Analysis of Artifacts from Marshall's General Education Assessment Repository

Academic Year 2015 – 2016

Summer Assessment Workgroup Members: Marie Archambault, Cam Brammer, Kim DeTardo-Bora, Robert Ellison, Victor Fet, Marty Laubach, Joan St. Germain, Anita Walz, Mary Welch, Mary Beth Reynolds (Office of Assessment), and Tim Melvin (Office of Assessment)

Executive Summary

Background

Recommendations from the 2015 Assessment Workgroup (with current status in red)

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 each trait and then describing four levels of increasingly sophisticated performance. 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 traits of the *Intercultural Thinking* rubric.

   (The Summer Assessment Workgroup revised the three rubrics (as drafts) used to assess this year's outcomes, *Information Literacy*, *Integrative Thinking,* and *Metacognitive Thinking,* using the format described above. Additionally, the Summer Workgroup suggested changing the *Information Literacy* outcome from "Students will revise their search strategies to find appropriate research tools, integrate relevant information from reliable sources, question and evaluate the complexity of the information environment, and use information in an ethical manner" to "Students will employ appropriate research tools, integrate relevant information from reliable sources, question and

evaluate information and its sources, and use information in an ethical manner." During academic year 2015-2016, we recommend soliciting feedback from the University Assessment Committee, the General Education Council and, through them, from Marshall University's faculty. Our goal is to shepherd these changes to the Information Literacy outcome through the appropriate committee structure at Marshall. Work will continue on revisions of rubrics for the other six outcomes.

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. (A committee has been formed to work on the *Intercultural Thinking rubric,* but the revisions are not complete).
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. (This recommendation has not been implemented).

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. (This has not been done).
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. (This recommendation was implemented at the beginning of fall 2015).
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. (This has not been done).
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. an entire course syllabus, GEAR upload instructions, etc. (This continues to be a part of GEAR training and it is not possible to create an assignment without uploading something in the assignment instruction section).

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 from summer 2015 suggested a need to redesign rubrics to be continuous, rather than categorical, in nature.

Finally, assessment data from 2013-2014 and 2014-2015 showed that Marshall's students improved their writing skills as they moved through the curriculum and, specifically, as they passed from 100/200 level writing intensive courses to 300/400 level writing intensive courses.

Procedures for 2016 Assessment

General Procedures

In summer 2016 we evaluated student artifacts produced in response to course assignments aligned to *Information Literacy*, *Integrative Thinking*, and *Metacognitive Thinking* that were uploaded to GEAR during academic year 2015-2016. Students enrolled in First Year Seminar (FYS), and courses with Critical Thinking (CT) and Writing Intensive (WI) designations uploaded artifacts aligned to these outcomes. It was possible for a single assignment to align to any number of outcomes and traits. However, we asked instructors to specify the primary outcome

to which the assignment aligned and all artifacts chosen randomly for assessment had indicated that the outcome in question was the primary outcome for the assignment/artifact. Although we have asked instructors teaching courses that have only multicultural (MC) or international (INT) designations to upload artifacts whose primary learning outcome is *Intercultural Thinking*, a small number of MC courses specified one of this cycle's outcomes as primary and were drawn for this sample.

In May 2016 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, as noted above, were revised prior to scoring, are included in the supporting documentation. Our sample initially consisted of 324 artifacts, 108 per outcome. However, during scoring we discovered that one artifact, *aligned* to *Integrative Thinking,* had been uploaded twice (once in PDF and once in Word format). The second was eliminated, leaving 107 artifacts aligned to *Integrative Thinking*. This resulted in a total of 323 unique artifacts in this sample. Each artifact was read by two independent reviewers. 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/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.

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

General Information about the Sample

One hundred seventy-one (171; 53%) of the artifacts in our sample were drawn from courses at the 100/200 level, with the remaining 152 (47%) drawn from courses at the 300/400 level. Thirty-seven (37%) percent of the students in the sample were freshmen, 15% were sophomores, 14% were juniors, and 33% were seniors.

Results and Analysis

One challenge in reporting results of GEAR assessment is that, although we assessed 323 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 (which the instructor or student had designated as its *primary* outcome), most of these artifacts aligned to more than one of the outcome's traits. The total number of traits across the three outcomes was 10 (4 each for *Information Literacy* and *Integrative Thinking*, and 2 for *Metacognitive Thinking*). A perusal of our supporting documentation shows that the artifacts evaluated by the Assessment Workgroup tagged to a total of 606 traits. However, scores for only 442 (73%) of those traits were usable for calculating means. One hundred sixty-four (164) were discarded either because they were judged not to align with the traits (128; 21%) or were not able to be assessed because of student upload or other type of error (36; 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 or other error), and the resulting number of scores able to be used for the analysis of means.

Outcome	Trait	Total Traits Aligned	# Misaligned (Scores of 100)	# Not Able to be Assessed (Score of 99)	Total # Excluded from Analysis of Means	Total Usable Traits
Information Literacy	Sources	59	11	9	**20**	39
	Relevance of Information	97	19	8	**27**	**70**
	Assumptions and Biases	33	12	2	**14**	**19**
	Citation	40	4	5	**9**	**31**
Integrative Thinking	Connections among Disciplines	91	24	2	**26**	**65**
	Relations among Domains of Thinking	32	8	2	**10**	**22**
	Transfer	32	7	0	**7**	**25**

	Connections to Experience	82	16	3	**19**	**63**
Metacognitive Thinking	Project Management	40	13	2	**15**	**25**
	Self-Evaluation	100	14	3	**17**	**83**
Totals		**606**	**125**	**36**	**164**	**442**

Results for *Information Literacy* showed that the mean score for the trait *citation* was significantly higher for students in 100/200 level courses than for those in 300/400 level courses. However, we had usable scores for only 9 students from 300/400 level courses as compared to usable scores for 22 students in 100/200 level courses. *Information Literacy* means did not differ significantly based on course level for any other trait; trait means also did not differ significantly based on class rank (freshman/sophomore compared to junior/senior). Students enrolled in courses at the 300/400 levels had significantly higher means for *Integrative Thinking*: connections among disciplines than did students enrolled in 100/200 level courses. Course level mean differences were not significant for any other trait of *Integrative Thinking* (note: there were no 300/400 level artifacts tagged to *domains* and only one tagged to *transfer*). Juniors and seniors also scored significantly higher than freshmen and sophomores in *Integrative Thinking*: connections among disciplines. For *Metacognitive* Thinking, mean differences did not differ based on course level, but freshmen and sophomores outperformed juniors and seniors on *Metacognitive Thinking*: self-evaluation.

Overall results showed mean performance for traits to range from 1.44 (*Integrative Thinking*: relations among domains of thinking) to 2.45 (*Information Literacy:* relevance of information). Mean performance for artifacts uploaded from freshmen and sophomores ranged from 1.32 (*Integrative Thinking*: connections among disciplines) to 2.4 (*Information Literacy:* relevance of information) and for juniors and seniors from 1.58 (*Integrative Thinking:* transfer) to 2.52 (*Information Literacy:* relevance of information). The overall strength for students in this sample was *Information Literacy*: relevance of information, while the overall weakness was *Integrative Thinking*.

Results for Course Type

Analyzing results by course type posed several challenges. First, the only course type that is unique, i.e. can have only one course type attribute, is First Year Seminar (FYS). Courses can have the other attributes (Critical Thinking [CT], Multicultural [MC], International [INT], Writing Intensive [WI], and Service Learning [SL]) in combination (and many do). So, when analyzing results by course type, we included all courses with the attribute we wanted to assess; this resulted in some courses being included in the analysis for more than one course type. Because the number of courses with MC and INT attributes in our sample was small, we did not conduct analyses of these course types. We also note that MC and INT courses have been asked to create assignments and ask students to upload artifacts whose primary alignment is to *Intercultural Thinking*, an outcome we did not assess this cycle. SL courses (which align to *Ethical and Civic Thinking*) were not included in our sample this year.

Critical Thinking (CT) Courses

CT courses in the assessment sample included those that aligned to each of the outcomes assessed: *Information Literacy, Intercultural Thinking,* and *Metacognitive Thinking.* All CT courses are at the 100/200 level. Results are below:

Information Literacy			**Integrative Thinking**			**Metacognitive Thinking**		
Trait	Number	Mean Score	**Trait**	Number	Mean Score	**Trait**	Number	Mean Score
Sources	6	2.67	**Connections among Disciplines**	16	1.36	**Project Management**	7	1.79
Relevance of Information	6	2.71	**Relations among Domains of Thinking**	14	1.55	**Self-Evaluation**	6	1.42
Assumptions and Biases	4	2.63	**Transfer**	15	1.83			
Citation	4	3.50	**Connections to Experience**	27	1.72			

These results must be interpreted with caution, as */n/*s are small. However, it appears that students in CT courses performed better on *Information Literacy* than on *Integrative and Metacognitive Thinking*. We note that all CT courses are at the 100 and 200 levels.

First Year Seminar (FYS) Courses

FYS courses in the assessment sample included those that aligned to each of the outcomes assessed: *Information Literacy, Intercultural Thinking,* and *Metacognitive Thinking.* FYS is, by definition, at the 100 level. Results are below:

Information Literacy			**Integrative Thinking**			**Metacognitive Thinking**		
Trait	Number	Mean Score	**Trait**	Number	Mean Score	**Trait**	Number	Mean Score
Sources	22	2.16	**Connections among Disciplines**	8	1.31	**Project Management**	7	2.36
Relevance of Information	20	2.35	**Relations among Domains of Thinking**	8	1.25	**Self-Evaluation**	36	2.29
Assumptions and Biases	4	2.25	**Transfer**	9	1.44			
Citation	18	2.03	**Connections to Experience**	12	1.33			

Most artifacts from FYS courses included in our sample aligned to *Metacognitive Thinking*: self-evaluation and to three of the four traits of *Information Literacy*. Strongest performance was in *Metacognitive Thinking* and *Information Literacy*.

Writing Intensive (WI) Courses

WI courses in the assessment sample aligned to all outcomes assessed: *Information Literacy, Intercultural Thinking,* and *Metacognitive Thinking.* Results are given below by course level for *Information Literacy*:

Trait	Course Level	Number	Mean Score
Sources	100/200	0	-----
	300/400	11	2.45
Relevance of Information	100/200	2	2.50
	300/400	44	2.46
Assumptions and Biases	100/200	0	-----
	300/400	11	1.86
Citation	100/200	0	-----
	300/400	9	1.22

All but two artifacts from WI courses aligned to *Information Literacy* in our sample came from courses at the 300/400 levels. With the exception of *Information Literacy*: relevance of information, */n/*s were low. Performance was stronger for "sources" and "relevance of information" than for the other two traits. However, performance in these 300/400 level courses does not appear to be significantly better than performance of students from 100/200 level FYS and CT courses.

WI results are given below by course level for *Integrative Thinking*:

Trait	Course Level	Number	Mean Score
Connections among Disciplines	100/200	6	1.79
	300/400	39	2.35
Relation among Domains of Thinking	100/200	8	1.59
	300/400	0	-----
Transfer	100/200	10	1.70
	300/400	0	-----
Connections to Experience	100/200	18	2.03
	300/400	19	2.00

Our sample did not contain any artifacts from WI courses that aligned to "relation among domains of thinking" or "transfer" at the 300/400 levels. On the other hand, a relatively large number (39) artifacts from 300/400 level WI courses aligned to "connections among disciplines" and 19 aligned to "connections to experience." The number of artifacts from WI courses at the 100/200 level was relatively small for each trait, with the largest being 18 that aligned to "connections to experience." There was essentially no difference in the mean scores for "connections to experience" based on course level. Students in 300/400 level courses did perform better than those in 100/200 level courses in "connections among disciplines," but the latter had a relatively small */n/* (6).

WI results are given below by course level for *Metacognitive Thinking*:

Trait	Course Level	Number	Mean Score
Project Management	100/200	5	1.90
	300/400	7	1.96
Self-Evaluation	100/200	5	2.20
	300/400	37	1.77

Although it appears that the mean score for WI courses from 100/200 level courses for "self-evaluation" was higher than that for courses from 300/400 level courses, only five artifacts from the former aligned, while there were 37 from the latter.

Misalignments

It is difficult to discern if misalignments occurred more often based on course type due to the differing */n/*s in each case. We refer the reader to the supporting documentation for additional detail.

Conclusion

Strongest performance among this group of students was for *Information Literacy*: relevance of information, while the weakest performances were scattered among the traits of *Integrative Thinking*. Of concern remains the number of assignments (and hence, student artifacts) that the Assessment Workgroup judged to not align to the *Outcomes*: traits to which they were tagged. Results for course type mirrored those of the overall analysis.

Recommendations from the 2016 Assessment Workgroup

Recommendations Concerning the General Process of Assignment Creation and Accurate Alignment to University Outcomes

We first note that, beginning with academic year 2016-2017, faculty will be asked to develop assignments that align to the outcomes as stated in Marshall University's Baccalaureate Degree Profile. In other words, we will abandon the former practice of asking instructors to indicate which performance level on the rubric they used when creating assignments. The reason for this decision is that the former rubric level descriptions were essentially different outcome statements. The Assessment Workgroup began the process of redeveloping the rubrics so that performance levels now specify *how well* each student demonstrates mastery of the university's *outcomes*, not whether or not the student achieves progressively more complex outcomes. Outlined below are concerns and recommendations from the Assessment Workgroup.

