Assessment Report - University Studies Cluster 2

Student Work Collection

2013-2014

This report describes the collection and analysis of student work in Cluster 2 that was conducted in 2013-2014. This assessment project was approved by the General Education Committee and follows the timetable for University Studies assessment approved by the Faculty Senate in 2010.

Courses, Sections, Seats and Assessment Participation

Cluster 2A

Course	Name	Seats	Sections Offered	Sections Providing Data
BIO 101	General Biology I	413	5	2
BIO 103	Topics in Biology	429	8	1
BIO 112	The Ocean Environment	107	1	1
BIO 122	Biology of Organisms II	157	1	1
BIO 221	Anatomy & Physiology	168	2	
CHM 130	Chemistry and the Environment	201	3	1
CHM 132	Chemistry and Nutrition	98	1	1
CHM 155	Modern Chemical Principles I	49	2	1
MAR 110	Natural Hazards and the Ocean	92	2	1
MLS 105	Contemporary Topics in Human Ecology I	205	1	1
MLS 211	Introduction to Human Physiology	56	1	
PHY 101	Introduction to Physics I	139	2	
PHY 102	Introduction to Physics II	89	1	
PHY 111	Physics for Science and Engineering	179	3	
PHY 112	Physics for Applied Science and Engineering II	143	3	
PHY 114	Classical Physics II	151	2	
PHY 115	Introduction to Classical Physics*	19	1	
PHY 151	Introductory Astronomy	196	5	1
PHY 171	Planet Earth & Its Resources I	38	2	
PHY 182	Introduction to the Weather	63	3	

Cluster 2B

Course	Name	Seats	Sections Offered	Sections Providing Data
BIO 102	General Biology II	145	3	2
BIO 108	Cancer Biology	141	3	1
BNG 162	Current Topics in Bioengineering	30	1	
BNG 255	Biology for Engineers	71	2	
CHM 170	Chemistry of the Elements	29	1	
CIS 381	Social and Ethical Aspects of Computing	39	1	
IST 111	Science of Kriyayoga	26	1	1
MLS 106	Contemporary Topics in Human Ecology II	191	3	
MNE 220	Thermodynamics I	93	1	
NUR 105	Human Nutrition	276	5	
PHY 115	Introduction to Classical Physics	19	1	
PHY 162	Science Technology & Society II: Environment	173	2	

Discussion

In 2013-2014, there were 73 sections offered across 32 different courses in Cluster 3 with over 4,200 total seats available. However, participation in the student-work collection assessment was low. Overall, only 21% of the sections in Cluster 3 (15 out of 73) participated in the assessment. Though this participation rate is low, the sections account for a large number of students. The participating sections in 2A enrolled just over 1,000 students, and the 2B sections enrolled just over 200. Of course, not all enrolled students in these sections participated in each assignment for which we collected student outcomes. That being said, the sample size for 2A is large enough to make relatively sound inferences to the full undergraduate population, assuming it is representative. The 2B sample, on the other hand, carries with it a relatively large margin of error.

Of course, calculating sampling confidence intervals assumes random sampling. Students were not chosen at random, because the process was designed to collect all students work (that is, it was not designed as a sampling process). Our sample resulted from the decisions of certain faculty members to participate (or not). The sample of work we ended up collecting, however, could be reasonably used to make inferences about the population to the extent that the sample could be viewed as lacking any obvious source of bias. One potential reason for bias comes from the fact that participation rates were not equal across departments, as the tables above make clear. Biology and Chemistry were the most likely to participate, and their sections provide the bulk of the data. None of the courses in Engineering participated, except for one Physics section.

Do these patterns of participation suggest a bias? The biology and chemistry (and physics) courses that participated tend to be taken by students in non-STEM majors who are strictly meeting a University Studies or College Studies requirement. The courses that are typically used by STEM students to meet their Cluster 2 requirements (e.g., MLS 211, PHY 111, MNE 220, BNG 255) were

not included in the data. If one assumes STEM students will generally perform better on Cluster 2 outcomes than non-STEM students (a reasonable assumption), then the participation patterns suggest the data will underestimate how well our student body as a whole performs on these outcomes.

