Peirce College 2017-18 Institutional Learning Outcomes Assessment Report

2017-2018 Goals Review

In sum, the original agenda for SLOAC for 2016-17 was tabled in light of the fact that 2017 marked the last occasion to draw direct evidence (i.e. student papers) from the Pearson / eLearning LMS. This fact made it imperative that we make the best possible use of that resource while it still remained available to us, particularly given that it will take a number of years for us to build up a store of student work equivalent in the new LMS.

1. Evaluate the prospect of revising the Peirce ILO statements - this was tabled in light of the new incoming presidential administration and the impact that resultant changes might have on foundational issues of mission and vision from which such a review should stem.
2. Conduct an ILO direct assessment project - see below
3. Conduct a writing prerequisite review - a review was begun in concert with the General Education division, but not completed. The review will be completed in the 2018 academic year.

Information Literacy Direct Assessment Project

Introduction

Since 2008 it has been Peirce's practice to assess one of its six institutional learning outcomes (available in Appendix A below) each year using direct evidence of student learning. 2017-18 was the year of information literacy assessment, which had last been assessed by similar means in 2011-12. The following summarizes the methods, results and implications of that assessment.

Methods

Sample Identification

First we identified our student cohort of interest, namely, the 127 students graduating with their bachelor's degrees between the dates of May 31, 2017 and 12/31 2017 (the most recent year for which data were available). Next we retrieved the course histories and transcripts of those students.

In parallel, the subset of courses that would yield papers appropriate for assessing information literacy outcomes was identified. See Appendix A for the specific courses.

Next we identified the subset of students who had taken and passed one or more of the courses included in our sample within 18 months of graduation. Then we identified the portion of those students who had also taken and passed one or more of those courses at least two sessions prior to that as well. As a result, we had a sample of recent graduates who had taken an assessable course close to graduation and another at some time at least six months prior to that. (The average age between the early paper and the late one was 2.7 years.) This gave us a set of students and courses that we attempted to then identify sample papers for.

A subset of identified student papers were either unavailable or inappropriate for assessed (if, for example, they had failed the intended paper). Any such student was eliminated from the cohort. A total of 77 students had available papers from both near graduation and earlier in their academic careers. The student IDs of those students are indicated in Appendix B.

The sample included students from each undergraduate division at Pierce, in the following numbers:

Division	N
Business	82
Healthcare	36
Information Technology	14
Legal Studies	22

We had for each student in our sample a close-to-graduation paper and a paper from earlier on. Each paper was anonymized and rated by three faculty assessors. The evaluators were provided with the student instructions as well as any grading rubric associated with each paper they were assigned, but the status of each paper as an 'old' or 'new' one was unknown, as was the specific course for which the paper was completed (outside of any unavoidable contextual clues from the nature of the assignment).

The papers that made up the sample can be found at https://drive.google.com/drive/folders/1atLoi-CBT4bl1TsOiQqaxh6oBwWItQP4 (access restricted).

Evaluation Rubric
With feedback provided by the faculty during a faculty-wide meeting in fall 2017, we developed our information literacy rubric based on a review of the rubric we developed when we last assessed our information literacy ILO back in 2011-12. The resultant rubric can be found online at https://docs.google.com/document/d/1lxvqHw_E3BiAnTghAZF8TSllFAMeCPYK1vuWP9UR1bY/edit?usp=sharing. In reviewing our options for developing an instrument for assessing information literacy, we turned first to the AACU's IL VALUE Rubric (available at https://drive.google.com/open?id=1gTEVd8_b3HjcjfkH4at2uVTjWk74tn6W) , but found large portions of it problematic when the sole source of evidence was a completed research paper. For example, the extent to which a student determined the nature and scope of appropriate information or actually accessed that information successfully or evaluated its utility successfully is evident in a finished research paper in only a very roundabout and indirect way. Thus we sought more concrete, prescriptive guidance in the rubric we developed ourselves. We also found considerable inspiration in the AACRL's excellent information literacy framework (http://www.ala.org/acrl/standards/ilframework) and the rubrics proposed for it (i.e. http://sandbox.acrl.org/library-collection/acrl-framework-rubric), but found the principles there too abstract and process-directed to be reliable in assessing the end product of the research process.

Reviewers were drawn from across the Peirce full-time faculty and the professional library staff. Results were gathered via the Google form available at https://docs.google.com/forms/d/e/1FAIpQLSeqJ48mSEO275SXCw_wszxT3n7p4e95voHOL8NoMMtRXezbGQ/viewform . Each reviewer read and scored either 15 or 16 papers total.

Norming
All evaluators participated in a norming exercise prior to conducting their actual assessments. For the norming exercise every evaluator assessed the same two papers and their decision making was reported back to them relative to the judgment of the group as a whole in a personalized scoring report. We then discussed impressions and suggested final revisions to the rubric in light of the insight gained from the norming. A sample of the norming report provided to all assessors can be found at https://docs.google.com/document/d/1QnerKZ76gl_N4emmWDkK4WQ0_3Dvk2P60-OSVY6A6fs/edit?usp=sharing and contains a discussion of the extent of our inter-rater agreement for the norming exercise.