1. A major concern among the members of the Assessment Workgroup was the large number of assignments/artifacts that the Workgroup judged to be misaligned to the outcomes/traits to which they were tagged. Several recommendations were made to improve this situation. These included:
   - Work with faculty to create assignments that align with the university outcomes addressed in Critical Thinking (CT), First Year Seminar (FYS) and Writing Intensive (WI) courses during the faculty development sessions that prepare instructors to teach these courses, as follows:
     - Center for Teaching and Learning for CT courses
     - Center for Teaching and Learning in conjunction with the Director of FYS for FYS courses
     - Center for Teaching and Learning in conjunction with the Director of Writing across the Curriculum for WI courses
   - Identify model assignments from those already uploaded to GEAR and create a repository of these assignments. This repository can function as both a resource for faculty developing new assignments and a teaching tool during faculty preparation to teach the aforementioned course types.
   - Ask the Center for Teaching and Learning to consider offering faculty development sessions focusing on alignment of assignments to Marshall University's outcomes.
   - Ask the General Education Council to require that all CT, INT, and MC courses include the assignment that will be used for general education assessment (i.e. GEAR upload) in course application and renewal materials and to explain explicitly how this assignment addresses the university outcome/trait to which it is aligned.
   - Ask that each assignment created with student artifacts uploaded into GEAR include an explicit explanation from the instructor as to how the assignment addresses the university outcome/trait(s) to which it is aligned.
   - Members of the Assessment Workgroup will submit a proposal for a session to be presented at the August 2016 iPED: Inquiring Pedagogies Conference. The purpose of this session will be to overview the general education assessment process and findings, and to discuss with faculty the importance of careful assignment alignment to university outcomes.

2. To reduce the number of artifacts from the assessment pool that must be discarded due to the Assessment Workgroup's judgment that the assignment itself does not align to the university outcome to which it was tagged, the Assessment Workgroup recommended that, in future, it evaluate each assignment for accuracy of alignment before the sample of artifacts is selected.

Recommendations regarding Marshall's Transition from GEAR to Blackboard Outcomes for Assessing Student Work

Marshall will begin to use Blackboard Outcomes for general education assessment during academic year 2016-2017. This will have some advantages over GEAR, but will pose some challenges as well. Advantages will include:

1. Faculty will have to create an assignment and align it to university outcomes only once if the assignment and alignment is completed in their master course shell. Unless something changes, i.e. assignments are changed or updated, once alignments are made in Blackboard, they will simply be copied the next time the course is offered.

2. Faculty will ask students to submit artifacts for the aligned assignment using the assignment module in Blackboard Learn. This will allow the faculty member to assess the artifact for course grading purposes and the student and faculty member will need to do nothing else to support university assessment. For the latter purpose, Blackboard Outcomes will make a copy of the artifact (which will not include any instructor grading or comments, i.e. it will be a *clean* copy) for later assessment.
3. As is the case with GEAR, when artifacts are randomly chosen for assessment in Blackboard Outcomes, course information will not be available to assessors.

Blackboard Outcomes also presents challenges. These include:

1. Faculty will align assignments to a university outcome and assessors will use that outcome's rubric, which will include all of the outcome's traits. Because not all assignments will align to every trait of the outcome, instructors will have to indicate in their assignment instructions (and/or explicit explanation regarding alignment) the traits to which the assignment aligns.

To help facilitate the transition from GEAR to Blackboard Outcomes, the following plans are in place.

1. Marshall's Baccalaureate Degree Profile outcomes have been entered into Blackboard.
2. Several faculty teaching FYS, Anthropology, and Sociology courses during summer 2016 will create assignments and align them to University outcomes within Blackboard. They will use the Blackboard assignment tool and the Office of Assessment will set up artifact collection through Blackboard Outcomes. The Office of Assessment will test the Blackboard Outcomes assessment process at the end of the summer.
3. Fall 2016 will be a semester set aside to prepare faculty to begin using Blackboard as an artifact repository for assessment purposes. To facilitate this process, the following steps will be taken:
    - The Office of Assessment will administer a survey to all faculty teaching FYS, CT, WI, MC, INT, and SL courses. The survey will ask a series of questions that will allow us to divide the group into three cohorts (seasoned Blackboard users who routinely use the Blackboard assignment tool, Blackboard users who have not used the assignment tool, non-Blackboard users).
    - After the survey has been completed, the Assessment Office will develop three online tutorials, one geared to each group of faculty identified above.
    - The Office of Assessment also will work with the Center for Teaching and Learning, the MU Design Center, and the Associate Vice President for Libraries and Online Learning to develop a schedule of training sessions for each cohort of faculty.
3. During spring 2017, our hope is that all faculty teaching general education courses will begin to use Blackboard for assignment creation and student artifact collection. They will have access to the online tutorials and to training sessions as they did during the fall semester.

Supporting Documentation

MARSHALL UNIVERSITY®

General Education Assessment Repository (GEAR) Artifact Assessment

Academic Year 2015 – 2016

Outcomes Assessed

Outcome	Abbreviation	Traits	Abbreviations
Information Literacy	IL	Sources	Sources
		Relevance of Information	Relevance
		Assumptions and Biases	A & B
		Citation	Citation
Integrative Thinking	IT	Connections among Disciplines	Discipline
		Relation among Domains of Thinking	Domain
		Transfer	Transfer
		Connections to Experience	Experience
Metacognitive Thinking	MT	Project Management	Project
		Self-Evaluation	Self

Course Types

Course Type	Abbreviation
Critical Thinking	CT
Multicultural	MC
International	INT
Writing Intensive	WI
Service Learning	SL (Not included in this year's assessment)
First Year Seminar	FYS

Course Types in IL, IT, and MT Primary Outcome Population

Course Type	Population *n*	Sample *n*	Percent
CT	354	23	6%
CT and INT	107	15	14%
CT and MC	52	4	8%
CT, MC, and WI	131	7	5%
CT and WI	224	16	7%
FYS	898	98	11%
MC	19	12	63%
MC and WI	13	13	100%
WI	263	135	51%
Total	**2,061**	**323**	**16%**

Course Types in IL, IT, and MT Primary Outcome Population: Each Course Counted Separately for Each Category

Course Type	Population *n*	Sample *n*	Percent
CT	868	65	7%
INT	107	15	14%
MC	215	36	17%
WI	631	171	27%
FYS	898	98	11%
Total	**2,719**	**385**	**14%**

Population/Sample Comparisons for Marshall's Course Types by Course Level (sample *n* does not = 323)

Course Type	**Course Level = 100/200**			**Course Level = 300/400**		
	Population	Sample	Percent	Population	Sample	Percent
Critical Thinking	868	65	7%	0	N/A	N/A
Multicultural	183	11	6%	32	25	78%
International	107	15	14%	0	0	N/A
Writing Intensive	418	31	7%	213	140	66%
FYS	898	98	11%	0	N/A	N/A
Total	**2,474**	**220**	**9%**	**245**	**165**	**67%**

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
Information Literacy	564	54	10%	90	54	60%
Integrative Thinking	1,006	54	5%	97	53	55%
Metacognitive Thinking	259	63	24%	45	45	100%
Total	**1,829**	**171**	**9%**	**232**	**152**	**66%**

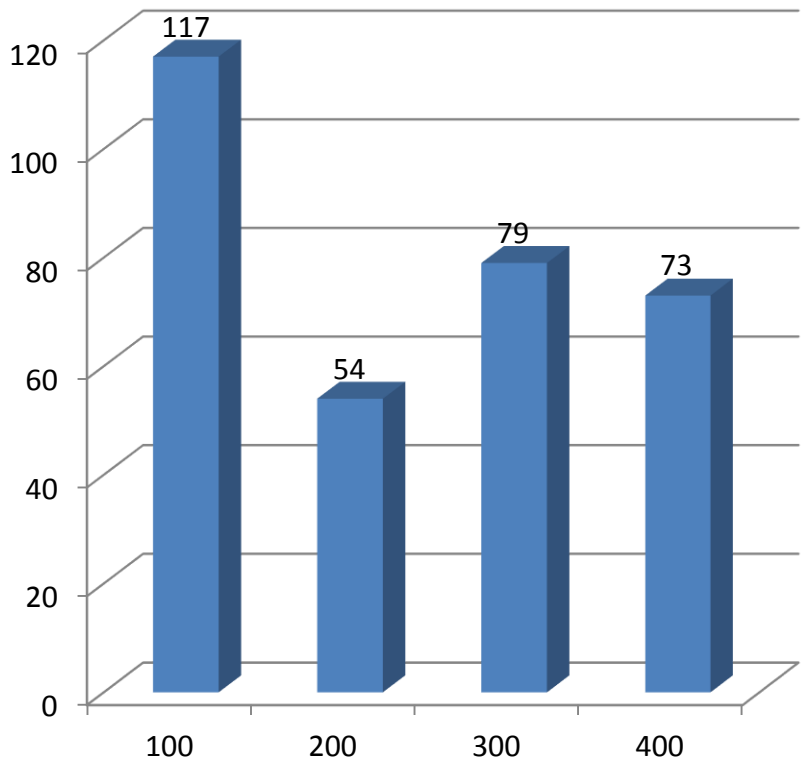
Distribution of GEAR Artifacts among the traits of Marshall's IL, IT, and MT Outcomes

Marshall Outcome	Outcome Traits (Primary)	# Uploaded Traits Tagged (Primary)	# Traits in Sample (Primary)	% of Total
Information Literacy	Sources	477	59	12%
	Relevance of Information	528	97	18%
	Assumptions and Biases	154	33	21%
	Citation	318	40	13%
	Information Literacy Total	**1,477**	**229**	**16%**
Integrative Thinking	Connections among Disciplines	836	91	11%
	Relation among Domains of Thinking	570	32	6%
	Transfer	445	32	7%
	Connections to Experience	929	82	9%
	Integrative Thinking Total	**2,780**	**237**	**9%**
Metacognitive Thinking	Project Management	139	40	29%
	Self-Evaluation	270	100	37%
	Metacognitive Thinking Total	**409**	**140**	**34%**

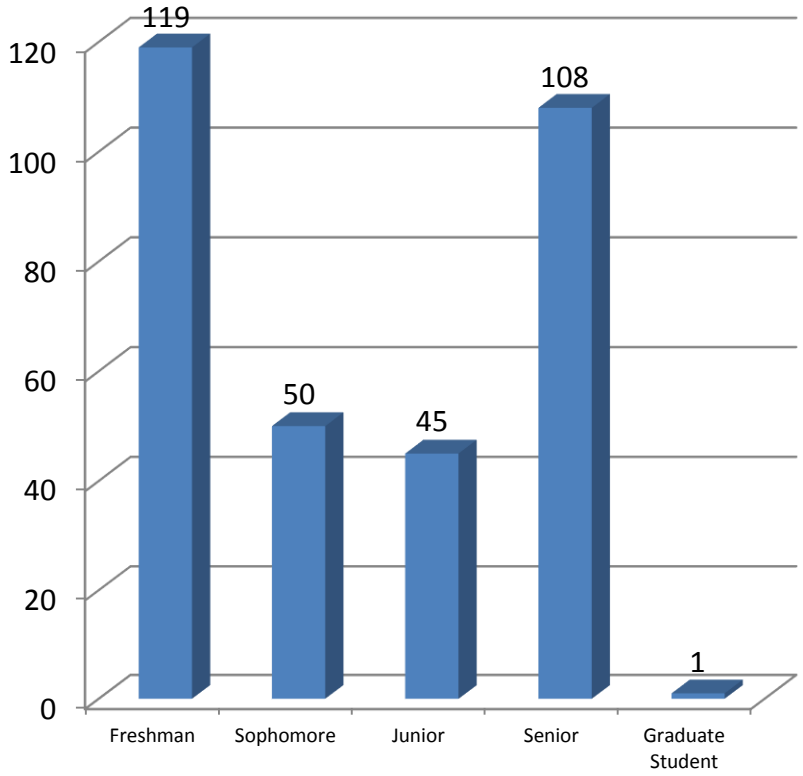
Sample Frequencies

Total # of artifacts assessed = 323

Course Level Frequencies



Class Rank Frequencies



Review Procedures

- Each artifact had two independent raters and usable scores on the 0 – 4 scale 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.

Review Procedures

- We also allowed reviewers to assign a score of 100 when they judged the assignment to not be correctly aligned with the outcome, or a score of 99 is there was a student upload error or other technical issue which prevented them from scoring the artifact. When one rater assigned a score of 100 or 99 and the second rater assigned a score of 0 – 4 or when one rater assigned a score of 100 and the second a score of 99, they also met to discuss the rationale for their scores to see if they could agree on a score or, at a minimum, scores on the 0 – 4 scale that differed by not more than one point. If they could not agree, a third reader was assigned.

Third Readers for this Year's Review

1. We had only three trait tags that required third reviews. In each case, the disagreement was between a score of either a score of 99 and a score on the 0 – 4 scale or between a score of 100 and a score on the 0 – 4 scale. Reconciliations were handled as follows:
   - Scenario 1: Rater 1 assigned a score of 99; rater 2 assigned a score of 2. Following discussion, rater 1 changed the score to 2 and rater 2 changed the score to 100. A third reader then assigned a score of 2. The final score was 2.
   - Scenario 2: Rater 1 assigned a score of 100; rater 2 assigned a score of 2. Following discussion, rater 1 maintained the score of 100 and rater 2 changed the score to 1. A third reader then assigned a score of 0. The final score was 1 (mean of the second rater's first and second scores and the third rater's score).
   - Scenario 3: Rater 1 assigned a score of 100; rater 2 assigned a score of 1. Following discussion, rater 1 maintained the score of 100 and rater 2 changed the score to 0. A third reader then assigned a score of 2. The final score was 1 (mean of the second rater's first and second scores and the third rater's score).

