Analysis of Artifacts from Marshall University's General Education Assessment Repository (GEAR)

Spring Semester 2014

Summer Assessment Workgroup Members: Harold Blanco, Robert Ellison, Susan Thomas Frank, Marty Laubach, Michael Newsome, Joan St. Germain, Gregg Twietmeyer, Anita Walz, Mary Welch, and Mary Beth Reynolds (Office of Assessment)

Executive Summary

Background

Recommendations from 2013 Assessment

The Marshall University Degree Profile, consisting of nine outcomes, each with between two and five traits (for a total of 34), was approved by Marshall's Faculty Senate in January 2013. Following this approval, the University General Education Assessment Repository (GEAR) was redesigned to allow artifacts to be uploaded and aligned (tagged) to the newly approved outcomes/traits. For pilot testing, students in First-Year Seminar (FYS) uploaded artifacts during the spring semester of 2013. Following the Assessment Workgroup's evaluation of a sample if these artifacts, it made recommendations for changes to GEAR. In the points below, we outline the 2013 Workgroup's recommended changes and explain how these were implemented for assessment in 2014.

1. Encourage instructors to specify the outcome(s)/trait(s) to which the assignment their students complete and upload into GEAR aligned. The Assessment Workgroup felt that instructor alignment of assignments with outcome(s)/trait(s) should reduce the number of student artifacts that appeared to be misaligned with the outcome(s)/trait(s) to which they are tagged. However, some FYS instructors indicated that they gave assignments that allowed students the flexibility of emphasizing a variety of different outcomes; these instructors asked that their students continue to be responsible for selecting the outcome(s)/trait(s) for these artifacts. Therefore, during academic year 2013 – 2014 GEAR was redesigned so that either the instructor or the student (but not both) could align an artifact with outcome(s)/trait(s). If the instructor chose the outcome(s)/trait(s) for the assignment, these outcome(s)/trait(s) applied to all student artifacts uploaded for that assignment. If the instructor anticipated that alignment of artifacts with specific outcome(s)/trait(s) would be unique to each student upload, they asked students to make the alignment.
2. Allow students to upload more than one file for a given assignment. For example, an assignment might ask students to write a paper and give a presentation using PowerPoint or Prezi. In this case, the student could upload both the paper (in Word or PDF) and the PowerPoint or Prezi as a separate file. GEAR was redesigned to allow this possibility.

3. Allow instructors to create more than one assignment to be uploaded to GEAR for the same class. This became especially important for academic year 2013 – 2014, as we added additional course types to GEAR uploads. Specifically, in addition to FYS, courses with service learning, multicultural, international, and writing intensive designations uploaded artifacts to GEAR in the spring of 2014. GEAR was redesigned to make uploading multiple assignments possible.
4. Include the instructor's assignment instructions in the upload. The 2013 Assessment Workgroup noted that it was often difficult for them to evaluate student artifacts because the instructor's instructions to students were not included. GEAR was redesigned to allow instructors to upload a file with their complete assignment instructions, type assignment instructions into a free-field box on the GEAR site, or do both.
5. Provide an instructor-friendly view of all students in the class who had uploaded an artifact. Although information about artifacts uploaded had appeared before, it was cumulative information and not tied to a specific course. GEAR was redesigned so that each instructor could easily see whether or not his/her students had uploaded the assigned artifact for his/her class.
6. Add a free-text box in GEAR that requires students to explain why their uploaded artifact addresses the outcome(s)/trait(s) to which it was aligned. This box was added as a mandatory field in GEAR.

General Procedures for 2014 Assessment

Recommended changes were made to GEAR by the spring semester of 2014. During that semester students enrolled in FYS as well as in courses carrying multicultural, international, writing intensive, and service learning designations uploaded artifacts to GEAR. Instructors were asked to create assignments aligned to *Communication Fluency* (writing intensive courses), *Ethical and Civic Thinking* (service learning courses), *Intercultural Thinking* (multicultural and international courses). Instructors were told that it was not necessary to align the assignments to all traits for the specified learning outcome; that they should align them only to those traits the assignment specifically addressed. Since FYS addresses five of the University's outcomes (*Information Literacy* and *Inquiry-Based*, *Integrative*, *Intercultural*, and *Metacognitive Thinking*), it was left to instructors and/or students to decide to which of these outcome(s) their assignments aligned. It was possible for a single assignment to map to any number of outcomes and traits. The total number of unique artifacts uploaded into GEAR in spring 2014 was 3,399.

In May 2014 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 Degree Profile, we expect students to perform at Level 3 (capstone) by the time of graduation. Although we wanted to assess at least 500 artifacts (15% of the sample), time constraints necessitated reducing this number to 270 (8% of the sample). Each artifact was read by two independent reviewers. We note that artifacts were assigned by outcome tags. Eight artifacts in the sample were aligned to more than one outcome, causing them to appear more than once in the review sample. The number of unique artifacts assessed was 262, still roughly 8% of the sample. This project was coordinated by the Office of Assessment.

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/traits to which the instructor or student had tagged it.
99	In the opinion of the evaluator, the artifact did not include enough information to allow assessment or there was some technical error within GEAR, e.g. in one case the intended artifact was overwritten by another subsequently uploaded. This problem has been corrected.
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

Our sample of 262 unique artifacts consisted of approximately twice as many from 100/200 as from 300/400 level courses. There were approximately twice as many freshman uploads in the sample when comparing them individually at the sophomore, junior, or senior level. There were more artifacts in the sample from writing intensive than from any other type of course. Most of the artifacts in our sample were aligned to only one outcome, but three were aligned to all nine outcomes. Only 21 of the 262 artifacts in our sample did not include an assignment description of any kind. Students determined the outcome(s)/trait(s) alignments for only 16 of the 262 artifacts in the sample; for the remaining artifacts, alignments were determined by course instructors.

Results and Analysis

One challenge in reporting results of GEAR assessment is that, although we assessed 262 unique 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 most artifacts aligned to 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 to the *Intercultural Thinking* outcome, bringing the total number of traits across the nine outcomes to 35. A perusal of our supporting documentation shows that the artifacts evaluated by the Assessment Workgroup tagged to a total of 673 traits. However, scores for only 507 (75%) of those traits were usable for purposes of calculating means. One hundred sixty-six were discarded either because they were judged not to align with the traits (136; 20%) or were not able to be assessed for reasons noted earlier (30; 5%). The chart below shows the number of artifacts aligned to each trait, the number excluded from the analysis due to receiving scores of 99 or 100, and the resulting number of scores able to be used for the analysis of means. Of the nine outcomes, only two (*Communication Fluency* and *Inquiry-Based Thinking*) had *n*s of at least 10 for each trait. *Ns* for *Communication Fluency* ranged from 36 to 56, while those for *Inquiry-Based Thinking* ranged from 12 to 16. Twelve additional traits had *n*s of at least 10. That left 15 traits with fewer than 10 usable scores. These numbers were reduced further when we attempted to compare means based on course level (100/200 compared to 300/400 level courses). Due to these low numbers, results of our analyses must be interpreted with caution. * = $n \geq 10$.

Outcome	**Trait**	**Total Artifacts Aligned**	**# Excluded from Analysis of Means**	**Total Usable Artifacts**
Communication Fluency	Context/Audience	43	7	36 *
	Design/Organization	59	3	56 *
	Diction	53	2	51 *
	Communication Style	51	2	49 *
Creative Thinking	Ambiguities & Possibilities	9	1	8
	Risk Taking	2	1	1
	Innovations	19	3	16 *
Ethical and Civic Thinking	Ethical Self-Awareness	19	5	14 *
	Professional Rules and Standards of Conduct	12	4	8
	Civic Well-Being	11	5	6
	Complex Ethical Issues	10	5	5
Information Literacy	Tool Use	24	12	12 *

	Relevance of Information	28	3	25 *
	Complex Information Environment	22	19	3
	Legal/Ethical Issues	12	2	10 *
Inquiry-Based Thinking	Problem/Question	18	6	12 *
	Research of Existing Knowledge	19	3	16 *
	Method of Inquiry	21	9	12 *
	Data Analysis and Conclusions	18	6	12 *
Integrative Thinking	Connections among Disciplines	6	1	5
	Relation among Domains of Thinking	3	3	0
	Transfer	4	0	4
	Connections to Experience	12	3	9
Intercultural Thinking	Own Culture	18	10	8
	Other Cultures	34	6	28 *
	Communication with Others from Different Cultures	11	6	5
	Global Awareness	25	11	14 *
	Cultural Conflict	20	4	16 *
	Global Contexts	8	3	5
Metacognitive Thinking	Project Management	12	6	6
	Self-Evaluation	16	5	11 *
Quantitative Thinking	Context	19	1	18 *
	Estimation	5	5	0
	Visual Representations	15	2	13 *
	Statistics	15	2	13 *

In general, those outcome traits with *ns* of 10 or more split between 100/200 and 300/400 level courses showed a trend toward higher means for 300/400 level courses. This was true for *Communication Fluency* (all traits), *Ethical and Civic Thinking* (ethical self-awareness), *Information Literacy* (tool use), and *Intercultural Thinking* (other cultures, global awareness, and cultural conflict). These differences reached statistical significance for *Communication Fluency* (design/organization, diction, and communication style) and for *Intercultural Thinking* (other cultures). Our sample included no artifacts at the 300/400 level for three outcomes (*Integrative, Metacognitive,* and *Quantitative Thinking*) and *Inquiry-Based Thinking* had very few tags for upper level courses.

Overall results showed mean performance for traits with 10 or more tags to range from 1.1 (*Ethical and Civic Thinking*: ethical self-awareness and *Quantitative Thinking*: statistics) to 2.2 (*Communication Fluency:* context/audience and *Information Literacy*: relevance of information). Mean performance for artifacts uploaded from 300/400 level courses ranged from 1.3 (*Ethical and Civic Thinking*: ethical self-awareness) to 2.4 (*Communication Fluency*: diction).

Based on these results, *Communication Fluency* and some aspects of *Information Literacy* appear to be relative strengths for our students. For these outcomes, students are progressing toward the graduation benchmark of level 3 (capstone).

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
Context Audience	100/200	16	2.0
	300/400	16	2.3
Design/Organization	100/200	23	1.8
	300/400	25	2.2
Diction	100/200	19	1.8
	300/400	26	2.4
Communication Style	100/200	18	1.5
	300/400	25	2.1

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

Multicultural/International Courses

The primary outcome to which artifacts from multicultural and international courses aligned was *Intercultural Thinking.* Usable scores were obtained by trait as follows:

Trait	Course Level	Number	Mean Score
Own Culture	100/200	5	1.0
	300/400	3	1.2
Other Cultures	100/200	16	1.1
	300/400	9	1.6
Communication with Others from Different Cultures	100/200	3	0.5
	300/400	1	0.5
Global Awareness	100/200	6	1.2
	300/400	7	1.5
Cultural Conflict	100/200	11	1.2
	300/400	4	1.3
Global Contexts	100/200	2	1.8
	300/400	2	1.3

Although there were no significant differences between these means based on course level, we note the small number of alignments in each cell. Further assessment using a larger sample size is needed to determine if the relatively low performance means (all < 2.0) for 300/400 level courses remains.

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	1	1.0
	300/400	10	1.3
Professional Rules and Standards of Conduct	100/200	1	1.5
	300/400	6	1.5
Civic Well-Being	100/200	1	1.0
	300/400	3	1.8
Complex Ethical Issues	100/200	0	N/A

	300/400	0	N/A

Although there were no significant differences between these means based on course level, we note the small number of alignments in each cell. However, there were more tags for 300/400 level than for 100/200 level service learning courses, but only the mean for Civic Well-Being approached 2.0. Additional assessment using a larger sample will be needed to determine the generalizability of this finding.

First Year Seminar (FYS)

FYS artifacts in our sample were fairly equally distributed among three of the course's outcomes (Information Literacy and Inquiry-Based and Metacognitive Thinking). Several artifacts also were aligned to Integrative Thinking. However, no FYS artifacts in our sample were aligned to Intercultural Thinking, which is one of the course outcomes.

Analysis of Misalignments

Traits that were judged to be misaligned (given scores of 100) were more likely to be those that described "process" rather than "product." Although we had attempted to address this issue with the inclusion of a free-text box in GEAR that asked students to explain why their artifacts aligned to specific outcomes(s)/trait(s), the Assessment Workgroup did not find most of these explanations helpful. Specific recommendations regarding this issue will follow at the end of the executive summary.

Data analysis showed that misalignments were independent of the person making the alignment (instructor versus student).

Recommendations from the 2014 Assessment Workgroup

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 and 99 when the instructor had failed to upload the assignment instructions.
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.
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.

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.
5. The workgroup strongly recommended that uploaded artifacts be summative in nature.
6. The workgroup recommended that we continue to assess artifacts for one outcome (can have multiple traits tagged for outcome) at a time.
7. The workgroup recommended that uploaded artifacts include process papers when tagged to an outcome/trait/performance level that addresses process rather than product. A quick perusal of the university's outcomes suggests that process papers should be mandated for the following outcome/trait performance levels.

Information Literacy

	Performance Levels			
Traits	**Introductory**	**Milestone**	**Capstone**	**Advanced**
Tool Use	**Identifies** the information need and multiple tools/strategies to gain needed information.	**Selects** the most appropriate tools and investigative methods for accessing the needed information.	**Revises** the search strategy if necessary and **employs** appropriate tools.	**Accesses** specialized information from proprietary information sources. (The deep Web, conferences, professional contacts, *et al*.)
Relevance of Information				**Collaborates** to **generate** and **disseminate** new information.
Complex Information Environment	**Reports** on the complexity of the worldwide information environment.	**Interprets** the complexity of the worldwide information environment.	**Questions** and **evaluates** the worldwide information environment.	**Adapts** to the limits of the worldwide information environment.
Legal/Ethical Issues			**Evaluates** ethical, legal and socioeconomic issues surrounding Information and information technology.	**Recommends** modifications to institutional policies and regulations.

Intercultural Thinking

	Performance Levels			
Traits	**Introductory**	**Milestone**	**Capstone**	**Advanced**
Communication with Others from Different Cultures		**Employs** verbal/nonverbal communication cues with different cultures in mind**.**		**Develops** ongoing interactions with others in different cultures, respecting the human dimensions of more than one worldview.

Metacognitive Thinking – I believe that this outcome, by its very nature, suggests the need for self-reflective process papers.

	Performance Levels			
Traits	**Introductory**	**Milestone**	**Capstone**	**Advanced**
Project Management	**Identifies** and **reflects upon** project goals.	**Develops** and **implements** a feasible plan to meet project goals.	**Evaluates** the effectiveness of a project plan or strategy.	**Proposes** an improved process for future projects. **Experiments** with new strategies.
Self-evaluation	**Identifies** and **reflects upon** prior knowledge and skills.	**Pursues** resources to improve knowledge and skills.	**Determines** degree of improvement in knowledge and skills.	**Develops** life-long learning skills in response to ongoing self-monitoring.

8. The workgroup recommended that instructors be provided with clearer definitions of rubric traits, especially for those of Inquiry-Based Thinking.
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).

Supporting Documentation

General Education Assessment Repository (GEAR) Artifact Assessment

Spring 2014

Sample

- Artifacts were randomly sampled for assessment. Although our goal had been to assess at least 15% of artifacts uploaded during the spring semester of 2014, time constraints forced us reduce this number considerably.

Total # of Artifacts Uploaded	# of Artifacts Assessed	% of Total
3,399	262	8%

Distribution of GEAR Artifacts among Marshall's Learning Outcomes

Marshall Outcome	Outcome Traits	# Uploaded Traits Tagged	# Traits Assessed	% of Total
Communication Fluency	Context/Audience	1,601	44 (1 eliminated due to inability to arrive at a score after three reads) = 43	3%
	Design/Organization	1,347	59	4%
	Diction	1,126	53	5%
	Communication Style	1,253	51	4%
	Communication Fluency Total	**5,327**	**206**	**4%**
Creative Thinking	Ambiguities and Possibilities	173	9	5%
	Risk Taking	32	2	6%
	Innovation	257	19	7%
	Creative Thinking Total	**462**	**30**	**7%**
Ethical and Civic Thinking	Ethical Self Awareness	130	19	15%
	Professional Rules and Standards of Conduct	72	12	17%
	Civic Well-Being	90	11	12%
	Complex Ethical issues	100	10	10%
	Ethical and Civic Thinking Total	**392**	**52**	**15%**

Distribution of GEAR Artifacts among Marshall's Learning Outcomes

Marshall Outcome	Outcome Traits	# Uploaded Traits Tagged	# Traits Assessed	% of Total
Information Literacy	Tool Use	209	24	12%
	Relevance of Information	283	28	10%
	Complex Information Environment	146	22	15%
	Legal/Ethical Issues	134	12	9%
	Information Literacy Total	**772**	**86**	**11%**
Inquiry-Based Thinking	Problem/Question	257	18	7%
	Research of Existing Knowledge	328	19	6%
	Method of Inquiry	306	21	7%
	Data Analysis and Conclusions	268	18	7%
	Inquiry-Based Thinking Total	**1,159**	**76**	**7%**
Integrative Thinking	Connections among Disciplines	86	6	7%
	Relation among Domains of Thinking	53	3	6%
	Transfer	121	4	3%
	Connections to Experience	146	12	8%
	Integrative Thinking Total	**406**	**25**	**6%**

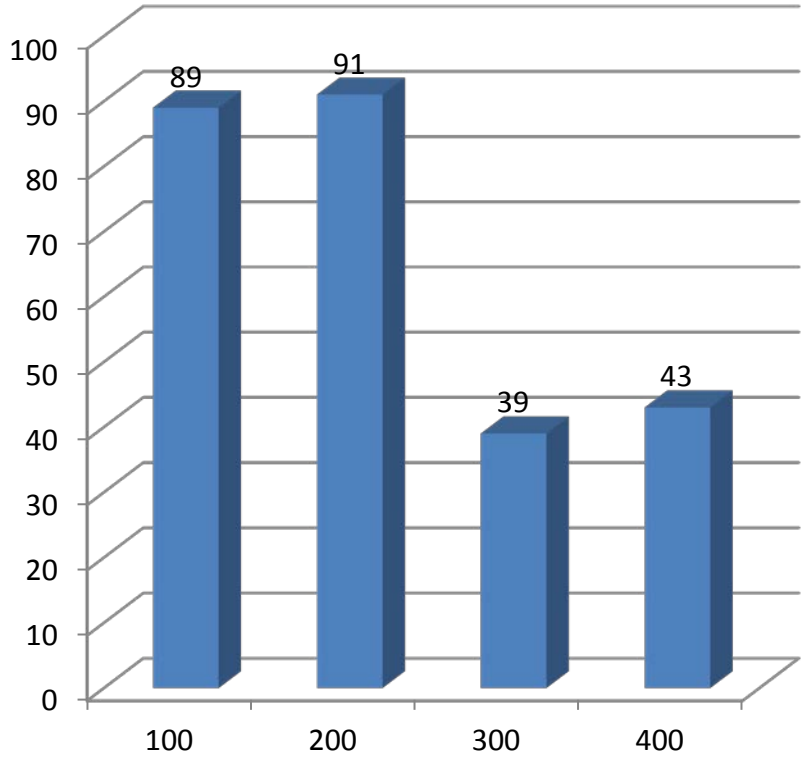
Distribution of GEAR Artifacts among Marshall's Learning Outcomes

Marshall Outcome	Outcome Traits	# Uploaded Traits Tagged	# Traits Assessed	% of Total
Intercultural Thinking	Own Culture	726	18	3%
	Other Cultures	1,000	34	3%
	Communication with Others from Different Cultures	363	11	3%
	Global Awareness	641	25	4%
	Cultural Conflict	484	20	4%
	Global Contexts	239	8	3%
	Intercultural Thinking Total	**3,453**	**116**	**3%**
Metacognitive Thinking	Project Management	101	12	12%
	Self-Evaluation	154	16	10%
	Metacognitive Thinking Total	**255**	**28**	**11%**
Quantitative Thinking	Context	61	19	31%
	Estimation	19	5	26%
	Visual Representations	42	15	36%
	Statistics	42	15	36%
	Quantitative Thinking Total	**164**	**54**	**33%**
Grand Total of Traits Tagged		**11,850**	**673**	**6%**

