric for *Ethical and Civic Thinking* should consist of the Director of Service Learning, faculty who teach Service Learning courses, and additional faculty from across the University. Faculty should critically examine course assignments to help inform rubric development.
3. Before *Multicultural* and *International* courses are recertified by the General Education Council, faculty teaching these courses should attend a minimum of a one-hour workshop to develop assignments that align to one or more of the traits of the *Intercultural* rubric.

General Recommendations

1. The Assessment Office should provide a list of students who did not complete GEAR uploads to course instructors and a list of instructors who did not create assignments in GEAR to department chairs.
2. The Assessment Office should provide the GEAR shell to instructors several weeks before the beginning of the semester and update the student roster for each course the second week of the semester.
3. The Assessment Office should communicate with instructors that student work uploaded to GEAR should have enough substance to permit evaluation, i.e. should be summative, rather than formative, in nature. This recommendation was repeated from last year.
4. Instructors should be reminded of the importance up uploading assignment instructions to GEAR. This recommendation was made again because, despite the fact that an assignment file must be uploaded for an assignment to be created, a few instructors uploaded other types of file, e.g. entire course syllabus, GEAR upload instructions.

Longitudinal Analysis

For the initial assessment of artifacts uploaded to GEAR (summer 2013), all artifacts assessed were drawn from the university's First Year Seminar (FYS) course and we used these artifacts to assess all nine university outcomes. Mean performance across students ranged from a low of 0 for *Intercultural Thinking* (communication with other cultures) to a high of 1.24 for *Communication Fluency* (design/organization and diction).   However, since artifacts were spread among so many outcomes, many traits had very small numbers (9 for communication with other cultures as compared to 24 for design/organization and 23 for diction). Other than the fact that all students included in the 2013 sample were freshmen, low means can be attributed to the fact that we had not yet settled on a score for misaligned artifacts, defaulting many of the scores to 0.

The second assessment of artifacts uploaded to GEAR (summer 2014) also included all nine outcomes, but we included artifacts from *Multicultural, International, Service Learning,* and *Writing Intensive* courses, in addition to those from FYS. The sample, however, continued to be skewed toward artifacts from lower level courses with freshman being the modal class rank for student artifacts in our sample. We decided to assign special codes to artifacts we felt to be misaligned to the outcomes or in cases of student upload or other technical issues that prevented assessment. This allowed us to see which outcomes/traits resulted in the greatest amount of confusion during the outcome/trait alignment process and resulted in recommendations to make sure instructors uploaded assignment instructions, specified the primary outcome to which their assignment aligned, and identified the performance level to which the assignment was written. Due to assessing all nine university outcomes again in 2014, we continued to have small numbers of artifacts aligned to each outcome, which led to the recommendation that we choose only three outcomes to assess in 2015, three more in 2016, and the last three in 2017 and continue to assess on a three-year cycle.

The third assessment of artifacts uploaded to GEAR (summer 2015) consisted of an in-depth assessment of artifacts that instructors aligned to the following outcomes as primary: *Intercultural Thinking* (due to sampling error, five of the alignments for *Intercultural Thinking* were secondary), *Ethical and Civic Thinking,* and *Communication Fluency*. One hundred eight artifacts were included for each outcome, resulting in a total of 324 artifacts. This sample resulted in higher numbers for each outcome trait.   Results this year suggested a need to redesign rubrics to be continuous, rather than categorical, in nature. We recommended that all assignments address the outcomes articulated in Marshall's Baccalaureate Degree Profile, rather than lower levels as articulated in present rubrics. To that end, workgroups will be formed during academic year 2015 – 2016 to revise the rubrics and we will communicate with course instructors regarding writing assignments to the University outcome statements. We feel that it is especially important to involve faculty who teach courses that align to the university's outcomes to be involved in revising the rubrics and outcomes (if necessary).

Finally, the past two years of assessment data have shown that Marshall's students improve their writing skills as they move through the curriculum and, specifically, as they pass from 100/200 level writing intensive courses to 300/400 level writing intensive courses.

Supporting Documentation

General Education Assessment Repository (GEAR) Artifact Assessment

Spring 2015

Due to a sampling error, five of the artifacts assessed for Intercultural Thinking had that outcome as a secondary, rather than a primary, tag. Three of these five artifacts were from First Year Seminar (FYS), which we did not intend to include in our sample. Because the five artifacts assessed for secondary tags did not differ significantly from the rest of the sample, I have included them in the analysis. Two additional artifacts with primary tags for Intercultural Thinking came from FYS, brining a total of 5 out of 108 artifacts that were assessed for Intercultural Thinking coming from FYS.

Course Type	Primary Domain	Domain Assessed	Traits Assessed	Scores	Result
FYS	Metacognitive Thinking	Intercultural Thinking	Own Culture	0.75	Score
			Other Cultures	100	Misaligned
			Communication	1.5	Score
FYS	Not Clear	Intercultural Thinking	Own Culture	99 for all traits	Student Error in Upload – Eliminated from Analysis
			Global Awareness		
			Cultural Conflict		
			Global Context		
FYS	Metacognitive Thinking	Intercultural Thinking	Own Culture	3.5	Score
MC and WI	Communication Fluency	Intercultural Thinking	Own Culture	100	Misaligned
			Other Cultures	4	Score
CT and MC	Communication Fluency	Intercultural Thinking	Own Culture	3.5	Score

Outcomes Assessed

Outcome	Abbreviation	Traits	Abbreviations
Communication Fluency	CF	Design/Organization	Design
		Diction	Diction
		Communication Style	Style
Ethical and Civic Thinking	ECT	Ethical Self-Awareness	ESA
		Professional Rules and Standards of Conduct	Rules
		Civic Well-Being	CWB
		Complex Ethical Issues	CEI
Intercultural Thinking	ICT	Own Culture	Own
		Other Cultures	Other
		Communication with Other Cultures	Communication
		Global Awareness	Global Awareness
		Cultural Conflict	Conflict
		Global Contexts	Global Contexts

Course Types

Course Type	Abbreviation
Critical Thinking	CT
Multicultural	MC
International	INT
Writing Intensive	WI
Service Learning	SL
First Year Seminar	FYS

Course Types in CF, ECT, and ICT Primary Outcome Population

Course Type	Population *n*	Sample *n*	Percent
CT and INT	43	19	44%
CT and MC	273	9	3%
CT, MC, and WI	141	14	10%
CT and WI	148	33	22%
FYS	141	5	4%
INT	18	6	33%
INT and WI	15	3	20%
MC	405	55	14%
MC and WI	54	3	6%
SL	126	97	77%
SL and WI	33	12	36%
WI	925	68	7%
Other	1	0	0%
Total	**2,323**	**324**	**14%**

Course Types in CF, ECT, and ICT Primary Outcome Population: Each Course Counted Separately for Each Category

Course Type	Population *n*	Sample *n*	Percent
CT	605	75	12%
INT	76	28	37%
MC	873	81	9%
WI	1,316	133	10%
SL	159	109	69%
FYS	141	5	4%
Total	**3,170**	**431**	**14%**

Population/Sample Comparisons for Marshall's Course Types by Course Level (*n* does not = 324)

Marshall Outcomes	Course Level = 100/200			Course Level = 300/400		
	Population	Sample	Percent	Population	Sample	Percent
Critical Thinking	605	75	12%	0	0	0%
Multicultural	731	60	8%	142	21	15%
International	43	19	44%	33	9	27%
Service Learning	18	18	100%	141	91	65%
Writing Intensive	818	50	7%	598	83	14%
FYS	141	5	4%	0	0	0%
Total	**2, 356**	**227**	**10%**	**914**	**204**	**22%**

Population/Sample Comparisons for Marshall's Learning Outcomes by Course Level

| Marshall Outcomes	Course Level = 100/200			Course Level = 300/400		
	Population	Sample	Percent	Population	Sample	Percent
Communication Fluency	627	38	6%	526	70	13%
Ethical and Civic Thinking	91	18	20%	141	90	64%
Intercultural Thinking	764	81	11%	174	27	16%
Total	**1,482**	**137**	**9%**	**841**	**187**	**22%**

Distribution of GEAR Artifacts among Marshall's Learning Outcomes

| Marshall Outcome	Outcome Traits (Primary)	# Uploaded Traits Tagged (Primary)	# Traits Assessed (Primary)	% of Total
Communication Fluency	Design/Organization	966	88	9%
	Diction	785	58	7%
	Communication Style	821	77	9%
	Communication Fluency Total	**2,572**	**223**	**9%**
Ethical and Civic Thinking	Ethical Self Awareness	157	87	55%
	Professional Rules and Standards of Conduct	109	59	54%
	Civic Well-Being	139	80	58%
	Complex Ethical issues	106	44	42%
	Ethical and Civic Thinking Total	**511**	**270**	**53%**
Intercultural Thinking	Own Culture	642	68	11%
	Other Cultures	715	81	11%
	Communication with Other Cultures	165	15	9%
	Global Awareness	381	58	15%
	Cultural Conflict	497	50	10%
	Global Contexts	246	34	14%
	Intercultural Thinking Total	**2,646**	**306**	**12%**