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

Outcome	Trait	Total Tags	# Not Able to be Assessed (Score of 99)	# Misaligned (Score of 100)	Total # Excluded from Analysis of Means	Tags Usable for Calculating Mean Scores
Information Literacy	Sources	59	9 (15%)	11 (19%)	20	39
	Relevance of Information	97	8 (8%)	19 (20%)	27	70
	Assumptions and Biases	33	2 (6%)	12 (36%)	14	19
	Citation	40	5 (13%)	4 (10%)	9	31
Integrative Thinking	Connections among Disciplines	91	2 (2%)	24 (26%)	26	65
	Relations among Domains of Thinking	32	2 (6%)	8 (25%)	10	22
	Transfer	32	0	7 (22%)	7	25
	Connections to Experience	82	3 (4%)	16 (20%)	19	63
Metacogniti ve Thinking Thinking	Project Management	40	2 (5%)	13 (33%)	15	25
	Self-Evaluation	100	3 (3%)	14 (14%)	17	83

Revised Information Literacy Rubric

Information Literacy: Students will **employ** appropriate research tools, **integrate** relevant information from reliable sources, **question and evaluate** information and its sources, and **cite sources** in an academic manner.

Traits: Performance Indicators/Performance Levels	**Level 0**	**Level 1**	**Level 2**	**Level 3**	**Level 4**
Sources: Employs appropriate research tools (i.e. sources).	Does not report sources.	**Reports** superficial information from questionable sources.	**Reports** basic information from a combination of reliable and questionable sources.	**Reports** both complex and basic information from mostly reliable sources.	**Reports** multiple and complex information from reliable sources.
Relevance of Information: Integrates relevant information from reliable sources.	Does not report sources.	**Reports** information that lacks relevance and quality.	**Reports** information from **limited** and **redundant** sources.	**Reports relevant** information from **reliable** sources.	**Integrates relevant** information from **reliable** sources.
Assumptions and Biases: Questions and **evaluates** information and its sources.	Does not report sources.	**Reports** information at face value.	**Shows** some evidence of questioning assumptions and biases of sources.	**Shows** evidence of working with multiple assumptions and biases from all sources.	**Analyzes** assumptions and biases and **evaluates** the relevance of contexts when presenting a position.
Citation: Cites sources in an academic manner.	Does not cite sources.	**Inappropriate** use of paraphrasing, quotes, and/or citations.	**Inconsistent** use of paraphrasing, quotes, and/or citations.	**Acceptable** use of paraphrasing, quotes, and/or citations.	**Correct** use of paraphrasing, quotes, and citations using the style manual of the discipline (as appropriate).

Revised Integrative Thinking Rubric

Integrative Thinking: Students will **make connections** and **transfer** skills and learning among varied disciplines, domains of thinking, experiences, and situations.

Traits: Performance Indicators/Performance Levels	Level 0	Level 1	Level 2	Level 3	Level 4
Connections among Disciplines: Connects examples, facts, or theories from more than one discipline.	**Makes no attempt** to connect.	**Connects** in a rudimentary manner.	**Connects** in an emerging manner.	**Connects** in a thorough manner.	**Creates** wholes out of multiple parts (synthesizes) or draws conclusions.
Relation among Domains of Thinking: Connects examples, facts, or theories from more than one of Marshall's Domains of Thinking.	**Makes no attempt** to connect.	**Connects** in a rudimentary manner.	**Connects** in an emerging manner.	**Connects** in a thorough manner.	**Creates** wholes out of multiple parts (synthesizes) or draws conclusions.
Transfer: Adapts and applies skills, abilities, theories, or methodologies gained in one situation and/or discipline to other situations and/or other disciplines.	**Makes no attempt** to adapt or apply.	**Adapts or applies** in a rudimentary manner.	**Adapts or applies** in an emerging manner.	**Adapts and applies** in a thorough manner.	**Adapts and applies** in an original or complex manner.
Connections to Experience: Connects relevant experience and academic knowledge.	**Makes no attempt** to connect.	**Connects** in a rudimentary manner.	**Connects** in an emerging manner.	**Connects** in a thorough manner.	**Creates** wholes out of multiple parts (synthesizes) or draws conclusions.

Revised Metacognitive Thinking Rubric

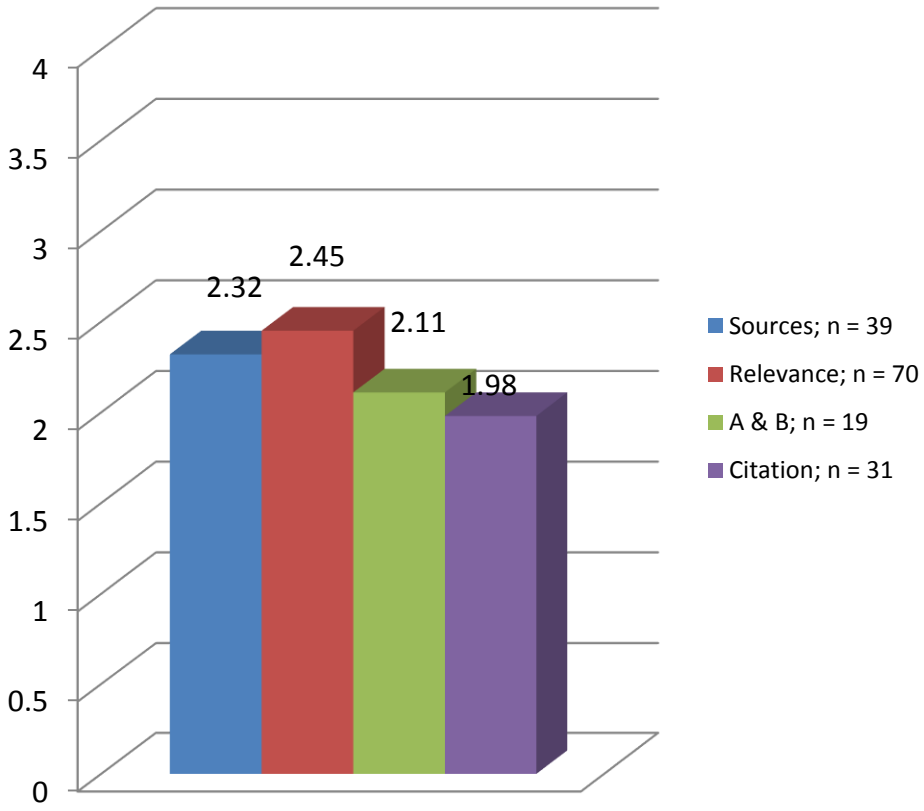
Metacognitive Thinking: Students will **evaluate** the effectiveness of a project plan or strategy and their improvement in knowledge and skills.

Traits: Performance Indicators/**Performance Levels**	**Level 0**	**Level 1**	**Level 2**	**Level 3**	**Level 4**
Project Management: **Evaluates** the effectiveness of a project plan or strategy.	**Makes no attempt at** evaluation.	**Evaluation:** Superficial	**Evaluation:** Clearly identifies strengths and weaknesses.	**Evaluation:** • Clearly identifies strengths and weaknesses. • Evidence of continual reflection or improvement. • Proposes an improvement plan that is not specific and/or feasible.	**Evaluation:** • Clearly identifies strengths and weaknesses. • Evidence of continual reflection or improvement. • Proposes an improvement plan that is specific and feasible.
Self-evaluation: **Evaluates** improvement in knowledge and skills.	**Makes no attempt at** evaluation.	**Reflects:** In a superficial manner.	**Reflects:** • With some depth • Without evidence of continual reflection or improvement.	**Reflects:** • In depth • Evidence of continual reflection or improvement.	**Reflects:** • Thoroughly. • Evidence of continual reflection or improvement. • Addresses changes in perspectives regarding his/her own learning.

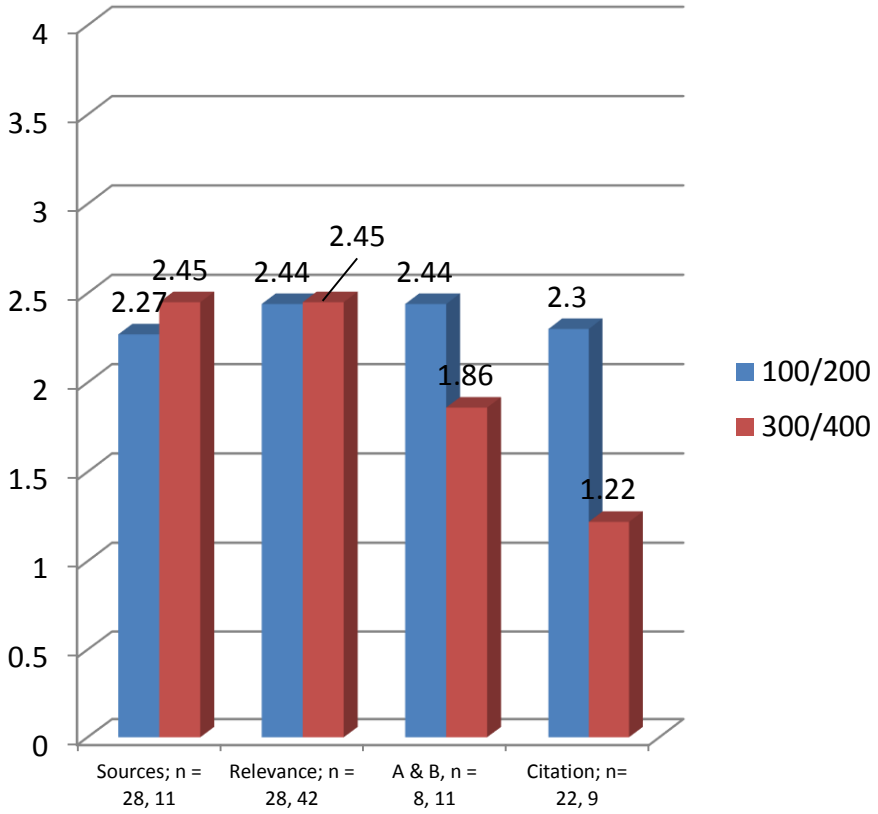
Information Literacy

Mean Scores on a scale of 0 – 4, with 4 being the highest possible score
Mean differences based on course level were significant only for *citation*.

