

Peirce College 2016-17 Institutional Learning Outcomes Assessment Report

Introduction & Context

The report that follows summarizes assessment research that was conducted as part of Peirce College's ongoing annual institutional learning outcome (ILO) assessment process, which was established in 2010. Following a six-year cycle, every year the College focuses on one of its ILOs for intensive direct assessment as well as college-wide inquiry and professional development. Peirce's institution-wide student learning outcome statements can be reviewed at <https://www.peirce.edu/docs/default-source/pdf/Catalogs/17-18-undergraduate-10-31-17.pdf?#page=8>

Data

(Note that links to all original data sets and source documents may be found in Appendix A, Data Sources, found below.)

Student Sample Selection

We sought to identify a writing sample from each student graduating with a BS degree in the most recent academic year prior to the study (i.e. 2015-16). There were 124 such students. Because we wanted to assess the students' writing ability as close to graduation as possible, we limited our sample to writing produced within the most recent year prior to graduation. This resulted in the elimination of four students who had not enrolled in any appropriate course in their final year before graduation (those having only enrolled in quantitative courses or who transferred their remaining credits, for example).

Writing Sample Selection

We then obtained the transcripts of each student, eliminated any grades of 'W' or 'F', and then removed any classes taken more than a year before graduation. Among those classes, we limited to courses that involved an assignment that would make an appropriate target for our assessment. Appropriate writing assignments were evaluated according to the following criteria:

- Sufficiently broad and non-technical to be accessible to faculty member outside of the discipline for which the paper was prepared: To better understand how our expectations of student writing might vary from one department to the next, we wanted extra-departmental evaluation to be practicable. Every paper was scored by at least one evaluator from outside of the corresponding course's department.
- Manageable length: To make most efficient use of evaluator's time and effort, we sought to choose papers that were long enough to provide sound basis for evaluating writing, but no longer. A length of 1,000 - 1,500 words was considered ideal.

- Involving research: While Peirce does have a separate institutional outcome for information literacy, full-scope assessment of writing nevertheless requires an assessment of the use of the research literature.

After identifying candidate courses and assignments according to these criteria, we began drawing sample papers, starting with the sections that were most densely enrolled by the target student population for efficiency’s sake and continuing through the ranked course list until we had obtained at least one paper for every student in the sample. This resulted in papers from the following courses:

ACC 450	ENG 202	HUM 300	MGT 425
BUS 440	HCA 480	ISC 310	MIS 205
CJS 104	HIS 105	ISC 425	MKT 208
CJS 304	HRM 310	ITN 200	MKT 401
COM 202	HUM 102	LGL 430	PSY 101
COM 345	HUM 107	MGT 310	SCI 330
ENG 103	HUM 275	MGT 404	SOC 240

The resultant sample represents a blend of courses from the College’s general education core curriculum plus upper-level courses from all of the bachelor’s degree-programs. Note that papers from capstone courses, normally an attractive assessment locus, were excluded here per the first two criteria detailed above.

All papers were then anonymized with regard to student, section, class (to the extent not self-evident to the assessor from the nature and topic of the writing), and instructor.

Method

Evaluators

All full-time faculty participated in the assessment. Each paper was assigned at random to three faculty evaluators.

Evaluation Instrument & Process

We used the American Association of Colleges and Universities’ VALUE (Valid Assessment of Learning in Undergraduate Education) written communication rubric for this assessment. The AACU VALUE Written Communication Rubric can be accessed at

<https://drive.google.com/open?id=0B7Y7z0UrG0TbTHlyTkFvbXY1aVE>.