Sample Frequencies
Total # of artifacts assessed = 324

Course Level Frequencies
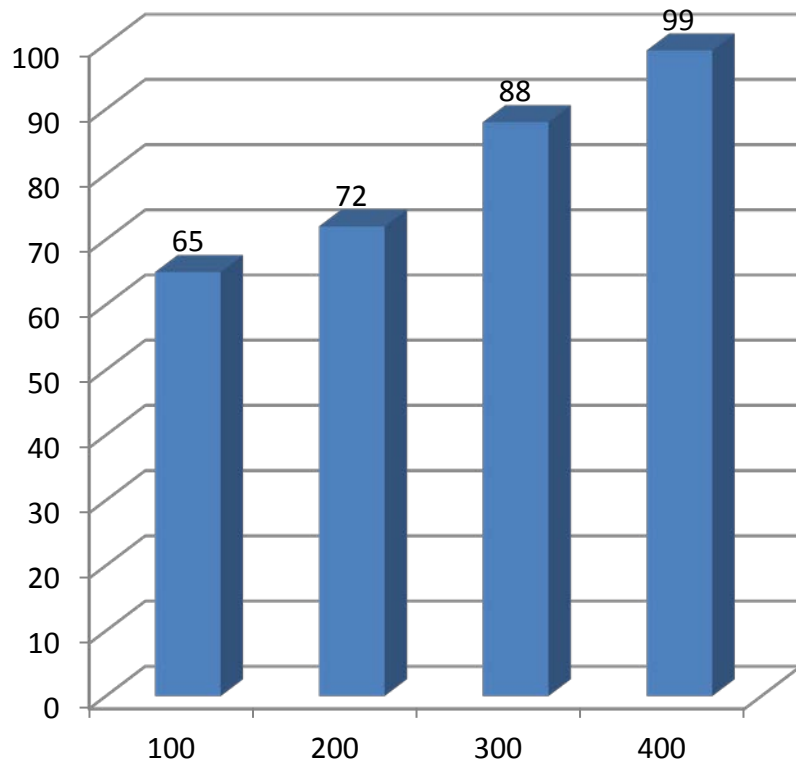
Class Rank Frequencies
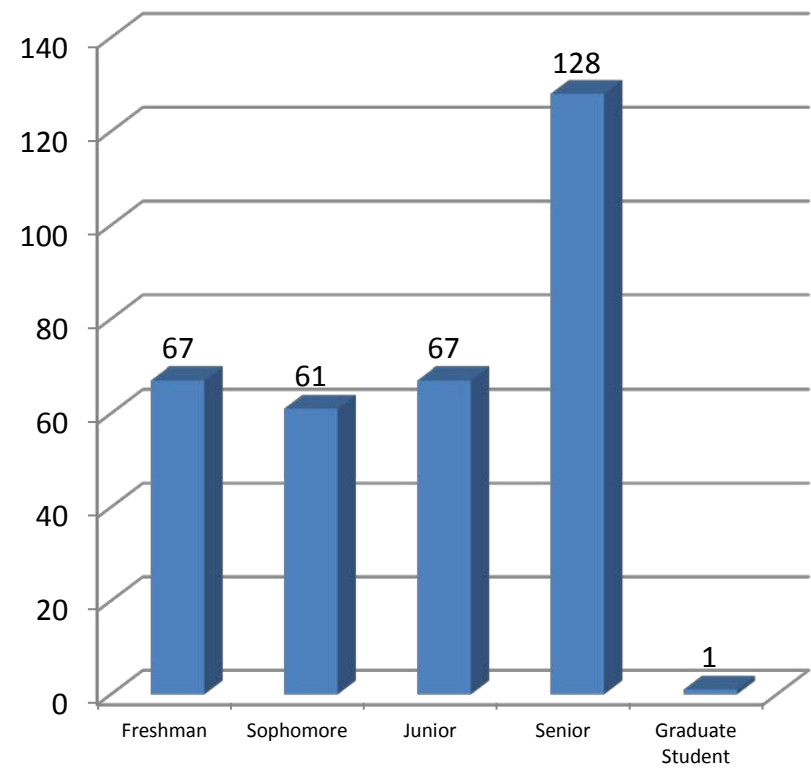
Review Procedures

- Please access muwww-new.marshall.edu/assessment/LearningOutcomes.aspx and click on the links for each Domain of Critical Thinking to access rubrics used for this assessment.

- Each artifact 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 artifact.

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.
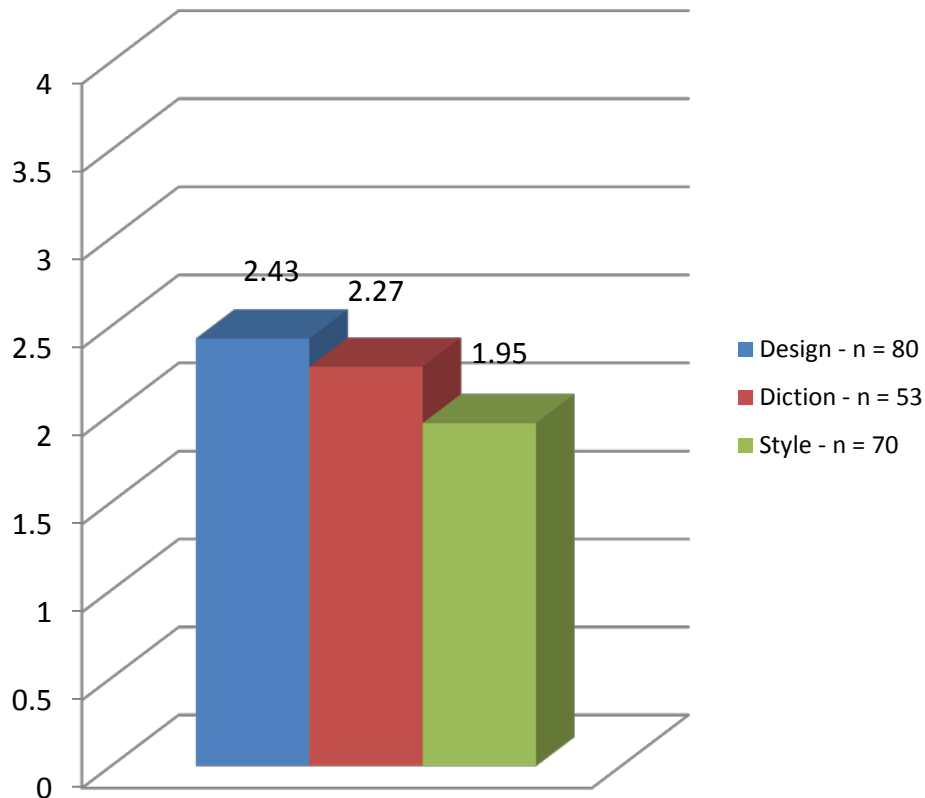
Artifacts Excluded Due to Inability to Assess or Misalignment with Tagged Outcomes/Traits

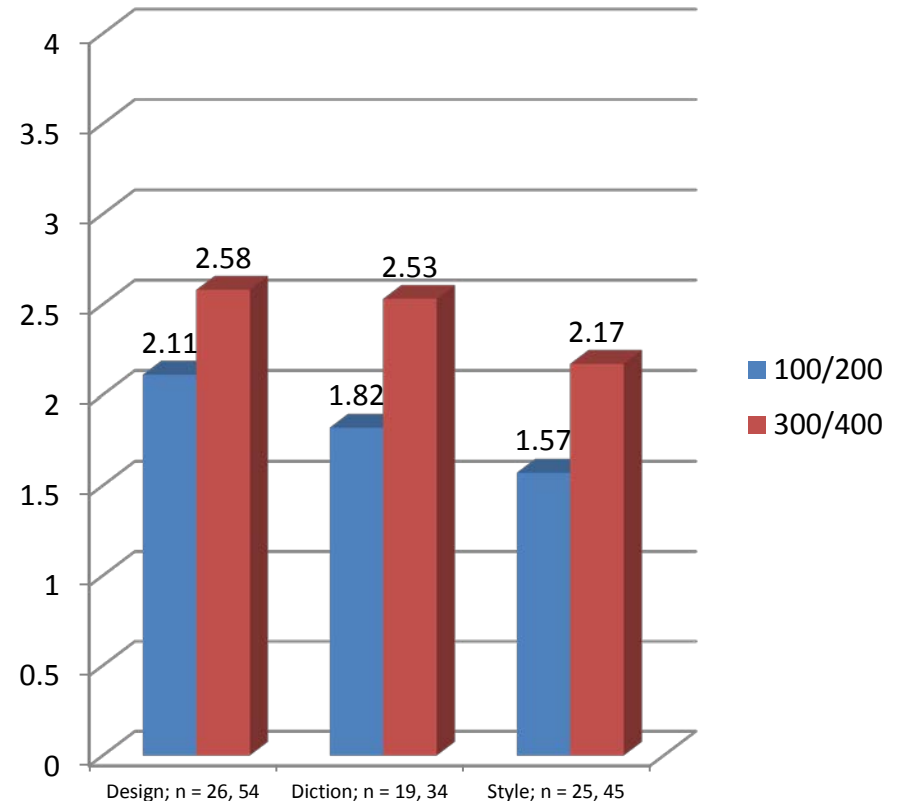| Outcome	Trait	Total Tags	# Not Able to be Assessed	# Misaligned
Communication Fluency	Design/Organization	88	8 (9%)	0
	Diction	58	5 (9%)	0
	Communication Style	77	7 (9%)	0
Ethical and Civic Thinking	Ethical Self-Awareness	87	1 (1%)	7 (8%)
	Professional Rules and Standards of Conduct	59	1 (2%)	9 (15%)
	Civic Well-Being	80	1 (1%)	2 (3%)
	Complex Ethical Issues	44	1 (2%)	8 (18%)
Intercultural Thinking	Own Culture	68	6 (9%)	6 (9%)
	Other Cultures	81	5 (6%)	5 (6%)
	Communication with Other Cultures	15	2 (13%)	4 (27%)
	Global Awareness	58	3 (5%)	21 (36%)
	Cultural Conflict	50	4 (8%)	7 (14%)
	Global Contexts	34	3 (9%)	22 (65%)

Communication Fluency

Mean Scores on a scale of 0 – 4, with 4 being the highest possible score
Mean differences were significant for course level for diction and communication style.

Overall Analysis

- Design - n = 80
- Diction - n = 53
- Style - n = 70

Bar values: 2.43, 2.27, 1.95

Analysis by Course Level

- 100/200
- 300/400

Category	100/200	300/400
Design; n = 26, 54	2.11	2.58
Diction; n = 19, 34	1.82	2.53
Style; n = 25, 45	1.57	2.17