Assessment Results

Cluster 2A

Section #	LO 1	LO 2	LO 3	LO 4	LO 1 data	LO 2 data	LO 3 data	LO 4 data	Notes
1	78	84	76	92	2 MC	5 MC	2 MC	1 MC	
2	80	77	63	72	9 MC	4 MC	3 MC	3 MC	
3	56	45	65	53	1 SA	1 MC	3 SA	2 SA	
4	63	83	92	84	58 MC	1 MC	2 MC	4 MC	
5	68				25 MC				
6	60				40 MC				
7	76	66	51	59	3 MC	3 MC	4 MC	5 MC	
8	73		73	86	8 MC		4 MC	1 SA	
9	22		56	25	1 SA		4 SA	2 SA	N = 3; based on homework questions, worth 0-3 pts
10	63	97	43	60	7 MC	1 MC	2 MC	1 SA	
11	76	69	54	59	3 MC	3 MC	4 MC	5 MC	
MEAN	65	74	64	66					
1*	45	42	33	58	2 MC	5 MC	2 MC	1 MC	Pre-test data for Section 1

Section #	LO 1	LO 2	LO 3	LO 1 data	LO 2 data	LO 3 data	Notes
1	88	96	80	1 SA	1 SA	1 SA	data are instructor scores
2	54	78		3 MC	5 MC		essay for LO 3 was not collected
3	57	77		3 MC	5 MC		essay for LO 3 was not collected
4	60	50	97	poster	essay	essay	data are instructor scores on relevant criteria
MEAN	65	75	89				

Cluster 2 Learning Outcomes

A. Science of the Natural World

After completing this course, students will be able to:

1. Recount the fundamental concepts and methods in one or more specific fields of science.
2. Explain how the scientific method is used to produce knowledge.
3. Successfully use quantitative information to communicate their understanding of scientific knowledge.
4. Use appropriate scientific knowledge to solve problems.

B. Science in the Engaged Community

After completing this course, students will be able to:

1. Analyze and evaluate the use of scientific information in the context of social, economic, environmental or political issues.
2. Apply scientific theories and knowledge to real-world problems.
3. Effectively communicate scientific information in writing.

Discussion

The data in the tables above reflect the average student performance on each learning outcome (LO 1, LO 2, etc.) in each section for which assessment data were collected. The sections have been placed in random order and assigned an anonymous section # for reference. Different courses provided different kinds of data to document student learning, and the "LO data" columns provide a key to what was used. Here, MC = Multiple Choice and the number represents how many questions were used; SA = Short Answer, which is used to describe any assignment or question that involved students writing an answer which was scored by the instructor and, again, the number represents how many questions were used. For MC data, the LO scores express what percentage of the students, on average across all the questions, answered correctly. For the SA data, the LO scores represent what percentage of the available points, on average, students received. Instructors worked with the Cluster 2 Coordinator, Tara Rajaniemi, to identify which questions mapped to each Learning Outcome.

For example, in Section #8 for Cluster 2A, there were 8 multiple-choice questions on the exams that measured Learning Outcome 1. On average across these questions, 73% of students answered correctly. Learning Outcome 4 was assessed with one short answer, and students on average earned 86% of the points on this question.

Before analyzing the data and drawing some conclusions, it is important to point out some of the limitations of the data. First, as noted in the first section of this report, all the numbers may be lower than the true population values because of the underrepresentation of STEM students in the sample. In addition, it is clear that for some sections we are relying on very slim pieces of evidence to assess student learning on these outcomes. Most outcomes were assessed with three or fewer questions. This may produce problems with both reliability and validity of the measurement—for example, if one question is very easy or very difficult, the data will not represent the true level of

competency among students in that section. Also, many sections did not provide student work to assess some of the outcomes. This was particularly problematic for LO 2 in 2A and LO 3 in 2B. Finally, there is an issue of comparability across sections. If one instructor provided student outcomes from easy questions and another from very difficult questions, or one graded short answer responses more harshly than another, there may be an incommensurability problem (for example, LO 4 in 2A ranged from 25 to 92—this range is unlikely to represent true variability in students' competency in these two sections).

With these caveats in mind, what do the assessment data tell us about student learning in Cluster 2? First, overall performance on each outcome appears to be passable, if barely so. The means for Cluster 2A Learning Outcomes—65, 74, 64, 66—would fall in the D or C category on most grading metrics. Cluster 2B means are similar, with the exception of LO 3, which is much higher, but based on only two data points. Adjusted upward to account for the sampling bias, it might be fair to conclude that students did moderately well on most outcomes.