Agreement results at the time of norming:

The following table summarizes the results of ever possible instance of pairwise agreement between all of the raters, each to the other:

Extent of Agreement	N	Percent of total
Perfect Agreement	1,531[1]	34% (25% expected by chance)
Off by 1	2,117	48% (37.5% expected by chance)
Off by 2	710	16% (25% expected by chance)
Perfect Disagreement	94	2% (12.5% expected by chance)

Scores of 'N/A', reported by two raters for the plagiarism category, are omitted here

While we did significantly better than would be expected by chance, it is difficult to claim that we did as well as one would like. Our main hope was to have better agreement *after* the norming exercise, when doing our actual assessment, than we did for the exercise itself. As discussed in our results section in detail, the norming debrief, together with suggestions for clarifying our rubric descriptors and an analysis of discussion comments submitted during the norming session, we hope, helped achieve this end.

Results

Before presenting the actual information literacy learning outcome performance reported for the student sample it is important to gain context by analyzing several underlying variables of our cohort. Those variables will be summarized in the following tables:

Career GPA Distribution

---
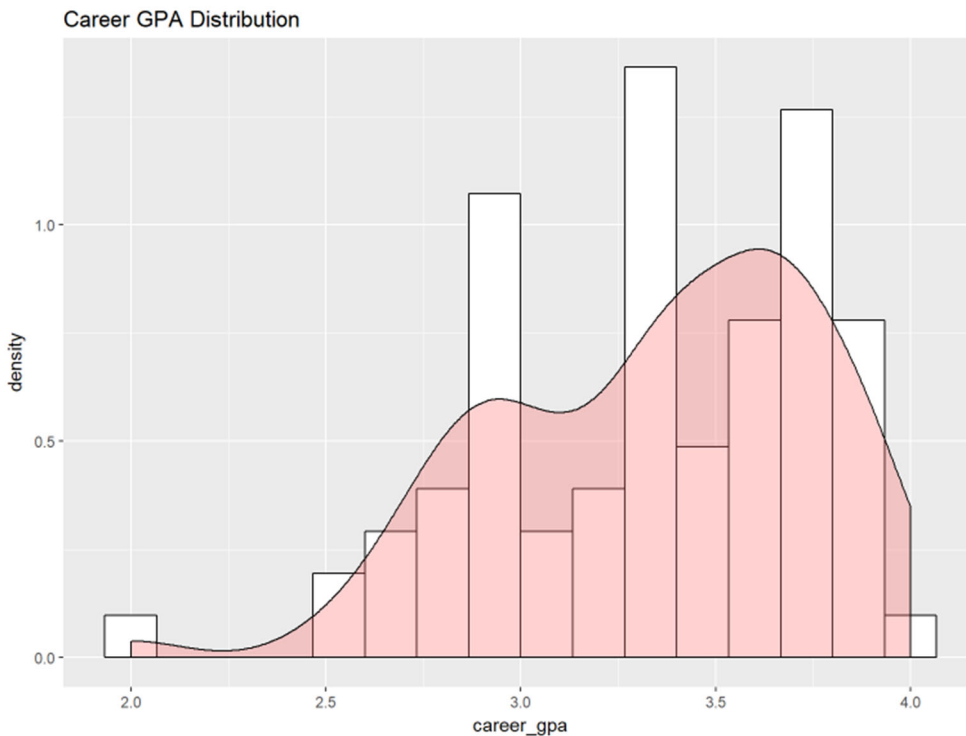
[1] I know these numbers look suspiciously huge, but pairwise-comparison of 26 raters involves 325 rater pairs times two papers times 7 criteria per paper equals 4,550 comparisons.

Outside of being skewed left by the minimum career GPA graduation requirement of 2.0 and having a perhaps surprisingly high mean and mode, there is nothing especially noteworthy here. When we bring focus to the grades earned on the specific courses from which the assessed papers were drawn, however, the shape of the distribution changes markedly:

Sampled Course GPA
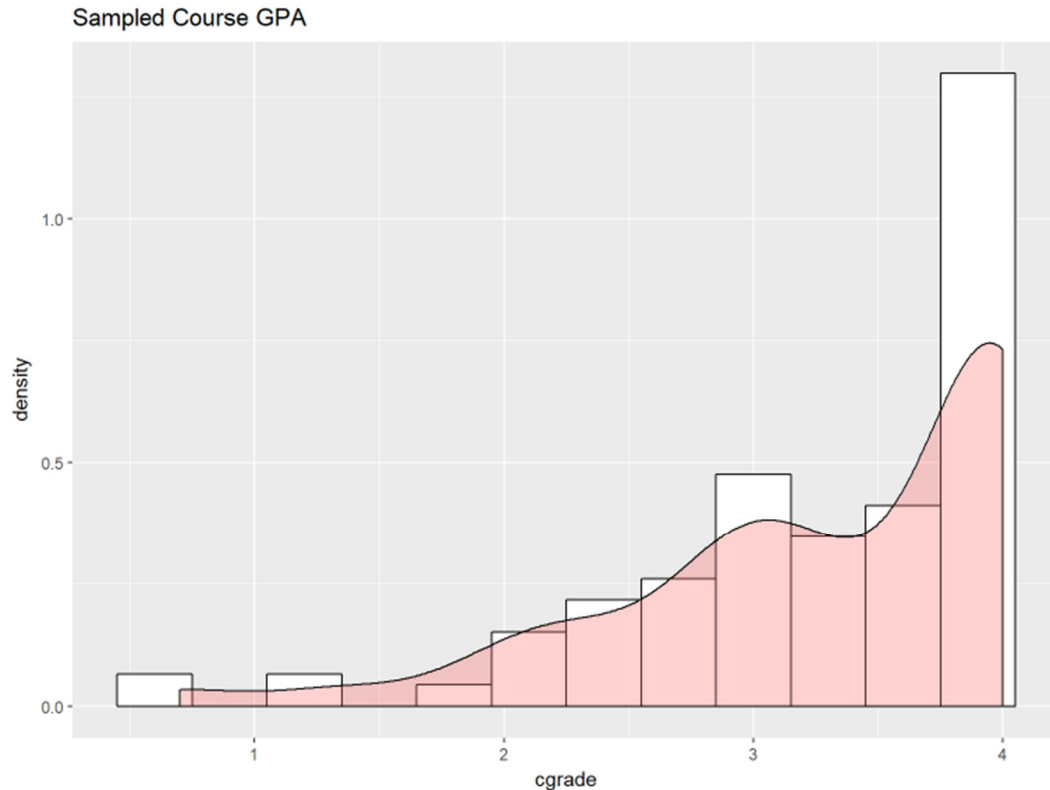
The modal grade for sampled students in the courses from which we drew sample papers was 4.0 or 'A', with almost every lower grade being less likely than the last. Even in light of the fact that we disqualified both courses and papers that the student cohort failed, this seems like an inexplicably strong collective performance. It is hard not to conclude that students excel in those courses for which research writing is central; this outcome seems counterintuitive knowing anecdotally the challenges student have with writing and especially with research writing.

An analysis of the grade distribution of the actual papers reviewed shows a very similar distribution, with more students receiving a grade of 'A' than all the other grades put together (56% 'A's), and one in four papers receiving a numeric grade of 100%! At the risk of editorializing, this represents nothing less than a dereliction of our collective duty as faculty, in addition to setting a spectacularly high expectation bar for the ILO performance of these papers in light of their stellar scores. It is difficult to even conceive of what a literally perfect research paper might mean or what an appropriate standard for such an outcome might look like.

Assessment Results

We start with an analysis of the scores from the papers prepared closest to graduation. This data subset parallels data collected in previous years and thus allows the best comparability to the results seen during our last information literacy ILO analysis conducted in 2012-13. Additionally, it reflects an estimate of the student cohort's performance very close to graduation, a desirable assessment locus.

2017-18 Results Summary (N = 196)

| Criteria	Mean	Median	Mode	Std. Dev	Score of 4	Score of 3	Score of 2	Score of 1	Score of N/A	Proficient+
Information Integration	2.5	2	2	0.87	15%	34%	41%	10%	0%	49%
In-Text Citation	2.6	3	3	0.95	17%	42%	25%	16%	0%	59%
Plagiarism	2.9	3	4	1.04	34%	27%	18%	12%	9%	61%
Support of Argument	2.6	3	3	0.82	12%	48%	30%	10%	0%	60%
Source Quality	2.6	3	3	0.84	12%	43%	34%	11%	0%	55%
Works Cited	2.8	3	3	0.92	22%	42%	23%	11%	1%	64%
Overall Score	2.6	3	3	0.83	11%	43%	35%	11%	1%	54%

Score Key:
4 (Green) - Exemplary
3 (Yellow) - Proficient
2 (Orange) - Developing
1 (Red) - Emerging
N/A (blank) - No score submitted