Overall Analysis



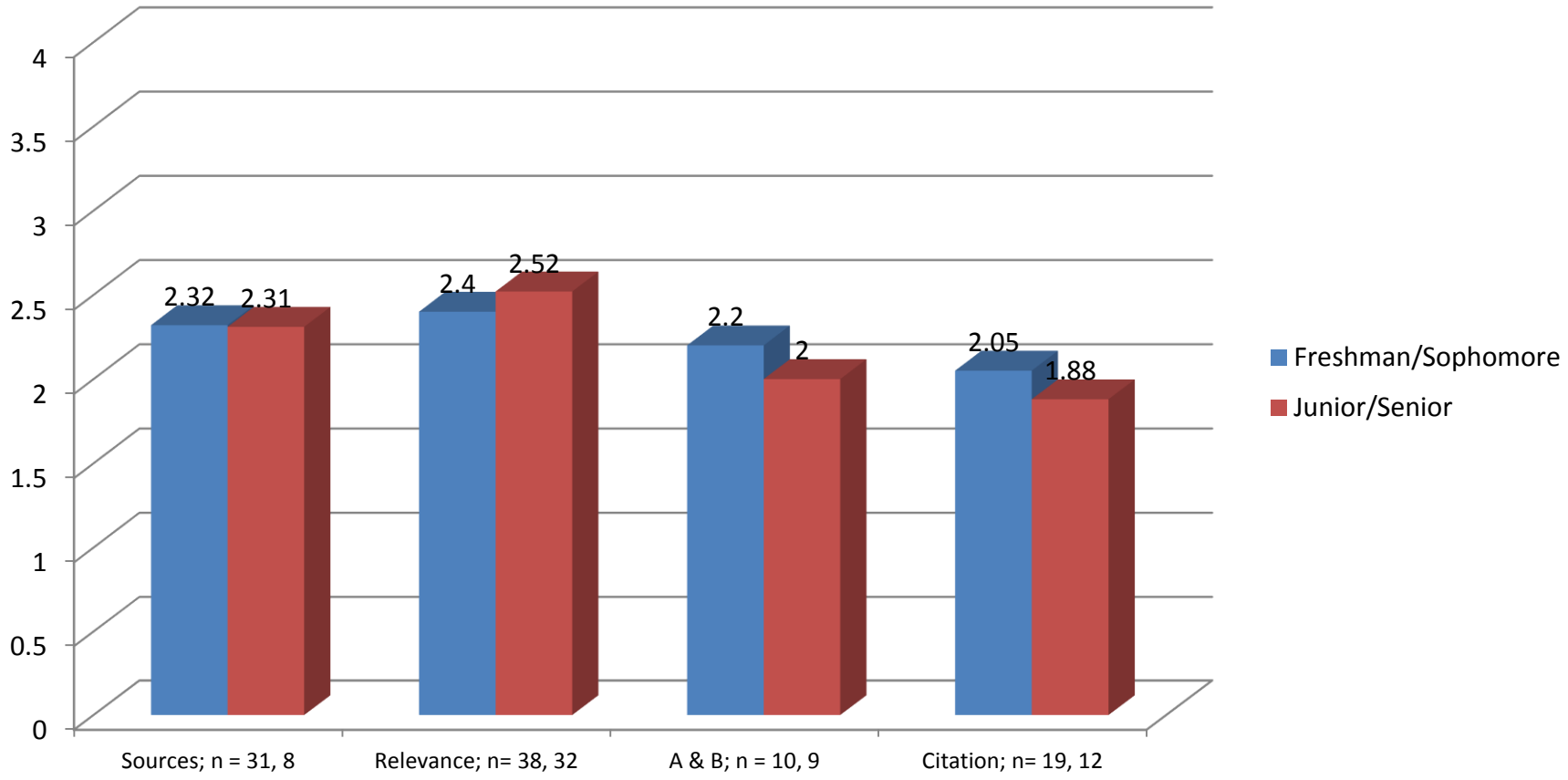
Analysis by Course Level



Information Literacy

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

Analysis by Class Rank

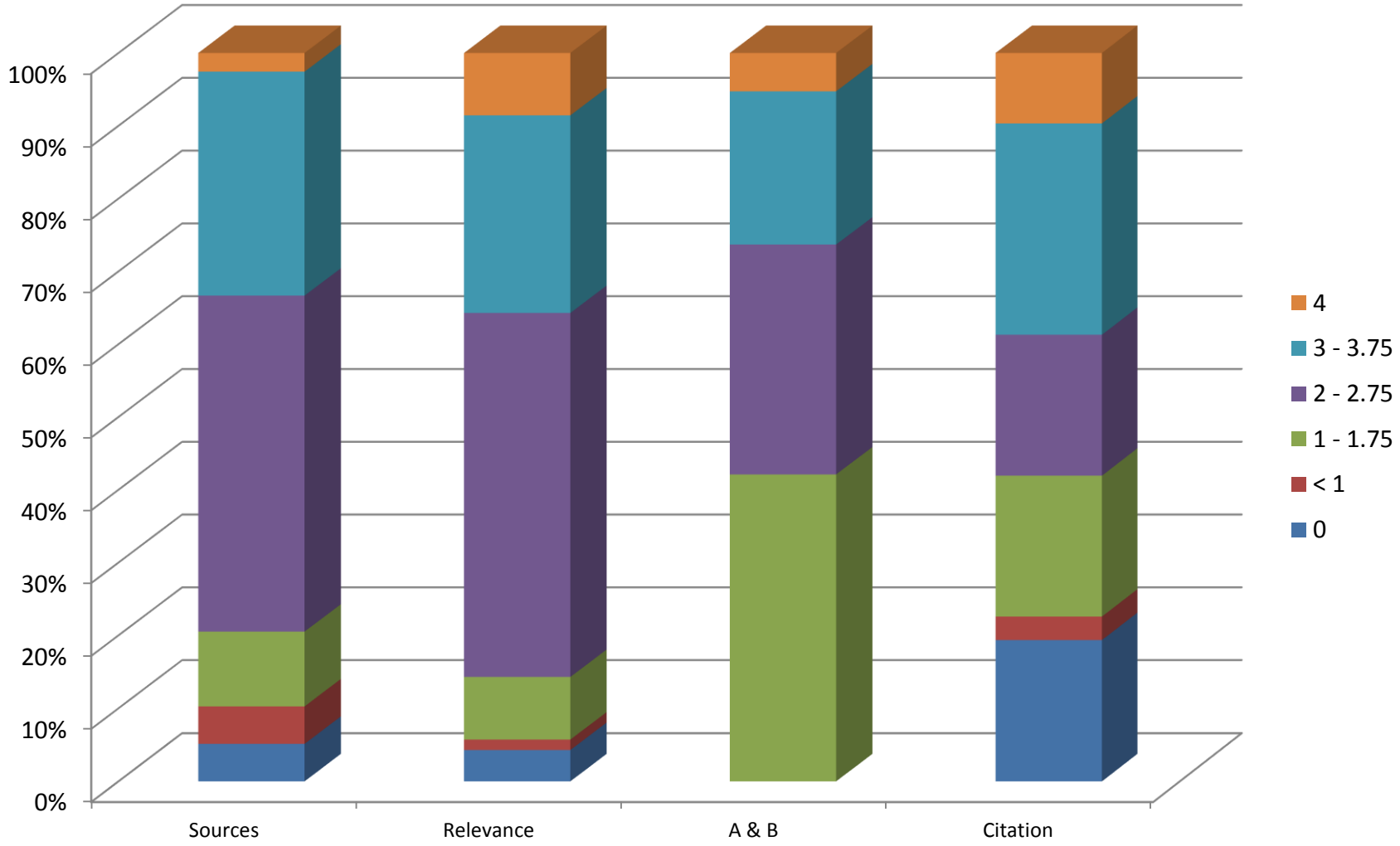


Information Literacy

Number of artifacts (with usable scores) scoring at each performance level

Trait/ Performance Level	Sources	Relevance	A & B	Citation	Total
0	2 (5%)	3 (4%)	0	6 (19%)	11 (7%)
> 0, but < 1	2 (5%)	1 (1%)	0	1 (3%)	4 (3%)
1 – 1.75	4 (10%)	6 (9%)	8 (42%)	6 (19%)	24 (15%)
2 – 2.75	18 (46%)	35 (50%)	6 (32%)	6 (19%)	65 (41%)
3 – 3.75	12 (31%)	19 (27%)	4 (21%)	9 (29%)	44 (28%)
4	1 (3%)	6 (9%)	1 (5%)	3 (10%)	11 (7%)
Total Tags with Usable Scores	39 (100%)	70 (100%)	19 (100%)	31 (100%)	159 (100%)

Information Literacy



Information Literacy

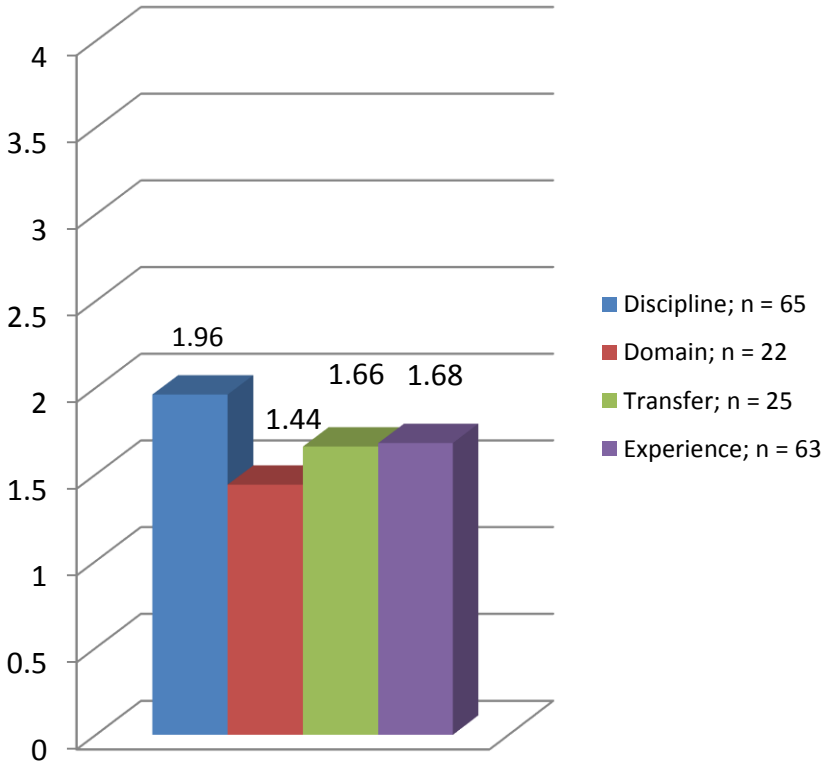
Inter-Rater Agreement Results

Trait/ Performance Level	Sources; Kappa = .095 (All Scores); Kappa = .005 (Misaligned and Unable to Score Excluded); Kappa Liberal = .425 (All scores); Kappa Liberal = .626 (Misaligned and Unable to Score Excluded)	Relevance of Information; Kappa = .262 (All Scores); Kappa = .169 (Misaligned and Unable to Score Excluded); Kappa Liberal = .624 (All Scores); Kappa = .752 (Misaligned and Unable to Score Excluded)	Assumptions and Biases; Kappa = .323 (All Scores); Kappa = .159 (Misaligned and Unable to Score Excluded); Kappa Liberal = .641 (All Scores); Kappa Liberal = .755 (Misaligned and Unable to Score Excluded)	Citations; Kappa = .374 (All Scores); Kappa = .431 (Misaligned and unable to score Excluded); Kappa Liberal = .572 (All Scores); Kappa Liberal = .740 (Misaligned and Unable to Score Excluded)
Agree on score	9 (15%)	27 (28%)	7 (21%)	16 (40%)
Difference = 1 point or less	15 (25%)	26 (27%)	8 (24%)	6 (15%)
Difference = 1.5 to 2 points	9 (15%)	12 (12%)	4 (12%)	5 (13%)
Difference = 2.5 to 3 points	2 (3%)	2 (2%)	0	1 (3%)
Agree on Misaligned	3 (5%)	9 (9%)	8 (24%)	0
Agree on Unable to Score	2 (3%)	3 (3%)	0	3 (8%)
Score + Misaligned	13 (22%)	13 (13%)	3 (9%)	8 (20%)
Score + Unable to Score	1 (2%)	2 (2%)	1 (3%)	0
Misaligned + Unable to Score	5 (8%)	3 (3%)	2 (6%)	1 (3%)
Total	59 (100%)	97 (100%)	33 (100%)	40 (100%)

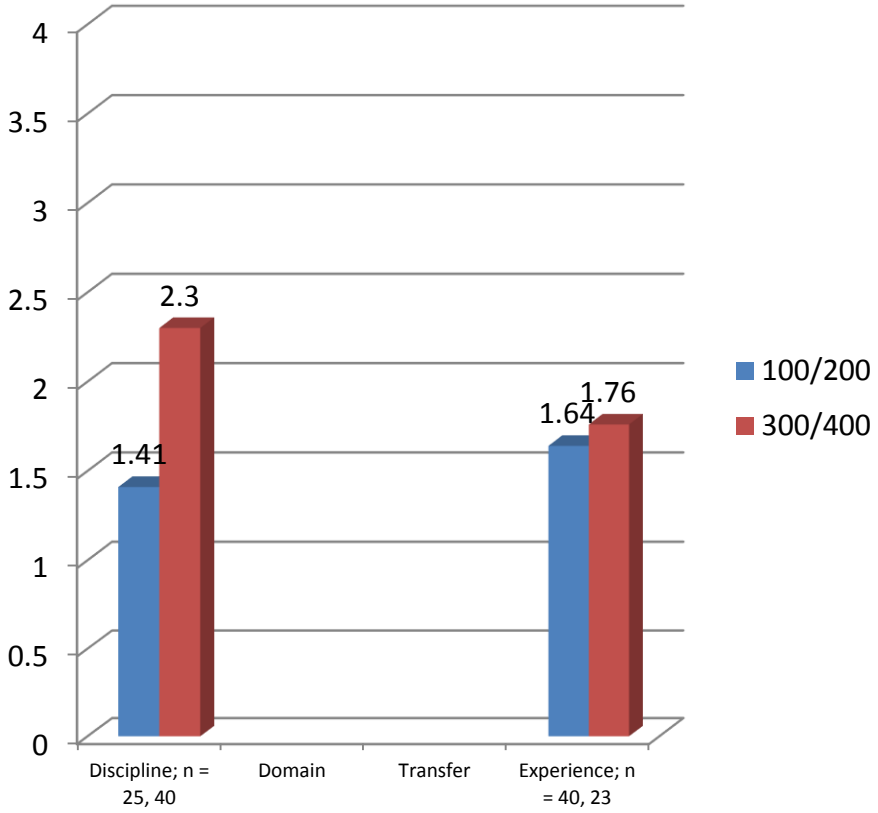
Integrative Thinking

Mean Scores on a scale of 0 – 4, with 4 being the highest possible score.
Mean differences for course level were significant for *Discipline*. We note that there were no artifacts from 300/400 level courses aligned to *Domain*, and there was only one artifact with a usable score (1) aligned to *Transfer.*