It is perhaps more interesting to observe some of the intra-sectional patterns. Most notable is that in six of the nine 2A sections for which more than one outcome was assessed, LO 3 had the lowest (or tied for lowest) scores. This outcome deals with the use of quantitative information in scientific contexts. This pattern suggests courses and faculty in 2A ought to focus more attention on improving student learning in this area. In contrast, LO 4 was never the lowest score in any section, suggesting our 2A courses are doing relatively well at fostering students' scientific problem-solving skills.

With far fewer data points, the patterns in Cluster 2B are more difficult to discern. If anything, the data suggest the weakest area may be LO 1, which involves the analysis and evaluation of scientific information in broader contexts. If Cluster 2A faculty should focus on any area for improvement, the data suggest this may be the place to put their efforts.

Assessment Report - University Studies Cluster 3

Student Work Collection

2014-2015

This report describes the collection and analysis of student work in Cluster 3 that was conducted in 2014-2015. This assessment project was approved by the General Education Committee and follows the timetable for University Studies assessment approved by the Faculty Senate in 2010.

Assessment Methods

Different approaches were taken to sampling student work in 3A and 3B. A description of the sampling process is presented in each section below, along with summary data about the courses, sections and seats involved in the collection of student work.

Cluster 3A

The collection of student work in 3A presented an unusual circumstance, because virtually all the enrollments in 3A occur in courses from a single department, English. Consequently, it was possible to work with the Chair and the literature faculty in English to devise an assessment process that relied on the faculty teaching the 3A courses. The literature faculty expressed a preference to score their own students' work themselves, rather than pass it along to anonymous readers, and were willing to take on the work of completing rubrics to do so.

To choose a sample, we focused on the population of 3A enrollments in Fall 2014. In that semester, there were 878 students enrolled in 26 sections (and 2 cross-listed sections) of 4 different courses (it is important to note that one of those courses, ENL 200, is a variable topics course, so the 20 sections of ENL 200 encompass distinct syllabi).

Because faculty were scoring the work themselves, we decided to draw a sample in advance by choosing 200 students at random from among the 938 enrolled. Each instructor was provided a list of students in their course(s) who had been chosen for the sample. They were provided a rubric and asked to return the completed rubric along with copies of the student work. The components of the rubric were taken directly from the 3A learning outcomes. Each item of student work was single-scored by the instructor.

In total, completed rubrics were turned in for 121 students out of the 200 sampled. The following table provides a breakdown of participation by section.

Dept	#	Section	Title	Class Topic	Total Enrolled	Drawn for Sample	Total Scored
AAS	200	01	Studies In Literature	Hip-hop Cinema & History	10	1	0
ENL	200	01	Studies In Literature	Award Winning Literature	40	12	11
ENL	200	02	Studies In Literature	Award Winning Literature	37	9	9
ENL	200	03	Studies In Literature	Award Winning Literature	40	7	7
ENL	200	05	Studies In Literature	-	35	13	13
ENL	200	06	Studies In Literature	-	40	6	5
ENL	200	07	Studies In Literature	-	37	7	6
ENL	200	08	Studies In Literature	-	40	8	8
ENL	200	10	Studies In Literature	The Beat Generation	36	10	12
ENL	200	11	Studies In Literature	American Comedy & Satire	40	7	0
ENL	200	12	Studies In Literature	Place and Space in Lit	35	10	10
ENL	200	13	Studies In Literature	The Journey	40	8	8
ENL	200	15	Studies In Literature	French Philosophy	13	2	0
ENL	200	16	Studies In Literature	Hip-hop Cinema & History	30	9	0
ENL	200	17	Studies In Literature	WORD!	35	7	6
ENL	200	18	Studies In Literature	Women in Medieval Poetry	40	8	7
ENL	200	19	Studies In Literature	Anti Heroes	35	9	0
ENL	200	20	Studies In Literature	-	40	9	0
ENL	200	21	Studies In Literature	Intro to Short Story	35	7	0
ENL	200	22	Studies In Literature	Intro to Short Story	35	7	0
ENL	200	23	Studies In Literature	Intro to Short Story	35	9	0
ENL	205	01	Travel Literature	-	35	6	4
ENL	205	02	Travel Literature	-	40	7	4
ENL	258	01	Literary Studies	-	20	5	4