2017-18 Scores Summary by Criterion, Color-Coded
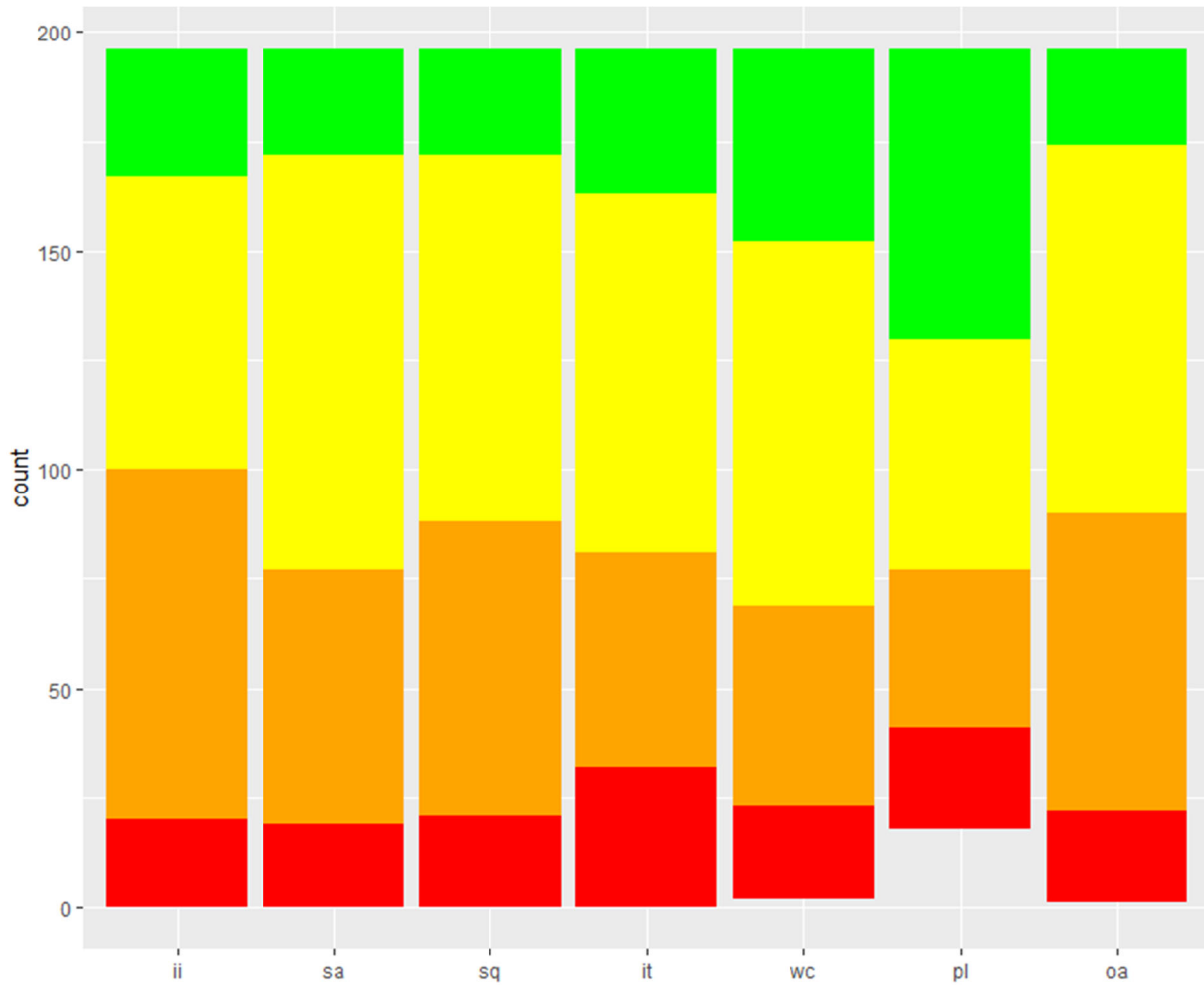
ii - Integration of information
sa - Support of Argument
sq - Source Quality
it - In-text citation
wc - Works cited
Pl - Plagiarism
oa - Overall

Results

Absolute scores

Peirce has established an ILO performance benchmark of 75% proficient or better.   Overall scores here fall well short of this mark at 54% proficient or better. Similarly, so does each individual scoring criterion, ranging from a low of 49% for integration of information to a high of 64% for works cited. The plagiarism criterion is noteworthy for its high proportions of top scores (34% Exemplary), although its binary nature and its moral imperative make this something of a modest win. In-text citation stands out from the other criteria with a higher than normal proportion of the lowest score (emerging), with 16% of

scores in that category relative to the the 10-12% range for all other categories. This appears to reflect a persistent reluctance even on the part of students close to graduation to include in-text citations to the work of others they employed in their papers. This is likely an area where it is possible to move the needle by repeated communication of the necessity of such citations wherever research writing is required, including upper-level courses where the need might be seen to go without saying, together with grading expectation adjustments that make it impossible for students to pass such assignments without including in-text citations consistent with our minimum graduation expectations.