Overall Analysis



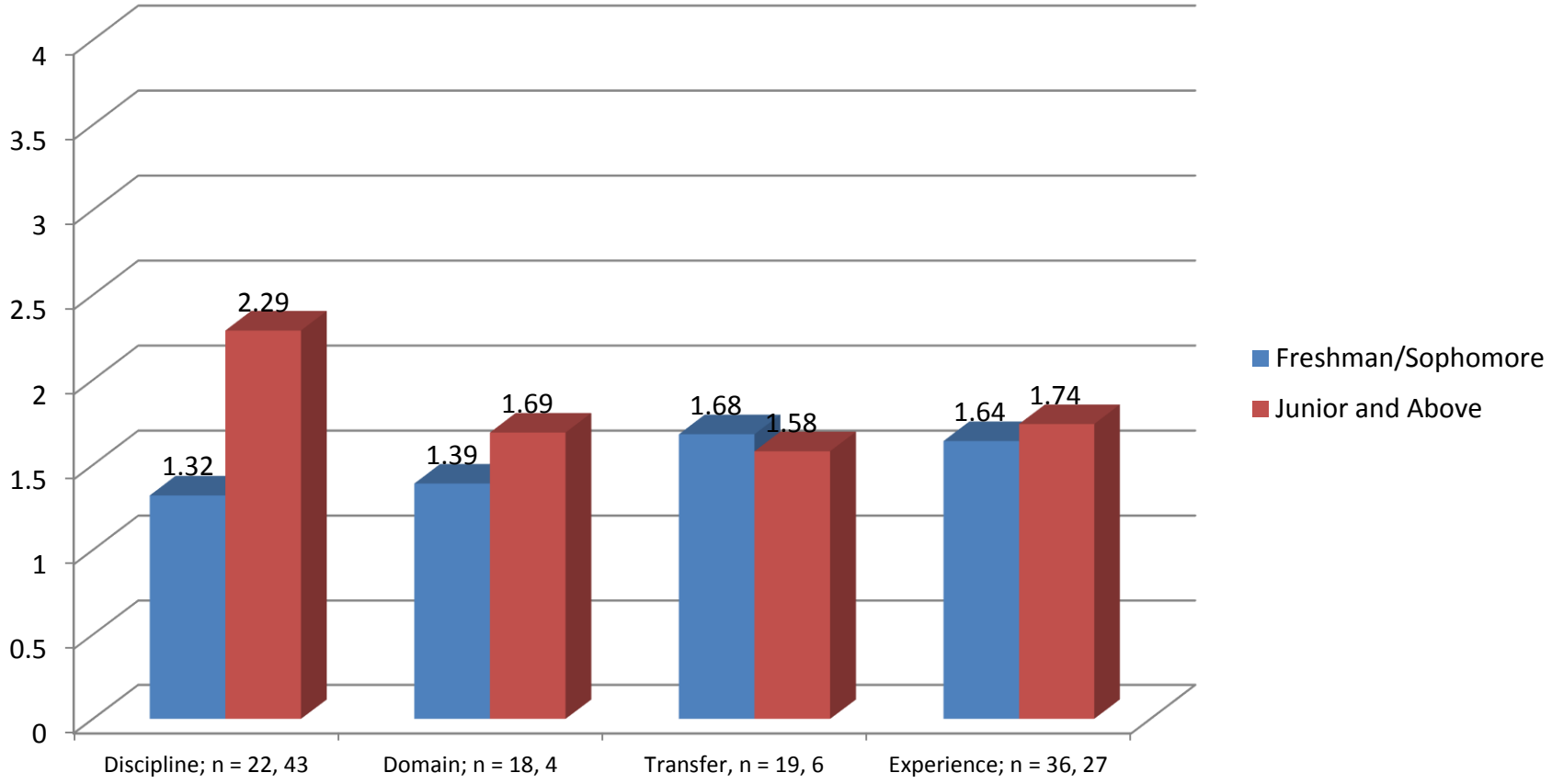
Analysis by Course Level



Integrative Thinking

Mean Scores on a scale of 0 – 4, with 4 being the highest possible score.
Mean difference was significant only for *Connections among Disciplines*.

Analysis by Class Rank

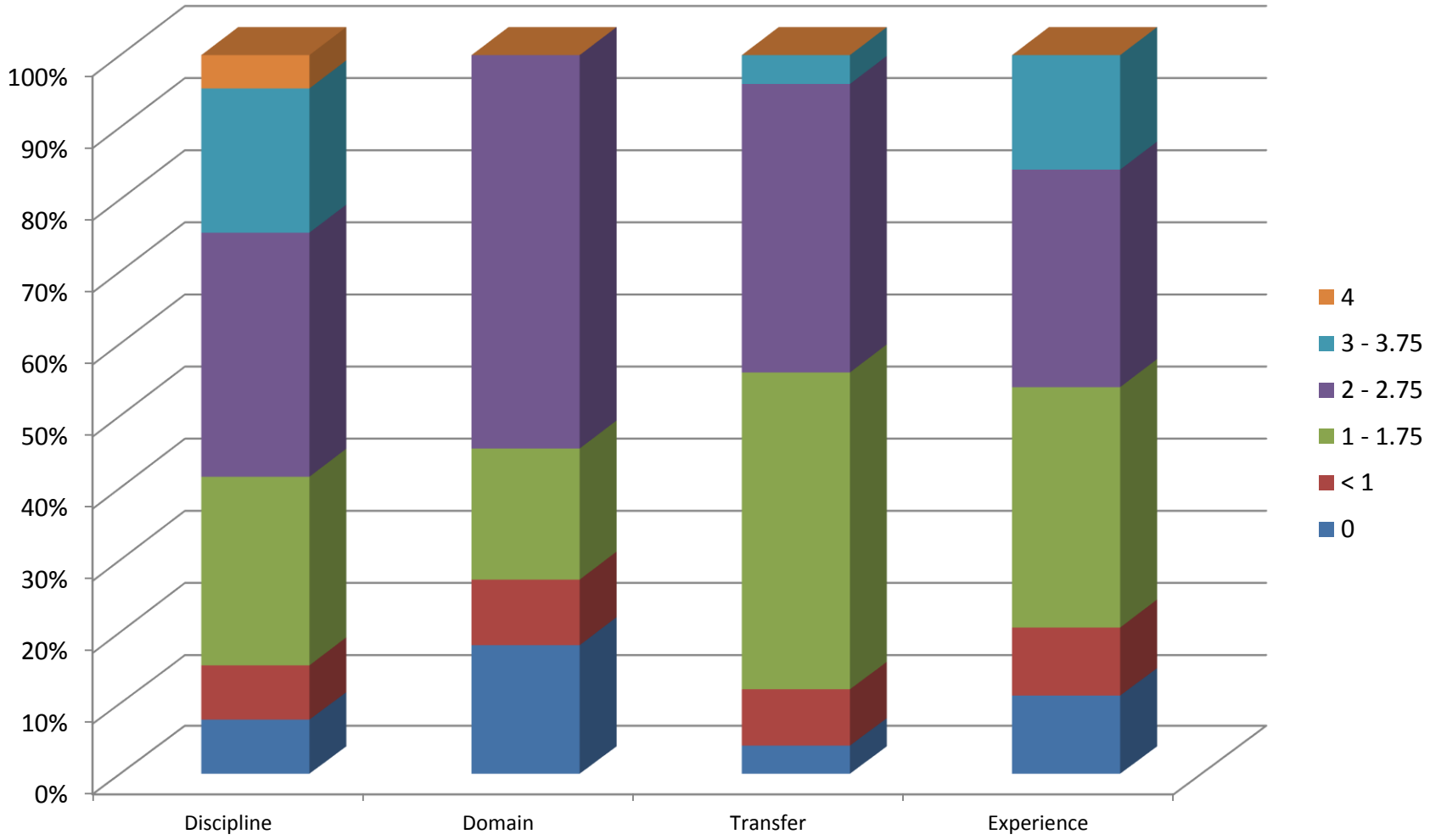


Integrative Thinking

Number of artifacts (with usable scores) scoring at each performance level

Trait/ Performance Level	Discipline	Domain	Transfer	Experience	Total
0	5 (8%)	4 (18%)	1 (4%)	7 (11%)	17 (10%)
> 0, but < 1	5 (8%)	2 (9%)	2 (8%)	6 (10%)	15 (9%)
1 – 1.75	17 (26%)	4 (18%)	11 (44%)	21 (33%)	53 (30%)
2 – 2.75	22 (34%)	12 (55%)	10 (40%)	19 (30%)	63 (36%)
3 – 3.75	13 (20%)	0	1 (4%)	10 (16%)	24 (14%)
4	3 (5%)	0	0	0	3 (2%)
Totals	65 (100%)	22 (100%)	25 (100%)	63 (100%)	175 (100%)

Integrative Thinking



Integrative Thinking

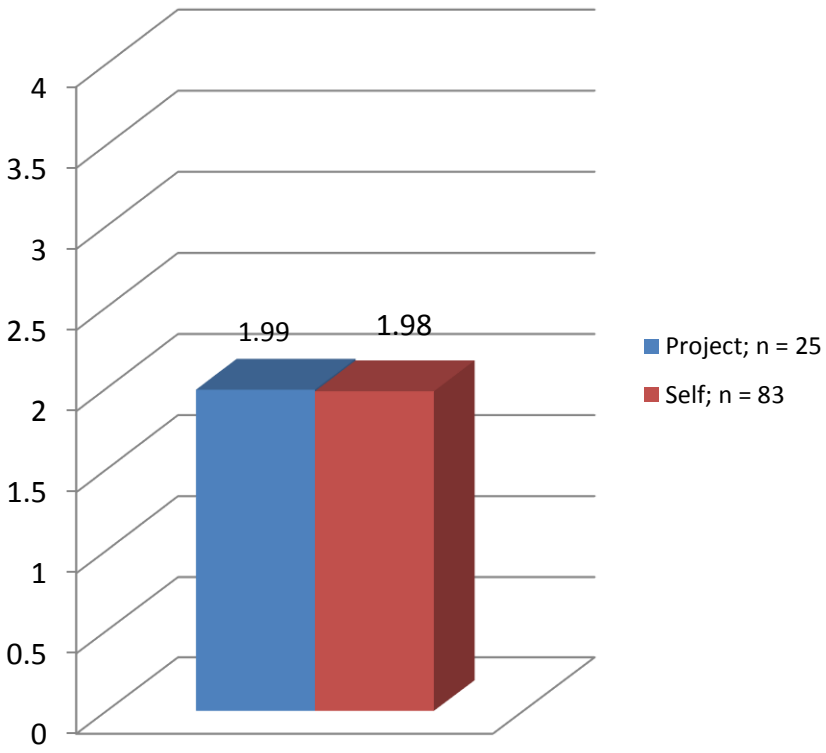
Inter-Rater Agreement Results

Trait/ Performance Level	Disciplines; Kappa = .118 (All Scores); Kappa = -.052 (Misaligned and Unable to Score Excluded) ; Kappa Liberal = .530 (All Scores); Kappa Liberal = .812 (Misaligned and Unable to Score Excluded)	Domains; Kappa = -.017 (All Scores; Kappa = -.058 (Misaligned and Unable to Score Excluded) ; Kappa Liberal = .118 (All Scores); Kappa Liberal = .423 (Misaligned and Unable to Score Excluded)	Transfer; Kappa = -.167 (All Scores); Kappa = -.157 (Misaligned and Unable to Score Excluded); Kappa Liberal = .384 (All Scores); Kappa Liberal = .741 (Misaligned and Unable to Score Excluded)	Experience; Kappa = .065 (All Scores); Kappa = .026 (Misaligned and Unable to Score Excluded); Kappa Liberal = .480 (All Scores); Kappa Liberal = .696 (Misaligned and Unable to Score Excluded)
Agree on score	14 (15%)	4 (13%)	3 (9%)	13 (16%)
Difference = 1 point or less	28 (31%)	4 (13%)	12 (38%)	26 (32%)
Difference = 1.5 to 2 points	7 (8%)	7 (22%)	2 (6%)	11 (13%)
Difference = 2.5 to 4 points	1 (1%)	0	2 (6%)	3 (4%)
Agree on Misaligned	10 (11%)	1 (3%)	0	3 (4%)
Agree on Unable to Score	2 (2%)	2 (6%)	0	2 (2%)
Score + Misaligned	28 (31%)	14 (44%)	13 (41%)	21 (26%)
Score + Unable to Score	1 (1%)	0	0	3 (4%)
Misaligned + Unable to Score	0	0	0	0
Total	91 (100%)	32 (100%)	32 (100%)	82 (100%)

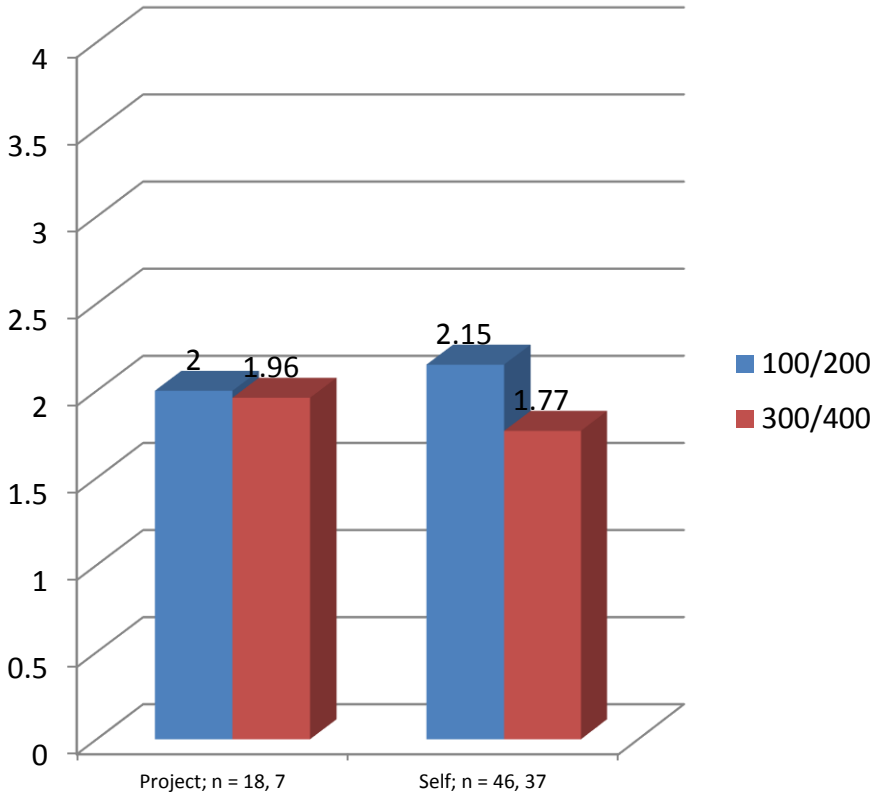
Metacognitive Thinking

Mean Scores on a scale of 0 – 4, with 4 being the highest possible score.
No mean differences for course level were significant.