Dept	#	Section	Title	Class Topic	Total Enrolled	Drawn for Sample	Total Scored
ENL	258	02	Literary Studies	-	15	4	3
ENL	258	03	Literary Studies	-	13	4	4
ENL	258	7101	Literary Studies	-	12	2	0
FRN	204	01	French Lit in Translation	French Philosophy	15	7	0

Although the total sample size was 121, the valid sample for each learning outcome is less because not every artifact was applicable to each outcome (some assignments did not prompt students to write in ways that encouraged demonstration of each outcome). Ideally, some *combination* of assignments should cover all the learning outcomes in every course, but for this assessment process faculty chose to pick one key assignment and use only that one in their scoring (it is unclear in those cases whether there were other assignments that did cover the missing outcomes).

Cluster 3B

At the time of assessment, the courses in 3B were all offered within the College of Visual and Performing Arts' departments of Art History and Music. To select student work for assessment, an effort was made to collect *all* relevant student work from 3B sections in the Fall. In the Spring, two additional courses were targeted. Some sections contributed multiple artifacts per student, in some cases because different assignments addressed different outcomes and in other cases because students were assessed more than once on particular outcomes.

The student work here consisted mainly of papers, exam essays, discussion board posts, and short writing assignments. In the end a total of 1775 separate student artifacts were collected from 49 different assignments. In the Fall, 1610 artifacts were collected from 9 sections of 6 different courses. In the Spring, 165 artifacts were collected from 2 sections in 2 different courses. Altogether, student work was collected from all the 3B courses that were approved at that time, with just two exceptions (MUS 103 and MUS 107).

Dept.	#	Section	Title	Semester	Students Enrolled	Artifacts Collected	Artifacts Sampled
ARH	105	1	Visual Imagery	Fall 2014	76	670	162
ARH	125	1	Renaissance to Modern Art	Fall 2014	65	53	14
ARH	125	2	Renaissance to Modern Art	Fall 2014	59	51	15
ARH	125	7101	Renaissance to Modern Art	Fall 2014	9	146	39

Dept.	#	Section	Title	Semester	Students Enrolled	Artifacts Collected	Artifacts Sampled
ARH	150	7101	Modern to Contemporary Art	Fall 2014	11	165	44
ARH	125	3	Renaissance to Modern Art	Fall 2014	57	44	12
MUS	101	7101	Introduction to Music I	Fall 2014	5	6	1
MUS	106	1	Art of Rock	Fall 2014	106	456	118
MUS	171	1	Music Theory I	Fall 2014	23	14	3
MUS	106	2101	Art of Rock	Fall 2014	20	5	1
ARH	150	1	Modern to Contemporary Art	Spring 2015	51	147	36
MUS	125	2	Jazz: A Listener's Guide	Spring 2015	21	18	5

After collecting all of the student work from these sections, a sample of 445 artifacts was drawn randomly, post hoc, from this collection of 1775 artifacts. In June, 2015, a group of six readers was trained to score the student work using the 3B rubric. Each artifact was assigned randomly to two different readers. Over the course of the summer, readers completed the rubrics for their assigned student work. Scores for each piece of student work were calculated by taking the average of the two readers' scores. Please see the Appendix for a discussion of the process used to determine which assignments did or did not address particular learning outcomes.

Assessment Results

Reported below are the mean scores across all the data collected for each of the 3A and 3B learning outcomes. In addition, the data present the valid number of cases (N) and the sampling margin of error at the 95% confidence level.

3A – Literature

Learning Outcome	Mean	Valid N	Margin of Error
1. Articulate how literature (fiction, poetry, drama and literary nonfiction) both reflects and helps shape culture, society and history.	2.29	108	+/-0.13
2. Explain how a text's literary form, style and content express its meanings using appropriate disciplinary terminology.	2.14	99	+/-0.15
3. Evaluate the rhetorical and contextual elements of ideas presented by literary texts and respond to them critically and analytically.	2.14	76	+/-0.21
4. Explain the ways in which literature expresses the values that humans attach to their experiences.	2.43	114	+/-0.11

Rubric Scoring Codes:
1 – Marginal: The marginal answer fails to demonstrate basic competency of the learning objective
2 – Emerging: The emerging answer demonstrates basic competency of the learning objective
3 – Proficient: The proficient answer demonstrates mastery of the learning objective