Communication Fluency

Mean Scores on a scale of 0 – 4, with 4 being the highest possible score
Mean differences were significant for class rank for diction and communication style.
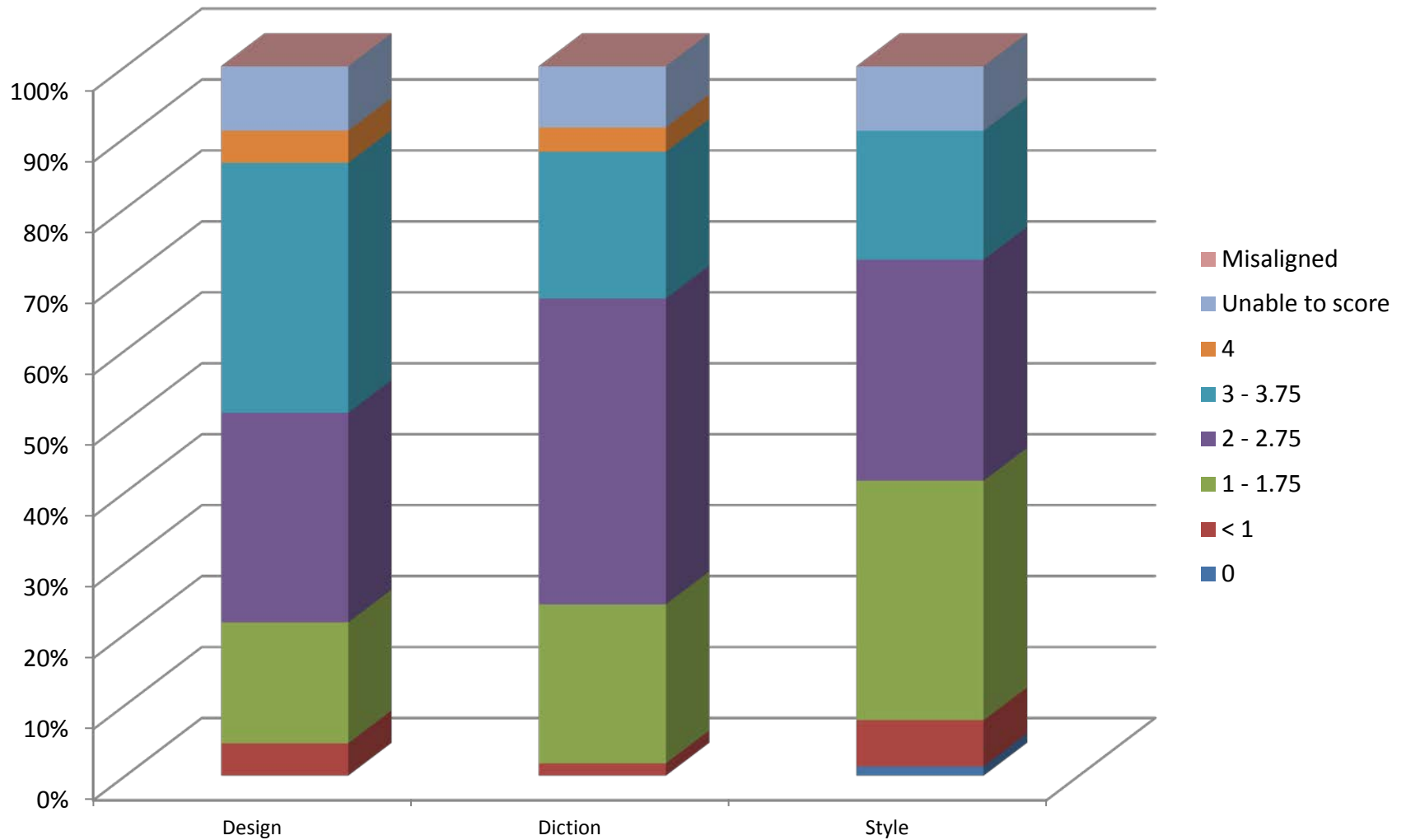
Analysis by Class Rank
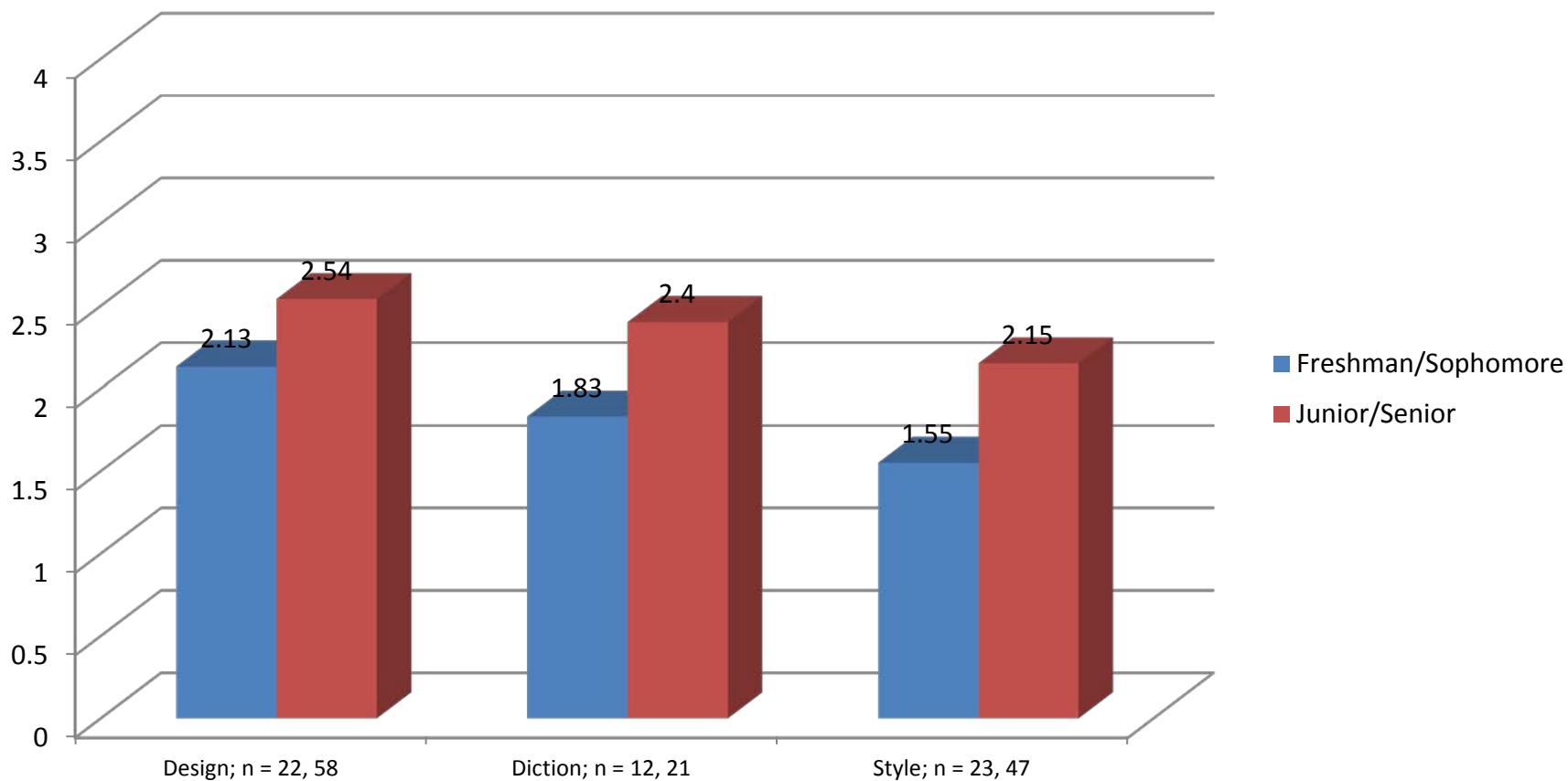
Communication Fluency
Number of artifacts scoring at each performance level

Trait/ Performance Level	Design	Diction	Style	Total
0	0	0	1 (1%)	1 (0%)
> 0, but < 1	4 (5%)	1 (2%)	5 (6%)	10 (4%)
1 – 1.75	15 (17%)	13 (22%)	26 (34%)	54 (24%)
2 – 2.75	26 (30%)	25 (43%)	24 (31%)	75 (34%)
3 – 3.75	31 (35%)	12 (21%)	14 (18%)	57 (26%)
4	4 (5%)	2 (3%)	0	6 (3%)
Unable to score	8 (9%)	5 (9%)	7 (9%)	20 (9%)
Misaligned	0	0	0	0
Totals	88 (100%)	58 (100%)	77 (100%)	223 (100%)

Communication Fluency

Legend:
- Misaligned
- Unable to score
- 4
- 3 - 3.75
- 2 - 2.75
- 1 - 1.75
- < 1
- 0

Categories: Design, Diction, Style
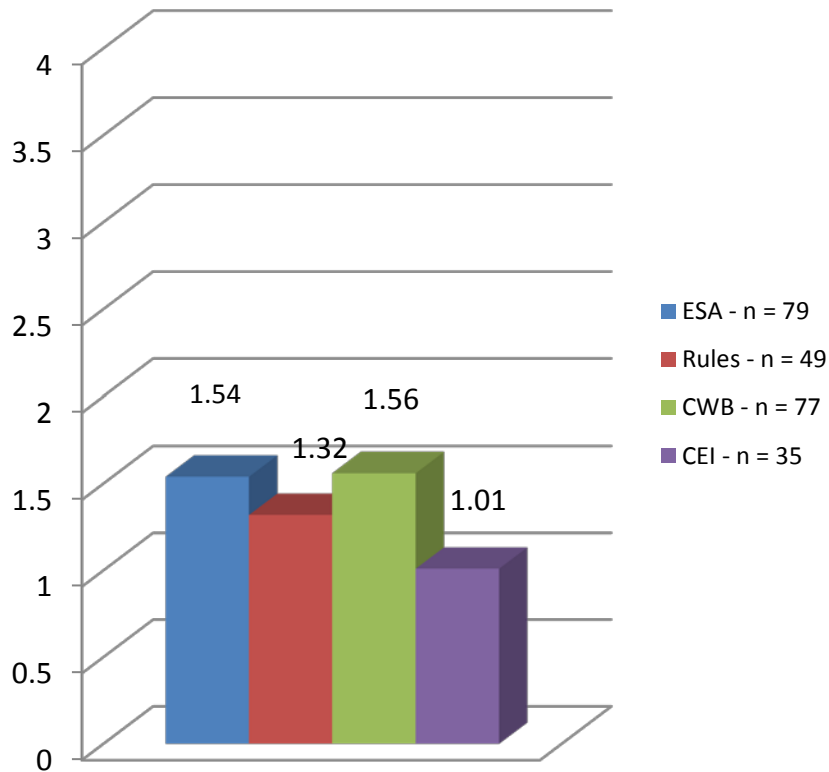
Communication Fluency