Overall Analysis



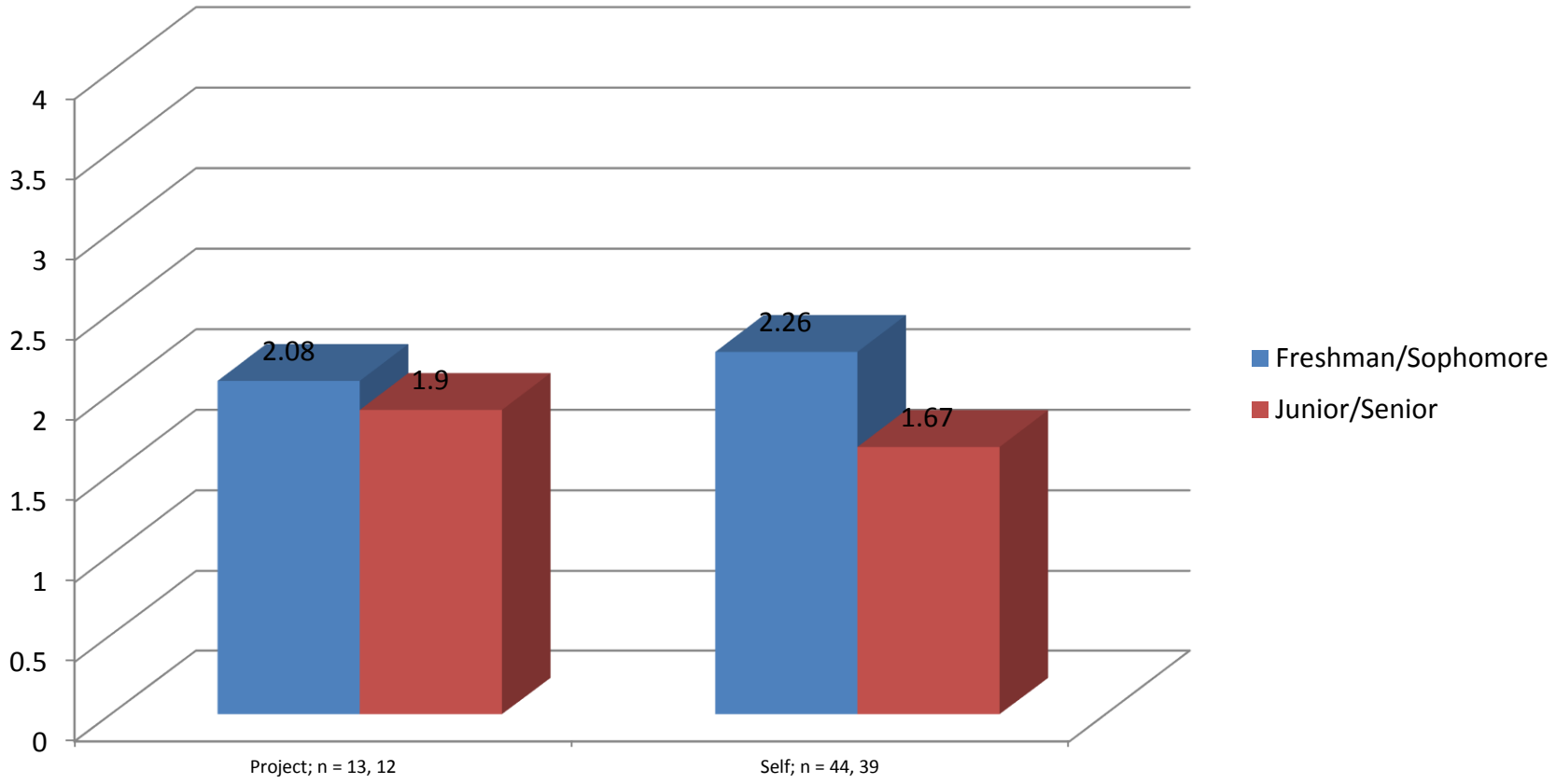
Analysis by Course Level



Metacognitive Thinking

Mean Scores on a scale of 0 – 4, with 4 being the highest possible score.
Mean difference was significant for Self-Evaluation, but not for Project Management.

Analysis by Class Rank

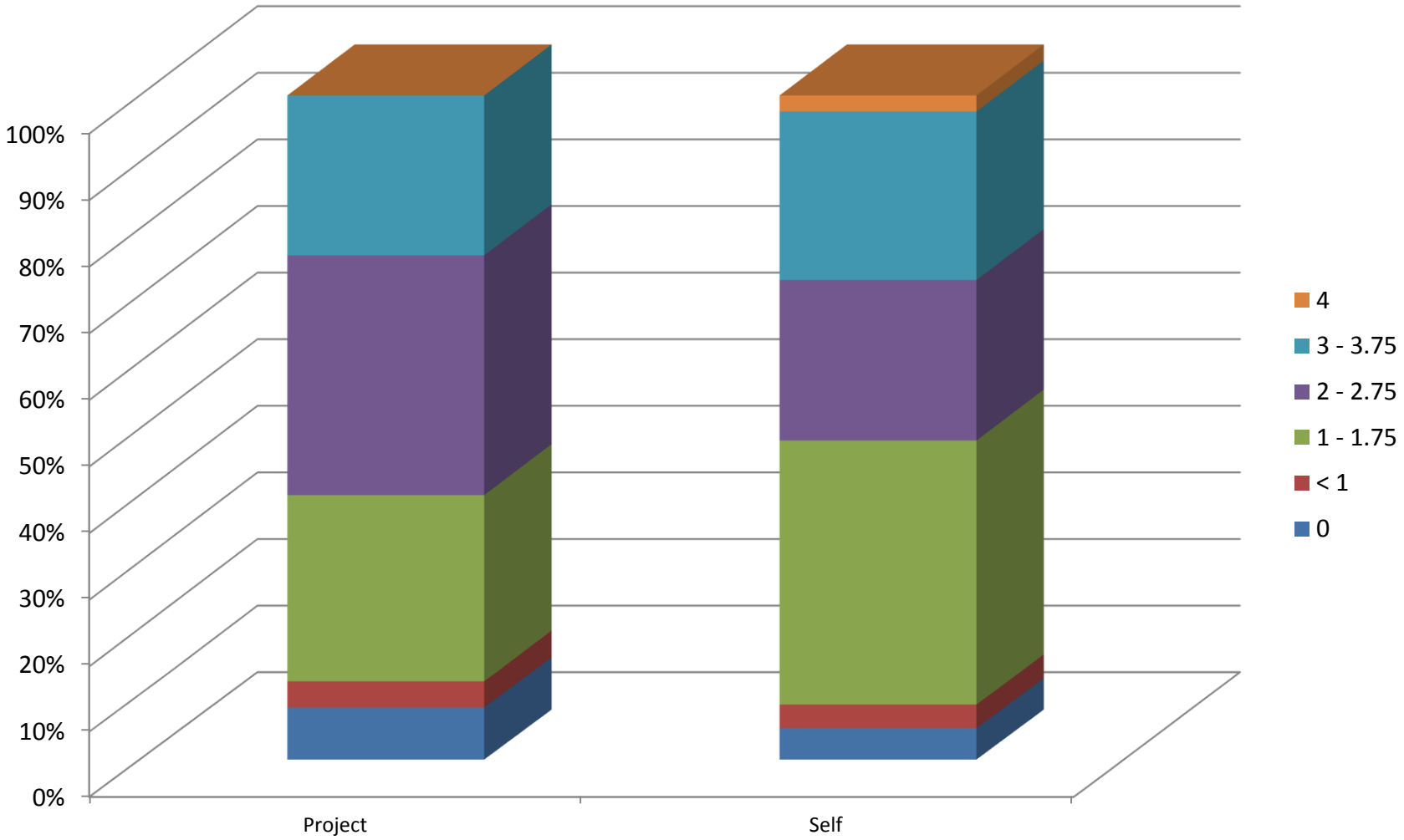


Metacognitive Thinking

Number of artifacts (with usable scores) scoring at each performance level

Trait/ Performance Level	Project	Self	Total
0	2 (8%)	4 (5%)	6 (6%)
> 0, but < 1	1 (4%)	3 (4%)	4 (4%)
1 – 1.75	7 (28%)	33 (40%)	40 (37%)
2 – 2.75	9 (36%)	20 (24%)	29 (27%
3 – 3.75	6 (24%)	21 (25%)	27 (25%)
4	0	2 (2%)	2 (2%)
Totals	25 (100%)	83 (100%)	108 (100%)

Metacognitive Thinking



Metacognitive Thinking

Inter-Rater Agreement Results

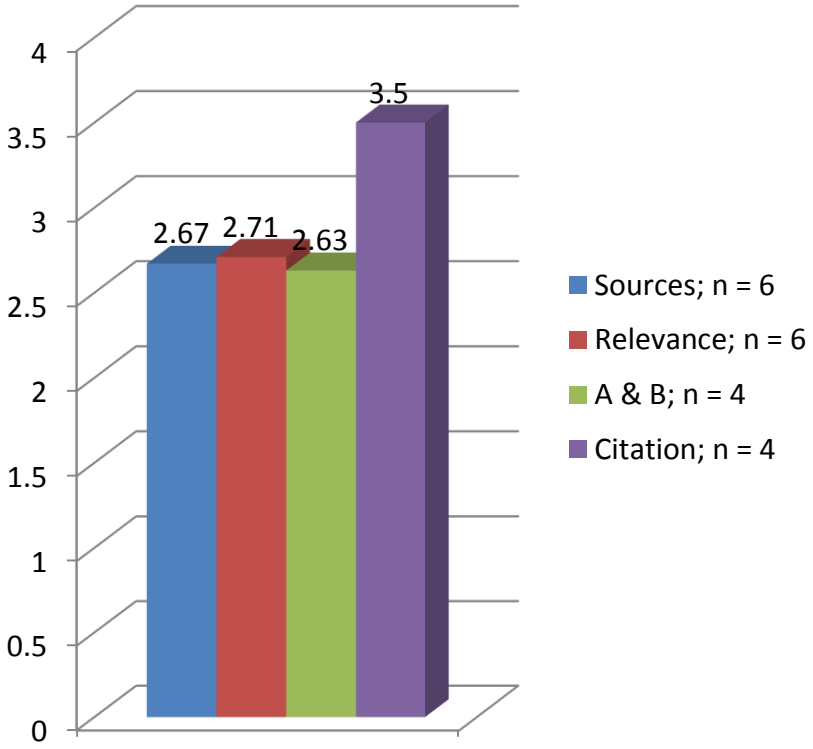
Trait/ Performance Level	Project; Kappa = .070 (All Scores); Kappa = -.089 (Misaligned and Unable to Score Excluded); Kappa Liberal = .413 (All Scores); Kappa Liberal = .784 (Misaligned and Unable to Score Excluded)	Self ; Kappa = .103 (All Scores); Kappa = .134 (Misaligned and Unable to Score Excluded); Kappa Liberal = .586 (All Scores); Kappa Liberal = .837 (Misaligned and Unable to Score Excluded)
Agree on score	4 (10%)	23 (23%)
Difference = 1 point or less	10 (25%)	37 (37%)
Difference = 1.5 to 2 points	3 (8%)	10 (10%)
Difference = 2.5 to 3 points	0	0
Agree on Misaligned	6 (15%)	2 (2%)
Agree on Unable to Score	0	1 (1%)
Score + Misaligned	14 (35%)	24 (24%)
Score + Unable to Score	2 (5%)	1 (1%)
Misaligned + Unable to Score	1 (3%)	2 (2%)
Total	40 (100%)	100 (100%)

Course Type Analysis

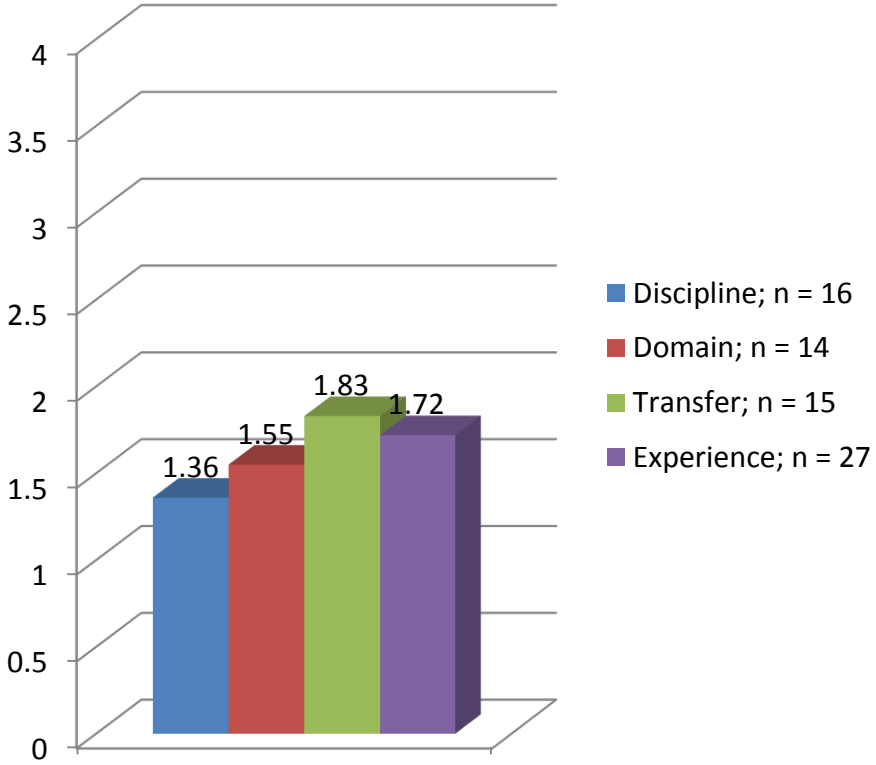
CT Courses

Mean Scores on a scale of 0 – 4, with 4 being the highest possible score. All CT courses are 100/200 Level. A few artifacts were from courses that, in addition to being CT, also were multicultural, international, or writing intensive.