Scores compared to past assessment results
The following table summarizes results from the most recent prior assessment of information literacy, conducted in 2012:

2011-12 Information Literacy ILO Assessment Results Summary

Criteria	Mean	Median	Mode	Std. Dev	Score of 4	Score of 3	Score of 2	Score of 1	Score of N/A	Proficient+
Information Integration	2.2	2	2	.95	27%	32%	31%	9%	0	40%
In-Text Citation	2.1	1	2	.94	33%	32%	28%	7%	0	35%
Plagiarism	2.4	2	2	1.08	26%	28%	25%	20%	0	46%
Support of Argument	2.4	3	3	.94	24%	25%	42%	9%	0	50%
Source Quality	2.3	2	2	.86	19%	42%	31%	8%	0	39%
Works Cited	2.5	3	3	.94	17%	32%	36%	15%	0	40%
Overall Score	2.2	2	2	.89	23%	38%	31%	8%	0	39%

* percentages may not sum due to rounding

The method for this study was very similar to that employed for the current year's study (although the 2012 study was not longitudinal, and all student papers assessed were drawn as close to graduation as practicable). While the rubric criteria were the same for 2012 and 2018, note that the language in the descriptors varied in certain cases.   (The 2012 version of the rubric can be viewed at https://drive.google.com/open?id=0B7Y7z0UrG0Tbc3loUE81aXFjTm8 .) Indeed, revisions to the rubric that occurred between the current assessment study and the previous one were explicitly precipitated by observations regarding the original rubric that came from the debrief we conducted with the

assessors after the initial study. For example, assessors in 2012 expressed puzzlement regarding the fact that the "support of argument" category was the category in which students performed best. Given that employing sources in support of one's argument seems qualitatively more demanding than documenting the use of those sources, this was seen as a counterintuitive outcome. A review of the rubric descriptors for that criterion made it clear that in drafting the instrument, the SLOA Committee had taken that fact into consideration ex ante and crafted low expectation descriptors. (This discussion goes to show how difficult it is to draw meaningful intra-categorical comparisons, never mind inferences). The integration of information descriptor language was thus revised to bring it more in line with the absolute expectations per performance echelon for the rest of the criteria.

Whatever the caveats, performance in 2018 is sufficiently better than that of 2012 to be strongly encouraging:

Comparison of 2017-18 Results to those of 2011-12

Criteria	Proficient + in 2018	Proficient + in 2012	Percentage Improvement
Information Integration	49%	40%	+9%
In-Text Citation	59%	35%	+24%
Plagiarism	61%	46%	+15%
Support of Argument	60%	50%	+10%
Source Quality	55%	39%	+16%
Works Cited	64%	40%	+24%
Overall Score	54%	39%	+15%

* percentages may not sum due to rounding

Scores are up in all criteria and by an average of 16%. It is difficult not to see this performance change as a strongly encouraging sign for the curriculum changes undertaken in the intervening period to improve student performance in information literacy.

Given the slight methodological differences, different rater sets and different scoring instruments employed between the two studies, however, ample potentially confounding factors exist, which is why in the current year's study we opted to design for a longitudinal analysis. As described in the methods section above, each student in the assessment sample had a paper drawn from early in their studies at Peirce as well as late. Papers were scored by the same raters using the same rubric, and were anonymized in terms of student, class and date, so raters had no basis for knowing if they were scoring

an early or a later paper. We believe this longitudinal design creates the conditions for an unbiased comparison that resists gaming, conscious or otherwise.

Longitudinal 2017-18 Results

Criteria	Near Graduation Papers		Early Career Papers		New - Old
	Mean	Std.Dev	Mean	Std. Dev	Difference in Means
Information Integration	2.54	0.95	2.40	0.88	+0.14
In-Text Citation	2.59	0.94	2.36	0.98	+0.23
Plagiarism	2.91	1.08	2.68	1.08	+0.23
Support of Argument	2.63	0.94	2.62	0.85	+0.01
Source Quality	2.57	0.86	2.61	0.89	-0.05
Works Cited	2.77	0.94	2.59	0.94	+0.18
Overall Score	2.55	0.89	2.50	0.84	+0.05

All of these categories except for "Source Quality" show improvement over time, and many of the categories (especially "In-text Citation", "Plagiarism", and "Works Cited") show improvement that informally appears as though it could be of practical significance.

Are the gains from old to new significant statistically?
Before we consider the practical significance of the difference, we need to assess statistical significance by examining how unlikely it would be for such differences to arise solely by chance. To do so we employ the nonparametric Wilcoxon rank sum test with continuity correction rather than a simple t-test because our data here are ordinal and thus not normally distributed. While the experiment lends itself to a paired test, because a few assessments were left uncompleted, together with the fact that any score of "N/A" needed to be removed ffrom the tested sample and thus resulted in uneven sample counts, we apply an unpaired version of the test. Results are as follows:

H_0: True location shift from early sample to recent = 0
H_a: True location shift from early to recent is greater than 0
W = 981320, p-value < .001

We would expect results like these to arise from chance less than 1 time out of 1,000 and thus find the difference to be strongly statistically significant. Given the ordinal scale of the data, assessing practical significance is challenging, but imagining an improvement on average that moved performance just shy of one quarter the way from "Developing" to "Proficient" (whatever that could mean, and with the caution that it is an interpretation fraught with peril) seems a valid accomplishment over the roughly 2.7 calendar years that on average separated the two paper's completion dates. (A better yardstick would have been credit accumulation rather than time passed, although these data were not available for this sample. Peirce students on average accumulate roughly 14 credits per year, so this equates to 40 or so credits, although credit pace varies widely from one student to the next.) If we take an even more dubious leap and imagine the benefit to ILO achievement to accrue uniformly per credit, we can imagine a gain just shy of three quarters of the way from one performance echelon to the next for differences on the magnitude of those for in-text citation and plagiarism from enrollment to graduation for a student completing a 120-credit bachelors. While a minimum benchmark intention would be to take students from "Developing" to "Proficient" (or indeed, from wherever they present at initially through to somewhere proficient or better), this result still seems encouraging. These figures are entirely speculative, of course.