Inter-Rater Agreement Results

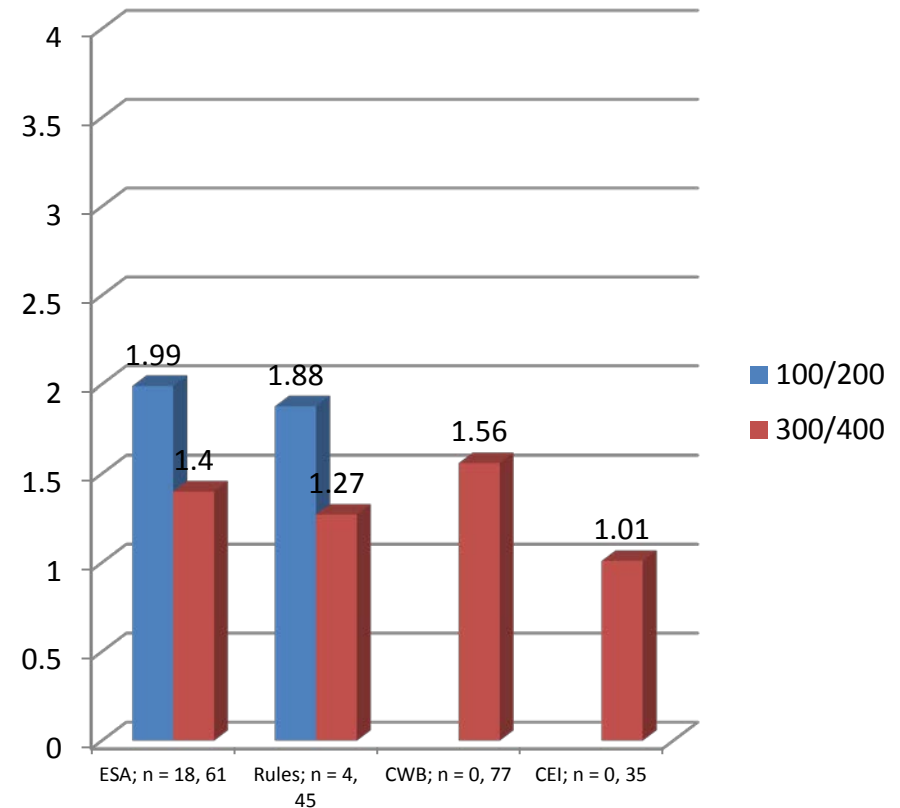| Trait/ Performance Level	Design; Kappa = .295 (All Scores); Kappa = .248 (Misaligned and Unable to Score Excluded)	Diction; Kappa = .285 (All Scores); Kappa = .224 (Misaligned and Unable to Score Excluded)	Style; Kappa = .137 (All Scores); Kappa = .082 (Misaligned and Unable to Score Excluded)
Agree on score	39 (44%)	25 (43%)	25 (32%)
Difference = 1 point or less	33 (38%)	27 (47%)	39 (51%)
Difference = 1.5 to 2 points	6 (7%)	2 (3%)	8 (10%)
Difference = 2.5 to 3 points	2 (2%)	0	0
Agree on Misaligned	0	0	0
Agree on Unable to Score	4 (5%)	3 (5%)	3 (4%)
Score + Misaligned	1 (1%)	0	0
Score + Unable to Score	3 (3%)	0	1 (1%)
Misaligned + Unable to Score	0	1 (2%)	1 (1%)
Total	88 (100%)	58 (100%)	77 (100%)

Ethical and Civic Thinking

Mean Scores on a scale of 0 – 4, with 4 being the highest possible score; For Ethical Self-Awareness, the mean for artifacts from 100/200 level courses was significantly higher than that for 300/400 level courses. All artifacts assessed for Civic Well Being and Complex Ethical Issues were from 300/400 level courses.

Overall Analysis

- ESA - n = 79
- Rules - n = 49
- CWB - n = 77
- CEI - n = 35

ESA: 1.54
Rules: 1.32
CWB: 1.56
CEI: 1.01

Analysis by Course Level

- 100/200
- 300/400

ESA; n = 18, 61: 1.99 (100/200), 1.4 (300/400)
Rules; n = 4, 45: 1.88 (100/200), 1.27 (300/400)
CWB; n = 0, 77: 1.56 (300/400)
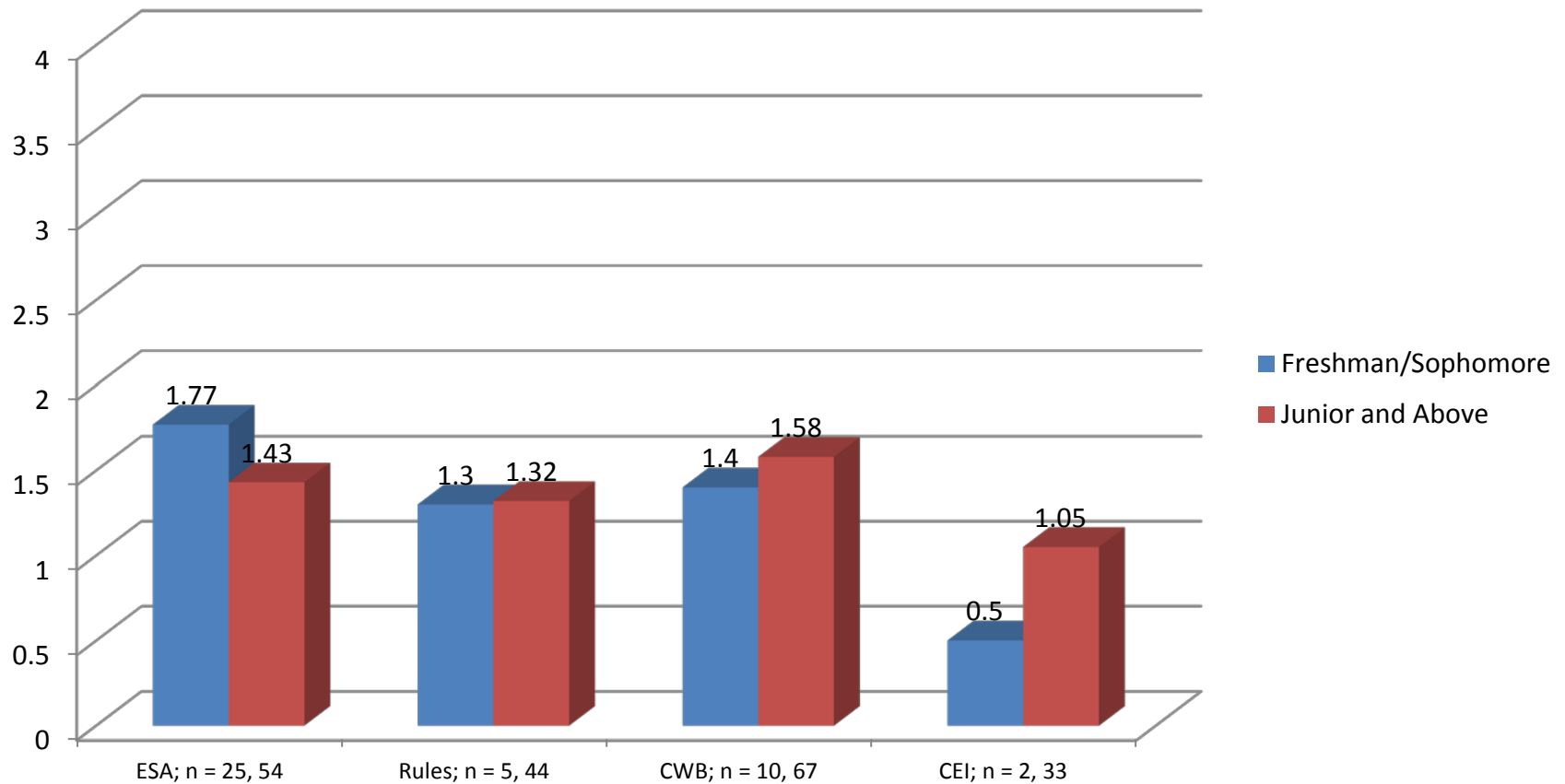CEI; n = 0, 35: 1.01 (300/400)

Ethical and Civic Thinking

Mean Scores on a scale of 0 – 4, with 4 being the highest possible score.
Mean differences were not significant for class rank.
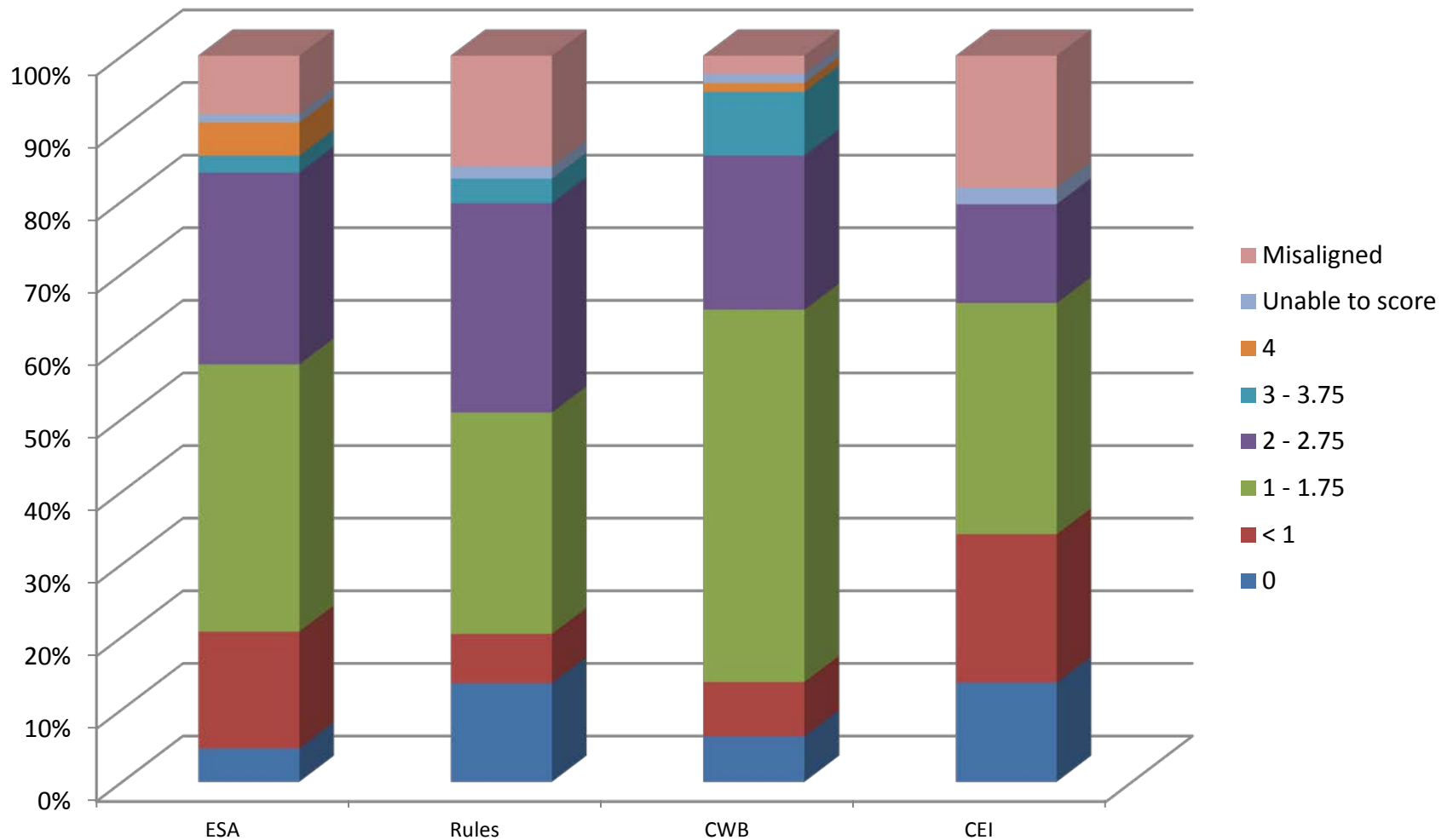
Analysis by Class Rank