Information Literacy



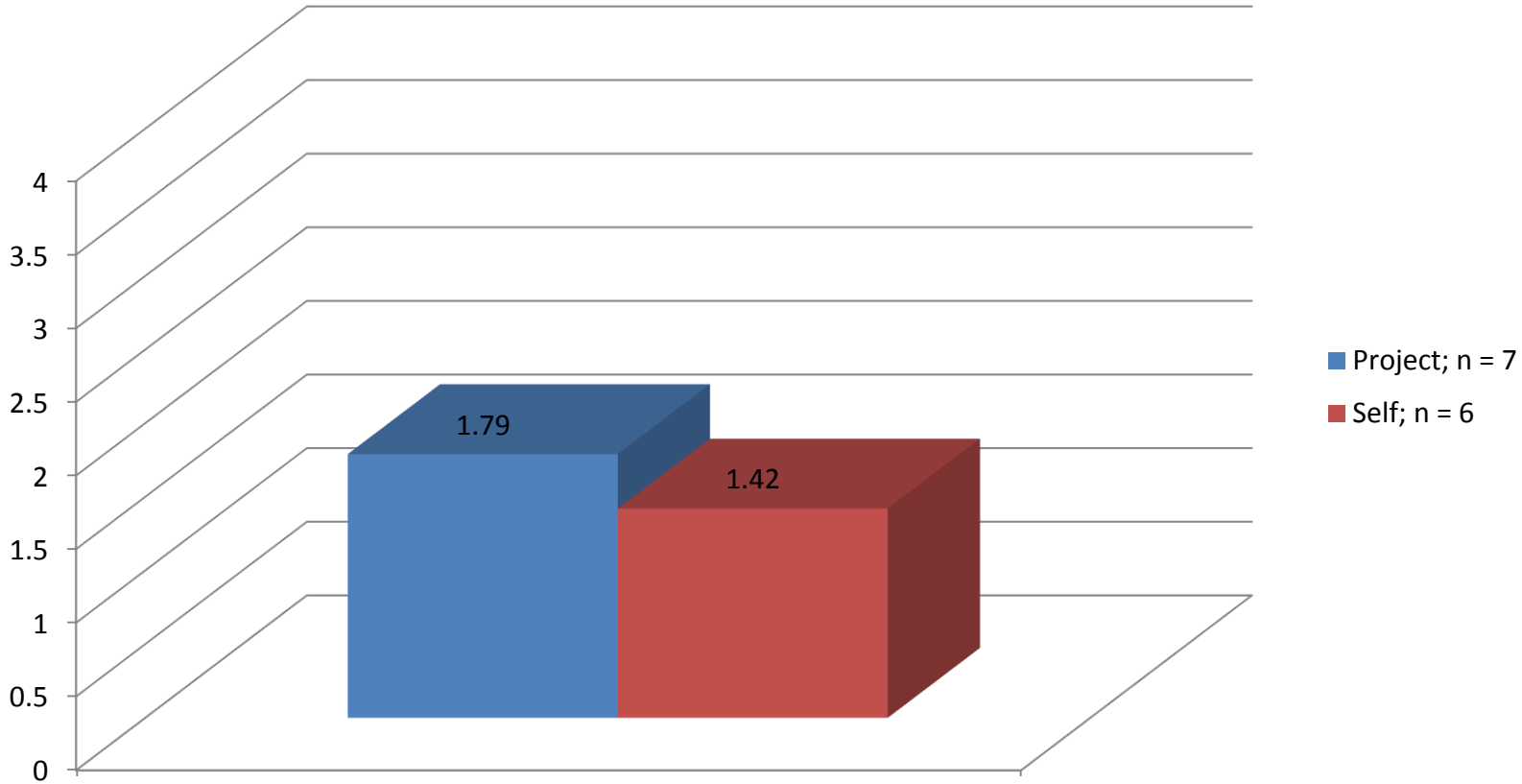
Integrative Thinking



CT Courses

Mean Scores on a scale of 0 – 4, with 4 being the highest possible score. All CT courses are 100/200 Level. A few artifacts were from courses that, in addition to being CT, also were multicultural, international, or writing intensive.

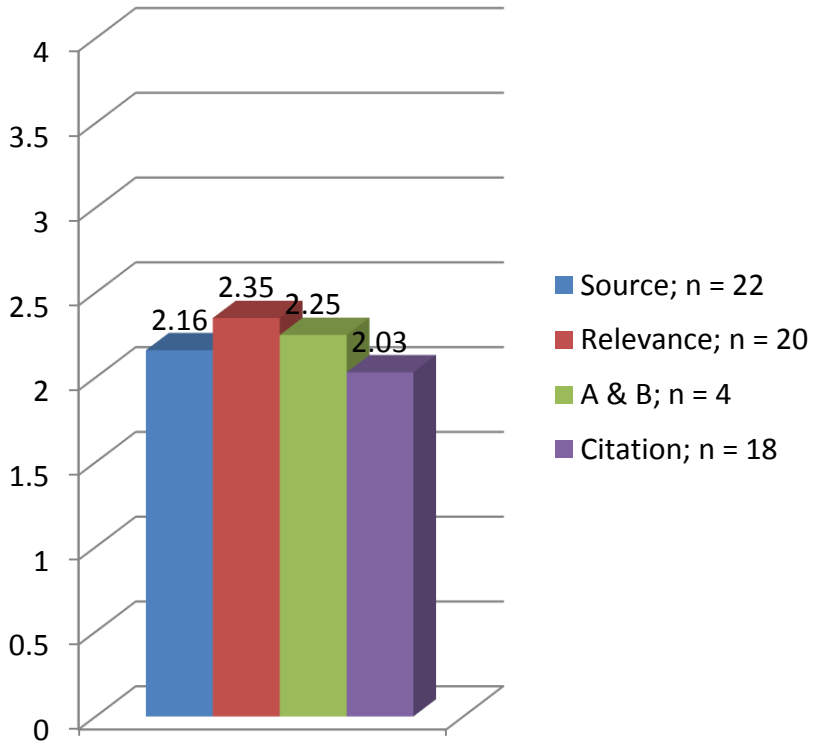
Metacognitive Thinking



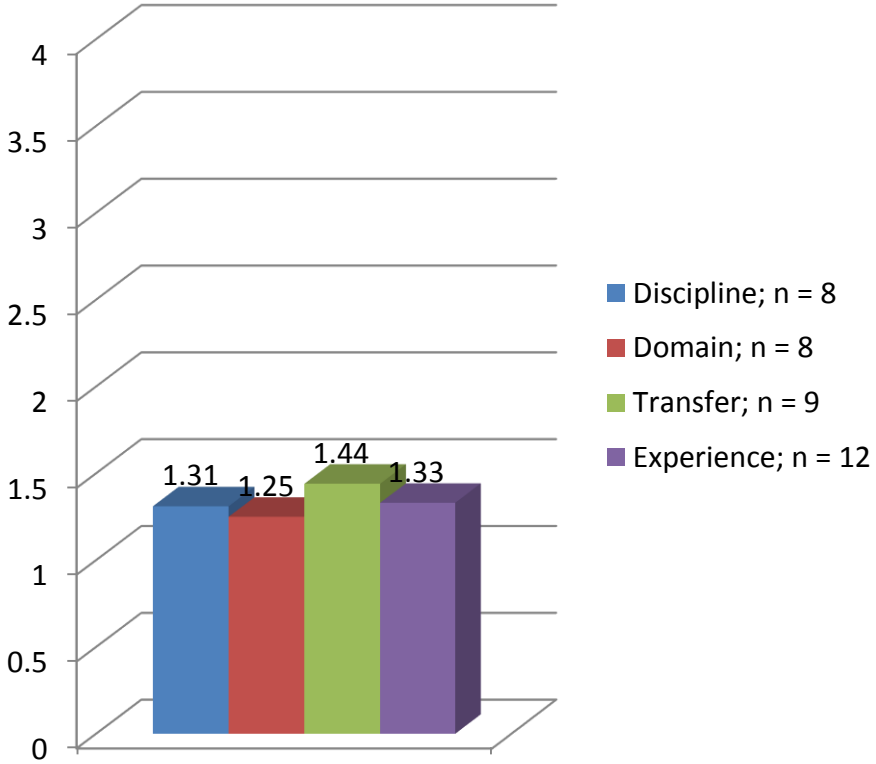
FYS Courses

Mean Scores on a scale of 0 – 4, with 4 being the highest possible score.

Information Literacy



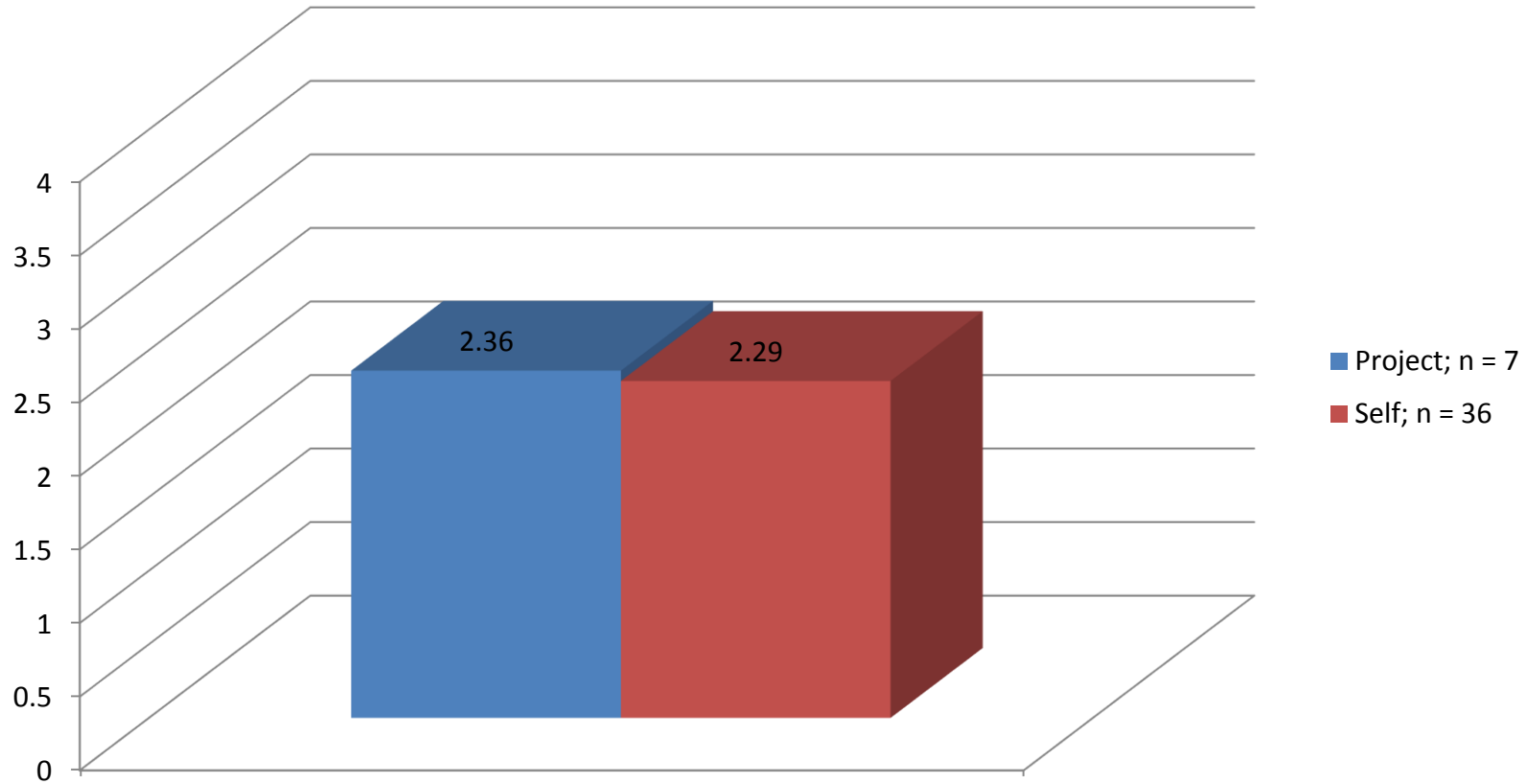
Integrative Thinking



FYS Courses

Mean Scores on a scale of 0 – 4, with 4 being the highest possible score.