Further, if we create a composite index of assessment scores and calculate the percentage change old to new and then plot that change as a function of time, we see the following:
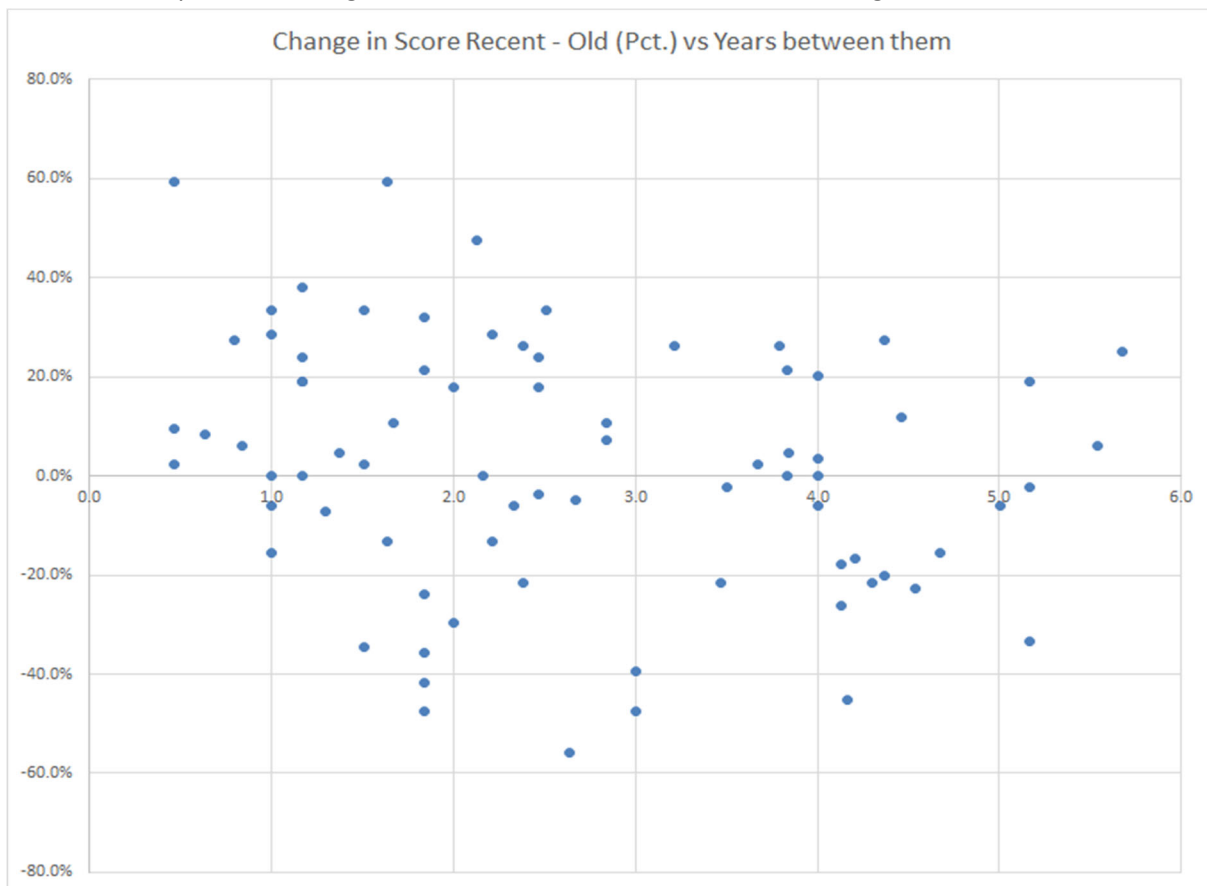
While it is statistically negligible, it is surprising to see a negative correlation between the among of time between recent and old classes and the extent of improvement (r = -0.22). It is also worth noting that while students with less than 1 year between papers showed 100% improvement old to new, those with 4 or more years between only improved 40% of the time. It is possible that there is a degree progress rate below which it becomes difficult to benefit from prior studies, perhaps as a lack of momentum makes building on past learning challenging. This certainly would not be surprising after some point. Efforts to improve time to degree are ongoing independent of this result in any case.