Legend:
- ■ Freshman/Sophomore
- ■ Junior and Above

Category	Freshman/Sophomore	Junior and Above
ESA; n = 25, 54	1.77	1.43
Rules; n = 5, 44	1.3	1.32
CWB; n = 10, 67	1.4	1.58
CEI; n = 2, 33	0.5	1.05

Ethical and Civic Thinking
Number of artifacts scoring at each performance level

Trait/ Performance Level	ESA	Rules	CWB	CEI	Total
0	4 (5%)	8 (14%)	5 (6%)	6 (14%)	23 (9%)
> 0, but < 1	14 (16%)	4 (7%)	6 (8%)	9 (20%)	33 (12%)
1 – 1.75	32 (37%)	18 (31%)	41 (51%)	14 (32%)	106 (39%)
2 – 2.75	23 (26%)	17 (29%)	17 (21%)	6 (14%)	63 (23%)
3 – 3.75	2 (2%)	2 (3%)	7 (9%)	0	10 (4%)
4	4 (5%)	0	1 (1%)	0	5 (2%)
Unable to score	1 (1%)	1 (2%)	1 (1%)	1 (2%)	4 (1%)
Misaligned	7 (8%)	9 (15%)	2 (3%)	8 (18%)	26 (10%)
Totals	87 (100%)	59 (100%)	80 (100%)	44 (100%)	270 (100%)

Ethical and Civic Thinking

Legend:
- Misaligned
- Unable to score
- 4
- 3 - 3.75
- 2 - 2.75
- 1 - 1.75
- < 1
- 0

Categories: ESA, Rules, CWB, CEI

Ethical and Civic Thinking

Inter-Rater Agreement Results

Trait/ Performance Level	ESA; Kappa = .070 (All Scores); Kappa = -.036 (Misaligned and Unable to Score Excluded)	Rules; Kappa = .185; Kappa = .267 (Misaligned and Unable to Score Excluded)	CWB; Kappa = .167; Kappa = .157 (Misaligned and Unable to Score Excluded)	CEI; Kappa = .188; Kappa = .057 (Misaligned and Unable to Score Excluded)
Agree on score	21 (24%)	17 (29%)	28 (35%)	11 (25%)
Difference = 1 point or less	33 (38%)	9 (15%)	34 (43%)	17 (39%)
Difference = 1.5 to 2 points	19 (22%)	10 (17%)	7 (9%)	5 (11%)
Difference = 2.5 to 4 points	3 (3%)	0	3 (4%)	2 (2%)
Agree on Misaligned	4 (5%)	2 (3%)	1 (1%)	4 (9%)
Agree on Unable to Score	1 (1%)	1 (2%)	1 (1%)	1 (2%)
Score + Misaligned	5 (6%)	19 (32%)	6 (8%)	5 (11%)
Score + Unable to Score	1 (1%)	1 (2%)	0	0
Misaligned + Unable to Score	0	0	0	0
Total	87 (100%)	59 (100%)	80 (100%)	44 (100%)

Intercultural Thinking

Mean Scores on a scale of 0 – 4, with 4 being the highest possible score; mean differences were not significant based on course level

Overall Analysis
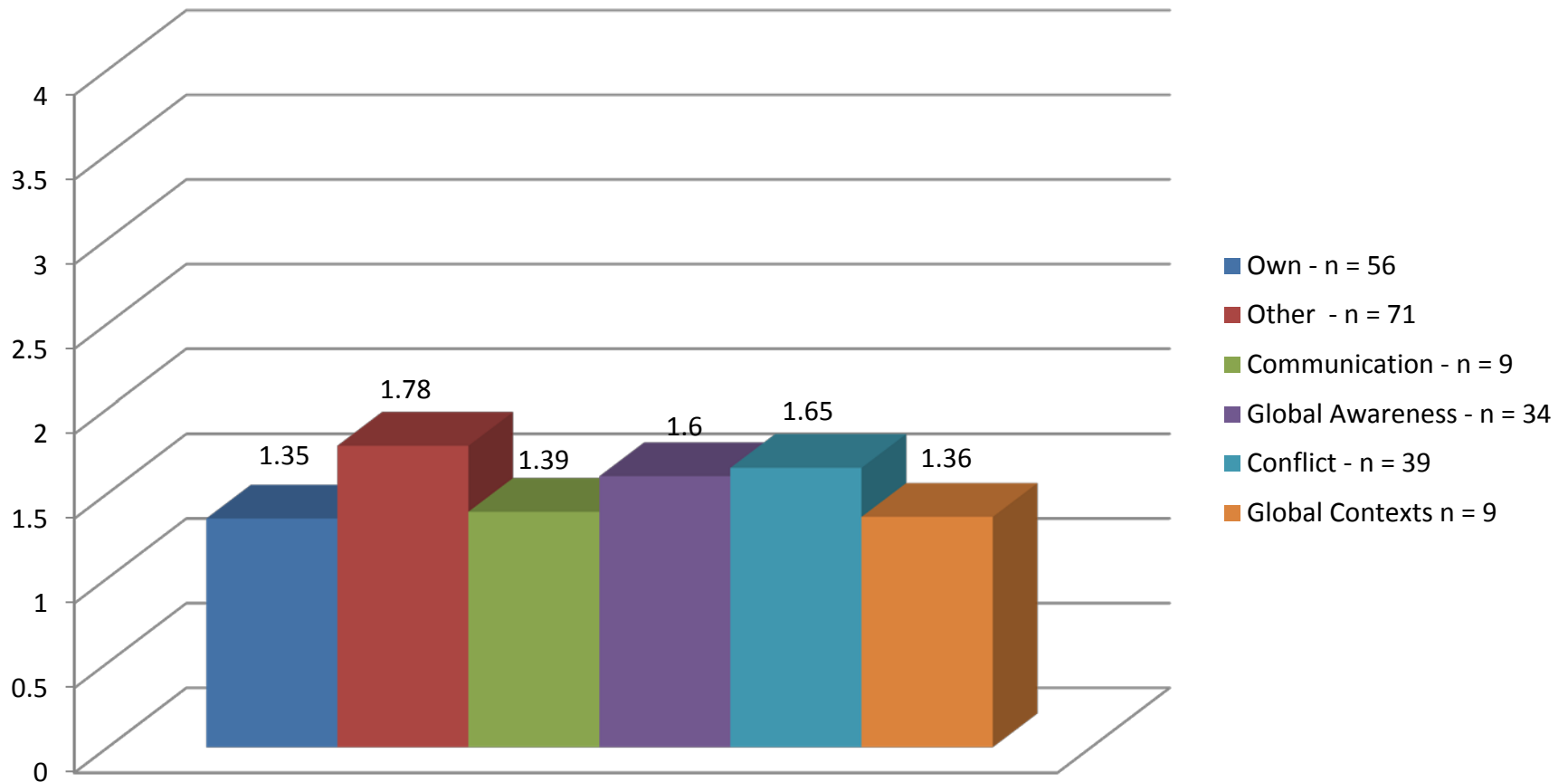
Intercultural Thinking

Mean Scores on a scale of 0 – 4, with 4 being the highest possible score; mean differences were not significant based on course level
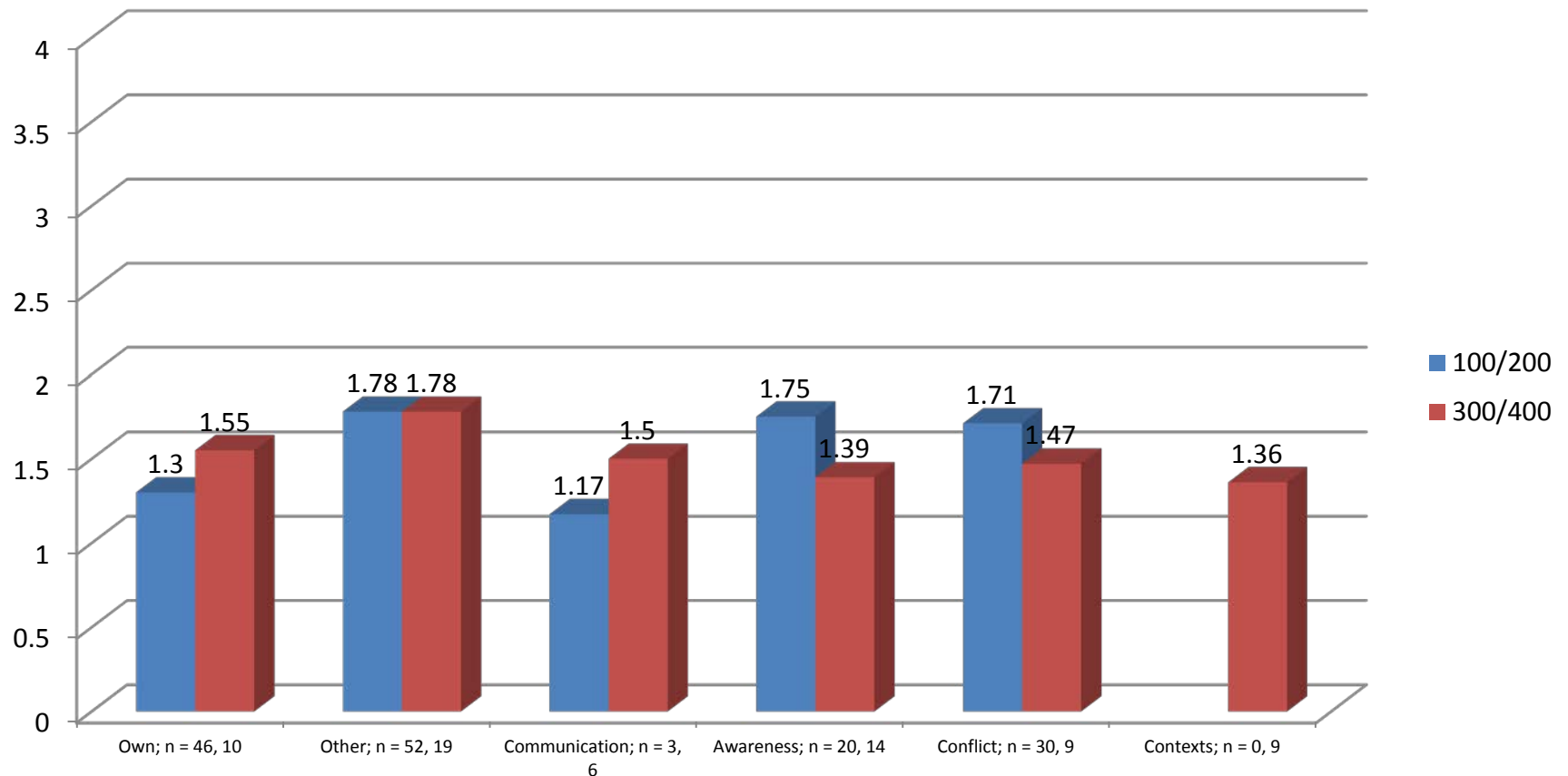
Analysis by Course Level