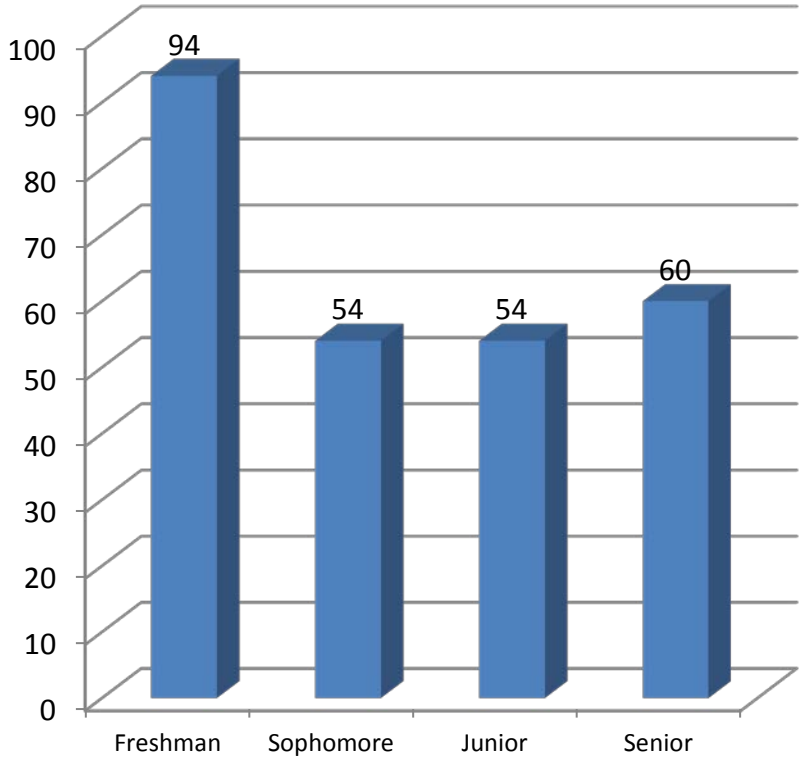
Sample Frequencies

Total # of artifacts assessed = 262

Course Level Frequencies



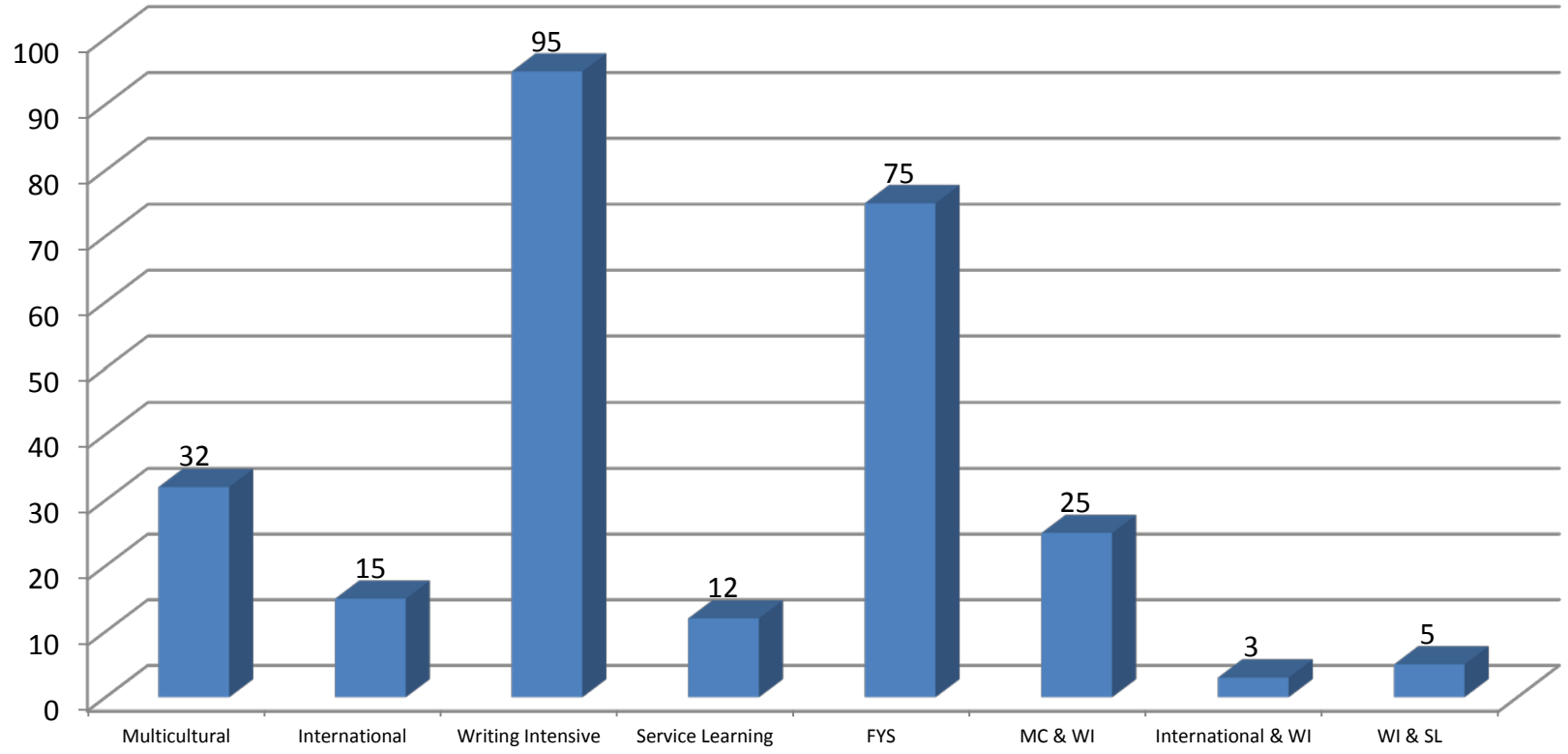
Class Rank Frequencies



Sample Frequencies

Total # of artifacts assessed = 262

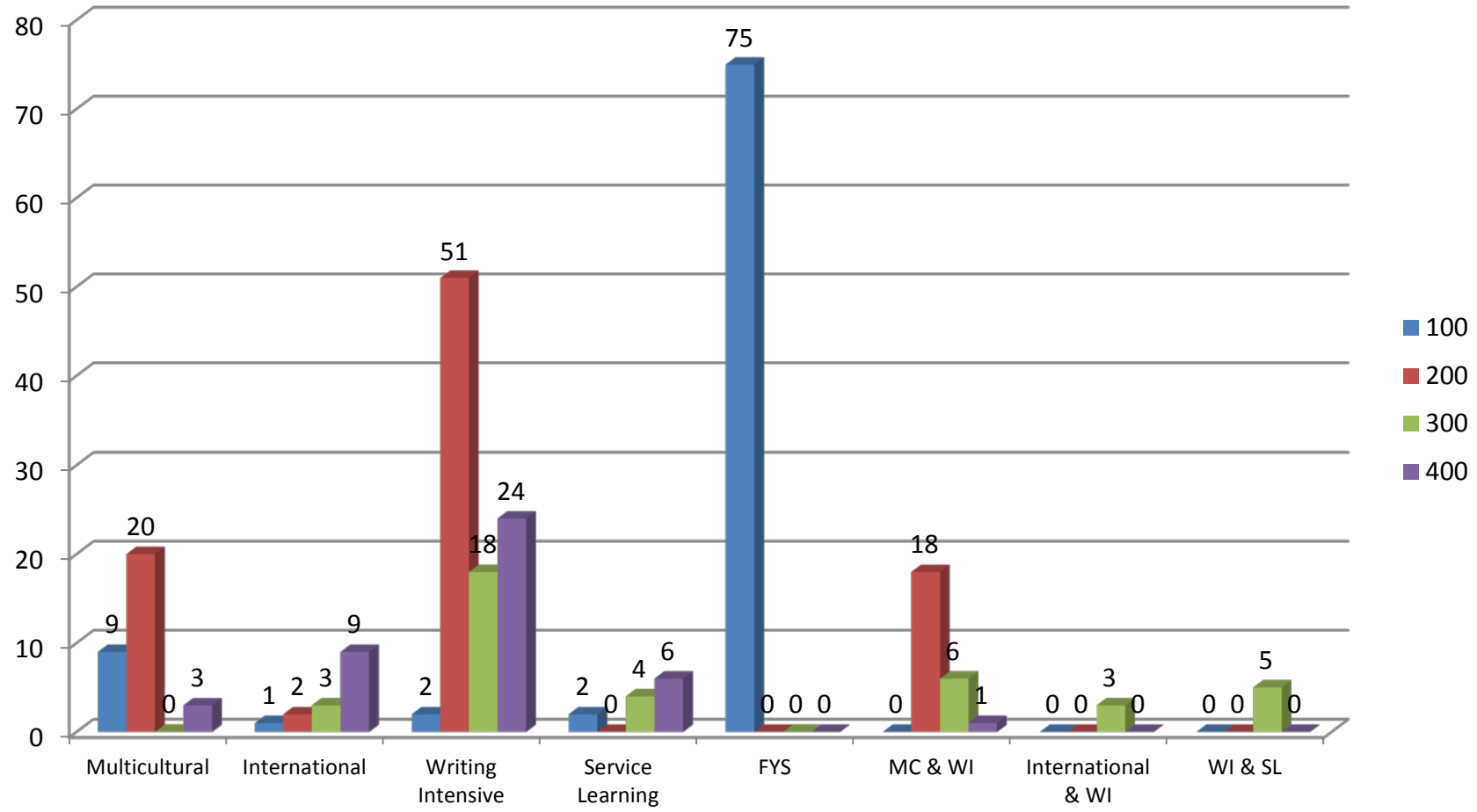
Course Type Frequencies



Sample Frequencies

Total # of artifacts assessed = 262

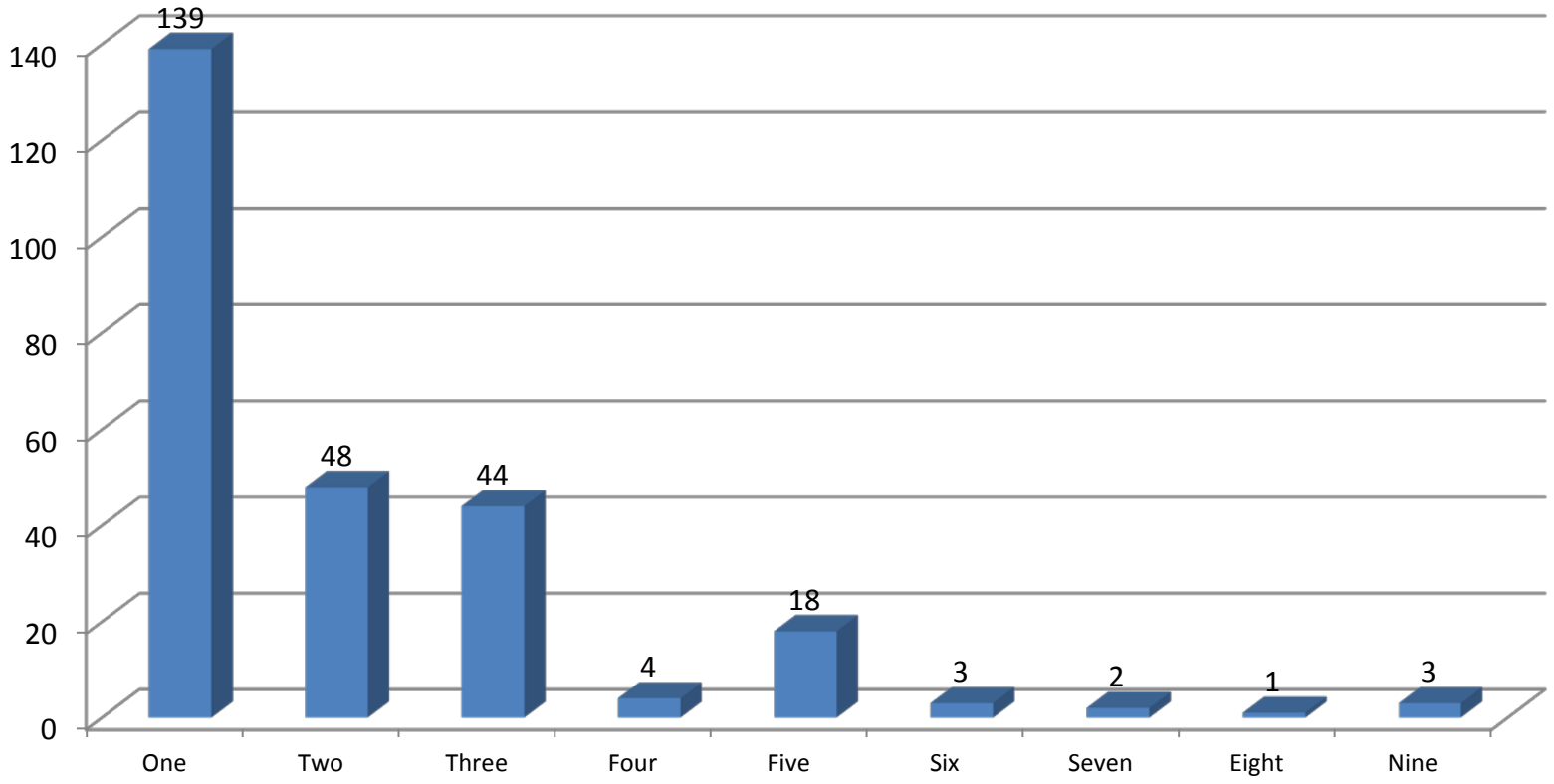
Course Type by Course Level



Sample Frequencies

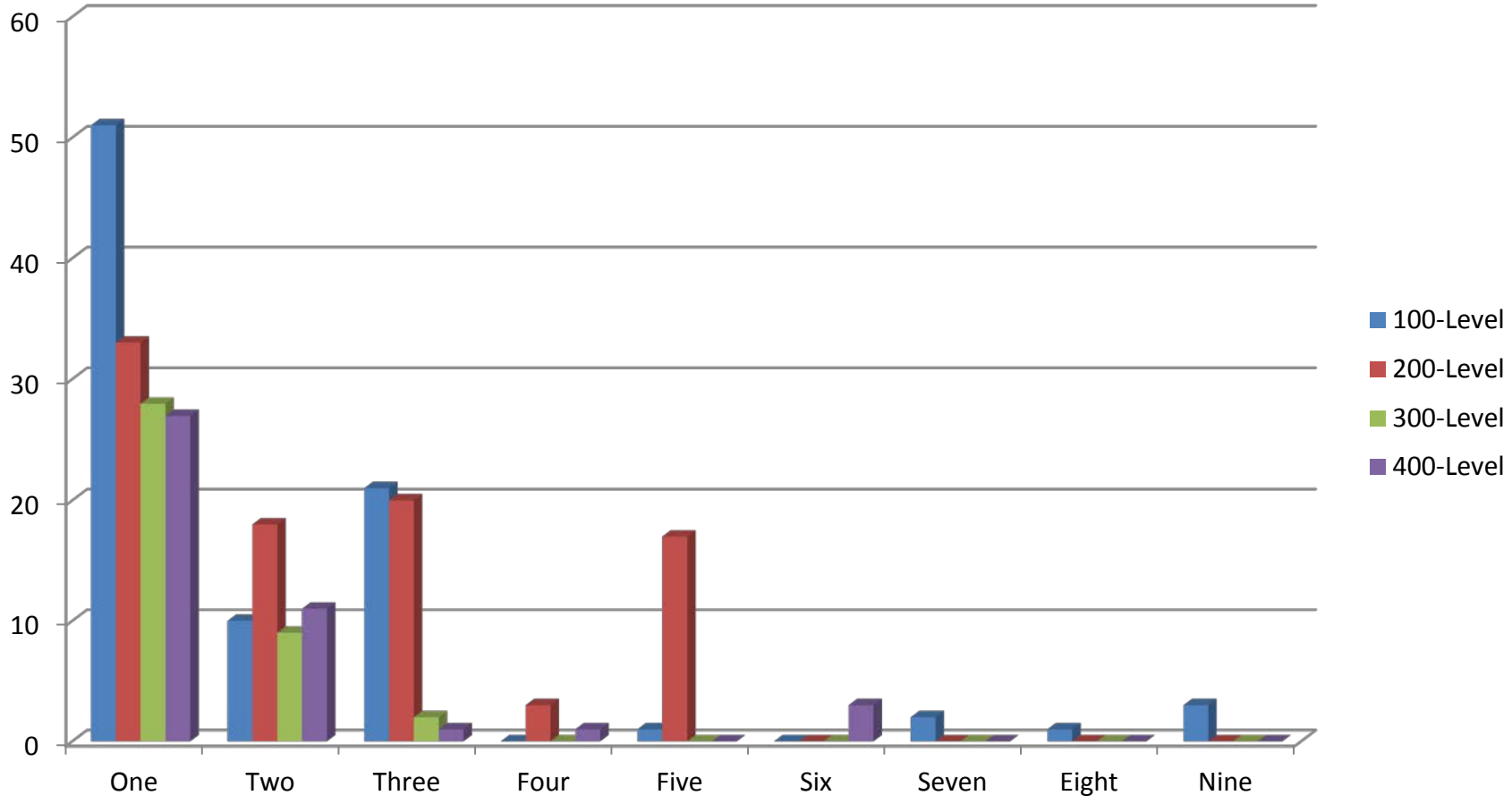
Total # of artifacts assessed = 262

 # Outcomes Tagged



of Outcomes Tagged by Course Level

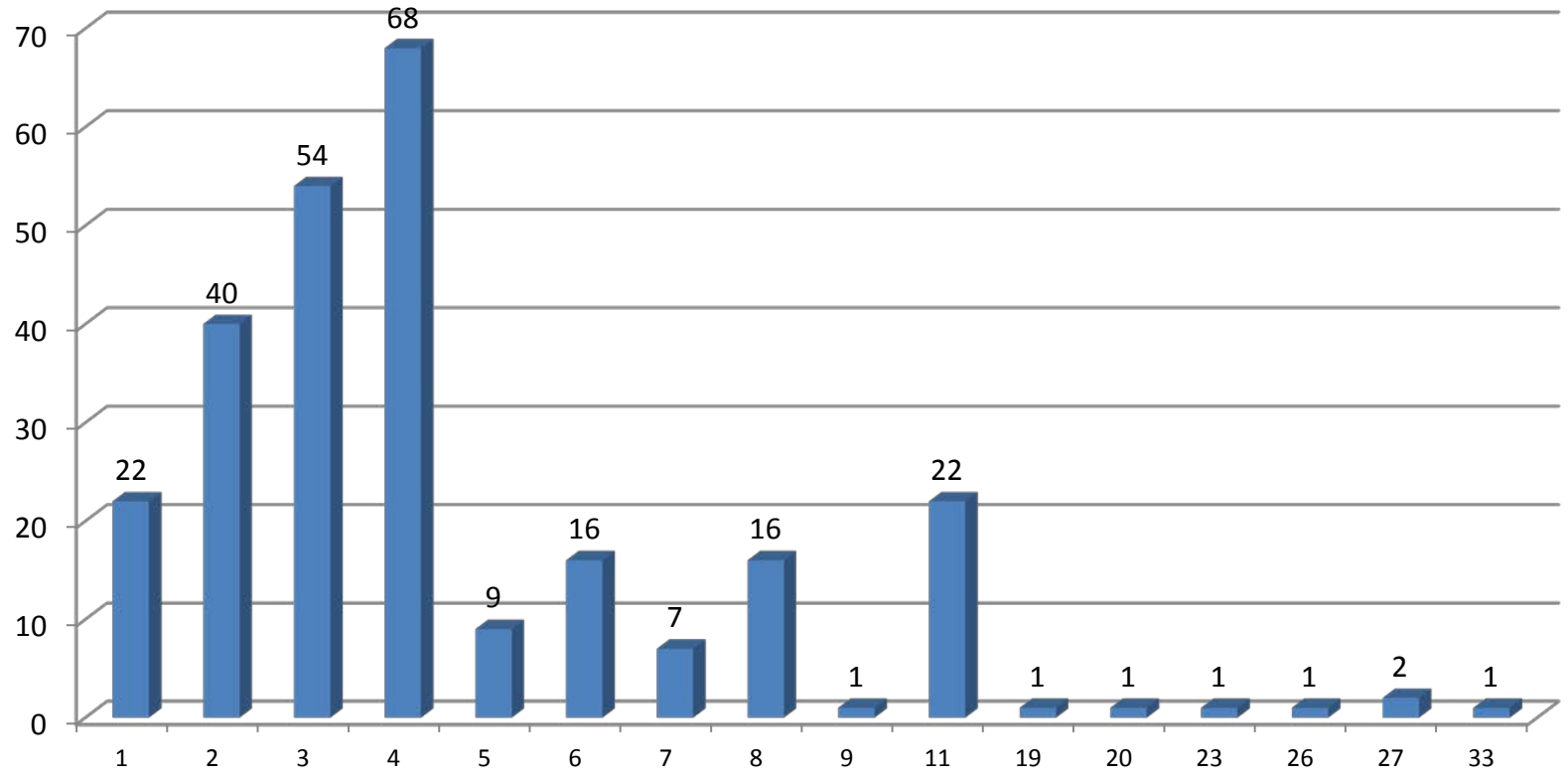
Total # of artifacts assessed = 262



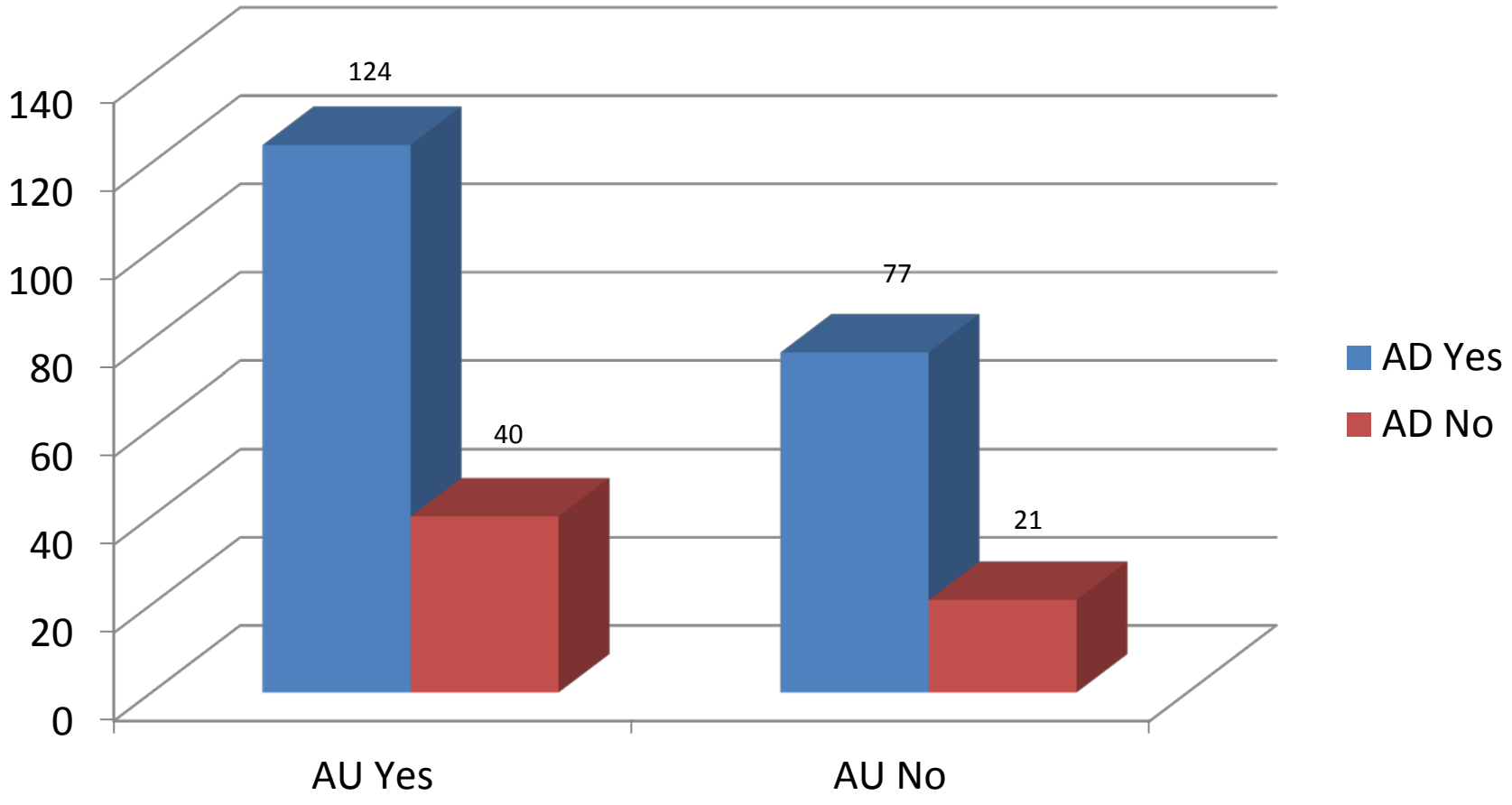
Sample Frequencies

Total # of artifacts assessed = 262

# of Traits Tagged



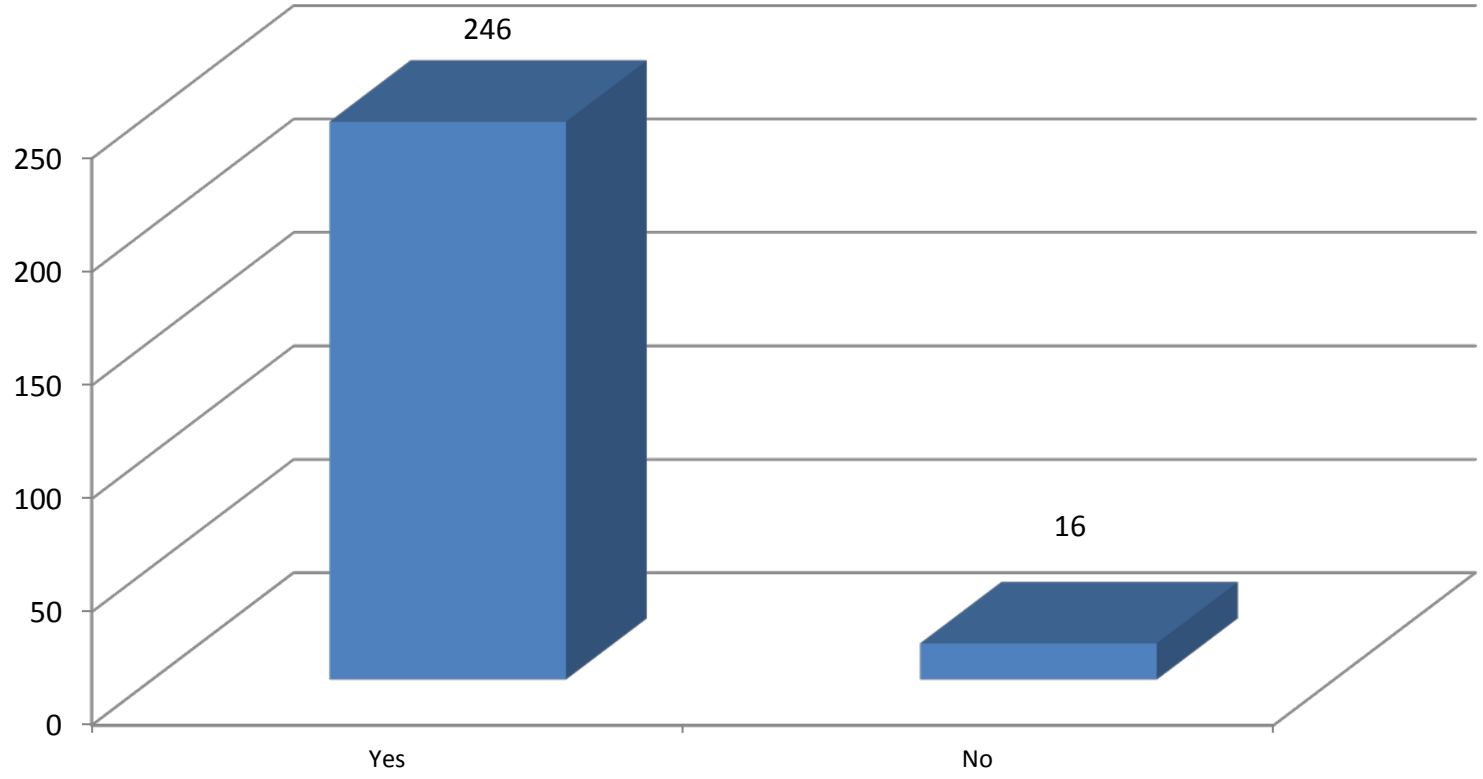
Assignment Instructions Uploaded by Assignment Description Included
Total # of artifacts assessed = 262



Sample Frequencies

Total # of artifacts assessed = 262

Instructor Tagged



Review Procedures

- Please access www.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	Context/Audience	43	2 (5%)	5 (12%)
	Design/Organization	59	2 (3%)	1 (2%)
	Diction	53	2 (4%)	0
	Communication Style	51	2 (4%)	0
Creative Thinking	Ambiguities & Possibilities	9	0	1 (11%)
	Risk Taking	2	0	1 (50%)
	Innovation	19	1 (5%)	2 (11%)
Ethical and Civic Thinking	Ethical Self-Awareness	19	0	5 (26%)
	Professional Rules and Standards of Conduct	12	0	4 (33%)
	Civic Well-Being	11	0	5 (46%)
	Complex Ethical Issues	10	0	5 (50%)

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

Outcome	Trait	Total Tags	# Not Able to be Assessed	# Misaligned
Information Literacy	Tool Use	24	3 (13%)	9 (38%)
	Relevance of Information	28	2 (7%)	1 (4%)
	Complex Information Environment	22	1 (5%)	18 (82%)
	Legal/Ethical Issues	12	2 (17%)	0
Inquiry-Based Thinking	Problem/Question	18	0	6 (33%)
	Research of Existing Knowledge	19	0	3 (16%)