3B – Visual and Performing Arts

Learning Outcome	Mean	Valid N	Margin of Error
1. Articulate the cultural context, history and formal and conceptual aspects of the art form studied.	1.77	339	+/-0.09
2. Interpret and create informed responses (via writing, presentation, performance or artifact) to the art form studied through the analysis of the form, content, context and methods of production using appropriate disciplinary terminology.	2.10	306	+/-0.08
3. Explain the ways in which the art form expresses the values that humans attach to their experiences.	1.88	148	+/-0.13

Rubric Scoring Codes:
0 – Missing: The answer provides no evidence of competency
1 – Marginal: The answer demonstrates only minimal competency
2 – Emerging: The answer demonstrates basic competency
3 – Proficient: The answer demonstrates mastery
NA: Learning Outcome Not Measured in Assignment

The first important point to make about the results is that the 3A and 3B data are not directly comparable because the rubric codes differed. In particular, the 3B readers felt strongly about having a "missing" code to distinguish from a "marginal" response, and as a result the 3B scores range from 0-3 while the 3A scores range from 1-3. For this reason, it is to be expected that the 3B scores would be lower, even if students were performing equally well on them compared to the 3A outcomes.

In the 3A data, the weakest student outcomes were 2 and 3, although the differences are not great across the four outcomes. Importantly, though, these were also the outcomes with the fewest scored cases, suggesting assignments may not be targeting these kinds of cognitive processes as robustly as outcomes 1 and 4. Within the context of the rubric scoring matrix, students on outcomes 2 and 3 performed barely above emerging. In contrast, the data on outcome 4 suggest fairly high student mastery, with the average score falling about halfway between emerging and proficient.

Interestingly, 3B students did not do particularly well on outcome 3, which mirrors 3A outcome 4. The average score falls just above the midpoint of the scale. This was also the outcome with the fewest scored artifacts, again suggesting assignments in 3B courses may not be targeting this outcome. Students did just a bit worse on outcome 1, and performed best on outcome 2, although again the differences are not great.

Appendix: A Technical Note on Determining Excluded Assignments

In the course of scoring the student work, one issue that arose was how to determine when a particular assignment did not prompt students to demonstrate their competency on a particular outcome. During the training and rubric calibration session, the readers expressed a desire to use a special "NA" code to denote when they thought a particular assignment did not address a particular outcome. Their concern was that a failure to find evidence of student competence on a particular outcome may not reflect the student's lack of ability but rather the failure of the assignment to prompt students to display that ability. Indeed, it is often the case that instructors intentionally assess the University Studies outcomes with different assignments.

Determining the "intention" of an assignment is not straightforward, however. In most cases, the instructors included a copy of the assignment directions, and this was helpful (although readers could, and did, disagree about whether some of these directions prompted particular outcomes). In other cases, no directions were included and readers had to interpolate, based on their readings of several artifacts from that assignment, whether an NA code was appropriate.

After all the data were collected and aggregated, a judgment had to be made about when a particular assignment did or did not prompt each outcome. This judgment could be based on the extent to which different readers agreed to give assignments an NA code for a particular outcome (and because the artifacts were randomly assigned to readers, several readers would have scored artifacts from each assignment). In cases where all readers marked all the artifacts from a particular assignment NA, it was clear that those artifacts did not address that outcome. But often readers disagreed and many of the assignments were marked NA by some fraction of the readers. Sometimes that fraction was very small, and it would be unreasonable to exclude the entire assignment (for example, if an assignment generated 15 artifacts and only one reader marked only one artifact NA for an Outcome). In other cases the percentage of NA responses was much higher. It was clear that a cutoff would need to be established such that if a certain percentage of the artifacts were marked NA for a particular outcome, the assignment would be deemed not to address that outcome.

In the end, choosing that cutoff must be arbitrary, but the decision was guided by an inductive process of recalculating the overall average assessment scores for each outcome at various cutoffs. At one extreme, all assignments were considered valid and just those specific artifacts marked NA were excluded as missing. At the other extreme, artifacts from a given assignment were only included if there were no NA scores at all for any artifact from that assignment. In between those extremes, cutoffs of 50%, 75% and 90% were used (i.e., an assignment was included only if 50% [or 75% or 90%] of the artifacts were not scored NA).