Threats to the Validity of Longitudinal Results

Likely the biggest factor muddying the results here is the fact that the courses drawn from for the recent papers are largely different from those for the early sample. This is of course not at all surprising; indeed the extent of the overlap between the two is more noteworthy:

Sample Paper Composition Per Course, New versus Old

(Courses with overlap are highlighted in yellow)

Course	% New Sample	% Old Sample
ACC101	0%	2%
ACC222	0%	2%
ACC450	4%	0%
BIS109	0%	16%
BUS100	0%	11%
BUS250	0%	8%
CJS101	0%	3%
CJS201	0%	3%
CJS401	2%	0%
COM312	33%	0%
ENG103	1%	21%
HCA210	0%	7%
HCA320	3%	2%
HCA480	5%	0%
HIA310	9%	0%
HIA360	0%	2%

|		
HRM201	1%	9%
HRM306	1%	0%
HRM350	5%	0%
HRM425	2%	0%
HUM102	3%	5%
HUM275	12%	0%
ISC325	3%	3%
ISC425	0%	2%
LAW103	3%	2%
MGT425	10%	0%
MIS205	2%	2%
MKT401	3%	3%
SCI220	0%	4%
Grand Total	**100%**	**100%**

If capstone courses, or higher level courses generally, motivate students to do research at levels not expected for lower-level work, this could skew longitudinal comparisons. This would be a larger concern if there were more top echelon (i.e. "4") scores, however.

Examining Grades between the new and old paper samples
While it is difficult to explain exactly how grades earned on the papers comprising the samples intersects with expectations regarding assessment results, it is nonetheless worthwhile to examine the two samples for differences in grade means, as well as to examine the relationship between grades earned and assessment scores more generally.

The average paper score for the early subsample was 90% while the score for the recent papers was 85%. This does little to confound the improvement in the more recent papers, however, given that one expects lower grades to correlate with lower outcome achievement. And outside of longitudinal comparison issues, our expectations of graduates are irrespective of their grades, provided that those grades are adequate to graduate.

In fact, the grade assigned to a paper and its average score in the assessment study had no obvious linear correlation at all (correlation coefficient r = .093). The lack of relationship is borne out visually in a scatter plot of the two variables:
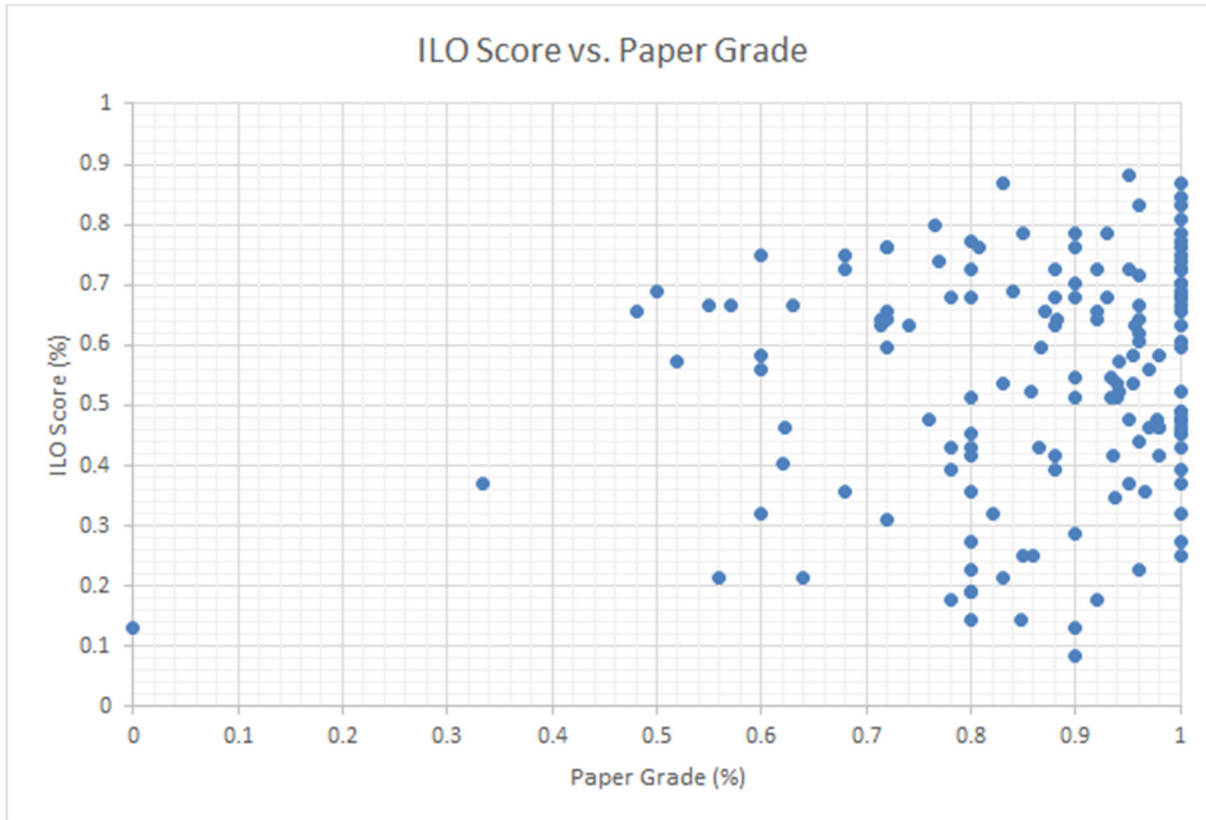
Note that this chart shows failing paper grades because, while we excluded from the sample any class that graduates had failed, there were several instances of students who failed the research paper in a candidate class but managed to pass the course nonetheless. Such students are included in the sample here.