	Method of Inquiry	21	0	9 (43%)
	Data Analysis and Conclusions	18	0	6 (33%)
Integrative Thinking	Connections among Disciplines	6	0	1 (17%)
	Relation among Domains of Thinking	3	0	3 (100%)
	Transfer	4	0	0
	Connections to Experience	12	0	3 (25%)

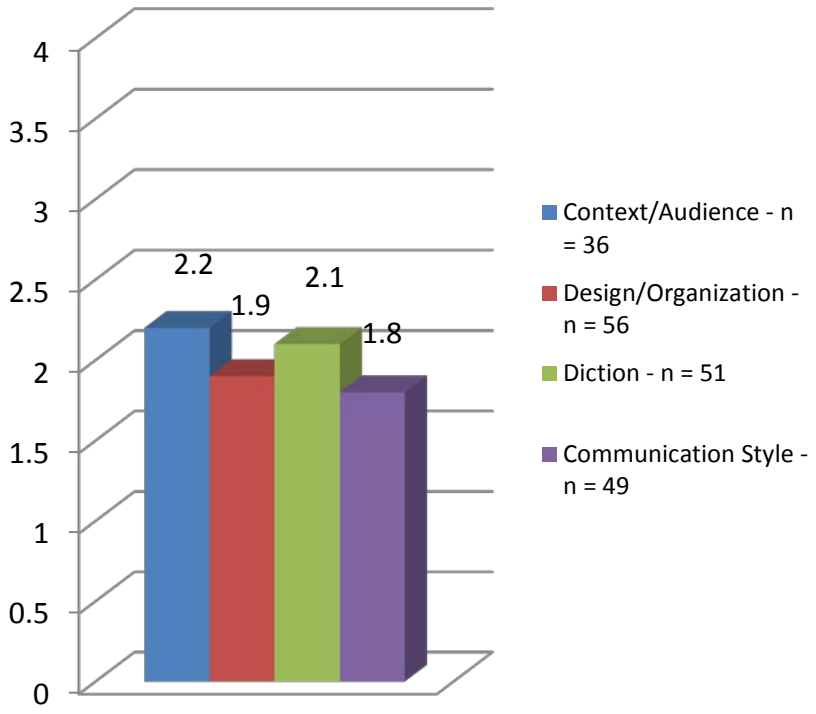
Artifacts Excluded Due to Inability to Assess or Misalignment with Tagged Outcomes/Traits

Outcome	Trait	Total Tags	# Not Able to be Assessed	# Misaligned
Intercultural Thinking	Own Culture	18	1 (6%)	9 (50%)
	Other Cultures	34	3 (9%)	3 (9%)
	Communication with Others from Different Cultures	11	1 (9%)	5 (46%)
	Global Awareness	25	1 (4%)	10 (40%)
	Cultural Conflict	20	1 (5%)	3 (15%)
	Cultural Contexts	8	0	3 (38%)
Metacognitive Thinking	Project Management	12	2 (17%)	4 (33%)
	Self-Evaluation	16	2 (13%)	3 (19%)
Quantitative Thinking	Context	19	1 (5%)	0
	Estimation	5	1 (20%)	4 (80%)
	Visual Representations	15	0	2 (13%)
	Statistics	15	0	2 (13%)

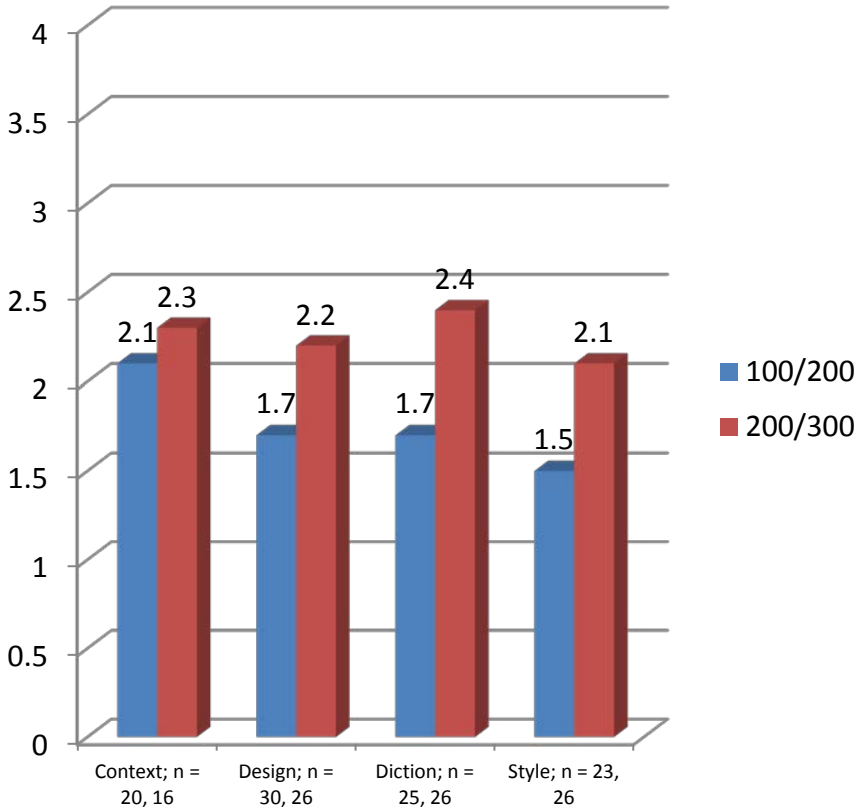
Communication Fluency

Mean Scores on a scale of 0 – 4, with 4 being the highest possible score
Mean differences based on class rank were not significant. Differences were significant for course level for design/organization, diction and communication style.

Overall Analysis



Analysis by Course Level

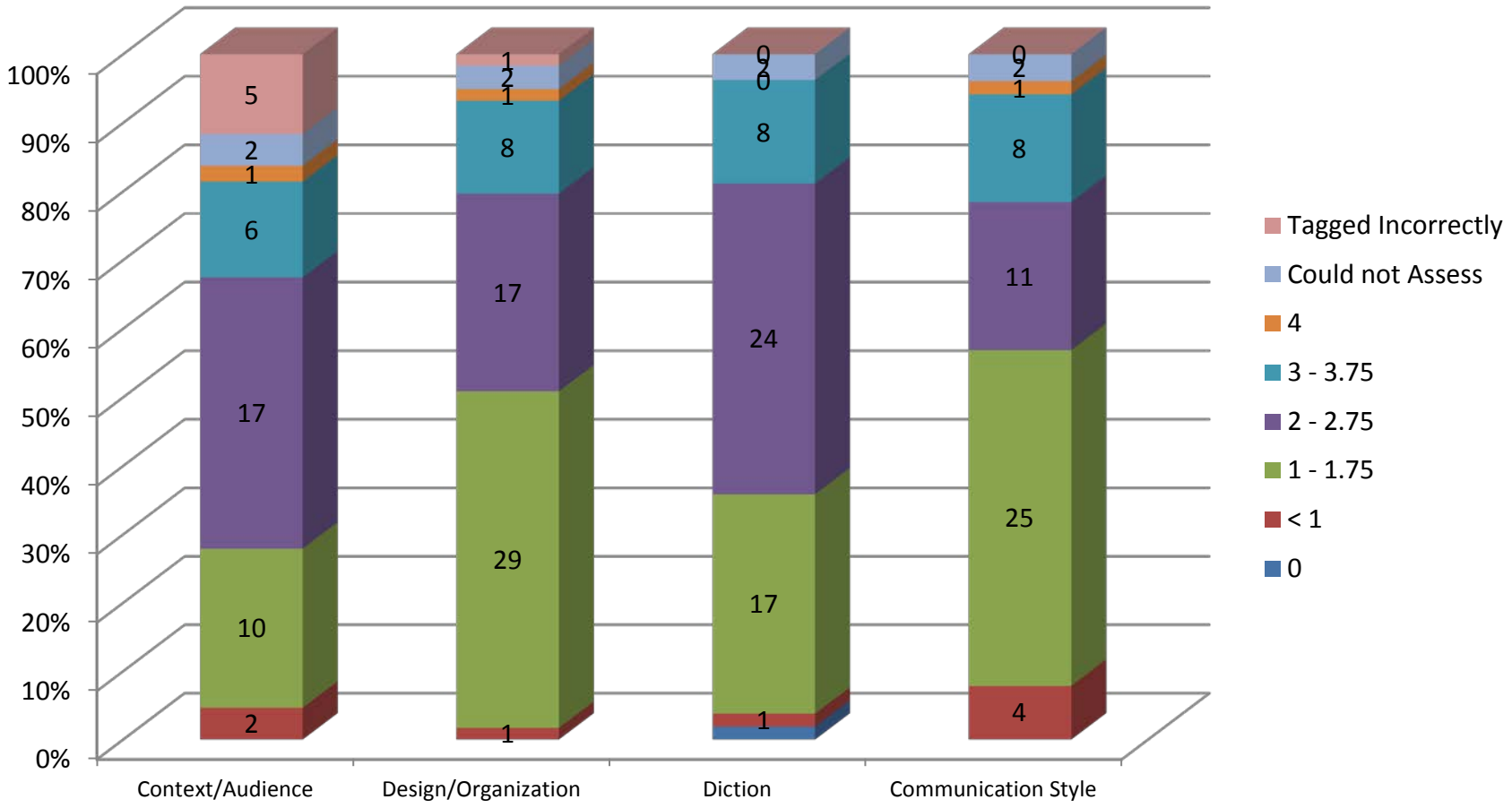


Communication Fluency

Number of artifacts scoring at each performance level

Trait/ Performance Level	Context/ Audience	Design/ Organization	Diction	CMM Style	Total
0	0	0	1 (2%)	0	1 (0%)
> 0, but < 1	2 (5%)	1 (2%)	1 (2%)	4 (8%)	8 (4%)
1 – 1.75	10 (23%)	29 (49%)	17 (32%)	25 (49%)	81 (39%)
2 – 2.75	17 (40%)	17 (29%)	24 (45%)	11 (22%)	69 (33%)
3 – 3.75	6 (14%)	8 (14%)	8 (15%)	8 (16%)	30 (15%)
4	1 (2%)	1 (2%)	0	1 (2%)	3 (1%)
Unable to Assess	2 (5%)	2 (3%)	2 (4%)	2 (4%)	8 (4%)
Tagged Incorrectly	5 (12%)	1 (2%)	0	0	6 (3%)
Totals	43 (100%)	59 (100%)	53 (100%)	51 (100%)	206 (100%)