Legend:
- 100/200 (blue)
- 300/400 (red)

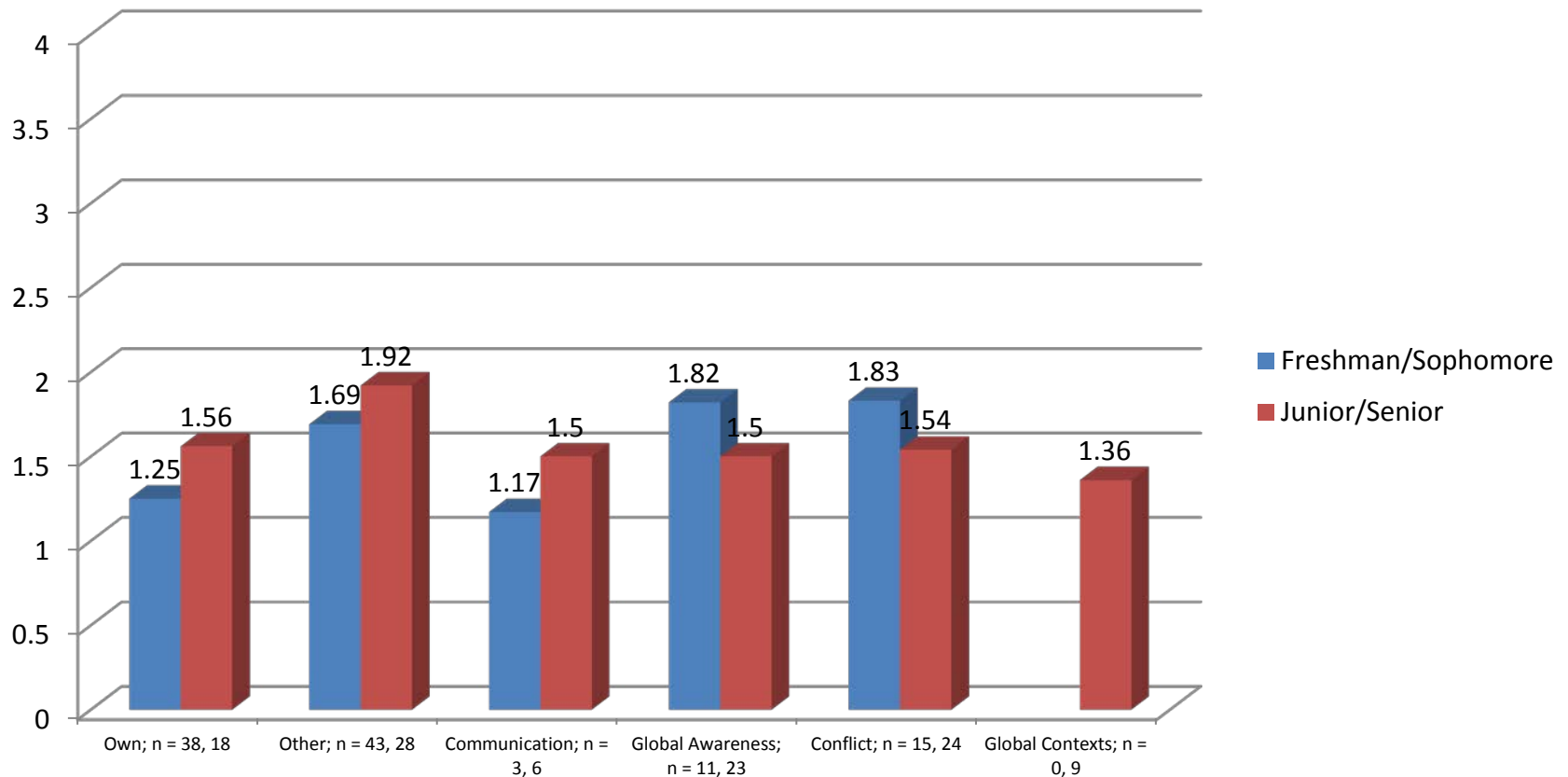| Category	100/200	300/400
Own; n = 46, 10	1.3	1.55
Other; n = 52, 19	1.78	1.78
Communication; n = 3, 6	1.17	1.5
Awareness; n = 20, 14	1.75	1.39
Conflict; n = 30, 9	1.71	1.47
Contexts; n = 0, 9		1.36

Intercultural Thinking

Mean Scores on a scale of 0 – 4, with 4 being the highest possible score.
Mean differences were not significant for class rank .
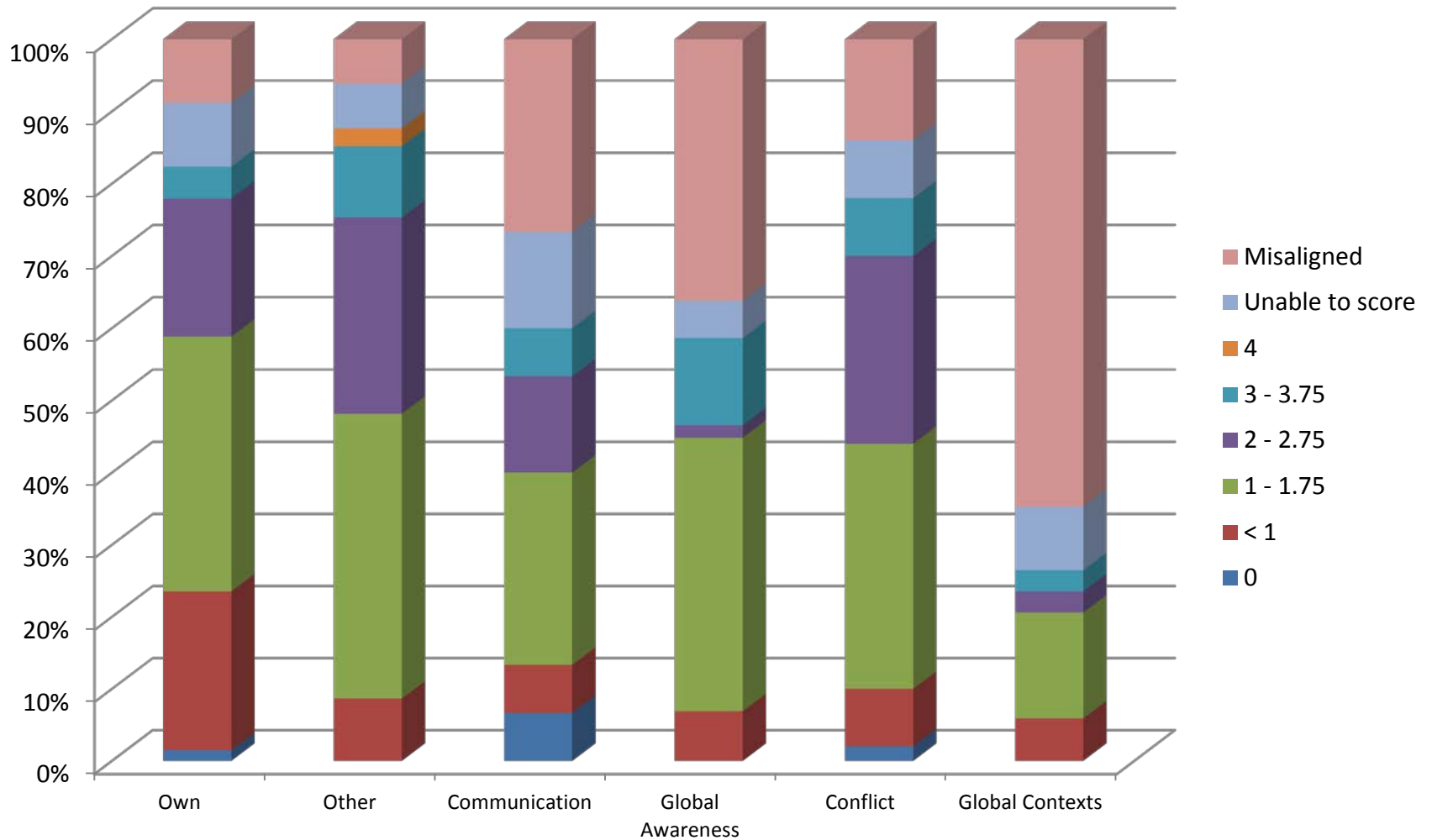

Analysis by Class Rank

Intercultural Thinking

Number of artifacts scoring at each performance level

Trait/ Performance Level	Own	Other	Communication	Global Awareness	Conflict	Global Contexts	Total
0	1 (1%)	0	1 (7%)	0	1 (2%)	0	3 (1%)
> 0, but < 1	15 (22%)	7 (9%)	1 (7%)	4 (7%)	4 (8%)	2 (6%)	33 (11%)
1 – 1.75	24 (35%)	32 (40%)	4 (27%)	22 (38%)	17 (34%)	5 (15%)	104 (34%)
2 – 2.75	13 (19%)	22 (27%)	2 (13%)	1 (2%)	13 (26%)	1 (3%)	52 (17%)
3 – 3.75	3 (4%)	8 (10%)	1 (7%)	7 (12%)	4 (8%)	1 (3%)	24 (8%)
4	0	2 (2%)	0	0	0	0	2 (1%)
Unable to score	6 (9%)	5 (6%)	2 (13%)	3 (5%)	4 (8%)	3 (9%)	23 (8%)
Misaligned	6 (9%)	5 (6%)	4 (27%)	21 (36%)	7 (14%)	22 (65%)	65 (21%)
Totals	68 (100%)	81 (100%)	15 (100%)	58 (100%)	50 (100%)	34 (100%)	306 (100%)

Intercultural Thinking

Legend:
- Misaligned
- Unable to score
- 4
- 3 - 3.75
- 2 - 2.75
- 1 - 1.75
- < 1
- 0

Categories: Own, Other, Communication, Global Awareness, Conflict, Global Contexts