Papers with grades of 'A' are spread widely across the spectrum of ILO score results, and do not look much different than the spread of scores of those papers with much lower grades.

Inter-Rater Reliability Pre- and Post-Norming Comparison

In addition to assessing student ILO performance data, it is also useful to examine rater performance in the main assessment study relative to that prior to our norming efforts. The hope here is that agreement is higher after the norming exercise than it was before

Pairwise Extent of Agreement	Before Norming (N=4,452)	After Norming (N = 2,277)
Perfect Agreement	34%	36% (+2%)
Off by 1	48%	49% (+1%)
Off by 2	16%	13% (-3%)
Perfect Disagreement	2%	1% (-1%)

* Scores may not sum to 100% due to rounding. Note that the N for the actual assessment study is significantly lower than for the norming session because, although there were many more papers assessed, those papers were each only read by 3 assessors, while for the norming exercise al papers were scored by all assessors, resulting in more pairwise comparison opportunities.

The agreement after the norming exercise is improved over that seen before, albeit fairly modestly. Seeing major disagreement (non-adjacent scores, i.e. those off by 2 or more) reduced from 18% to 14% seems like an encouraging result consistent with good faith efforts at reliable scoring.

Key Takeaways

1. Absolute ILO student performance at or near graduation is not yet where we want it.
   Our institutional benchmark for proficient-or-better scoring percentages is set at 75%. We landed generally in the 50 - 60% range for the individual information literacy criteria, and at 54% for the summative "overall" category.
2. There is credible evidence that we are improving over time.
   Through a comparison of the results from this year's study to a very similar one conducted in 2012, we have an indication that our success in instilling information literacy outcomes in our students is improving over time.
3. Grading is grossly misaligned with outcome achievement.
   56% of the sampled papers received a grade a "A". Over a quarter received a grade of 100%, defying the implausibility of a perfect research paper. Yet under 17% of the scores assigned for the ILO assessment study were in the "Exemplary" echelon. Grades are the primary means by which we communicate to students how well they are achieving the outcomes we have set out for them. We are misleading students through these grades. We are telling them, 26% of the time, that their work is without consequential room for improvement, yet we are assessing them short of the top tier in their information literacy outcome achievement. While this outcome is not surprising (it is harder to give discouraging news to students, plus faculty members are justifiably reluctant to disadvantage students with poor grades), it is troubling.

Action Items in Light of these Assessment Results

While the results from the assessment study detailed herein are on the whole encouraging, there are nonetheless several action items that result from the study and the discussions faculty have had around them. They include the following:

1. Incorporate an introduction to research skills into BIS 111.
   BIS 111 is required of all students and is typically taken in student's first session at Peirce. Given that it covers word processing skills, it is an ideal opportunity to introduce information literacy ideas and techniques, particularly related to the mechanics of in-text citation and works cited documentation.
2. Review prerequisite requirements to make sure that ENG 103 is required for all courses requiring research writing.
   While BIS 111 is a useful locus for introducing basic information literacy concepts and techniques, ENG 103 is the main event when it comes to achieving baseline outcomes in information literacy and research writing. No course that requires substantive research should allow students without the foundation of ENG 103 or its equivalent
3. Conduct during the 18-19 academic year a faculty-led professional development session dedicated to a discussion of grading and its implications.
4. In a year's time, revisit the grading patterns for the courses that made up the basis for the sampled papers here to check for a change in grading habits, particularly with regard to the assignment of grades of 100%.
   The idea that work in these courses or indeed any with significant research writing is regularly perfection is misleading and likely demotivating to students. Until our ex post assessment studies reflect such perfection, we should work to avoid this situation.
5. In accordance with our ongoing ILO assessment agenda (noted in the table below), focus on ILO 2. Given that when we last examined this outcome (in 2014-15) we investigated critical thinking, our focus over the coming year regarding ILO 2 will be on quantitative literacy skills.

ILO Assessment Timetable

Year	ILO
2018-2019 (current year)	1. Communicate clearly and effectively both orally and in writing
2019-2020	2. Solve problems using critical, analytical, and quantitative skills
2020-2021	4. Demonstrate information literacy
2021-2022	5. Use information technology proficiently and responsibly
2022-2023	6. Identify and respond to ethical issues in the workplace and the community

(Note that ILO 3, Demonstrate up-to-date knowledge, skills, and methods in one's discipline, given that it is by its nature specific to the specific degree programs' respective disciplines, does not lend itself to institution-wide assessment and is as such omitted from institution-wide assessment. Program level assessment happens in all degree programs every year.)