Communication Fluency



Communication Fluency

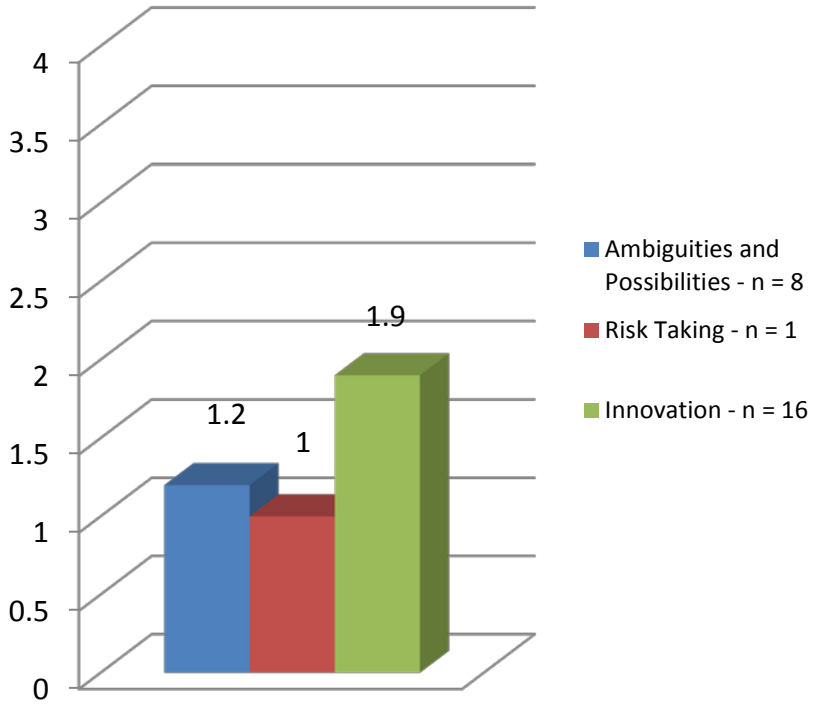
Inter-Rater Agreement Results

Trait/ Performance Level	Context/ Audience Kappa = .058	Design/ Organization Kappa = .118	Diction Kappa = .082	Style Kappa = .144	Total Kappa = .106
Agree	11 (25%)	21 (36%)	17 (32%)	19 (37%)	68 (33%)
Difference = 1 point or less	19 (43%)	24 (41%)	24 (45%)	20 (39%)	87 (42%)
Difference = 1.5 to 2 points	3 (7%)	10 (17%)	9 (17%)	10 (20%)	32 (15%)
Difference = 2.5 to 3 points	1 (2%)	0	0	0	1 (0%)
Agree on Misaligned	0	0	0	0	0
Agree on Unable to Score	0	0	0	0	0
Score + Misaligned	7 (16%)	1 (2%)	1 (2%)	0	9 (4%)
Score + Unable to Score	2 (5%)	2 (3%)	1 (2%)	1 (2%)	6 (3%)
Misaligned + Unable to Score	1 (2%)	1 (2%)	1 (2%)	1 (2%)	4 (2%)
Total	44 (100%)	59 (100%)	53 (100%)	51 (100%)	207 (100%)

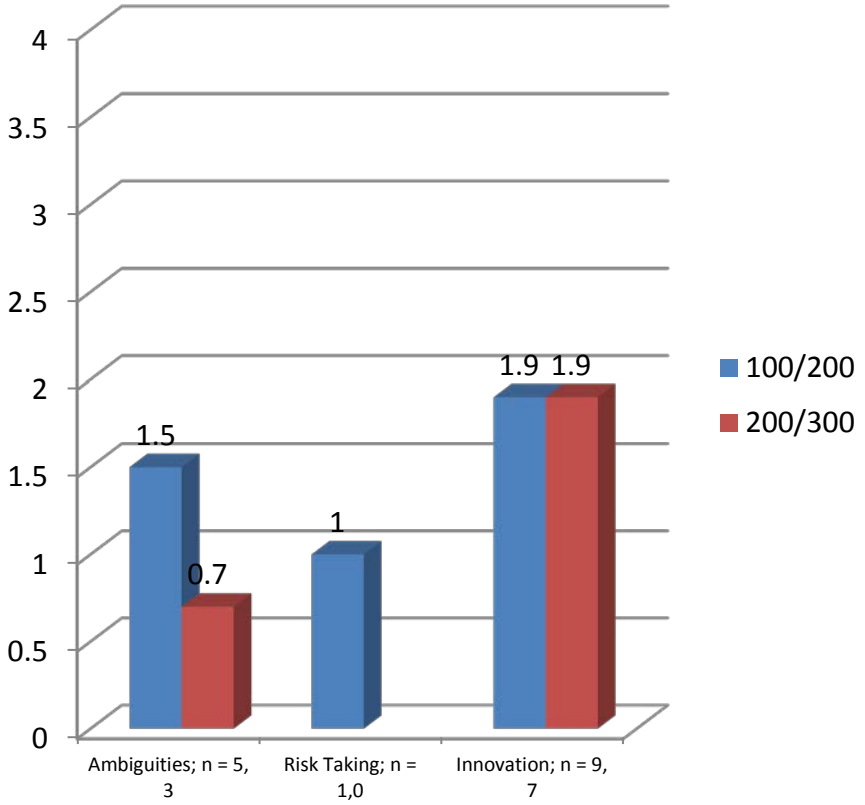
Creative Thinking

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

Overall Analysis



Analysis by Course Level

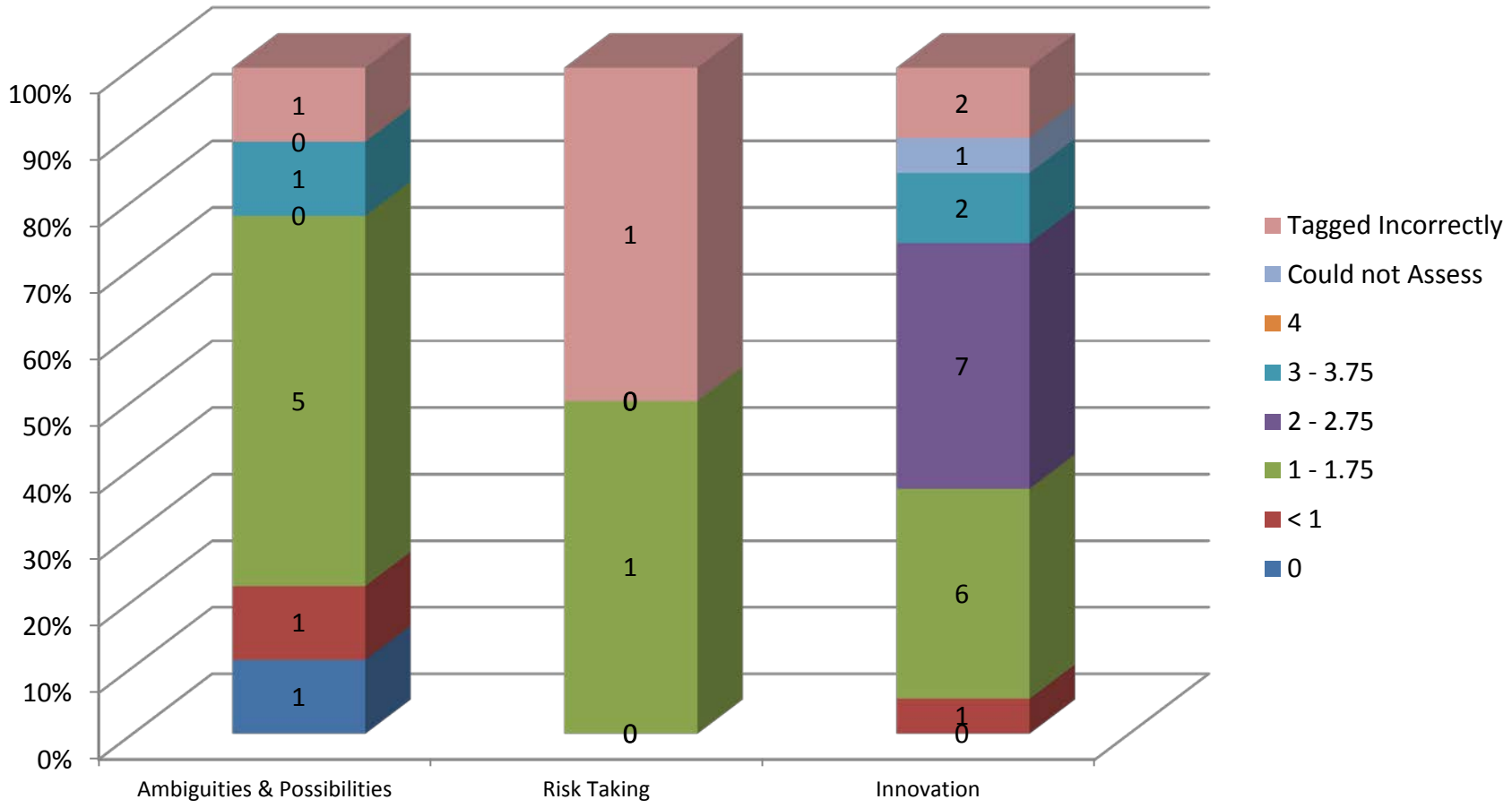


Creative Thinking

Number of artifacts scoring at each performance level

Trait/ Performance Level	Ambiguities and Possibilities	Risk Taking	Innovation	Total
0	1 (11%)	0	0	1 (3%)
> 0, but < 1	1 (11%)	0	1 (5%)	2 (7%)
1 – 1.75	5 (56%)	1 (50%)	6 (32%)	12 (40%)
2 – 2.75	0	0	7 (37%)	7 (23%)
3 – 3.75	1 (11%)	0	2 (11%)	3 (10%)
4	0	0	0	0
Unable to Assess	0	0	1 (5%)	1 (3%)
Tagged Incorrectly	1 (11%)	1 (50%)	2 (11%)	4 (13%)
Totals	9 (100%)	2 (100%)	19 (100%)	30 (100%)

Creative Thinking



Creative Thinking

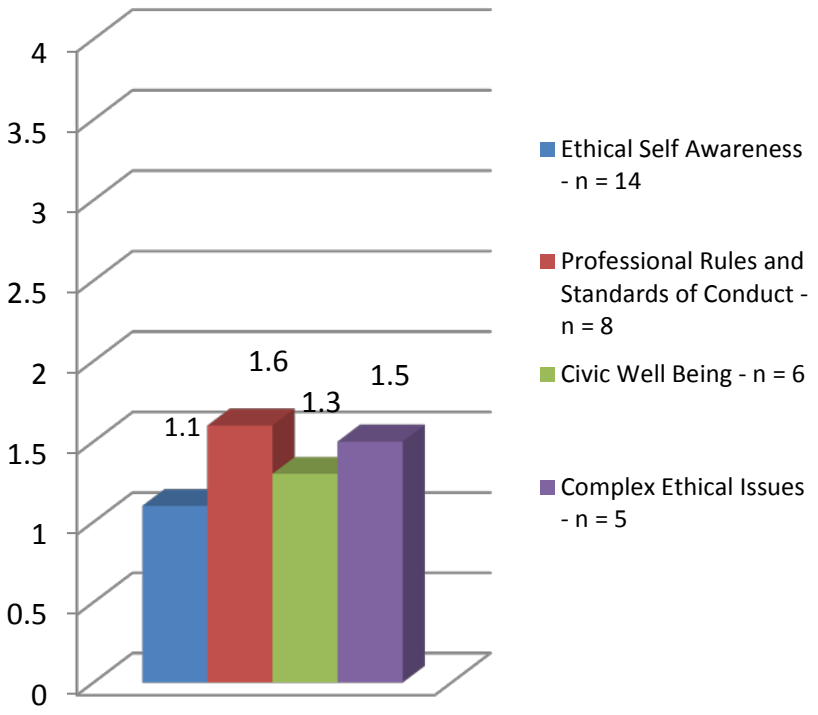
Inter-Rater Agreement Results

Trait/ Performance Level	Ambiguities and Possibilities Kappa = -.033	Risk Taking Kappa = -.333	Innovation Kappa = .163	Total Kappa = .127
Agree	2 (22%)	0	4 (21%)	6 (20%)
Difference = 1 point or less	1 (11%)	0	6 (32%)	7 (23%)
Difference = 1.5 to 2 points	2 (22%)	0	2 (11%)	4 (13%)
Difference = 2.5 to 3 points	0	0	0	0
Agree on Misaligned	0	0	1 (5%)	1 (3%)
Agree on Unable to Score	0	0	1 (5%)	1 (3%)
Score + Misaligned	2 (22%)	1 (50%)	2 (11%)	5 (17%)
Score + Unable to Score	2 (22%)	0	3 (16%)	5 (17%)
Misaligned + Unable to Score	0	1 (50%)	0	1 (3%)
Total	9 (100%)	2 (100%)	19 (100%)	30 (100%)

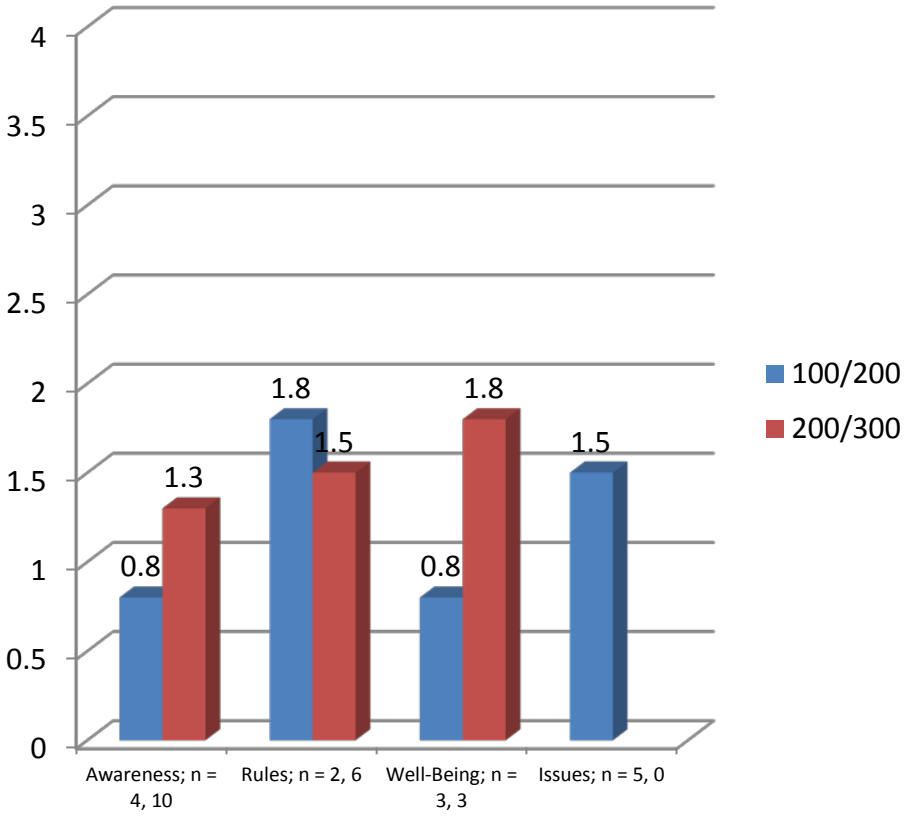
Ethical and Civic Thinking

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

Overall Analysis



Analysis by Course Level

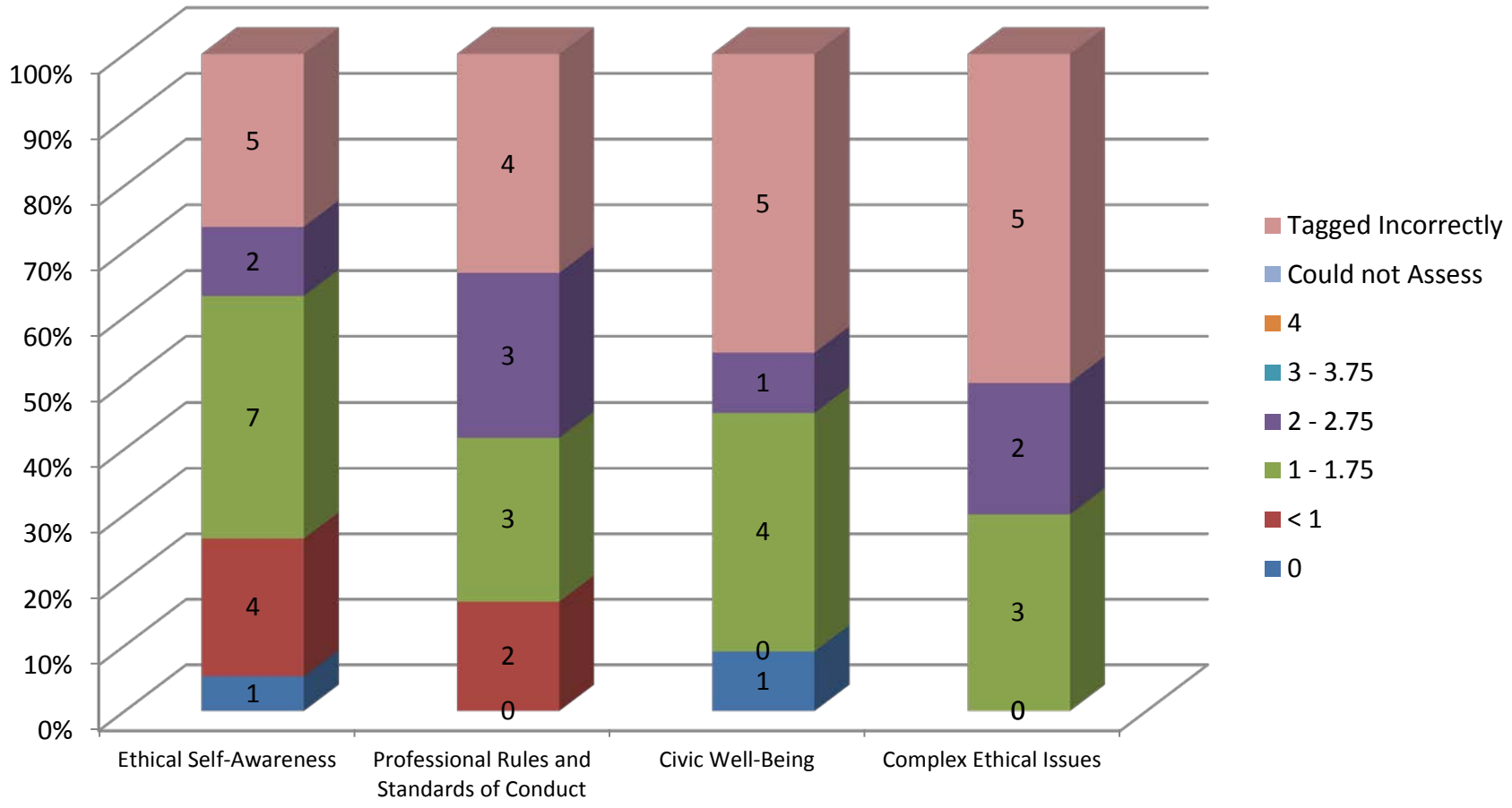


Ethical and Civic Thinking

Number of artifacts scoring at each performance level

Trait/ Performance Level	Ethical Self-Awareness	Professional Rules and Standards of Conduct	Civic Well-Being	Complex Ethical Issues	Total
0	1 (5%)	0	1 (9%)	0	2 (4%)
> 0, but < 1	4 (21%)	2 (17%)	0	0	6 (12%)
1 – 1.75	7 (37%)	3 (25%)	4 (36%)	3 (30%)	17 (33%)
2 – 2.75	2 (11%)	3 (25%)	1 (9%)	2 (20%)	8 (15%)
3 – 3.75	0	0	0	0	0
4	0	0	0	0	0
Unable to Assess	0	0	0	0	0
Tagged Incorrectly	5 (26%)	4 (33%)	5 (45%)	5 (50%)	19 (37%)
Totals	19 (100%)	12 (100%)	11 (100%)	10 (100%)	52 (100%)