Below is a table presenting the average score, 95% confidence interval (in parentheses), and valid sample size for each of the 3B outcomes for the various cutoff scenarios.

Cutoff	3B Outcome 1	3B Outcome 2	3B Outcome 3
All cases	1.70 (1.61-1.78) n=407	1.95 (1.87-2.03) n=401	1.67 (1.57-1.76) n=351
Over 50% valid	1.77 (1.69-1.85) n=375	2.08 (2.01-2.16) n=319	1.76 (1.65-1.87) n=220

Over 75% valid	1.77 (1.68-1.86) n=339	2.10 (2.03-2.18) n=306	1.88 (1.76-2.01) n=148
Over 90% valid	1.74 (1.65-1.84) n=283	2.10 (2.03-2.18) n=286	1.85 (1.73-1.98) n=134
Only 100% valid	1.80 (1.66-1.94) n=115	2.15 (2.07-2.22) n=265	1.75 (1.63-1.87) n=166

Based on this table, a decision was made to use a 75% cutoff. The sample sizes get smaller as the cutoffs get more restrictive (and the confidence intervals get bigger, accordingly), which is undesirable. But the means and even these confidence intervals are fairly consistent at the 50%, 75% and 90% cutoffs. At the extremes (when all cases are used or the data are restricted just to assignments with 100% valid scores), the numbers change a bit. Given the stability of the numbers within the middle cutoffs and the concerns about smaller samples at the more restrictive cutoffs, the 75% cutoff appears to be a reasonable choice.

To summarize what this choice means in terms of the final 3B data presented in this report, it is helpful to consider an example. Imagine a particular assignment produced 18 artifacts. A pair of readers scored each of these artifacts, so there are 36 scores for Outcome 1, Outcome 2 and Outcome 3. Of the 36 scores, 0 were marked NA for Outcome 1, 5 were marked NA for Outcome 2 and 14 were marked NA for Outcome 3. Using the 75% cutoff, this assignment would be used to assess Outcomes 1 and 2, but not Outcome 3, because less than 75% of the artifacts were given a valid score for Outcome 3 (14 were marked NA, so 22 were valid; 22 is only 61% of the 36 scores). As a consequence, *all* of the Outcome 3 scores from this assignment would be excluded, on the grounds that the assignment did not prompt students to display competency of that outcome.

Assessment Report - University Studies Cluster 4

Student Work Collection

2015-2016

This report describes the collection and analysis of student work in Cluster 4 that was conducted in 2015-2016. This assessment project was approved by the General Education Committee and follows the timetable for University Studies assessment approved by the Faculty Senate in 2010.

Approved Courses in Cluster 4 – The Social World: Humanity and Society

A. Human Questions and Contexts
ECO 231 - Principles of Microeconomics
EDU 452 - Sheltered English Immersion
HON 101 - Scholarship in Community
HON 201 - Knowing Ourselves
HST 101 - History of Western Civilization I
HST 102 - History of Western Civilization II
PHL 101 - Introduction to Philosophy
PHL 110 - Principles in Critical Thinking
PHL 215 - Introduction to Ethics
PHL 221 - History of Western Philosophy: Ancient
PHL 222 - History of Western Philosophy: Modern
PSC 213 - Law and Courts in Film
PSY 101 - General Psychology
REL 300 - Advanced Topics in Religious Studies
SUS 101 - Principles of Sustainability
SUS 202 - Topics in Sustainability

B. The Nature of US Society
ARH 349 - Architecture and Sustainability in American Post-Industrial Cities
CJS 190 - Introduction to Crime and Justice Studies
ECO 231 - Principles of Microeconomics
ECO 232 - Principles of Macroeconomics
EGR 303 - Engineering Economy
ENL 372 - Writing About Popular Culture
HON 202 - Transformative American Ideas
HST 115 - History of the United States I
HST 116 - History of the United States II
POL 102 - Introduction to Policy Studies
PSC 101 - Introduction to American Politics
PSC 212 - American Politics in Film

SOA 101 - Introduction to Sociology
SOA 111 - Introduction to Cultural Anthropology
URB 201 - City Life: Introduction to Urban Studies
WGS 101 - Introduction to Women's Studies
WGS/PHL 104 - Identities: Gender, Race, Sexuality
WGS 202 - Women's Health and Environment