Evaluators completed their ratings by completing the form available at

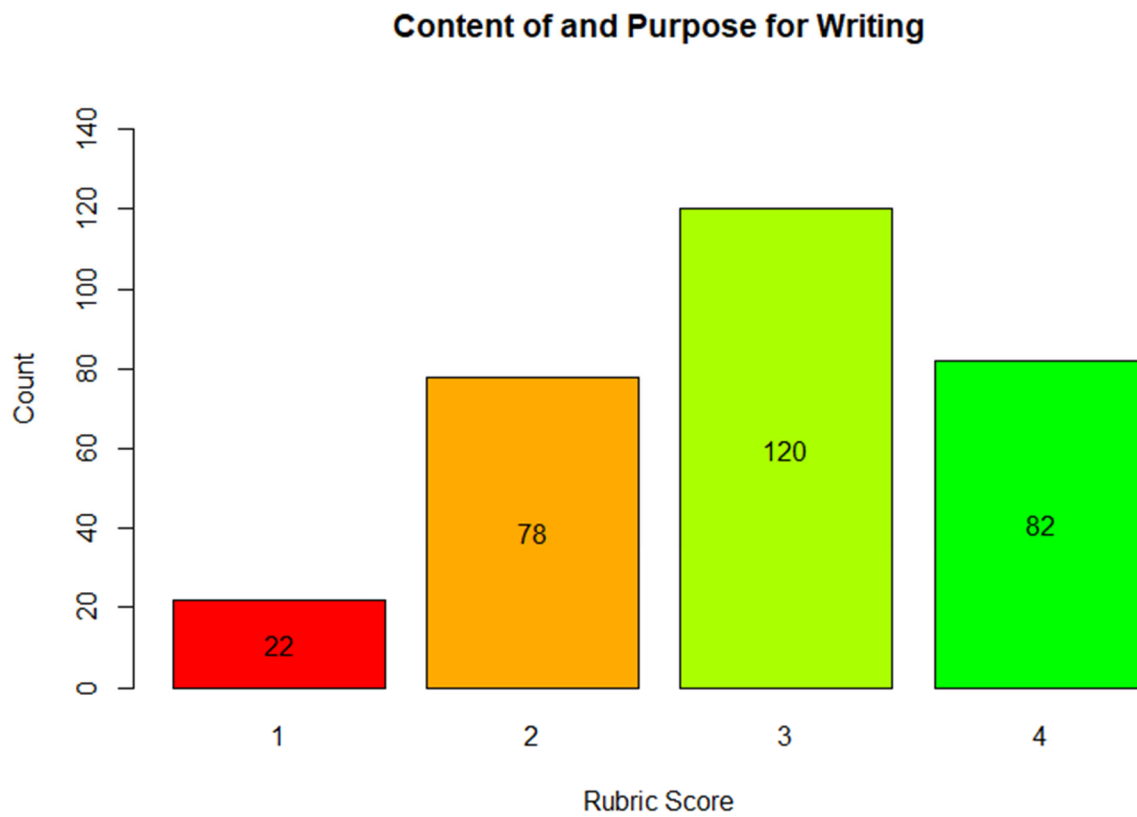
https://docs.google.com/forms/d/1EGKZNVHx_-RGskjJScEvJBepD7PgfftK-RBB1fFQ1tA/edit, ranking each paper on a scale from 1 to 4 for each of the 5 writing criteria .

Results

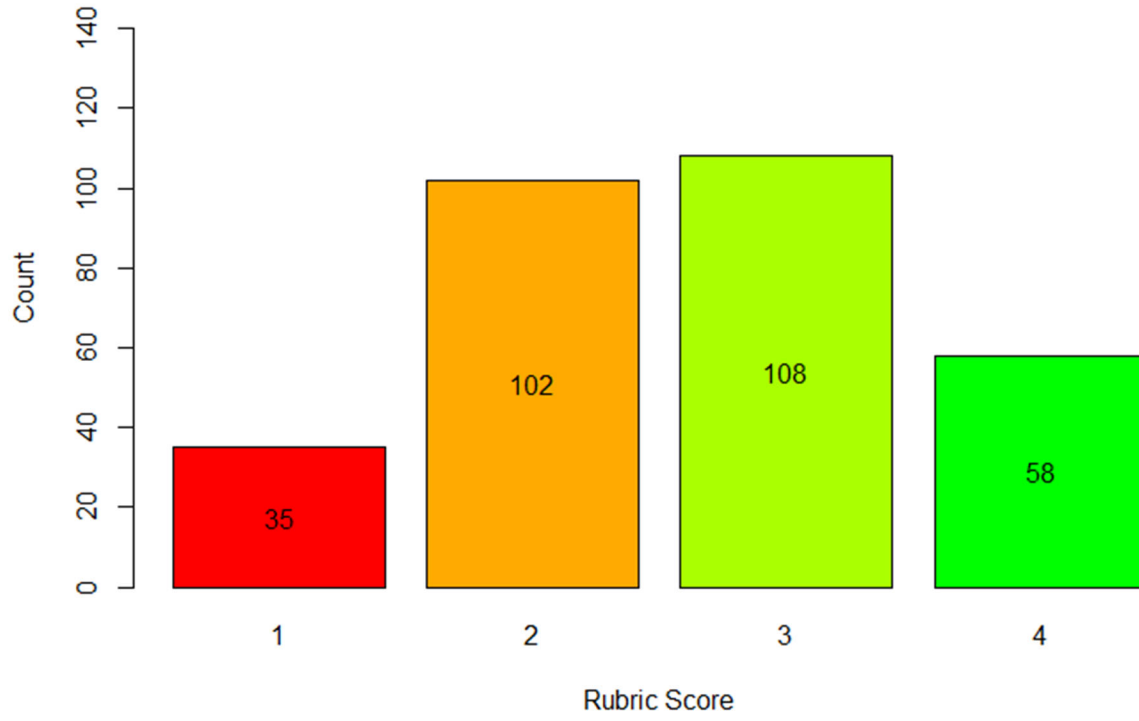
The writing criteria evaluated were as follows:

- Context of and Purpose for Writing (“CPW”)
- Content Development (“CDEV”)
- Genre and Disciplinary Conventions (“GDC”)
- Sources and Evidence - (“SEV”)
- Control of Syntax and Mechanics (“CSM”)

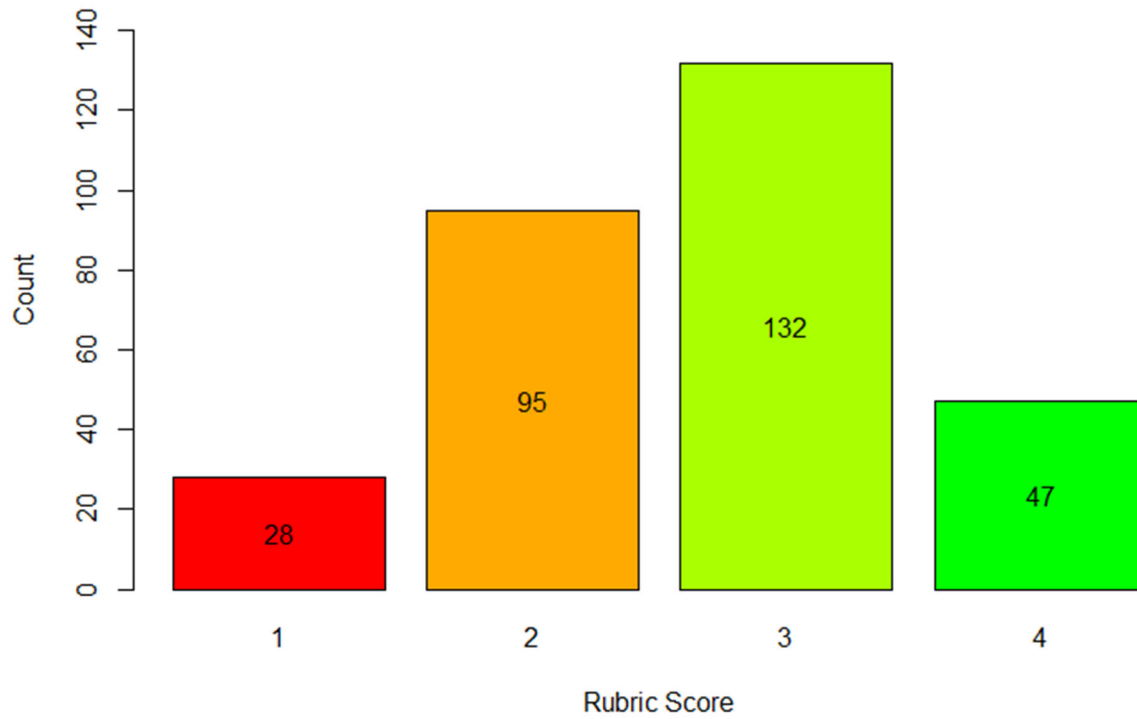
Each category was scored on a scale from a maximum of 4 (“Capstone”), to a minimum of 1 (“Benchmark”). A summary of the results per criterion is as follows:



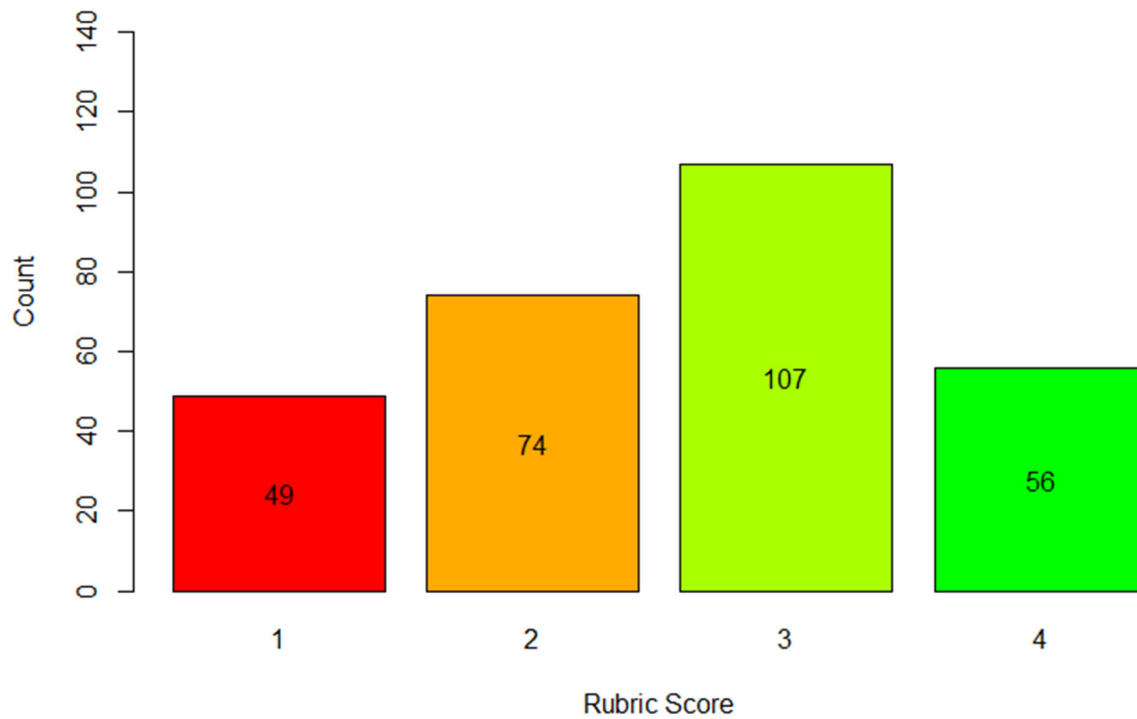
Content Development



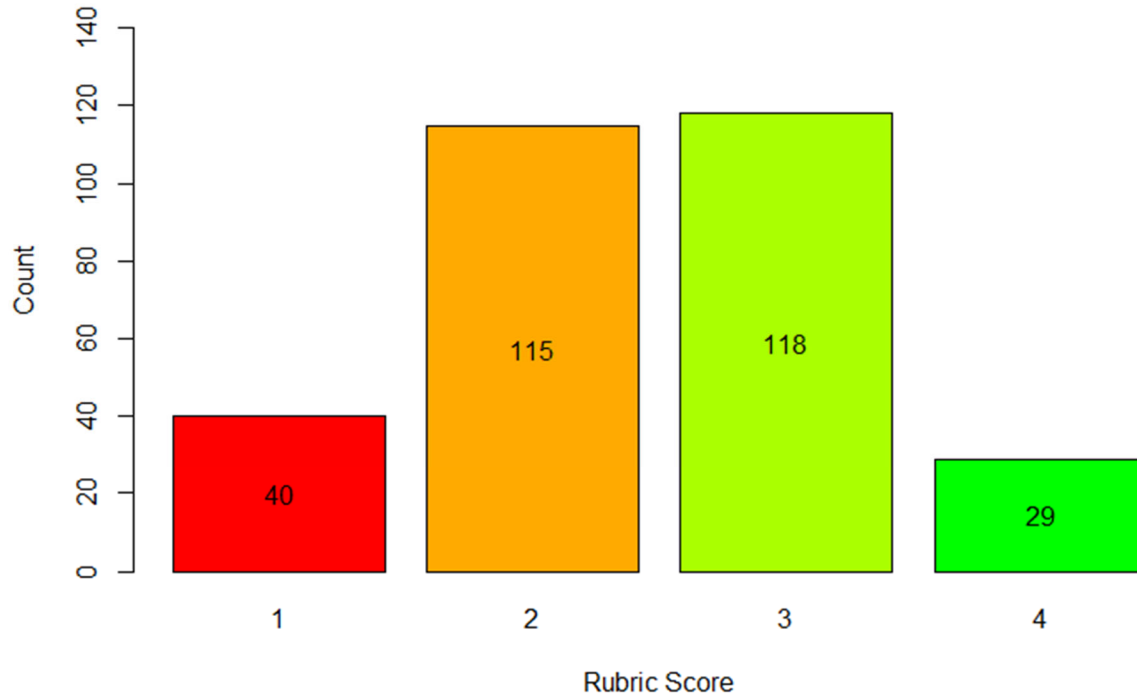
Genre and Disciplinary Conventions



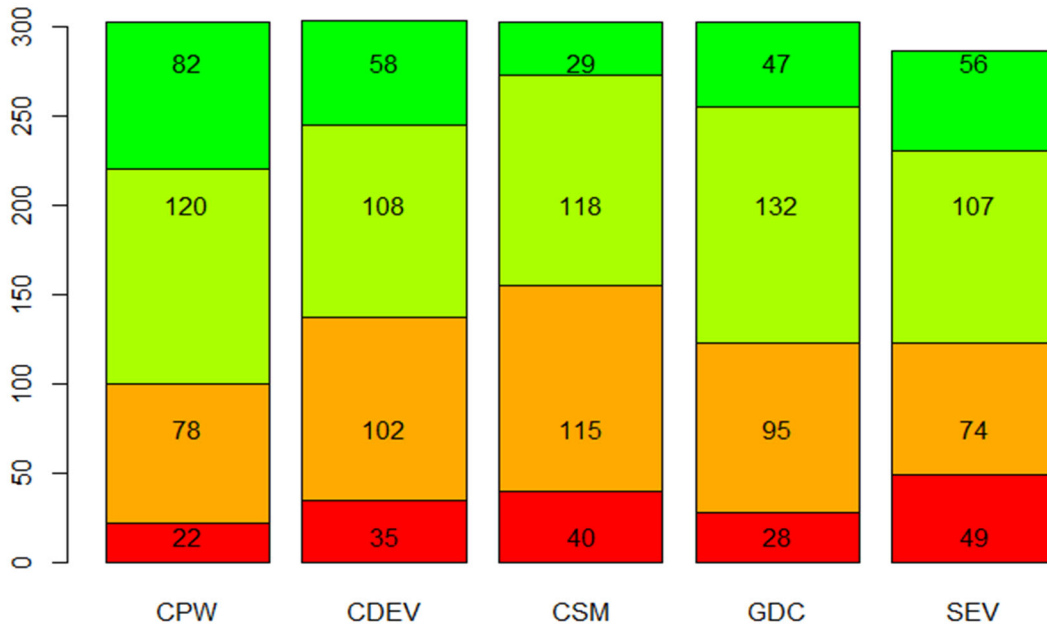
Sources and Evidence



Control of Syntax and Mechanics



Overall

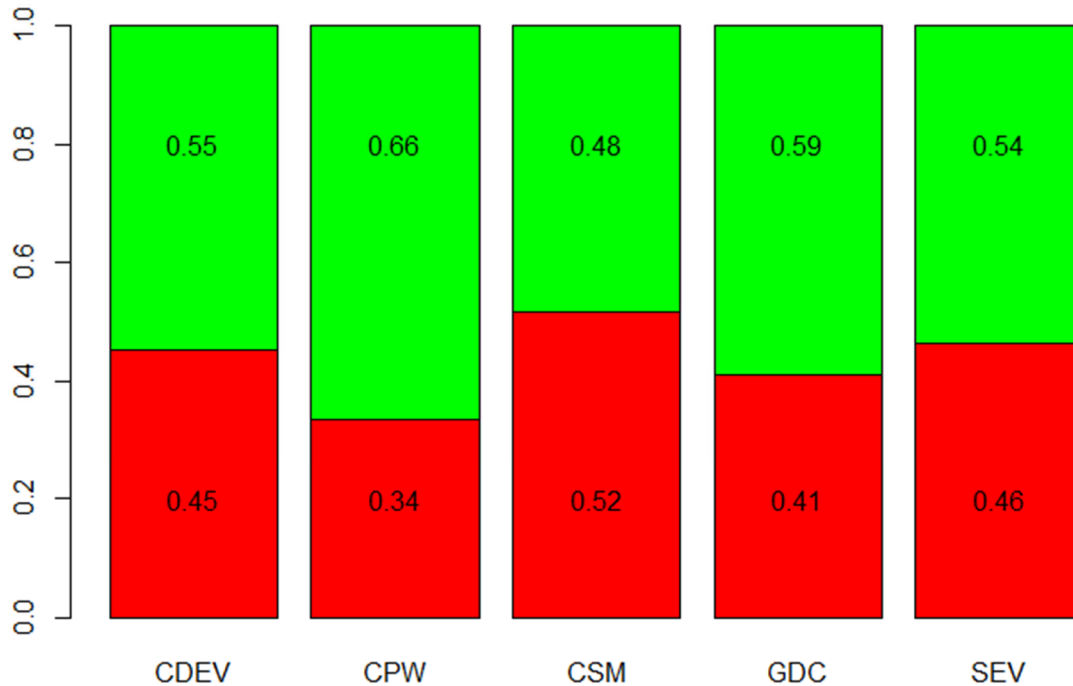


Note that disparity in the total counts from one criterion to the next are a result of scores of "N/A", which are omitted here

General Discussion

In each criterion, the modal score is 3, or proficient, which is encouraging. Further, In every criterion but “control of syntax and mechanics,” the number of papers earning a score of 4 is greater than those earning a score of 1.