Ethical and Civic Thinking



Ethical and Civic Thinking

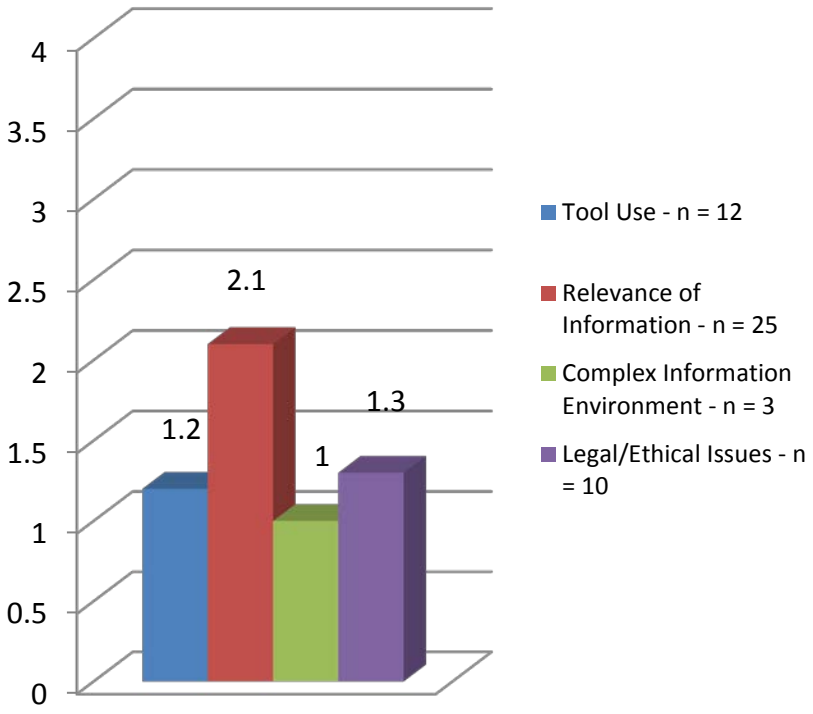
Inter-Rater Agreement Results

Trait/ Performance Level	Ethical Self-Awareness Kappa = .053	Professional Rules and Standards of Conduct Kappa = -.190	Civic Well-Being Kappa = -.019	Complex Ethical Issues Kappa = .231	Total Kappa = .006
Agree	3 (16%)	0	1 (9%)	4 (40%)	8 (15%)
Difference = 1 point or less	3 (16%)	5 (42%)	3 (27%)	1 (10%)	12 (23%)
Difference = 1.5 to 2 points	4 (21%)	2 (17%)	0	0	6 (12%)
Difference = 2.5 to 3 points	0	0	1 (9%)	0	1 (2%)
Agree on Misaligned	1 (5%)	0	0	0	1 (2%)
Agree on Unable to Score	0	0	0	0	0
Score + Misaligned	5 (26%)	2 (17%)	3 (27%)	2 (20%)	12 (23%)
Score + Unable to Score	2 (11%)	2 (17%)	2 (18%)	2 (20%)	8 (15%)
Misaligned + Unable to Score	1 (5%)	1 (8%)	1 (9%)	1 (10%)	4 (8%)
Total	19 (100%)	12 (100%)	11 (100%)	10 (100%)	52 (100%)

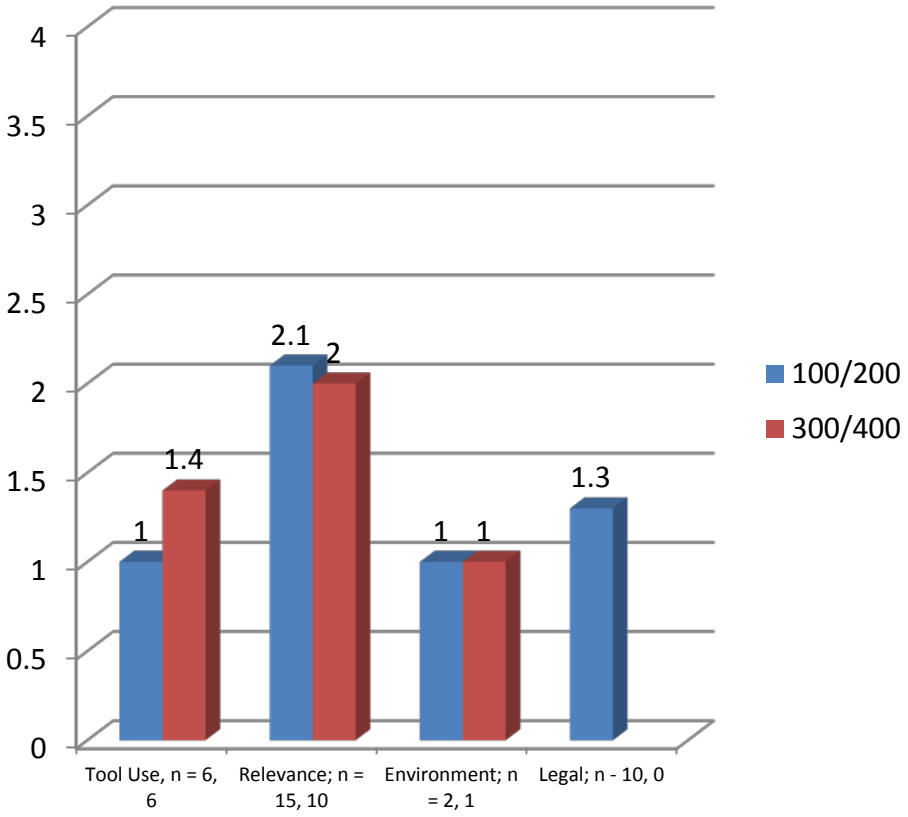
Information Literacy

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

Overall Analysis



Analysis by Course Level

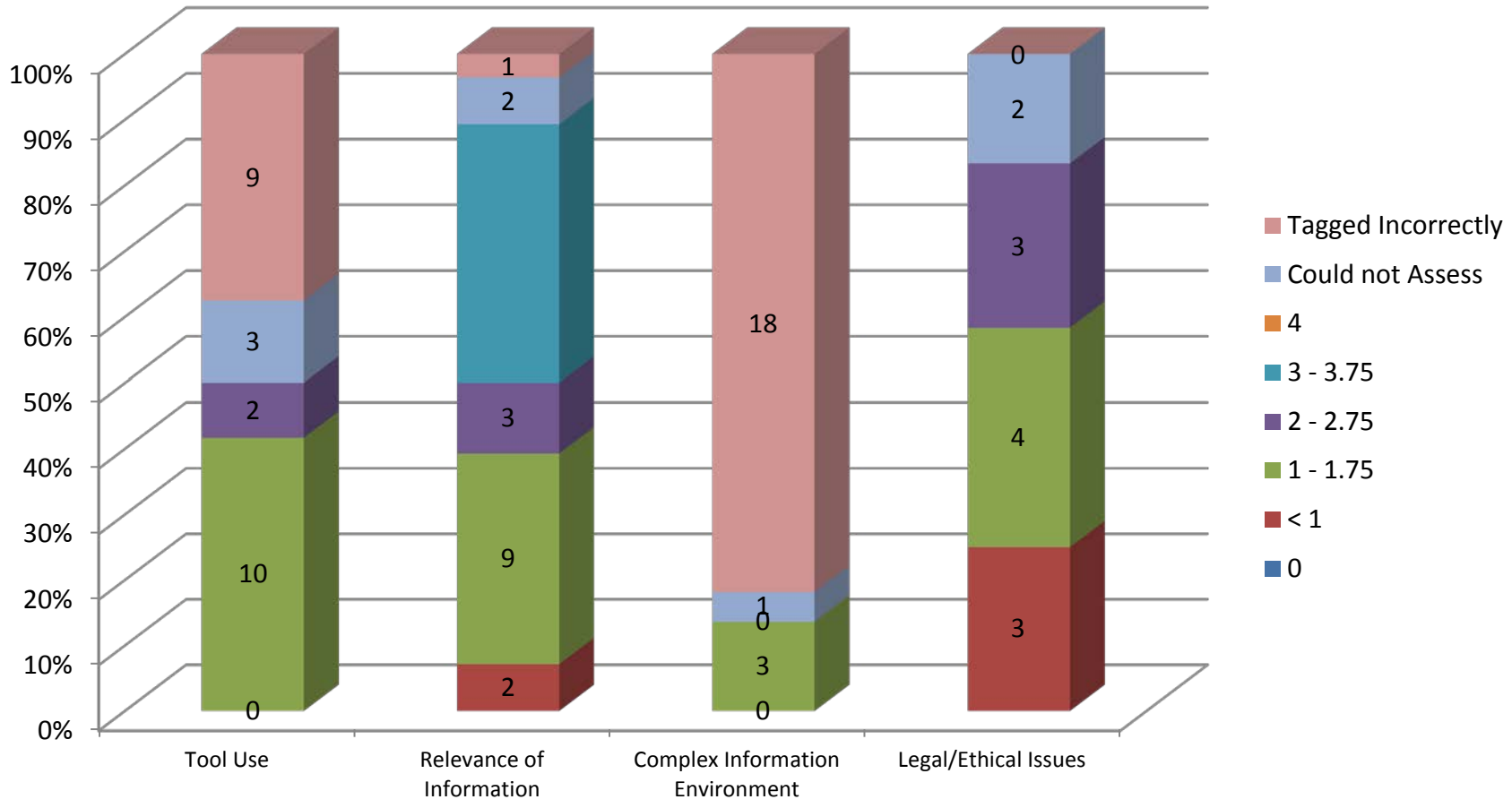


Information Literacy

Number of artifacts scoring at each performance level

Trait/ Performance Level	Tool Use	Relevance of Information	Complex Information Environment	Legal/Ethical Issues	Total
0	0	0	0	0	0
> 0, but < 1	0	2 (7%)	0	3 (25%)	5 (6%)
1 – 1.75	10 (42%)	9 (32%)	3 (14%)	4 (33%)	26 (30%)
2 – 2.75	2 (8%)	3 (11%)	0	3 (25%)	8 (9%)
3 – 3.75	0	11 (39%)	0	0	11 (13%)
4	0	0	0	0	0
Unable to Assess	3 (13%)	2 (7%)	1 (5%)	2 (17%)	8 (9%)
Tagged Incorrectly	9 (38%)	1 (4%)	18 (82%)	0	28 (33%)
Totals	24 (100%)	28 (100%)	22 (100%)	12 (100%)	86 (100%)

Information Literacy



Information Literacy

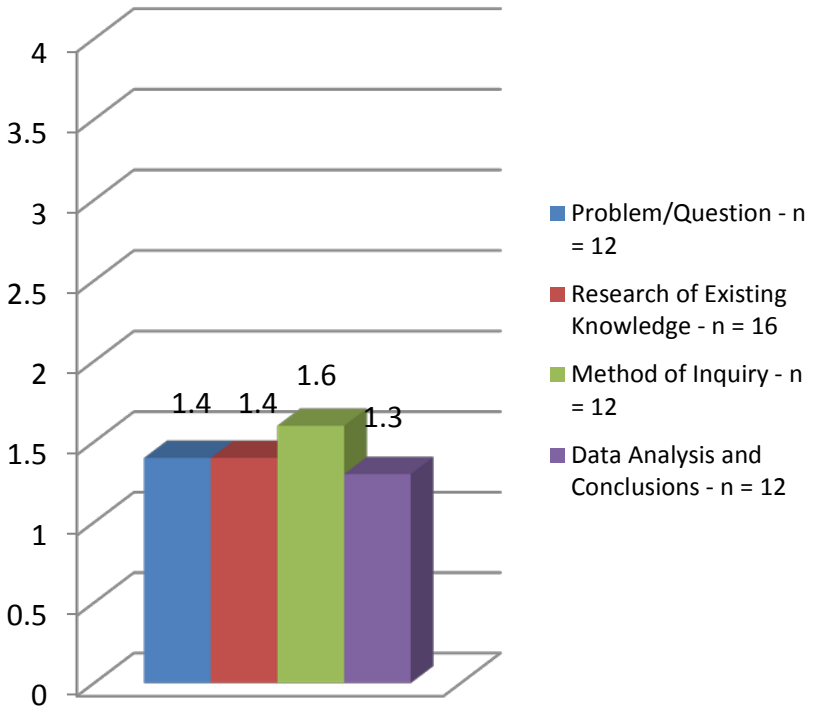
Inter-Rater Agreement Results

Trait/ Performance Level	Tool Use Kappa = .217	Relevance of Information Kappa = .353	Complex Information Environment Kappa = .229	Legal/Ethical Issues Kappa = .179	Total Kappa = .321
Agree on Rubric Score	7 (29%)	13 (46%)	3 (14%)	3 (25%)	26 (30%)
Difference = 1 point or less	1 (4%)	6 (21%)	0	7 (58%)	14 (16%)
Difference = 1.5 to 2 points	0	5 (18%)	0	0	5 (6%)
Difference = 2.5 to 3 points	0	0	0	0	0
Agree on Misaligned	2 (8%)	1 (4%)	7 (32%)	0	10 (12%)
Agree on Unable to Score	1 (4%)	1 (4%)	1 (5%)	1 (8%)	4 (5%)
Score + Misaligned	11 (46%)	1 (4%)	10 (45%)	0	22 (26%)
Score + Unable to Score	1 (4%)	1 (4%)	1 (5%)	0	3 (3%)
Misaligned + Unable to Score	1 (4%)	0	0	1 (8%)	2 (2%)
Total	24 (100%)	28 (100%)	22 (100%)	12 (100%)	86 (100%)

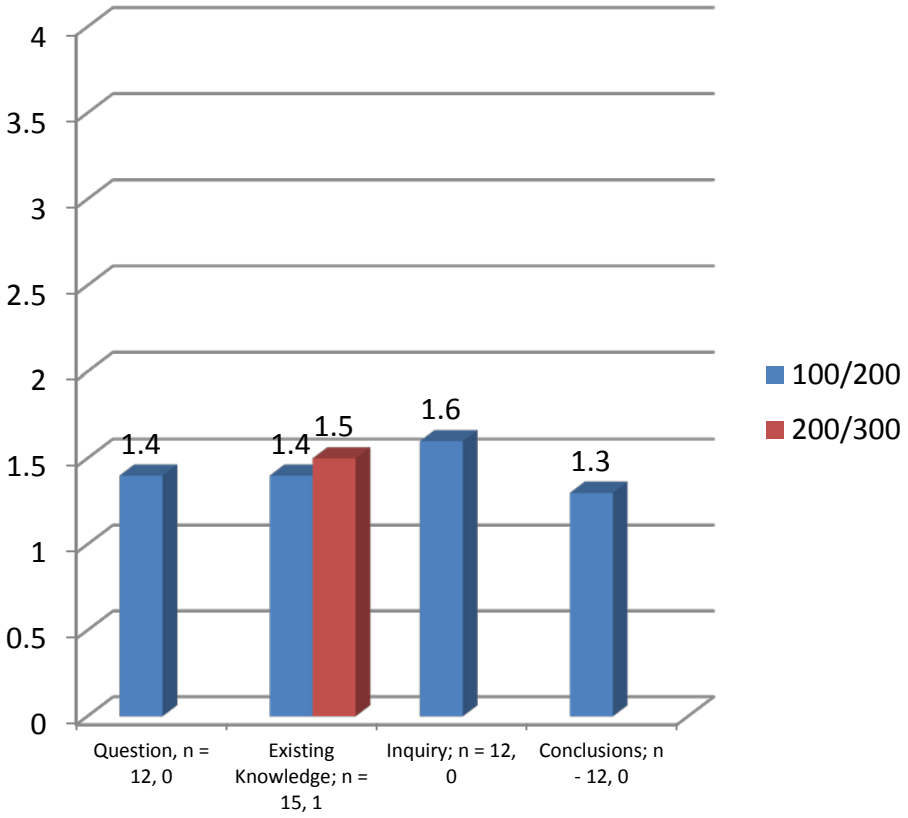
Inquiry-Based Thinking

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

Overall Analysis



Analysis by Course Level

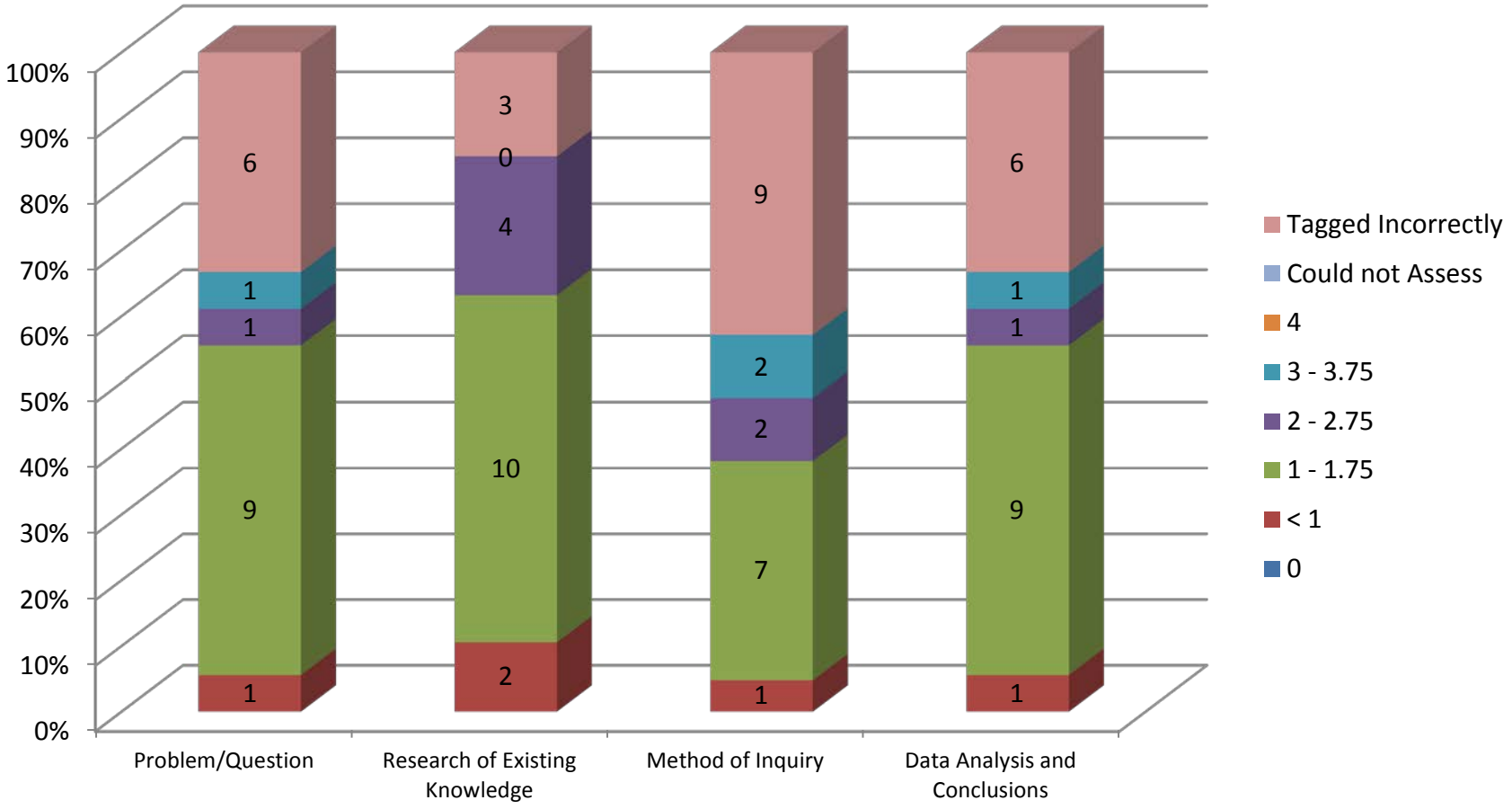


Inquiry-Based Thinking

Number of artifacts scoring at each performance level

Trait/ Performance Level	Problem/Question	Research of Existing Knowledge	Method of Inquiry	Data Analysis and Conclusions	Total
0	0	0	0	0	0
> 0, but < 1	1 (6%)	2 (11%)	1 (5%)	1 (6%)	5 (7%)
1 – 1.75	9 (50%)	10 (53%)	7 (33%)	9 (50%)	35 (46%)
2 – 2.75	1 (6%)	4 (21%)	2 (10%)	1 (6%)	8 (11%)
3 – 3.75	1 (6%)	0	2 (10%)	1 (6%)	4 (5%)
4	0	0	0	0	0
Unable to Assess	0	0	0	0	0
Tagged Incorrectly	6 (33%)	3 (16%)	9 (43%)	6 (33%)	24 (32%)
Totals	18 (100%)	19 (100%)	21 (100%)	18 (100%)	76 (100%)