C. The Nature of the Global Society
ARH 362 - Islamic Art
ATR 380 - History of Craft I
ATR 381 - History of Craft II
ECO 107 - Economics of Pollution
HON 203 - Creating Global Community
HON 301 - Honors Research Across Disciplines
HST 103 - World Civilizations I
HST 104 - World Civilizations II
IST 151 - Introduction to Indian Civilization
MGT 120 - Green Entrepreneurship
PHL 230 - Global Philosophies
PHL 341 - Philosophy of the Good Life
POR 120 - Introduction to the Portuguese-Speaking World
POR 220 - Introduction to Linguistics and European Languages
POR 371 - Gender and Society in Brazilian Cinema
PSC 151 - Introduction to Comparative Politics
PSC 161 - Introduction to International Relations
PSC 284 - Model United Nations
PSC 384 - International Law and Organization
PSC 385 - Politics of Global Climate Change
REL 201 - Introduction to Religious Studies
SOA 101 - Introduction to Sociology
SOA 111 - Introduction to Cultural Anthropology
SUS 101 - Principles of Sustainability
WGS 213 - Sex Workers, Nannies and Maids
WGS 352 - Disposable Women
WGS 369 - Global Women's Health Politics and Activism

Assessment Methods

Our approach was to identify a random sample of students enrolled in Cluster 4 courses in 2015-2016 and to collect their work by asking the instructors to submit these students' work artifacts (papers, essays, etc.) from assignments that demonstrated their competency on the Cluster learning

outcomes.[1] This work was then scored using a rubric by trained readers. These readers were drawn from volunteers among the faculty and were compensated for their work during Summer 2016.

To identify a sample of student work, 400 student names were drawn at random from all students enrolled in courses in each of the three requirements in Cluster 4. Collection of student work for 4B took place in Fall 2015 and collection for 4A and 4C was undertaken in Spring 2016. The following tables describe the sampling frames and the number of student work artifacts received.

Cluster 4A (collected Spring 2016)

Courses:	8
Sections:	36
Instructors:	17
Enrollment capacity:	1718
Actual enrollment:	1480
Students drawn for sample:	400
Student artifacts in final sample:	213
Response rate:	53%

Cluster 4B (collected Fall 2015)

Courses:	13
Sections:	56
Instructors:	30
Enrollment capacity:	2537
Actual enrollment:	2278
Students drawn for sample:	400
Student artifacts in final sample:	357
Response rate	89%

Cluster 4C (collected Spring 2016)

Courses:	15
Sections:	36
Instructors:	23
Enrollment capacity:	1629
Actual enrollment:	1331
Students drawn for sample:	400
Student artifacts in final sample:	247
Response rate:	62%

[1] The vast majority of the student work identified by instructors was from writing assignments. Only several sections produced multiple-choice exam data. Because these sections were all for the same course, they were excluded in our analysis in order to preserve the anonymity of the data.

First, it is important to note that there is an ample number of courses, sections and seats in all three Cluster requirements. Students have had no difficulty finding a seat in a Cluster 4 course and typically have a broad diversity of choices spanning many academic disciplines.

Second, the response rates for the student work collection are not as high as they might be. The rate for 4B was actually quite good, and in all cases we received work for over half the students drawn for the samples. But with sample sizes of 200-250, the sampling margin of error is relatively large, which can make precise inferences about the full population of students difficult. For that reason, margins of error are reported in the results below.

Assessment Results

Reported below are the scores across all the data collected for each of the Cluster 4 learning outcomes.[2] In addition, the data present the valid number of cases (N) and the sampling margin of error at the 95% confidence level.