If one aggregates the scores on proficient or better (3 & 4) versus less than proficient (1 & 2), the results are as follows:



This perspective makes evident that we are falling short of the expectations we set for ourselves in terms of ILO proficiency by graduation. Our bar for all six ILOs is a minimum of 75% proficiency, and we see in the above that performance is well short of that mark in all categories.

Inter-Criterion Analysis

The content development, genre and disciplinary conventions, and sources of evidence criteria cluster in the mid-50 percent range proficiency-wise. Context and purpose of writing is the bright spot, significantly higher at 66% proficient. Given that proficiency in this area is a precondition to any further success in any of the subsequent stages of writing, this is certainly a good spot for headway, although the results seen there are still short of our 75% benchmark.

The grammar criterion (control of syntax and mechanics) is the low point. Given that grammar and mechanics are more foundational and involve higher order cognition to a lesser degree than the other categories in general, this result may seem surprising. Usage flaws, however, may be more salient, in the sense that they are obvious and difficult to miss than shortcomings in the other criteria, which are

likely more ambient and cumulative in their impact. Still, when reporting on and discussing the results of this assessment at a faculty assembly on February 24th, 2017, the faculty collectively expressed little surprise at this outcome, and a collective consensus that efforts to improve student grammar and syntax need to be ongoing. See “Implications and Loop Closing” below for a discussion of proposed interventions on that front.

In terms of performance at the lowest of the four echelons, sources and evidence outweighs the other categories, including syntax/mechanics. As such, searching and using sources is its own priority as well. Luckily, information literacy will be the focus of the 2017-18 ILO focus.

Comparing Past Performance

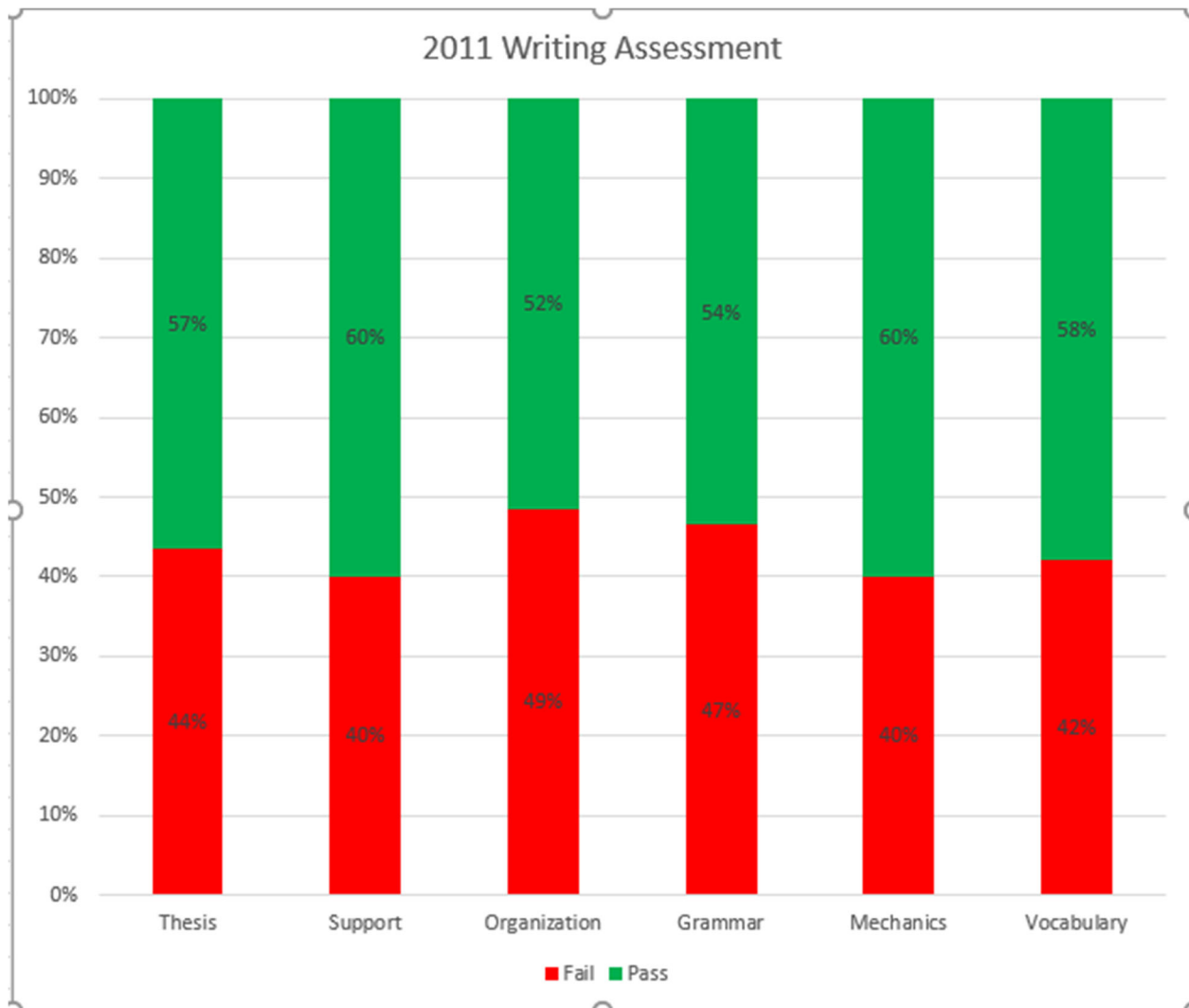
The Written Communication ILO was previously assessed in 2011 for the 2009-10 academic year. A summary of the results of this assessment is available at <https://drive.google.com/file/d/0B7Y7z0UrG0TbVmZraXpsM0RySVU/view?usp=sharing>. Clearly, one of the goals of conducting cyclic ILO assessments is to gauge changes in student achievement over time. Changes in our assessment practices over the intervening period, however, mitigate our ability to make such comparison meaningfully.