Inquiry-Based Thinking



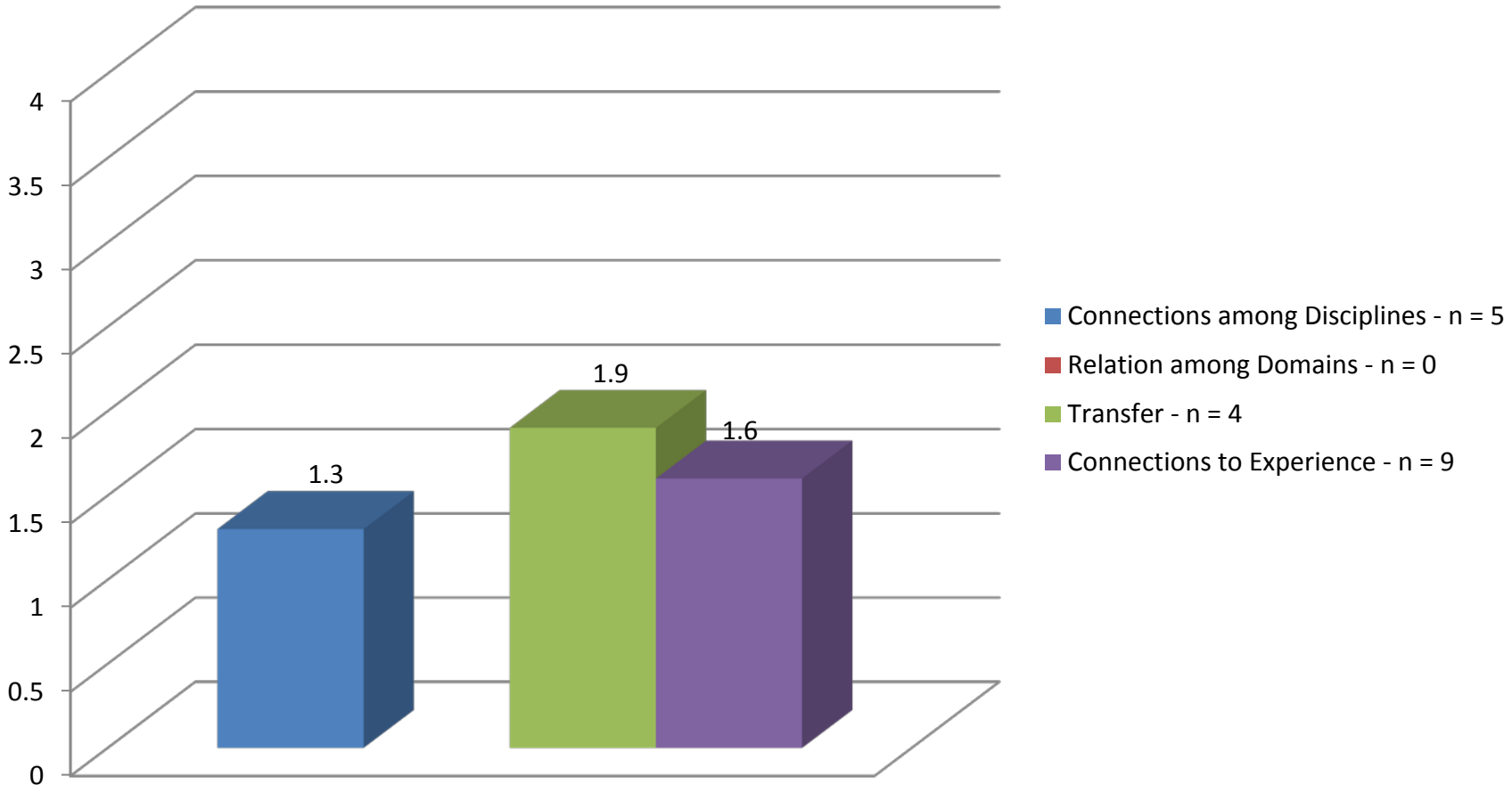
Inquiry-Based Thinking

Inter-Rater Agreement Results

Trait/ Performance Level	Problem/Question Kappa = -.033	Research of Existing Knowledge Kappa = -.073	Method of Inquiry Kappa = .138	Data Analysis and Conclusions Kappa = .036	Total Kappa = .024
Agree on Rubric Score	4 (22%)	5 (26%)	5 (24%)	5 (28%)	19 (25%)
Difference = 1 point or less	5 (28%)	6 (32%)	2 (10%)	3 (17%)	16 (21%)
Difference = 1.5 to 2 points	2 (11%)	3 (16%)	1 (5%)	2 (11%)	8 (11%)
Difference = 2.5 to 3 points	0	0	1 (5%)	0	1 (1%)
Agree on Misaligned	0	0	2 (10%)	1 (6%)	3 (4%)
Agree on Unable to Score	0	0	0	0	0
Score + Misaligned	6 (33%)	4 (21%)	8 (38%)	6 (33%)	24 (32%)
Score + Unable to Score	1 (6%)	1 (5%)	2 (10%)	1 (6%)	5 (7%)
Misaligned + Unable to Score	0	0	0	0	0
Total	18 (100%)	19 (100%)	21 (100%)	18 (100%)	76 (100%)

Integrative Thinking

Mean Scores on a scale of 0 – 4, with 4 being the highest possible score; mean differences were not significant based on either course level or class rank. Note: There were no 300/400 Level Courses Tagged to Integrative Thinking in this Sample

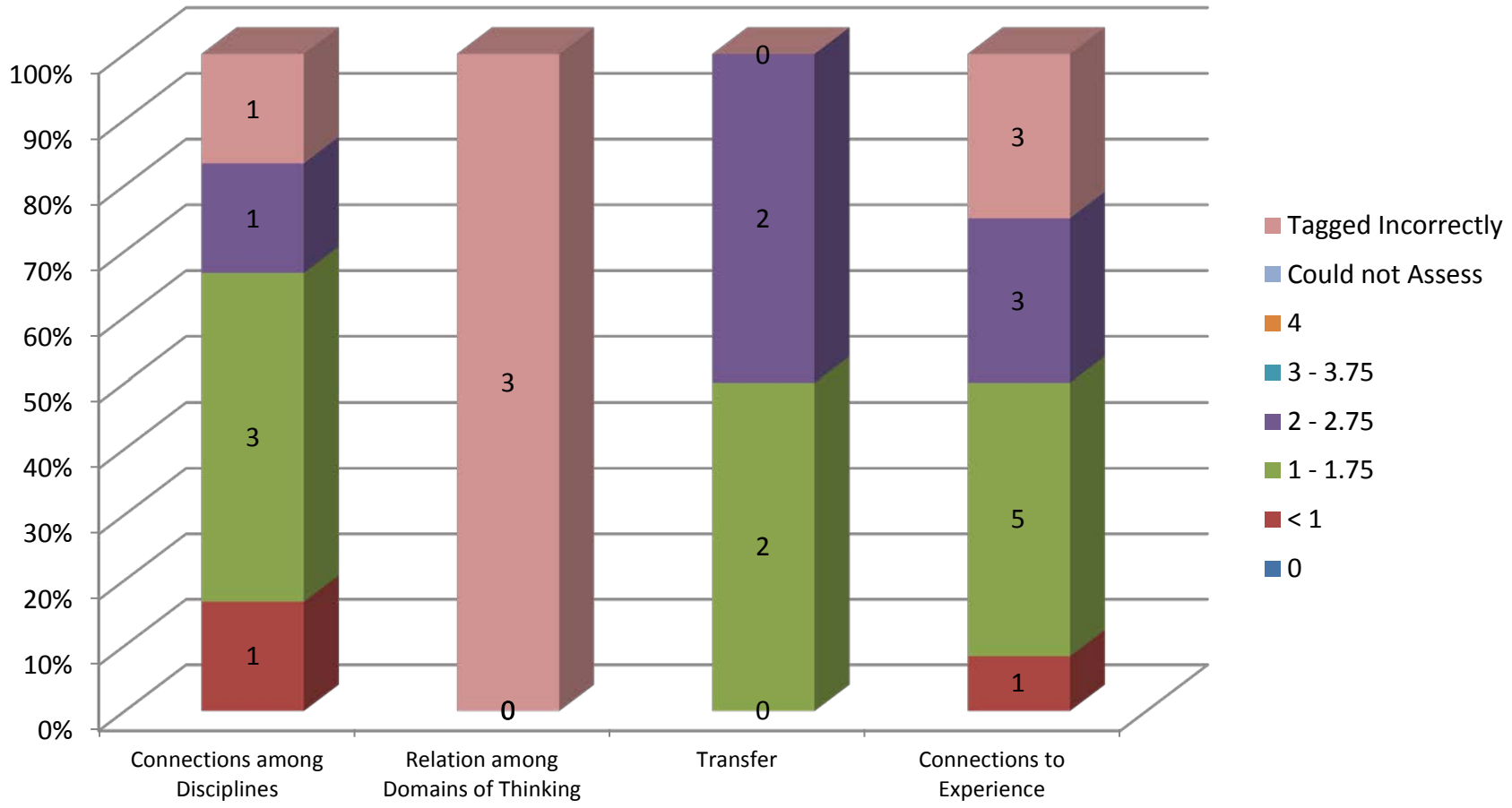


Integrative Thinking

Number of artifacts scoring at each performance level

Trait/ Performance Level	Connections among Disciplines	Relations among Domains of Thinking	Transfer	Connections to Experience	Total
0	0	0	0	0	0
> 0, but < 1	1 (17%)	0	0	1 (8%)	2 (8%)
1 – 1.75	3 (50%)	0	2 (50%)	5 (42%)	10 (40%)
2 – 2.75	1 (17%)	0	2 (50%)	3 (25%)	6 (24%)
3 – 3.75	0	0	0	0	0
4	0	0	0	0	0
Unable to Assess	0	0	0	0	0
Tagged Incorrectly	1 (17%)	3 (100%)	0	3 (25%)	7 (28%)
Totals	6 (100%)	3 (100%)	4 (100%)	12 (100%)	25 (100%)

Integrative Thinking



Integrative Thinking

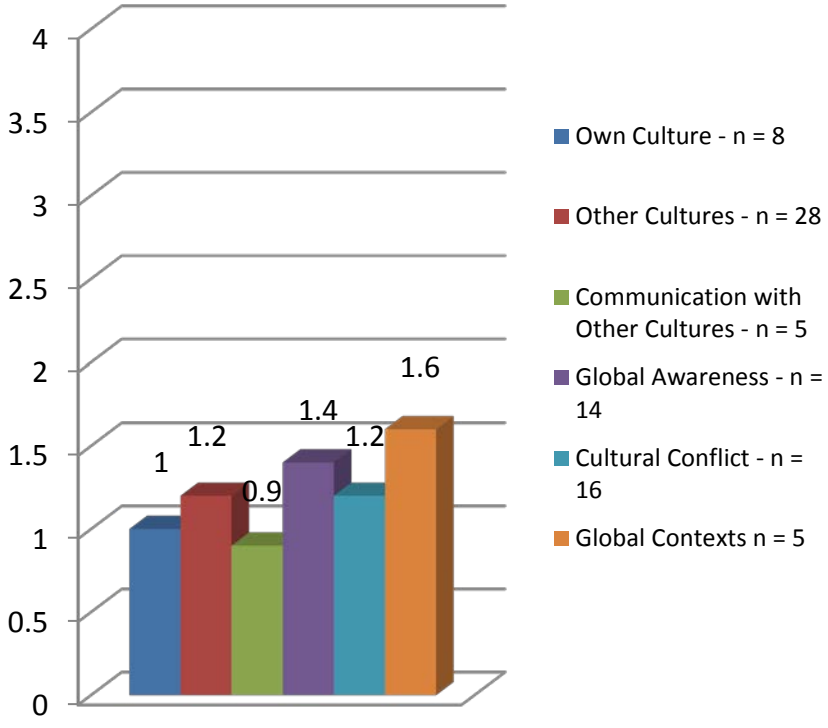
Inter-Rater Agreement Results

Trait/ Performance Level	Connections among Disciplines Kappa = .400	Relation among Domains of Thinking Kappa not computed	Transfer Kappa not computed	Connections to Experience Kappa = .068	Total Kappa = .170
Agree on Rubric Score	2 (33%)	0	1 (25%)	3 (25%)	6 (24%)
Difference = 1 point or less	2 (33%)	0	3 (75%)	3 (25%)	8 (32%)
Difference = 1.5 to 2 points	0	0	0	1 (8%)	1 (4%)
Difference = 2.5 to 3 points	0	0	0	0	0
Agree on Misaligned	1 (17%)	1 (33%)	0	1 (8%)	3 (12%)
Agree on Unable to Score	0	0	0	0	0
Score + Misaligned	0	2 (67%)	0	3 (25%)	5 (20%)
Score + Unable to Score	1 (17%)	0	0	1 (8%)	2 (8%)
Misaligned + Unable to Score	0	0	0	0	0
Total	6 (100%)	3 (100%)	4 (100%)	12 (100%)	25 (100%)

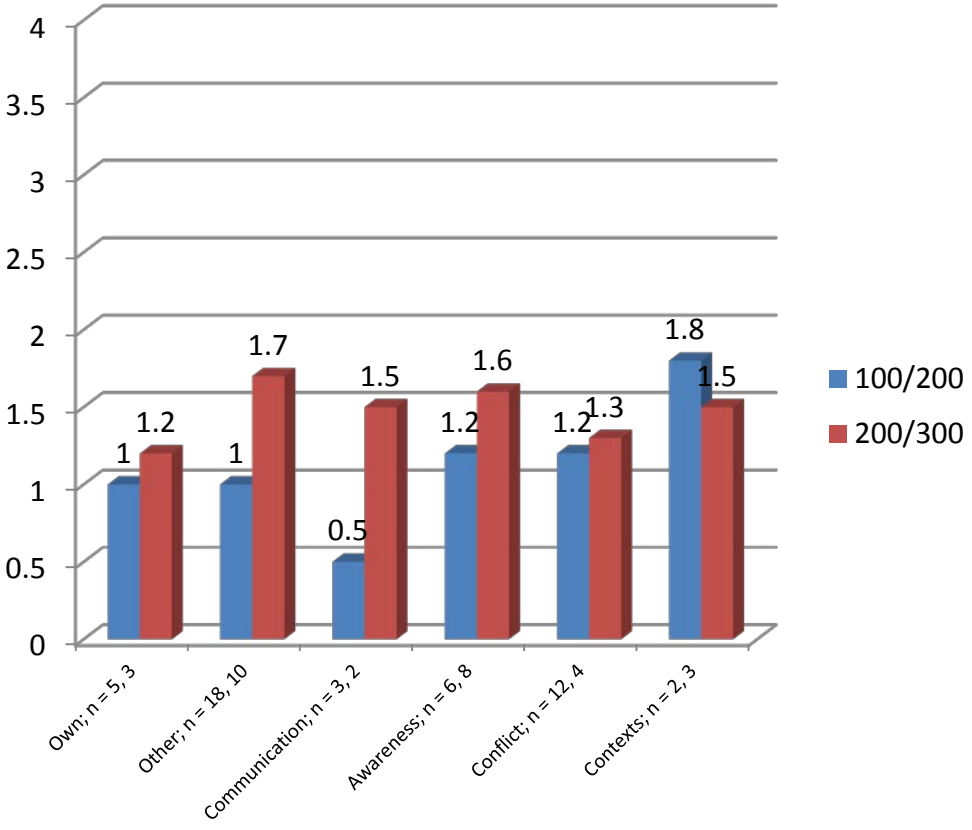
Intercultural Thinking

Mean Scores on a scale of 0 – 4, with 4 being the highest possible score; mean differences were not significant for class rank; for course level, they were significant only for “Other Cultures”

Overall Analysis



Analysis by Course Level

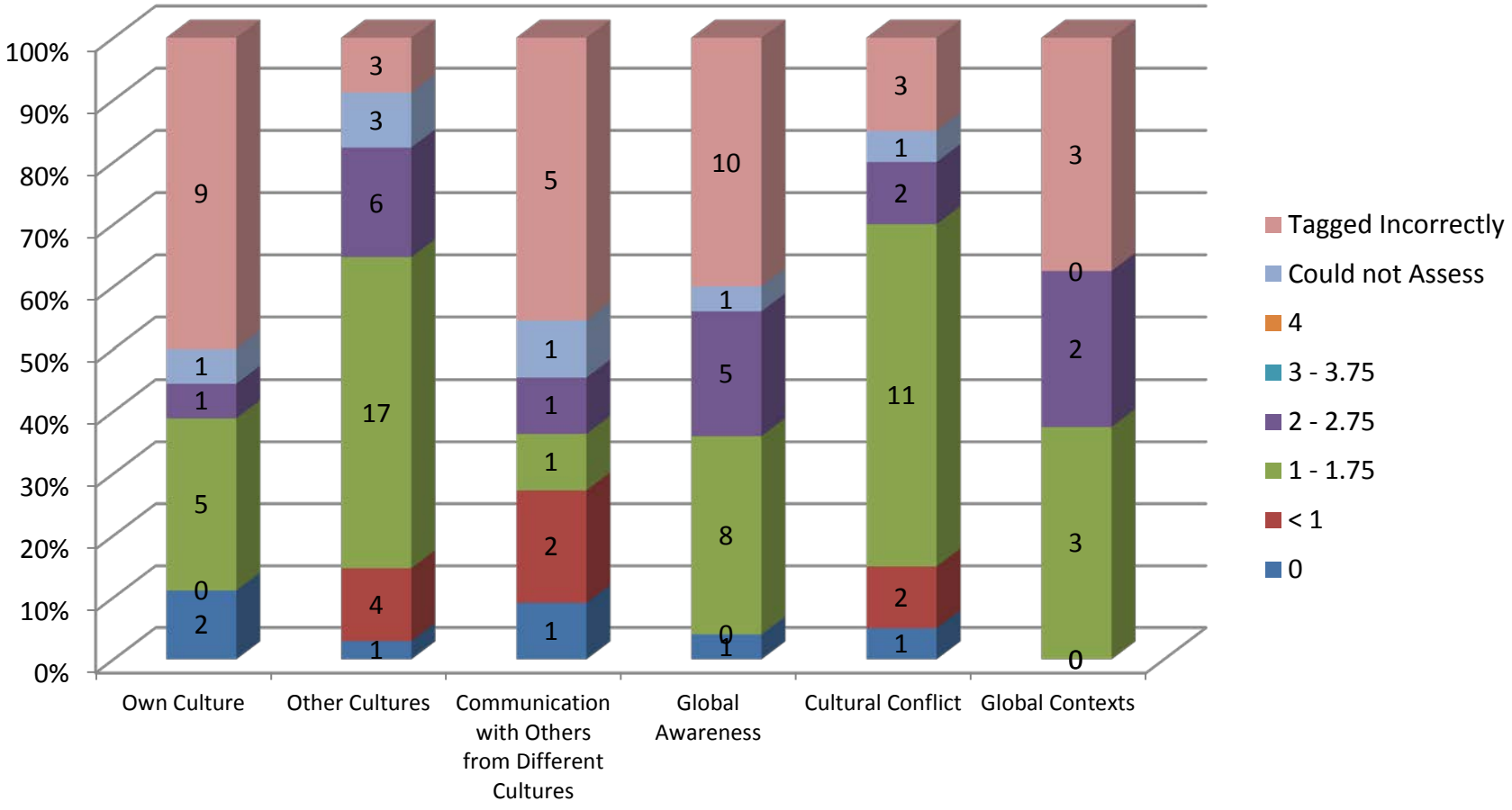


Intercultural Thinking

Number of artifacts scoring at each performance level

Trait/ Performance Level	Own Culture	Other Cultures	Communication with Others from Different Cultures	Global Awareness	Cultural Conflict	Global Contexts	Total
0	2 (11%)	1 (3%)	1 (9%)	1 (4%)	1 (5%)	0	6 (5%)
> 0, but < 1	0	4 (12%)	2 (18%)	0	2 (10%)	0	8 (7%)
1 – 1.75	5 (28%)	17 (50%)	1 (9%)	8 (32%)	11 (55%)	3 (38%)	45 (39%)
2 – 2.75	1 (6%)	6 (18%)	1 (9%)	5 (20%)	2 (10%)	2 (25%)	17 (15%)
3 – 3.75	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0
Unable to Assess	1 (6%)	3 (9%)	1 (9%)	1 (4%)	1 (5%)	0	7 (6%)
Tagged Incorrectly	9 (50%)	3 (9%)	5 (45%)	10 (40%)	3 (15%)	3 (38%)	33 (28%)
Totals	18 (100%)	34 (100%)	11 (100%)	25 (100%)	20 (100%)	8 (100%)	116 (100%)