Intercultural Thinking
Inter-Rater Agreement Results

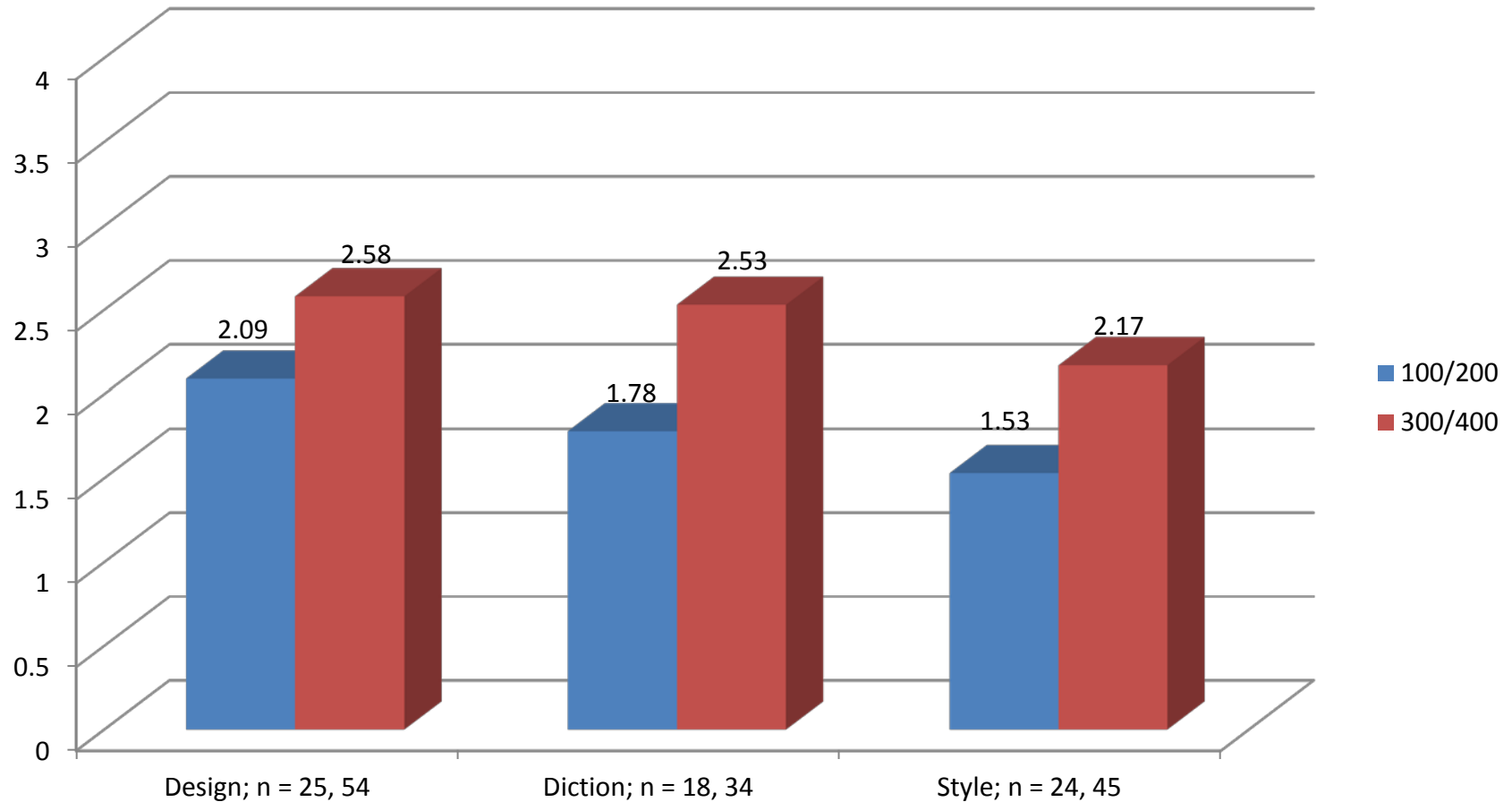| Trait/ Performance Level	Own; Kappa = .145 (All Scores); Kappa = -.023 (Misaligned and Unable to Score Excluded)	Other ; Kappa = .189 (All Scores); Kappa = .106 (Misaligned and Unable to Score Excluded)	Communication; Kappa = .286 (All Scores); Kappa = .360 (Misaligned and Unable to Score Excluded)	Global Awareness Kappa = .170 (All Scores); Kappa = -.098 (Misaligned and Unable to Score Excluded)	Conflict; Kappa = -.267 (All Scores); Kappa = .168 (Misaligned and Unable to Score Excluded)	Global Contexts ; Kappa = -.349 (All Scores); Kappa = .172 (Misaligned and Unable to Score Excluded)
Agree on score	17 (25%)	25 (31%)	4 (27%)	7 (12%)	13 (26%)	2 (6%)
Difference = 1 point or less	28 (41%)	31 (38%)	1 (7%)	11 (19%)	9 (18%)	3 (9%)
Difference = 1.5 to 2 points	8 (12%)	13 (16%)	2 (13%)	3 (5%)	9 (18%)	1 (3%)
Difference = 2.5 to 3 points	2 (3%)	0	1 (7%)	5 (9%)	1 (2%)	0
Agree on Misaligned	1 (1%)	1 (1%)	0	12 (21%)	4 (8%)	15 (44%)
Agree on Unable to Score	6 (9%)	4 (5%)	2 (13%)	3 (5%)	3 (6%)	3 (9%)
Score + Misaligned	6 (9%)	6 (7%)	5 (33%)	17 (29%)	10 (20%)	10 (29%)
Score + Unable to Score	0	1 (1%)	0	0	1 (2%)	0
Misaligned + Unable to Score	0	0	0	0	0	0
Total	68 (100%)	81 (100%)	15 (100%)	58 (100%)	50 (100%)	34 (100%)

Course Type Analysis

Writing Intensive Courses
Mean comparison by course level for Communication Fluency
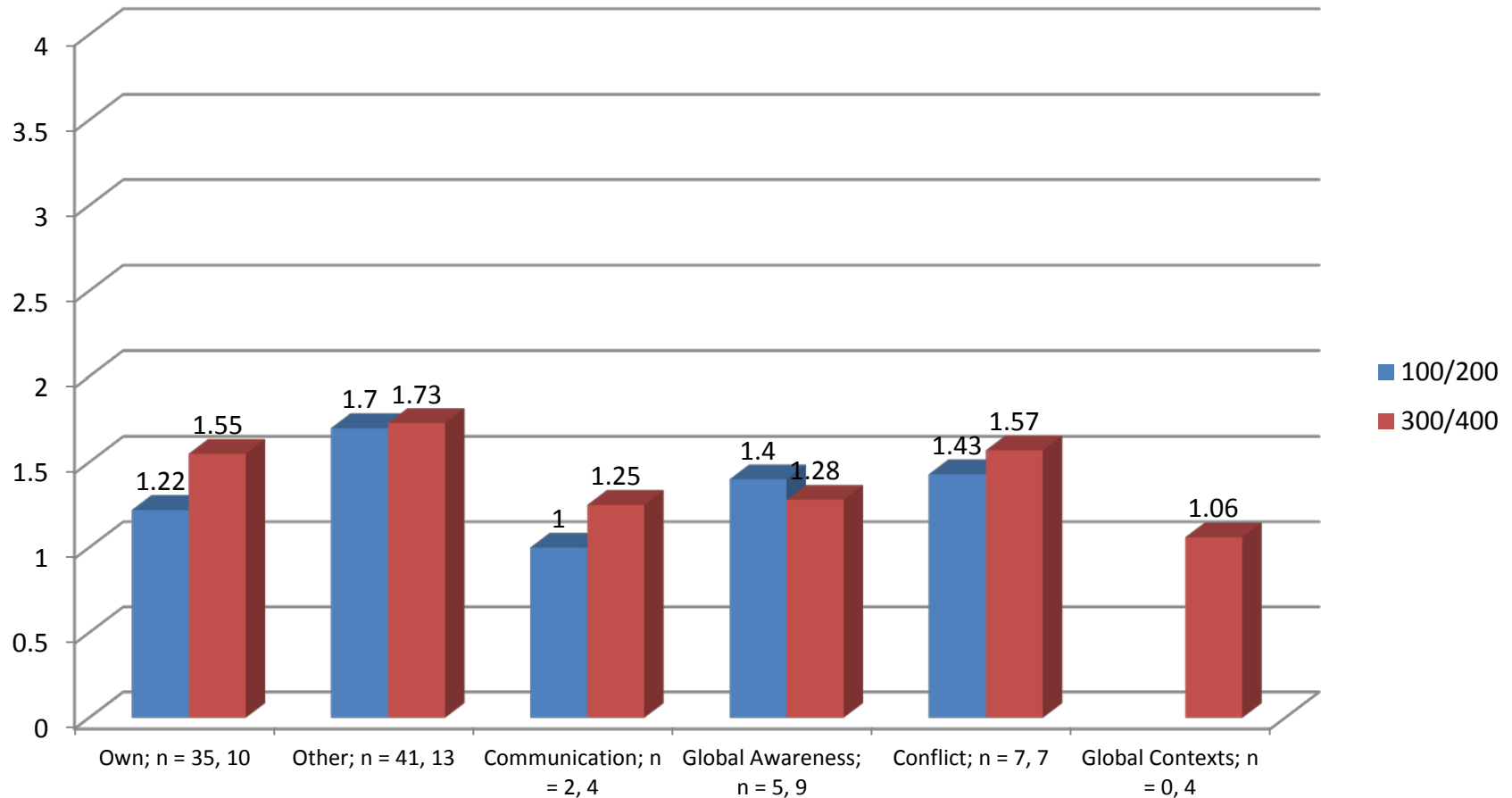### (100/200 compared to 300/400)
Means for Diction and Communication Style were significantly different based on course level.

Legend:
- 100/200 (blue)
- 300/400 (red)

Design; n = 25, 54: 2.09, 2.58
Diction; n = 18, 34: 1.78, 2.53
Style; n = 24, 45: 1.53, 2.17

Multicultural Courses

Mean comparison by course level; (100/200 compared to 300/400)
Mean differences were not significant based on course level.