For the 2011 assessment, we used both a different rubric and a different assessment scale. At the time, the AACU VALUE rubrics did not exist and we were unable to locate any cross-disciplinary assessment instruments at all. We examined what institutional rubrics were publicly available, and adopted our own instrument.

In 2017, we switched to the VALUE rubric because of the validity and credibility an instrument developed by a large, respected, multi-institutional organization brings. Further, the rubric did align well with what we seek to assess in writing, albeit in a more compact and somewhat more abstract form.

In comparing the two rubrics, one finds the 2011 instrument more granular, more concrete and more specific. (We suspect it is for this reason that the inter-rater reliability of this assessment effort was higher than the one conducted in 2017; further discussion of reliability occurs in a subsequent section of this document.)

Results from the 2010 data are summarized in terms of proficient or better (“pass” here) versus less than proficient (“fail”) are as follows:



Results here are on par with those in 2017, with slightly higher student performance on the whole. Interestingly, scores in mechanics are among the highest (although grammar and mechanics are treated separately and grammar scores are on the lower end) and organization, not addressed explicitly in the VALUE rubric, seems to be a pain point.

Ultimately, however, the only conclusion we can draw confidently from this historical comparison is that differences between the two rubrics preclude valid comparison. While we have not yet conducted a detailed analysis of the two rubrics relative to one another, a casual review reveals that, as is befitting an organization representing higher education broadly and publicly, the AACU's VALUE rubric has quite high expectations for student writing, with its lower scoring echelon ("benchmark") using terms considerably more optimistic than the lowest level in the rubric we used in 2011. For example, the AACU rubric's descriptor for the lowest level score for mechanics reads "Uses language that sometimes impedes meaning because of errors in usage," while the corresponding 2011 descriptor advises the score for papers having "frequent, intrusive errors..that hinder communication." In part this is a function of a

four-tier scale versus six, but it also seems the AACU rubric is simply calibrated higher, depressing the 2017 results relative to those of 2011.

As is discussed in further detail in the “Conclusions & Implications” section that follows, the challenges of inter-rubric comparisons are among the factors that call out for a longitudinal assessment approach.

Validity and Reliability

We are confident in the general validity of the VALUE rubric employed here both because it is a large-scale, collaborative inter-institutional effort by a respected body and also because the AACU itself has undertaken various validation studies (e.g. <https://www.aacu.org/publications-research/periodicals/increasing-validity-outcomes-assessment>).

In terms of reliability, we cite our norming efforts, discussed in the methods section above, as well as the following inter-rater reliability results:

Inter-Rater Reliability

Note that we strive here not for statistical rigor but rather for straightforward, simple to interpret, useful indications of reliability.