Intercultural Thinking



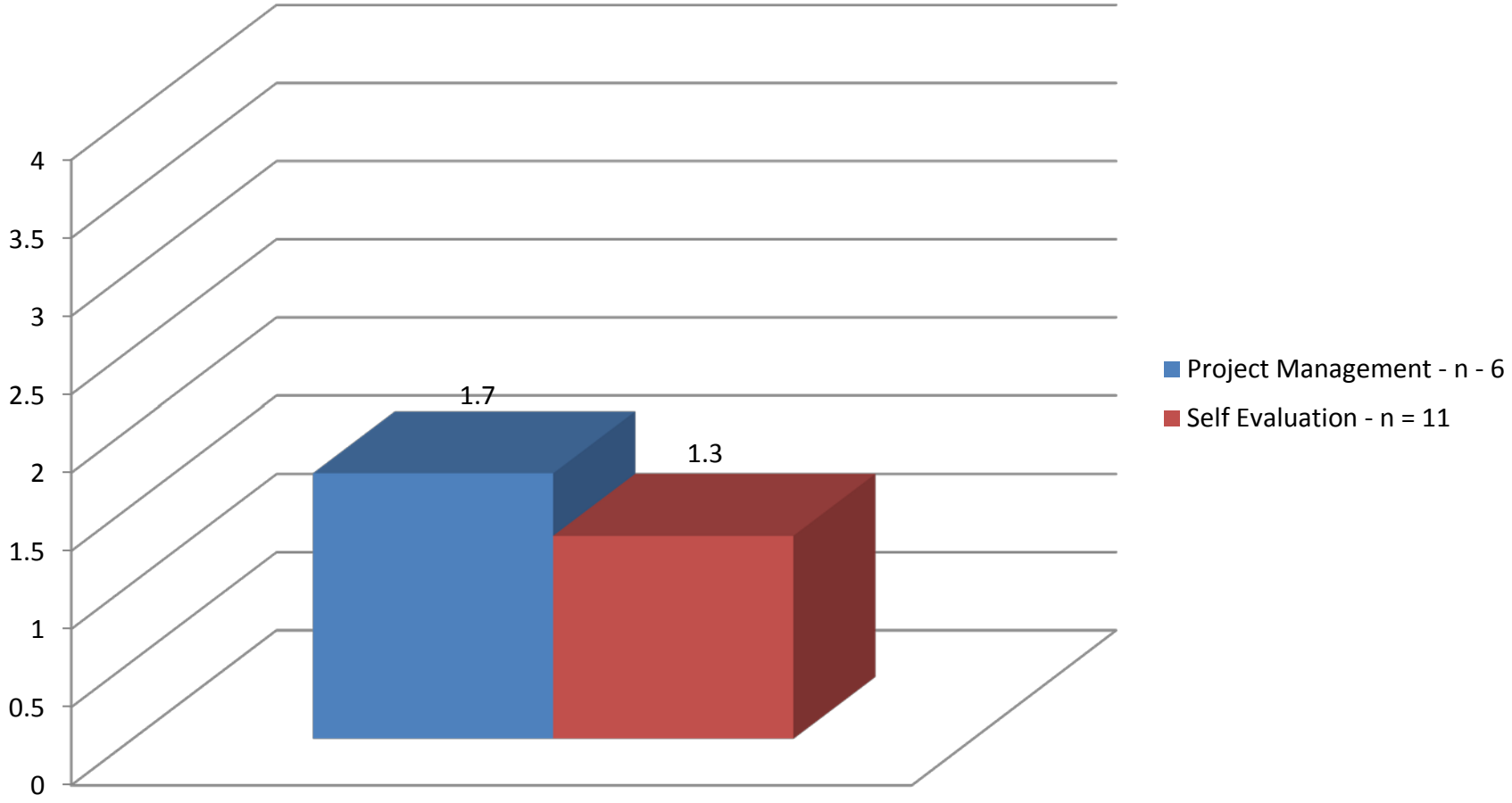
Intercultural Thinking

Inter-Rater Agreement Results

Trait/ Performance Level	Own Culture Kappa = .122	Other Cultures Kappa = .074	Communication with Others from Different Cultures Kappa = .305	Global Awareness Kappa = .175	Cultural Conflict Kappa = -.115	Global Contexts (Experimental) Kappa = -.037	Total Kappa = .099
Agree on Rubric Score	3 (17%)	9 (26%)	2 (18%)	6 (24%)	3 (15%)	1 (13%)	24 (21%)
Difference = 1 point or less	2 (11%)	6 (18%)	3 (27%)	4 (16%)	7 (35%)	2 (25%)	24 (21%)
Difference = 1.5 to 2 points	0	4 (12%)	0	0	2 (10%)	0	6 (5%)
Difference = 2.5 to 3 points	0	1 (3$)	0	0	0	0	1 (1%)
Agree on Misaligned	1 (6%)	0	3 (27%)	3 (12%)	0	0	7 (6%)
Agree on Unable to Score	0	1 (3%)	0	0	0	0	1 (1%)
Score + Misaligned	5 (28%)	8 (24%)	0	6 (24%)	5 (25%)	2 (25%)	26 (22%)
Score + Unable to Score	4 (22%)	4 (12%)	2 (18%)	3 (12%)	2 (10%)	1 (13%)	16 (14%)
Misaligned + Unable to Score	3 (17%)	1 (3%)	1 (9%)	3 (12%)	1 (5%)	2 (25%)	11 (9%)
Total	18 (100%)	34 (100%)	11 (100%)	25 (100%)	20 (100%)	8 (100%)	116 (100%)

Metacognitive Thinking

Mean Scores on a scale of 0 – 4, with 4 being the highest possible score; mean differences were not significant based on either course level or class rank. Note: There were no 300/400 level courses tagged for Metacognitive Thinking.

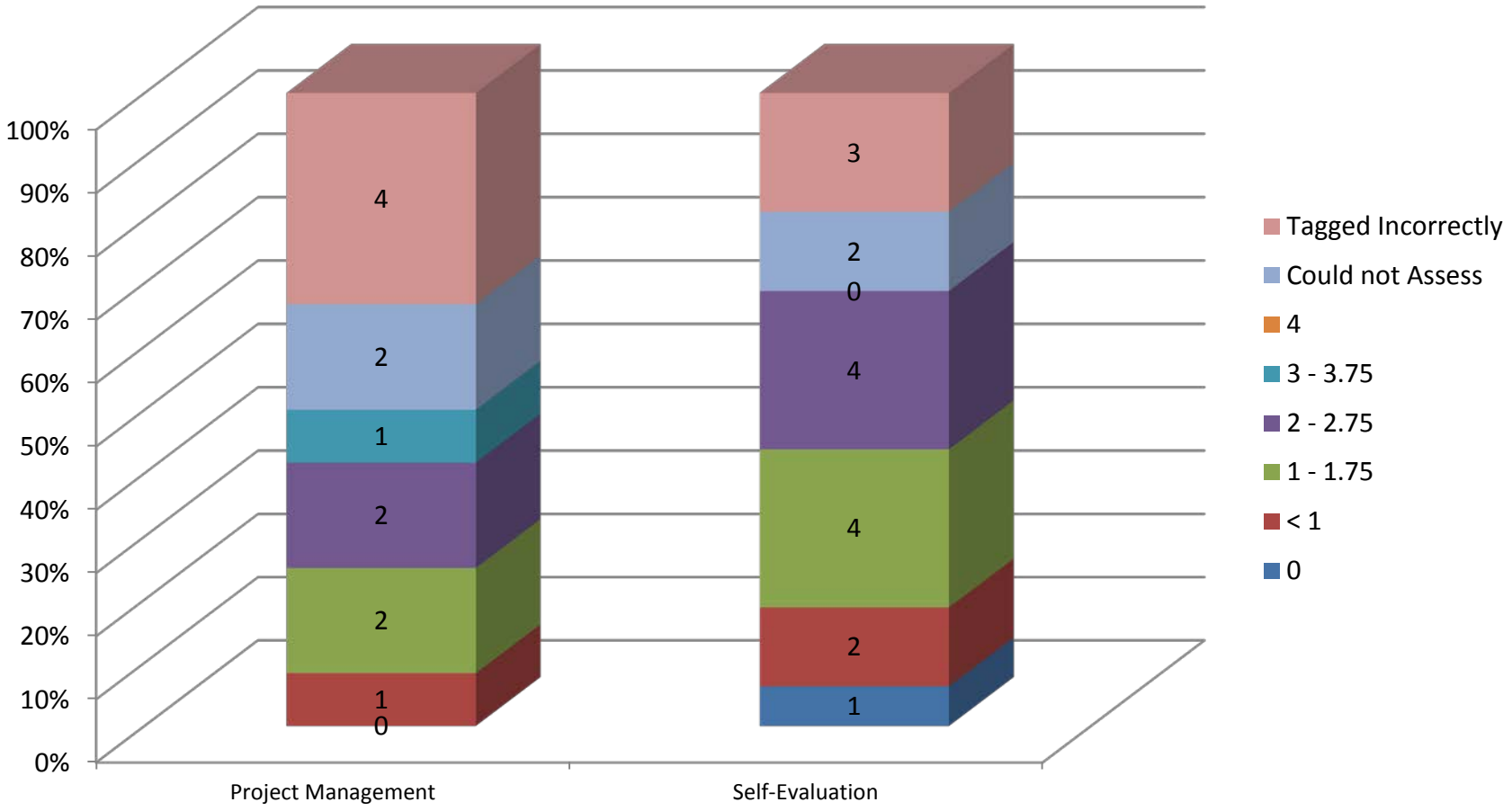


Metacognitive Thinking

Number of artifacts scoring at each performance level

Trait/ Performance Level	Project Management	Self-Evaluation	Total
0	0	1 (6%)	1 (4%)
> 0, but < 1	1 (8%)	2 (13%)	3 (11%)
1 – 1.75	2 (17%)	4 (25%)	6 (21%)
2 – 2.75	2 (17%)	4 (25%)	6 (21%)
3 – 3.75	1 (8%)	0	1 (4%)
4	0	0	0
Unable to Assess	2 (17%)	2 (13%)	4 (14%)
Tagged Incorrectly	4 (33%)	3 (19%)	7 (25%)
Totals	12 (100%)	16 (100%)	28 (100%)

Metacognitive Thinking



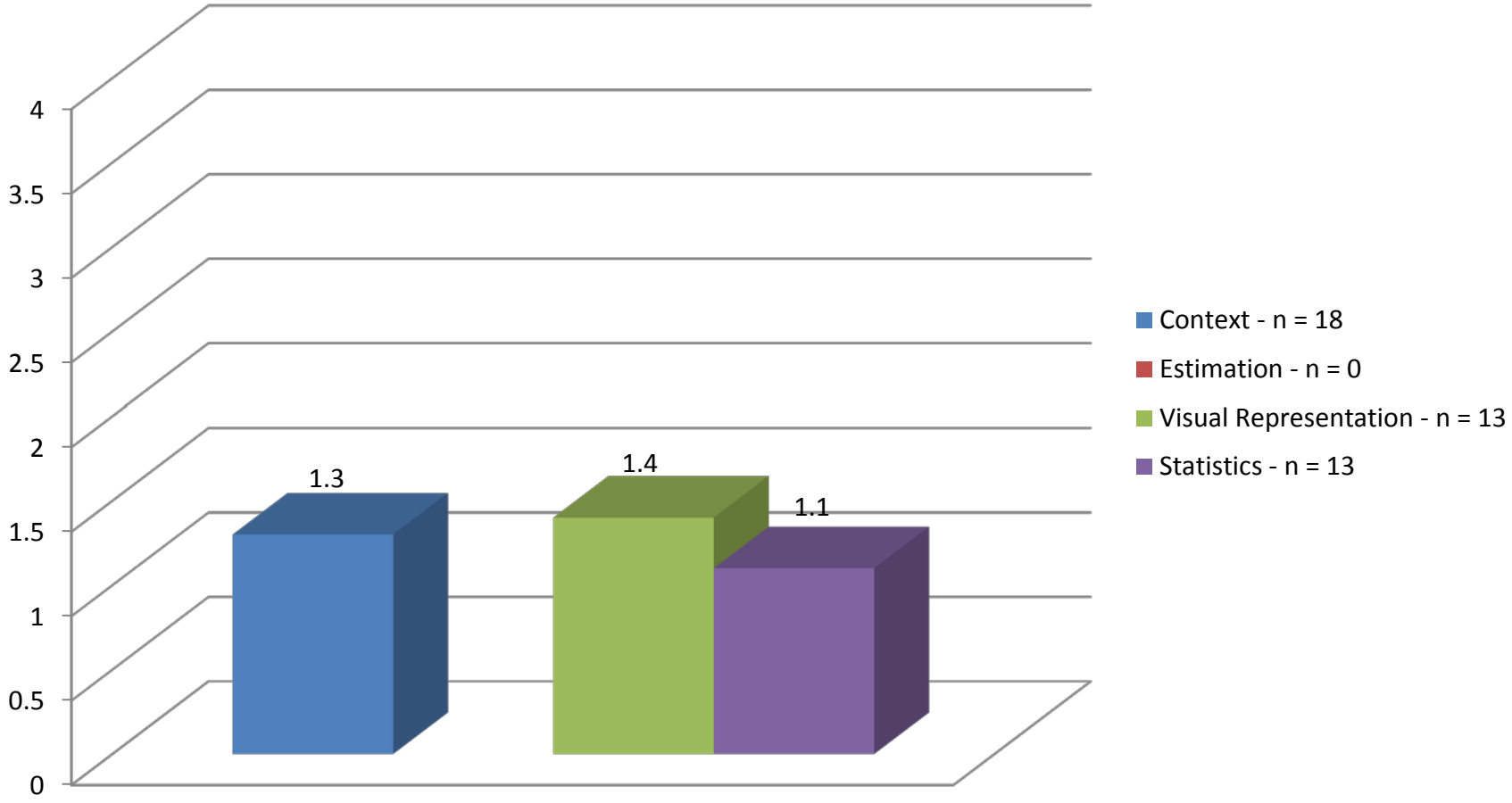
Metacognitive Thinking

Inter-Rater Agreement Results

Trait/ Performance Level	Project Management Kappa = .053	Self-Evaluation Kappa = .238	Total Kappa = .162
Agree on Rubric Score	2 (17%)	5 (31%)	7 (25%)
Difference = 1 point or less	1 (8%)	4 (25%)	5 (18%)
Difference = 1.5 to 2 points	0	0	0
Difference = 2.5 to 3 points	0	0	0
Agree on Misaligned	0	1 (6%)	1 (4%)
Agree on Unable to Score	1 (8%)	0	1 (4%)
Score + Misaligned	7 (58%)	4 (25%)	11 (39%)
Score + Unable to Score	0	0	0
Misaligned + Unable to Score	1 (8%)	2 (13%)	3 (11%)
Total	12 (100%)	16 (100%)	28 (100%)

Quantitative Thinking

Mean Scores on a scale of 0 – 4, with 4 being the highest possible score; mean differences were not significant based on either course level or class rank. Note: There were no artifacts from 300/400 level courses tagged to Quantitative Thinking

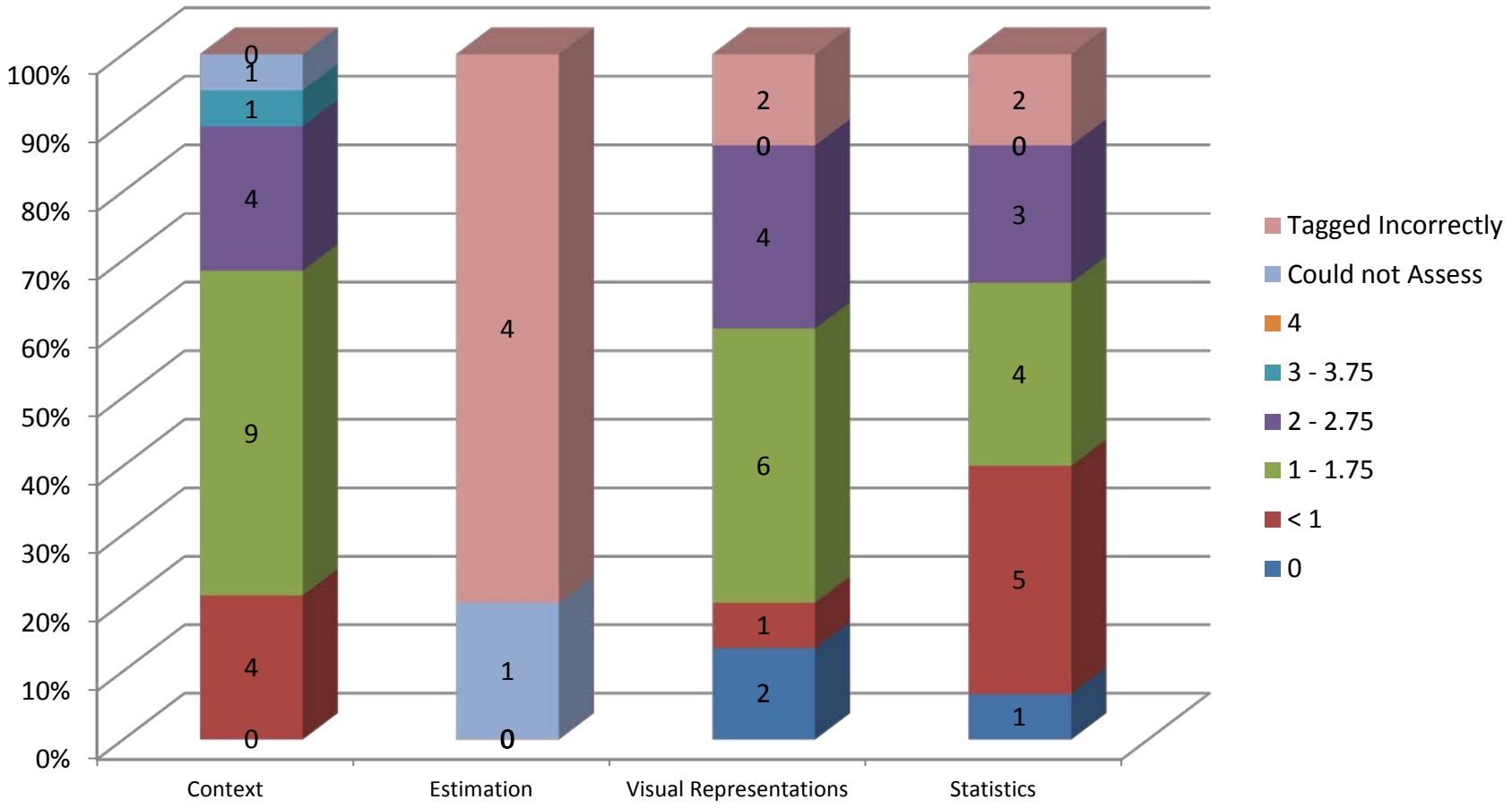


Quantitative Thinking

Number of artifacts scoring at each performance level

Trait/ Performance Level	Context	Estimation	Visual Representations	Statistics	Total
0	0	0	2 (13%)	1 (7%)	3 (6%)
> 0, but < 1	4 (21%)	0	1 (7%)	5 (33%)	10 (19%)
1 – 1.75	9 (47%)	0	6 (40%)	4 (27%)	19 (35%)
2 – 2.75	4 (21%)	0	4 (27%)	3 (20%)	11 (20%)
3 – 3.75	1 (5%)	0	0	0	1 (2%)
4	0	0	0	0	0
Unable to Assess	1 (5%)	1 (20%)	0	0	2 (4%)
Tagged Incorrectly	0	4 (80%)	2 (13%)	2 (13%)	8 (15%)
Totals	19 (100%)	5 (100%)	15 (100%)	15 (100%)	54 (100%)