Legend: ■ 100/200 ■ 300/400

Category	100/200	300/400
Own; n = 35, 10	1.22	1.55
Other; n = 41, 13	1.7	1.73
Communication; n = 2, 4	1	1.25
Global Awareness; n = 5, 9	1.4	1.28
Conflict; n = 7, 7	1.43	1.57
Global Contexts; n = 0, 4		1.06

International Courses

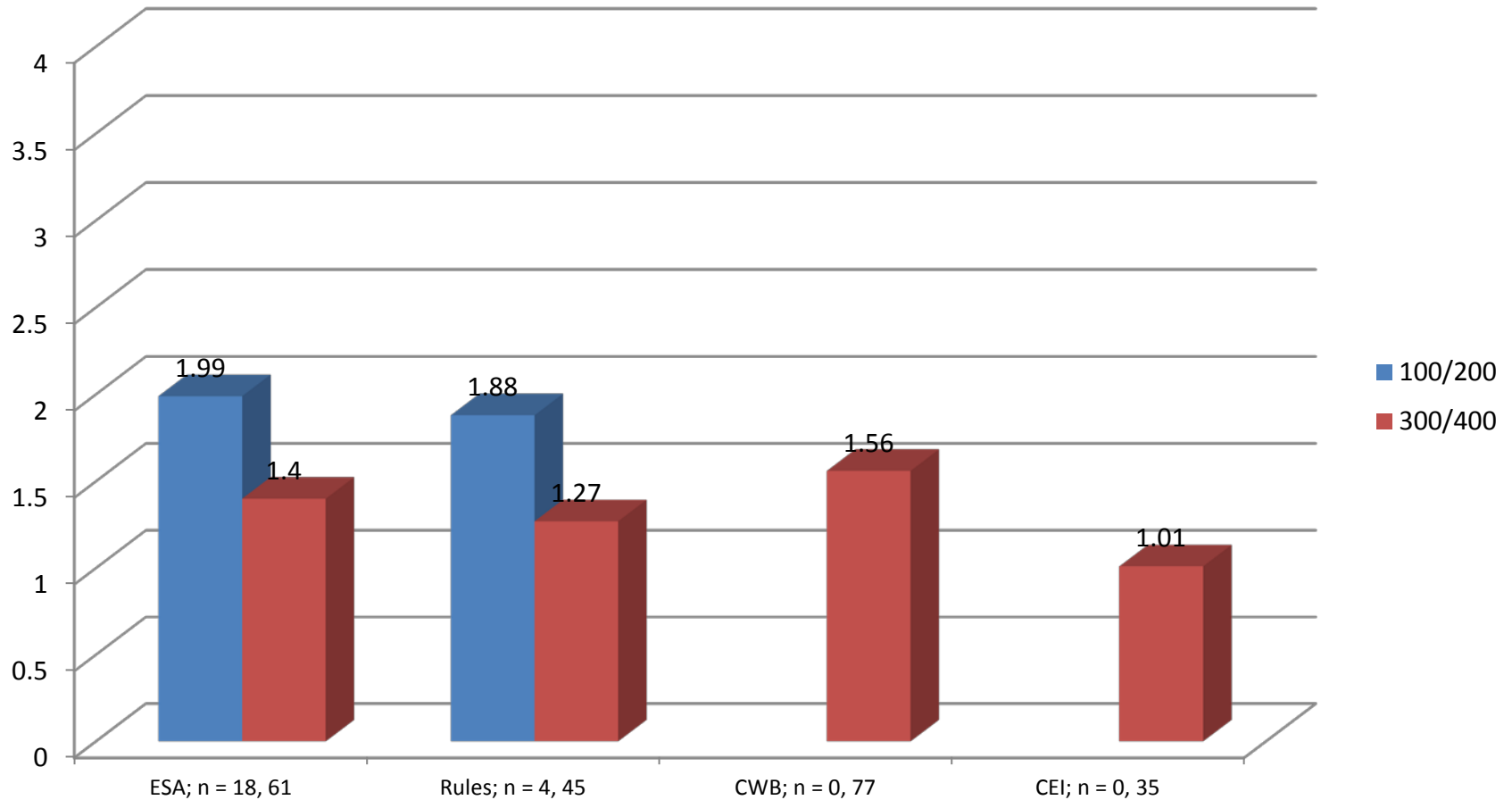Mean comparison by course level; (100/200 compared to 300/400)
Mean differences were not significant based on course level. However, *n/s* were small, with "0" for some traits.
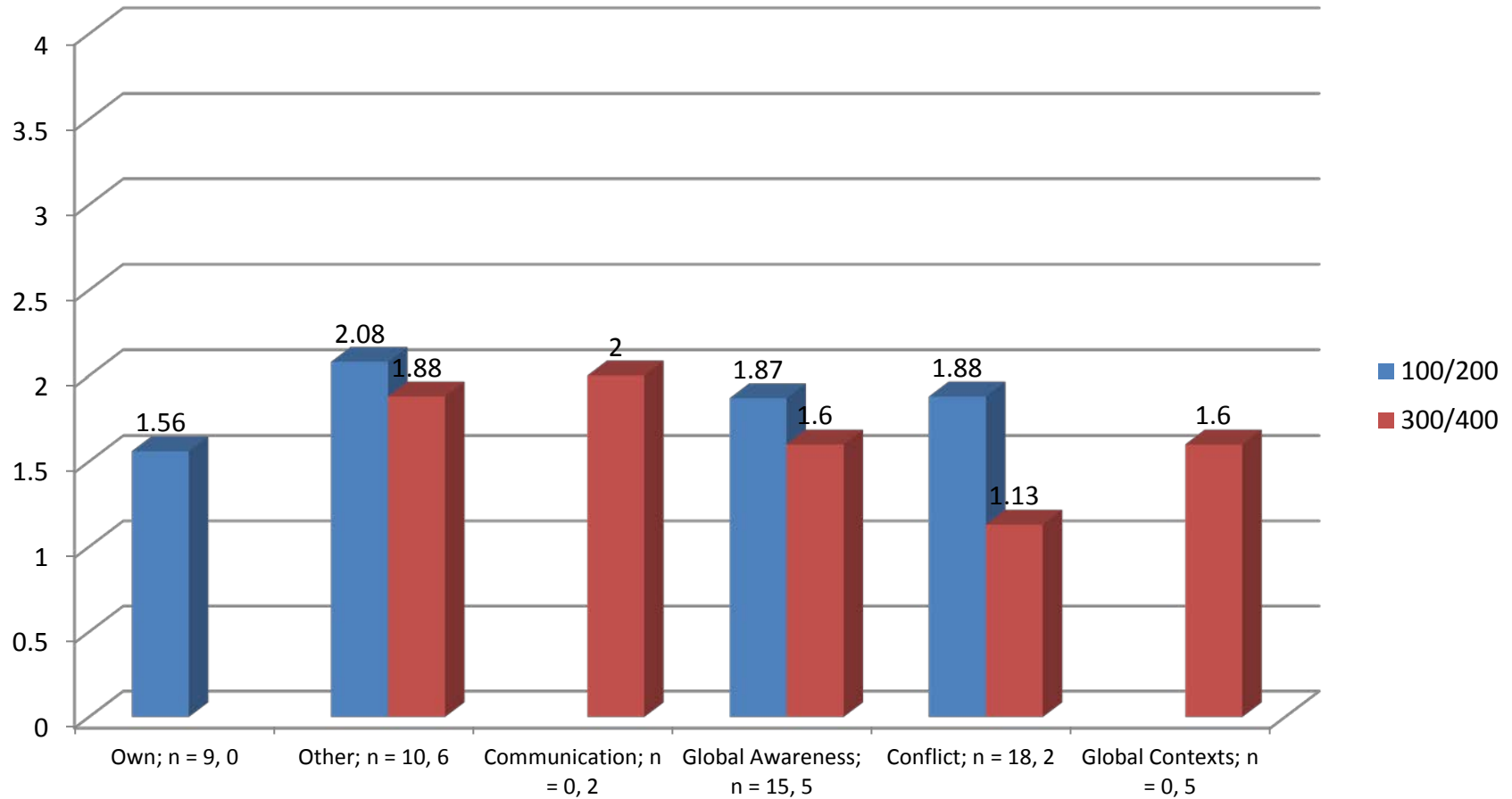
Service Learning Courses

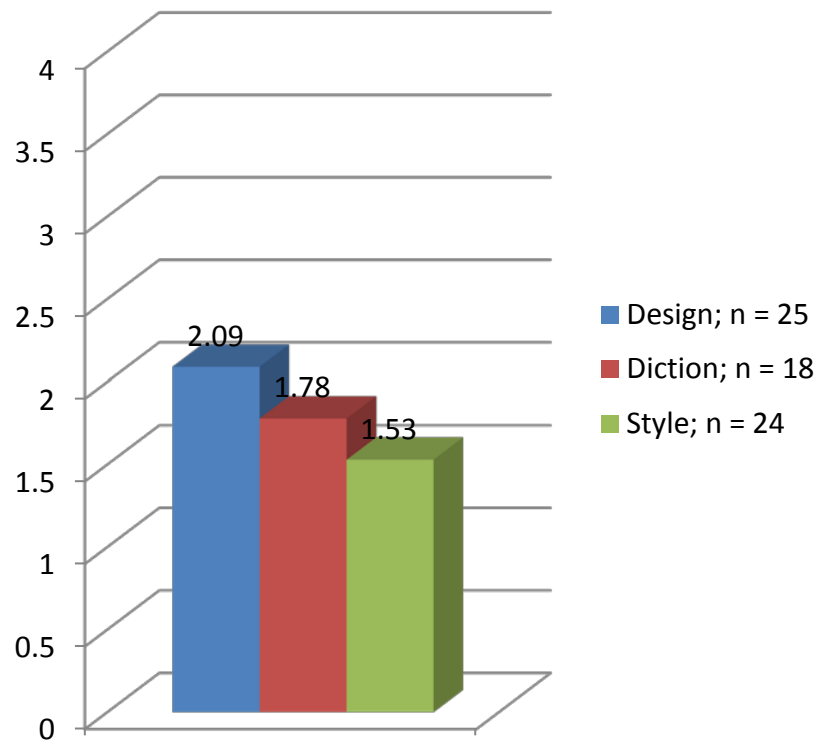Mean comparison by course level; (100/200 compared to 300/400)
Students in 100/200 level courses scored significantly higher than students in 300/400 level courses on Ethical Self-Awareness. There were no artifacts from 100/200 level courses tagged to Civic Well-Being or Complex Ethical Issues.

Legend:
- 100/200
- 300/400

Category	100/200	300/400
ESA; n = 18, 61	1.99	1.4
Rules; n = 4, 45	1.88	1.27
CWB; n = 0, 77		1.56
CEI; n = 0, 35		1.01

CT Courses

Mean Scores on a scale of 0 – 4, with 4 being the highest possible score. All CT courses are 100/200 Level. All uploading to Intercultural Thinking were 200-level courses.   For Communication Fluency, the only trait tagged to 100-level courses was "design (n = 7)." Means did not differ significantly for this trait based on course level.

Communication Fluency

- Design; n = 25
- Diction; n = 18
- Style; n = 24

2.09 1.78 1.53

Intercultural Thinking

- Own; n = 19
- Other; n = 20
- Communication; n = 2
- Global Awareness; n = 18
- Conflict; n = 26

1.71 1.88 1 1.83 1.78