The assessment exercise resulted in 1,166 pair-wise rating comparisons to assess for agreement. Of those, 34.7% agreed exactly, 48.3% were off by one order (e.g. 1 vs 2, 3 vs 4), 14.9% were off by two, with the remaining 2.1% of scores being diametrically opposed (1 vs. 4). While these agreement results were well above what one expect to arise from chance (35% exact matching versus 25% from chance, for example), the results leave ample room for improvement. As discussed in the past performance comparison section, we believe that the nature of the rubric may play a factor here. Future analysis will examine the specific rating criteria and student papers that precipitated extreme disagreement to see what might have precipitated disagreement. We’ll look for patterns in specific underlying assignments as well.

Inter-Program Comparisons

By averaging all five criteria together and computing a per-program average, we can obtain a broad perspective on inter-program achievement differences, which are as follows:

Program	N	Overall Mean
Criminal Justice	1	1.7
Information Technology	17	2.4
Human Resources	2	2.5
Business	39	2.6
Integrated Leadership	18	2.7
Accounting	14	2.7
Legal Studies	24	2.8
Health Care	12	3.0

Criminal Justice and Human Resources, with one and two observations respectively, need to be excluded from further analysis entirely. Interpreting the remaining results needs to be done cautiously as well. The fact that the Legal Studies program would rank near the top of the ranking is unsurprising, given that writing is the stock in trade of those in the legal field, but that the Health programs would outscore legal studies is unexpected and interesting. That the Information Technology programs would appear at the bottom of the list is also probably not surprising, given that a larger proportion of its curriculum is dedicated to technical pursuits.

Evaluating Interventions Resulting from 2011 Assessment

Comparing the scores obtained in 2010-11 to those obtained in 16-17 is the most straightforward way to gain some understanding of whether aggregate student performance has improved in the intervening years. Factors compromising our ability to do that confidently were discussed in the “Comparing Past Performance” section above.

In addition, several curricular changes were undertaken as a result of our 2011 findings, but not all students in the 2016 sample were impacted by them. Those changes included adding a second composition course, ENG 103, to the required general education core in lieu of our research and information literacy course, BIS 109. Our sense in making that change was that information literacy skills were best covered in the context of research writing.

Some students in the 2016 sample took the new ENG 103, others took BIS 109 instead, and some opted to do both. [The majority took neither, because the transcript data was drawn exclusively from students' final year of study. Most of the "neither" category probably did take BIS 109, but simply took it more than a year ago. This analysis needs to be redone using full transcripts]

Rubric performance averaged across all criteria per 103/109 status was as follows:

Courses	N	Mean Score
Both Courses	15	2.56
ENG 103 Only	19	2.56
BIS 109 Only	42	2.63
Neither	51	2.73

A preliminary analysis of variance showed no statistically significant differences among these groups but because of the transcript caveat noted above, those results are not reported here and will be replaced when a full transcript analysis is done.

An additional intervention in response to the 2011 assessment, involving a new writing intensive program mandating rigorous writing in a student's major courses, debuted in 2016, too late for students in this sample to be impacted by it. For a description of the Writing Intensive program, see the College Catalog.

Conclusions

It is very difficult to draw comfortable inferences about student performance in 2016 relative to that of 2011 beyond the fact that there is little reason to suspect that performance has improved measurably and that there is little reason to suspect that it has worsened measurably in the intervening period. Our last round of institution-wide writing assessment resulted in 2010 resulted in sweeping curricular changes, some of which impacted a subset of the students in the 2016 sample and some of which had not yet been fully implemented at that time. Ongoing assessment will be required to understand trends in student writing performance over time as well as the impact of the efforts we make to improve that performance.

As our institutional understanding of the practice of learning assessment improves over time, and as higher education as a whole continues to work collaboratively to do the same, it seems certain that our desire to alter our assessment instruments over time will remain strong. Further, even if we were to commit perpetually to a given rubric, incoming student preparedness can and does change over time.

As an open-admissions institution that collects neither high school performance nor standardized test score data, our opportunity for accounting for such variation is limited.

As such, among our most confident conclusions is that SLOA assessment is best done longitudinally, with assessments completed for all students as early in their career as possible and then as close to graduation as possible, and then the extent of gains over time assessed per student. We are in the process of conducting an initial longitudinal ILO assessment in 2017-18.

In addition to the recent implementation of our writing intensive course program, we have several other recent curricular changes, including a tighter and more systematic involvement of library resources and personnel in ENG 103, and improved tutoring resources. Improving student writing remains among our highest priorities and efforts to that end are continuous and urgent.