Quantitative Thinking



Quantitative Thinking

Inter-Rater Agreement Results

Trait/ Performance Level	Context Kappa = .116	Estimation Kappa = .583	Visual Representations Kappa = .112	Statistics Kappa = .176	Total Kappa = .265
Agree on Rubric Score	8 (42%)	0	3 (20%)	3 (20%)	14 (26%)
Difference = 1 point or less	5 (26%)	0	7 (47%)	8 (53%)	20 (37%)
Difference = 1.5 to 2 points	4 (21%)	0	1 (7%)	2 (13%)	7 (13%)
Difference = 2.5 to 3 points	1 (5%)	0	0	0	1 (2%)
Agree on Misaligned	0	3 (60%)	2 (13%)	2 (13%)	7 (13%)
Agree on Unable to Score	1 (5%)	1 (20%)	0	0	2 (4%)
Score + Misaligned	0	1 (20%)	2 (13%)	0	3 (6%)
Score + Unable to Score	0	0	0	0	0
Misaligned + Unable to Score	0	0	0	0	0
Total	19 (100%)	5 (100%)	15 (100%)	15 (100%)	54 (100%)

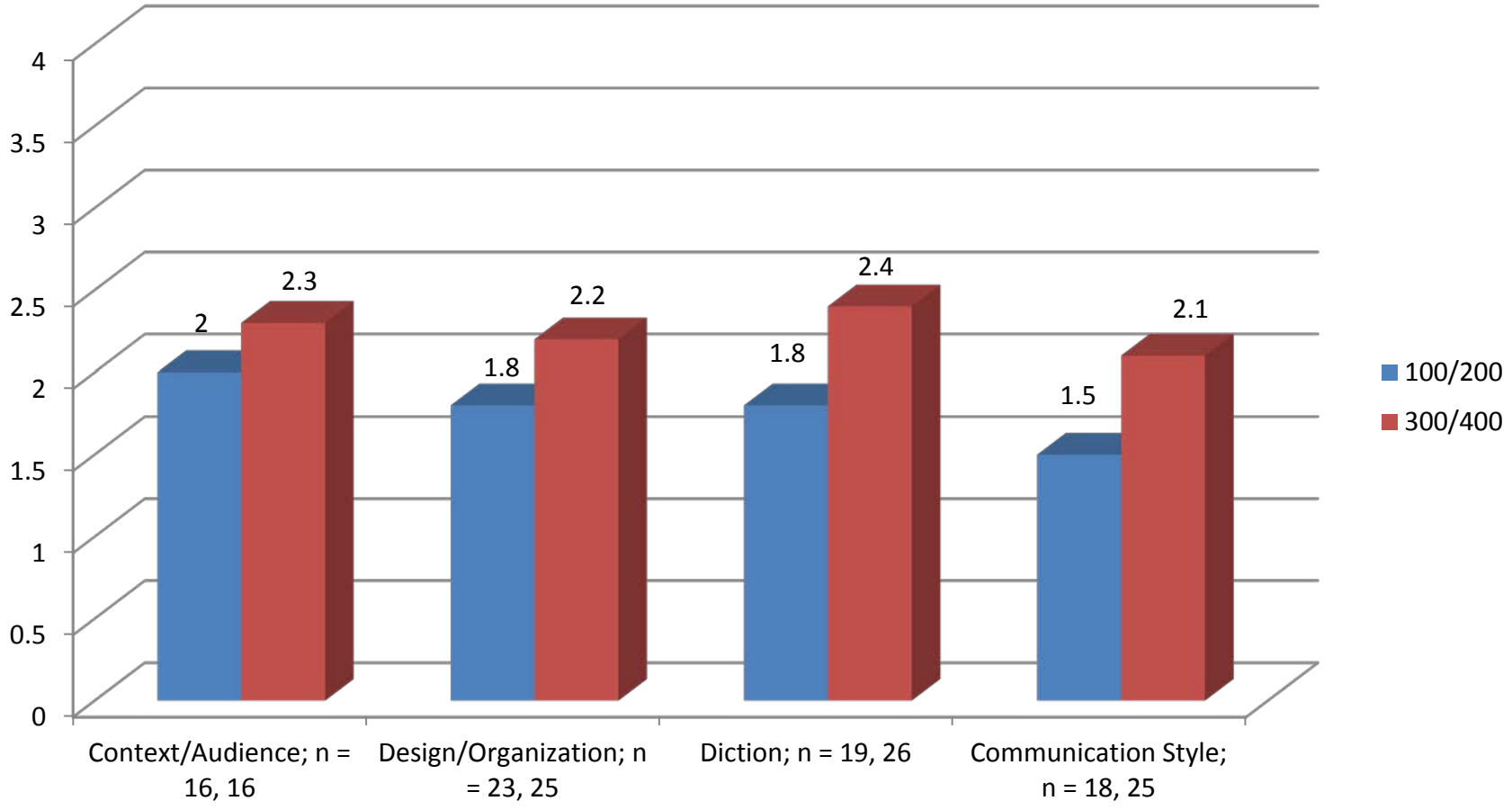
Course Type Analysis

Writing Intensive

Mean comparison by course level

(100/200 compared to 300/400)

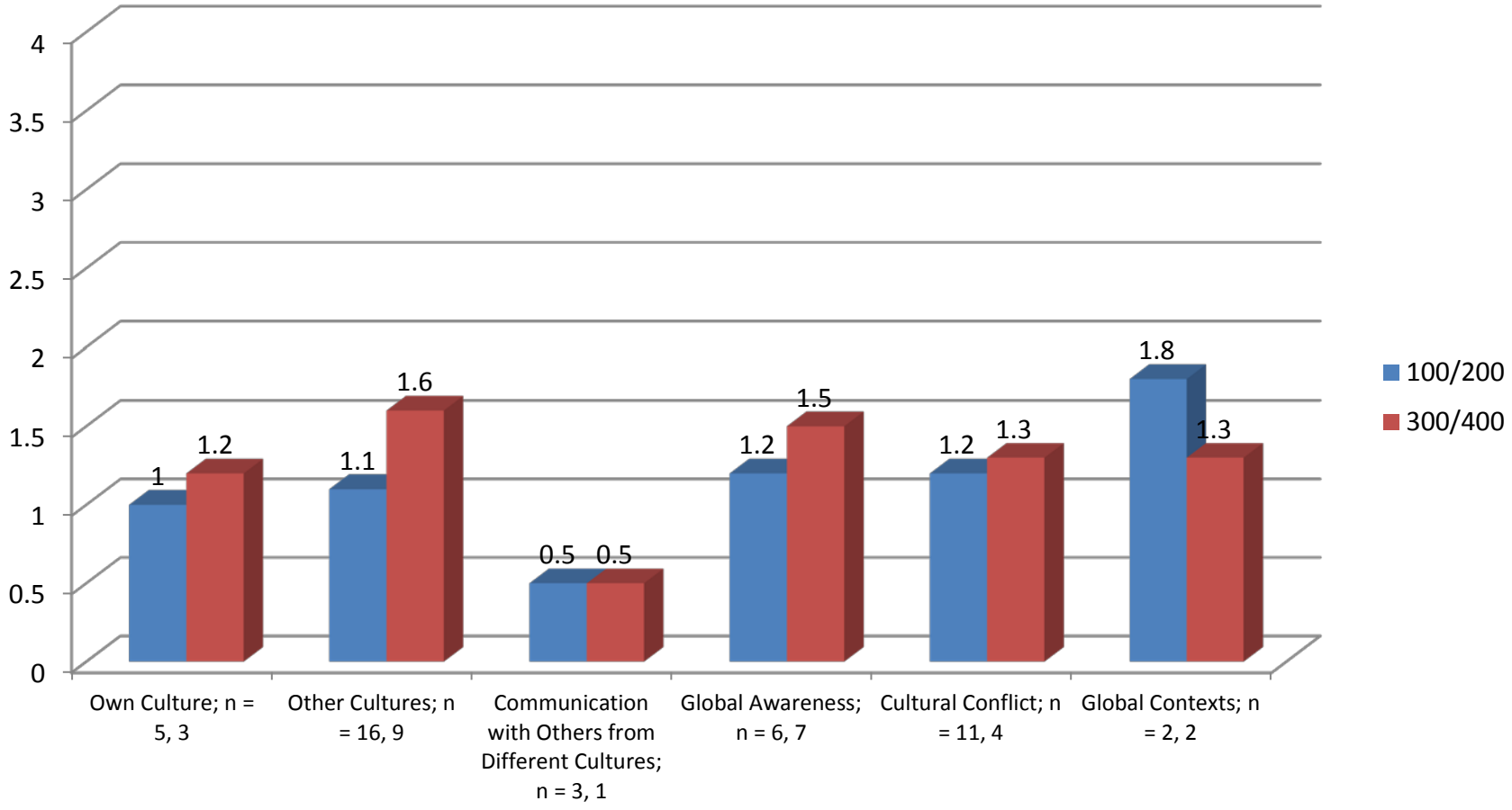
Means for Diction and Communication Style were significantly different based on course level.



Multicultural/International Courses

Mean comparison by course level; (100/200 compared to 300/400)

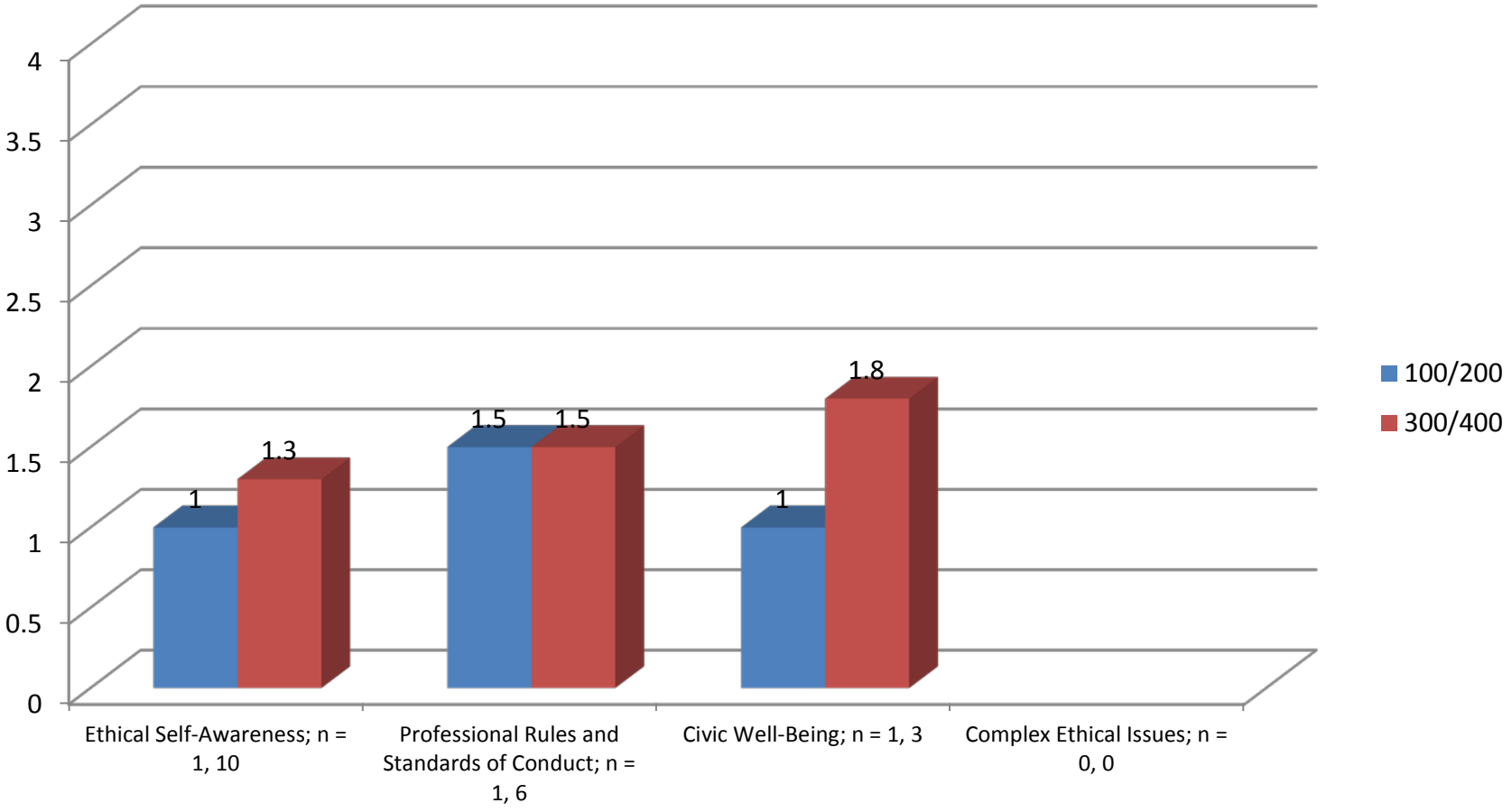
Mean differences were not significant based on course level.



Service Learning Courses

Mean comparison by course level; (100/200 compared to 300/400)

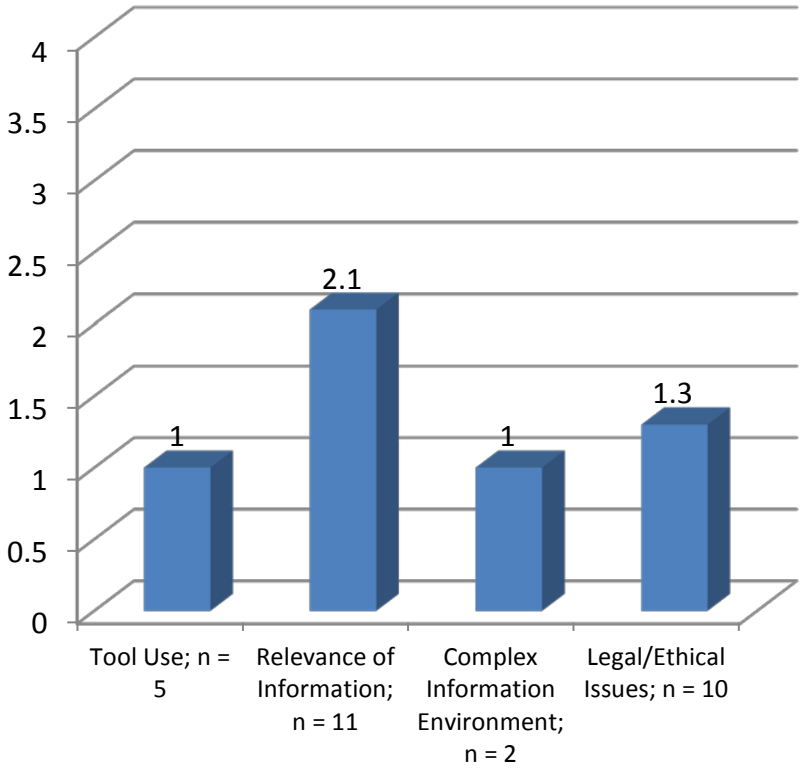
Mean differences were not significant based on course level.



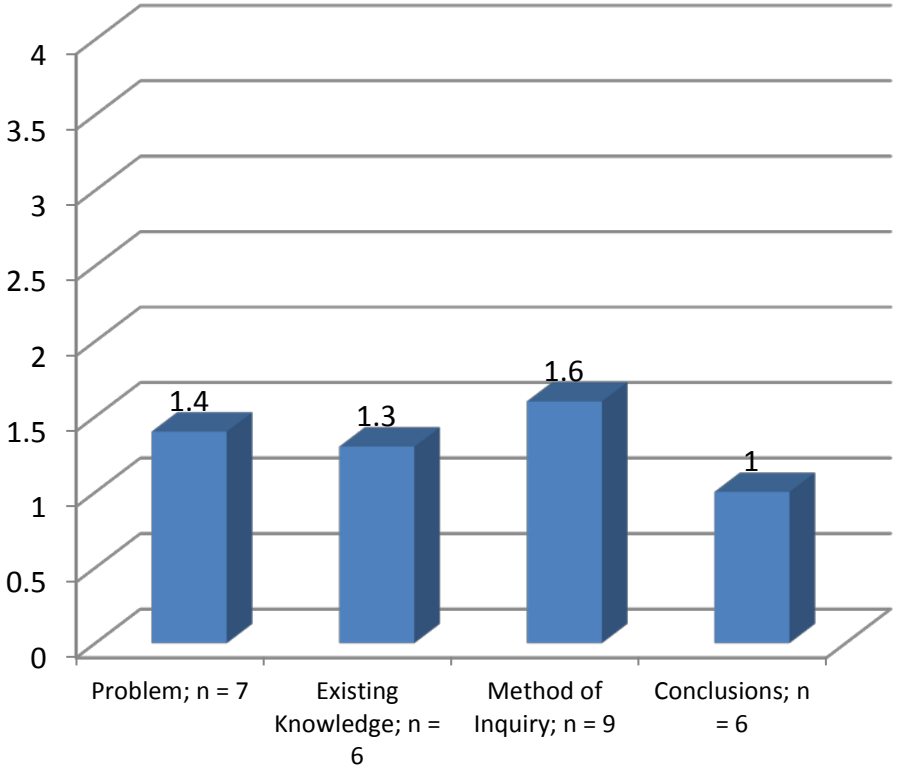
First Year Seminar

Mean comparison by outcome

Information Literacy



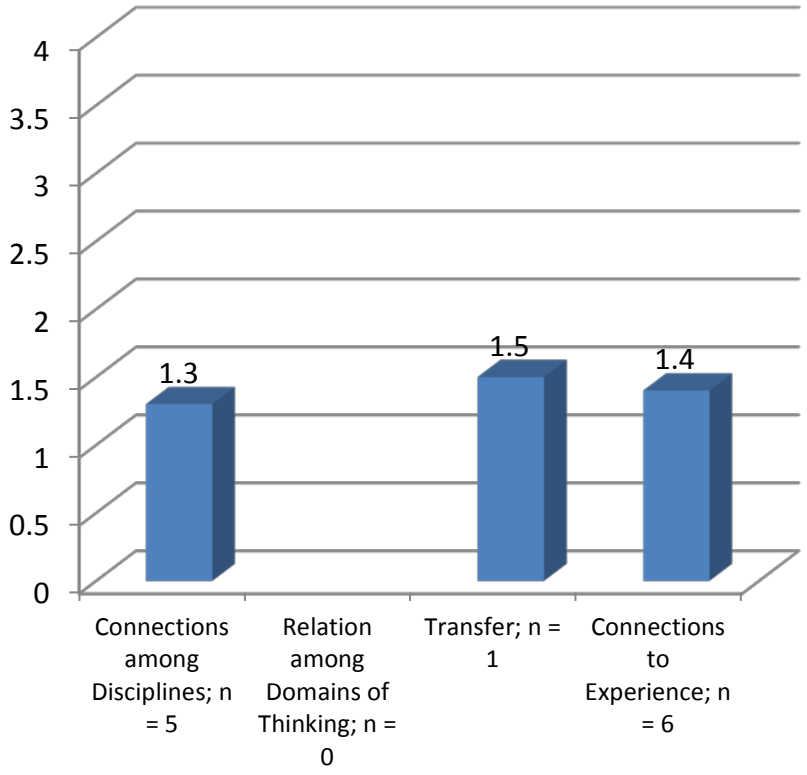
Inquiry-Based Thinking



First Year Seminar

Mean comparison by outcome

Integrative Thinking



Metacognitive Thinking

First Year Seminar

Mean comparison by outcome