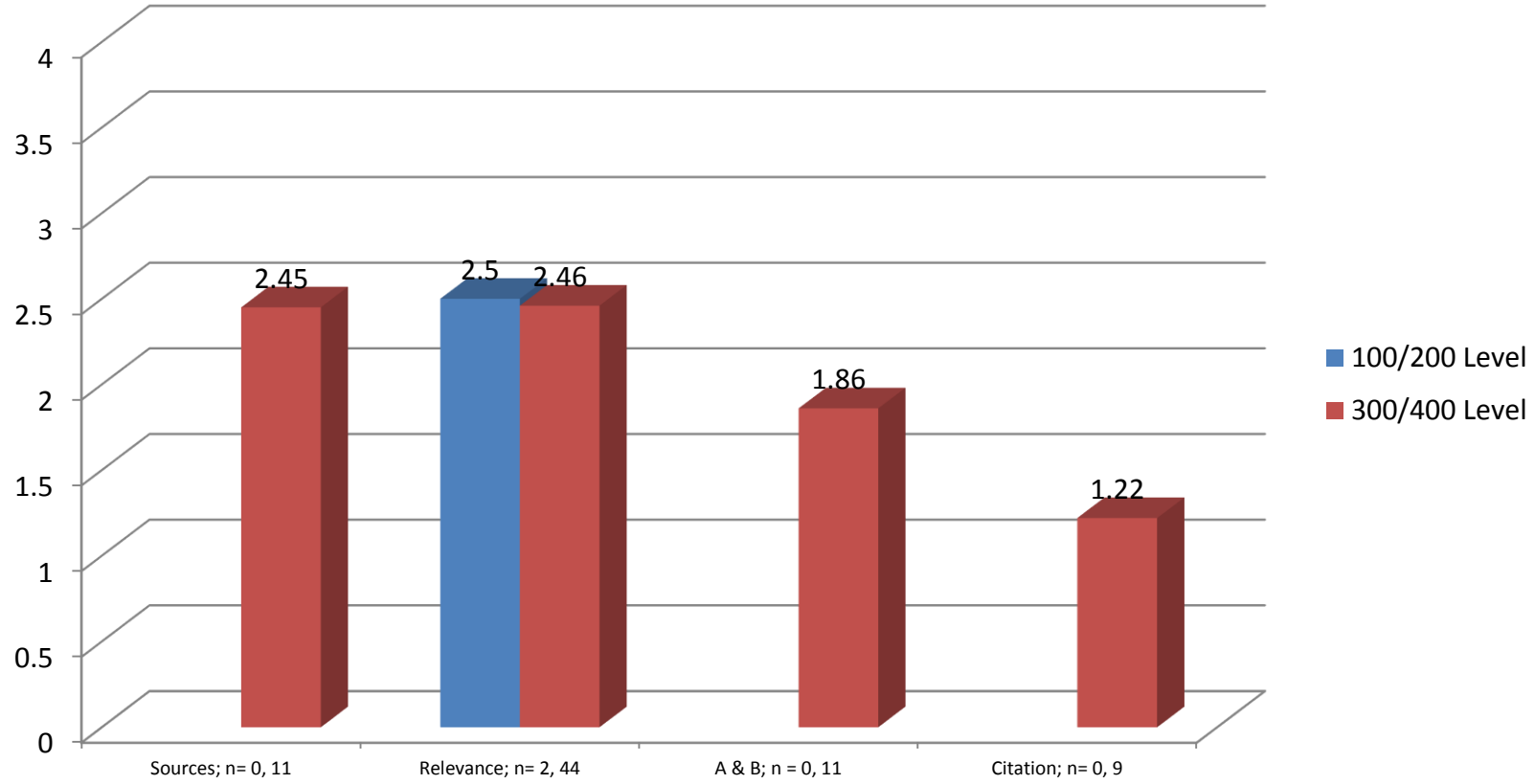
Metacognitive Thinking



Writing Intensive Courses

Mean Scores on a scale of 0 – 4, with 4 being the highest possible score. For this analysis all courses had WI as their only attribute and all but two artifacts were from 300/400 level courses.

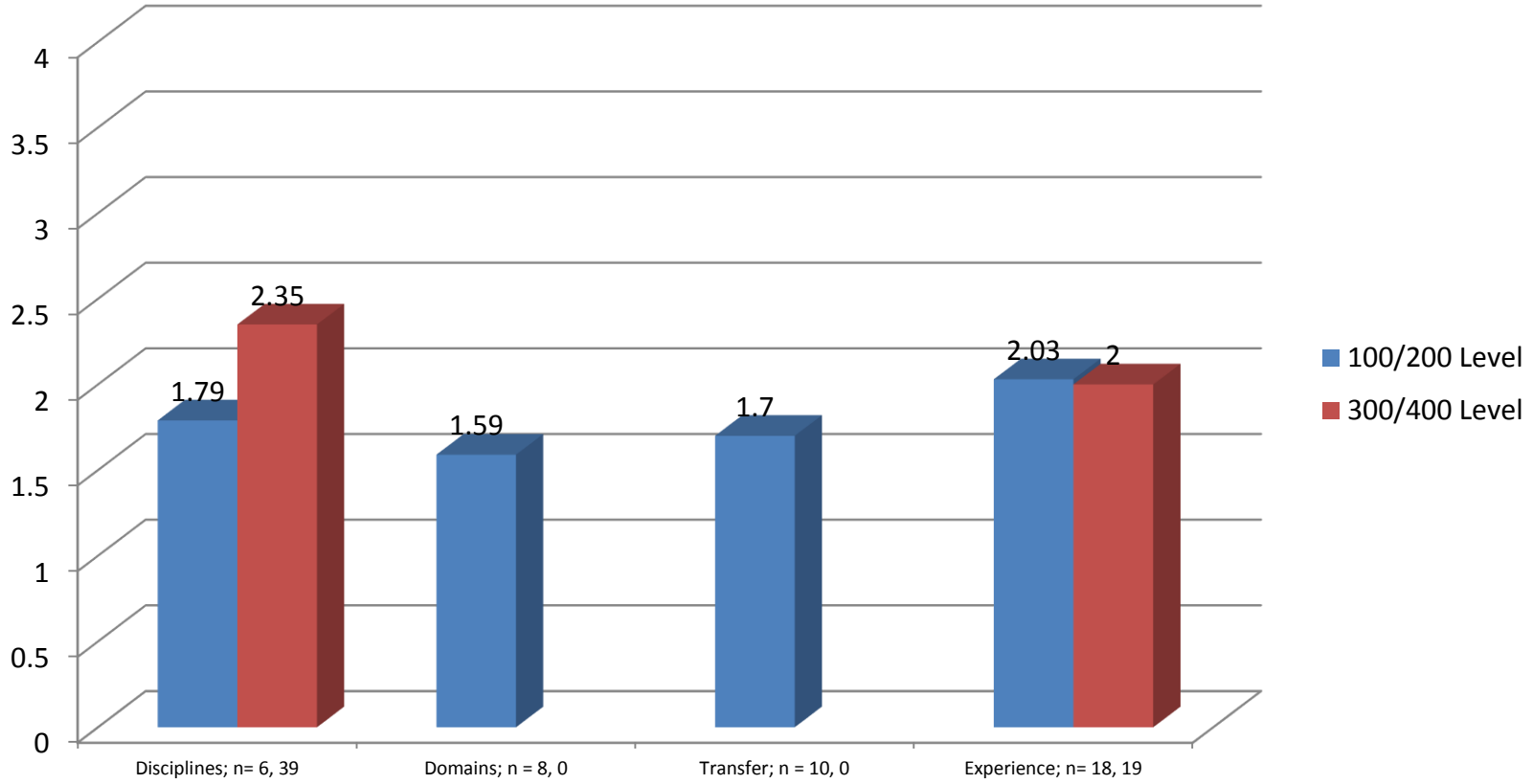
Information Literacy



Writing Intensive Courses

Mean Scores on a scale of 0 – 4, with 4 being the highest possible score. Some courses had attributes in addition to WI (CT, Multicultural, and/or International).

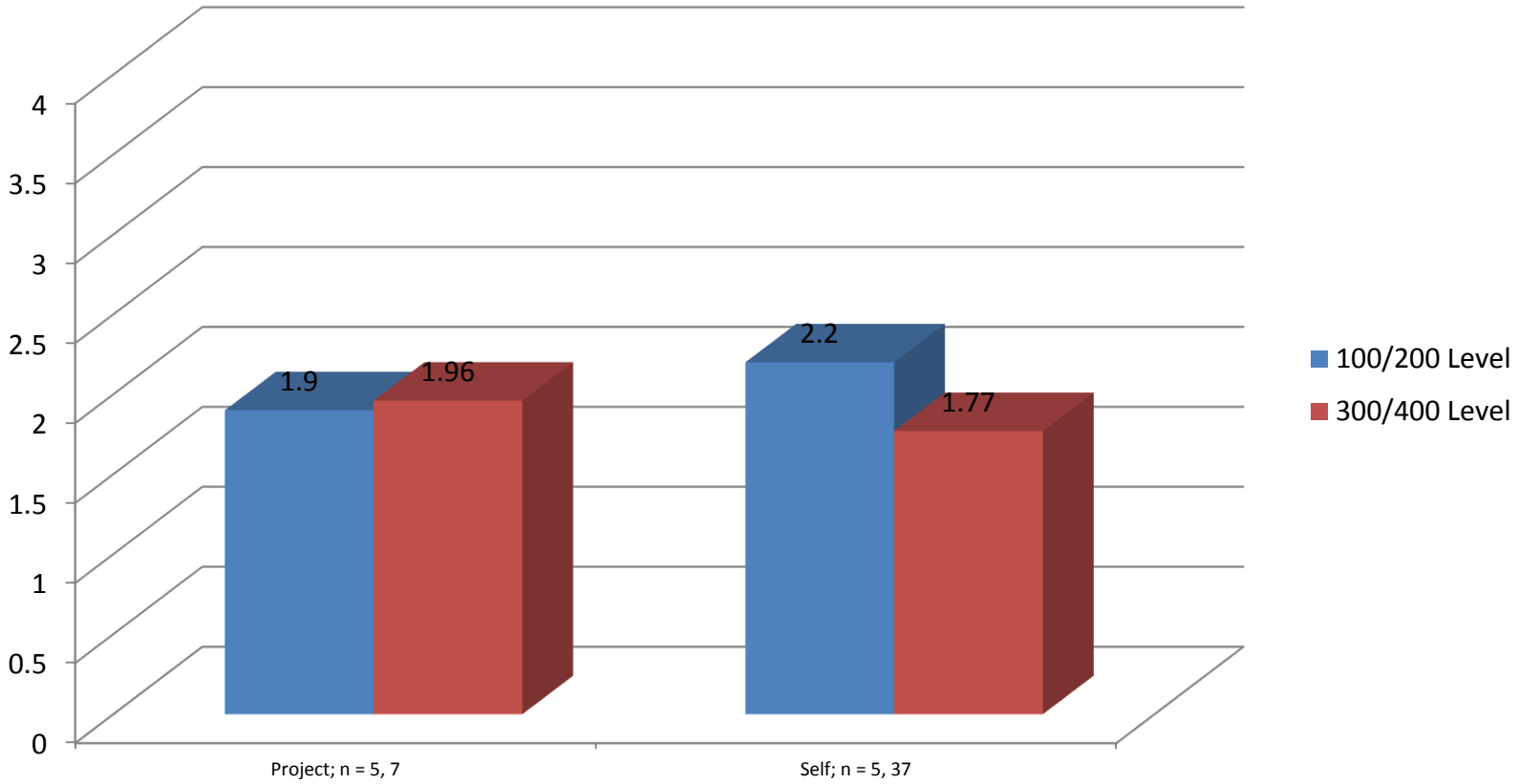
Integrative Thinking



Writing Intensive Courses

Mean Scores on a scale of 0 – 4, with 4 being the highest possible score. Some courses had attributes in addition to WI (CT, Multicultural, and/or International).

Metacognitive Thinking



MARSHALL UNIVERSITY®

Analysis of Misalignment by Course Type

Note: For these analyses, courses with multiple attributes are analyzed by *primary* course type. Rules for designating *primary* course type are as follows:

1. CT is always primary.

2. MC is primary when paired with WI.

3. WI is always secondary; therefore, the analysis of WI courses indicates that this was the *only* attribute.

Information Literacy

Trait/ # and % of artifacts misaligned	Sources	Relevance	A & B	Citation	Total
CT	3/9 = 33%	3/9 = 33%	0/4 = 0%	0/4 = 0%	6/26 = 23%
FYS	8/39 = 21%	8/32 (25%)	6/10 = 60%	3/25 = 12	25/106 = 24%
MC	N/A	6/8 = 75%	6/8 = 75%	N/A	12/16 = 75%
WI	11/59 = 19%	2/48 = 4%	0/11 = 0%	1/11 = 9%	14/129 = 11%
Total	22/107 = 21%	19/97 = 20%	12/33 = 36%	4/40 = 10%	57/277 = 21%

Integrative Thinking

Trait/ # and % of artifacts misaligned	Discipline	Domain	Transfer	Experience	Total
CT	7/25 = 28%	4/20 = 20%	3/18 = 17%	4/34 = 12%	18/97 = 19%
FYS	4/12 = 33%	4/12 = 33%	1/10 = 10%	1/13 = 8%	10/47 = 21%
MC	3/4 = 75%	N/A	3/4 = 75%	0/4 = 0%	6/12 = 50%
WI	10/50 = 20%	N/A	N/A	11/31 = 35%	21/81 = 26%
Total	24/91 = 26%	8/32 = 25%	7/32 = 22%	16/82 = 20%	55/237 = 23%

Metacognitive Thinking

Trait/ # and % of artifacts misaligned	Project	Self	Total
CT	7/16 = 44%	2/14 = 14%	9/30 = 30%
FYS	4/11 = 36%	0/36 = 0%	4/47 = 9%
MC	N/A	0/13 = 0%	0/13 = 0%
WI	2/13 = 15%	8/37 = 22%	10/50 = 20%
Total	13/40 = 33%	10/100 = 10%	23/140 = 16%