4A – Human Questions and Contexts

Learning Outcome	Mean	Margin of Error	Percent Marginal/ Proficient	Valid N
1. Explain different perspectives on: a) what it means to be human and how the significance of human existence has been understood; b) the nature of human relationships and how these relationships are evidenced in regard to the broader world; or c) how knowledge is obtained, maintained and changed, as well as how individuals come to understand and think about the world around them.	1.98	+/-.10	78	187
2. Recognize ethical issues in complex contexts and evaluate the ethical positions taken by themselves and others.	1.91	+/-.10	73	193
3. Locate, analyze, summarize, paraphrase and synthesize material from a variety of sources.	2.02	+/-.12	80	82
4. Evaluate arguments made in support of different perspectives on human questions and contexts.	1.73	+/-.11	58	152

[2] As was necessary in the Cluster 3 assessment, certain assignments had to be excluded for certain outcomes when "most" readers concluded the assignments did not prompt students to address a particular learning outcome. We conducted the same type of analysis here, which involved an examination of mean scores under different exclusion scenarios (that is, scenarios defining what exactly is meant by "most" readers). The mean scores were very consistent across the scenarios, and so a decision was made to use the same 75% exclusion scenario that was used for Cluster 3. In short, if more than 25% of the artifacts from a given assignment were identified as invalid (the reader did not think it addressed the outcome), the entire assignment was excluded.

4B – The Nature of US Society

Learning Outcome	Mean	Margin of Error	Percent Marginal/ Proficient	Valid N
1. Explain: a) the development of US culture and sub-culture from different perspectives; b) US social and cultural domains in relationship to other regions of the world; or c) the different facets of citizenship in the United States.	1.81	+/-.07	74	319
2. Locate, analyze, summarize, paraphrase and synthesize material from a variety of sources.	1.63	+/-.13	61	215
3. Evaluate arguments made in support of different perspectives on US society.	1.52	+/-.10	59	258

4C – The Nature of the Global Society

Learning Outcome	Mean	Margin of Error	Percent Marginal/ Proficient	Valid N
1. Explain basic problems faced by societies and cultures outside the US or issues that shape societies globally.	1.91	+/-.08	77	230
2. Locate, analyze, summarize, paraphrase and synthesize material from a variety of sources.	1.77	+/-.10	61	197
3. Evaluate arguments made in support of different perspectives on global society.	1.78	+/-.12	69	163

Rubric Scoring Codes:
0 – Missing: The answer demonstrates no competency
1 – Marginal: The answer demonstrates only minimal competency
2 – Emerging: The answer demonstrates basic competency
3 – Proficient: The answer demonstrates mastery
NA: Learning Outcome Not Measured in Assignment

The aggregate scores indicate students on average fell between marginal and emerging for almost all outcomes in all three of the requirements. The average student achieved a score of emerging (2) on only one outcome—4A-3, which deals with information literacy skills. On several other outcomes, students did nearly as well (4A-1, 4A-2, 4C-1). On no outcome did the average fall below 1.5, which would be the midpoint between marginal and emerging. Looked at differently, no fewer than 58%

of students scored emerging/proficient on any outcome, and students did much better on many of the outcomes. For example, 80% scored emerging/proficient on 4A-3, dealing with critical thinking, and over 70% scored emerging/proficient on a number of other outcomes (4A-1, 4A-2, 4B-1, 4C-1).

The learning outcomes for each requirement in Cluster 4 were intended to be parallel. Each requirement contains at least one "knowledge" outcome that gets at how well students can explain the content knowledge in that area (Cluster 4A also has an outcome focused on ethical reasoning). Additionally, they each have an "information literacy" outcome (locate, analyze, summarize, paraphrase and synthesize material from a variety of sources) and a "critical thinking" outcome (Evaluate arguments made in support of different perspectives on…).

Generally, scores on the knowledge outcomes were the strongest in each requirement. In contrast, scores tended to the weakest on the critical thinking outcomes. The information literacy scores were quite high in 4A, but were nearly as weak as the critical thinking outcomes in 4B and 4C.

…

In addition to collecting and scoring student work, a survey of the faculty teaching Cluster 4 courses in 2015-2016 was administered. The survey asked faculty about their teaching practices in their Cluster 4 courses, including the types of pedagogies used and the types and amounts of writing required. The purpose of the survey was to determine whether certain kinds of pedagogical practices were more effective in promoting student mastery of the Cluster 4 outcomes by statistically analyzing the relationship between those practices at the course-level and the outcomes of students within those classes.

Before presenting those relationships, it is useful to provide descriptive statistics about pedagogical practice in Cluster 4. There was a total of 64 faculty teaching 135 sections of 36 courses in Cluster 4 in the semesters we collected student work. We collected student work from 43 faculty teaching 28 different courses. An invitation was sent to all of those 43 faculty members, representing 53 different course-instructor combinations, and responses were received by 34 faculty covering 40 different course-instructor combinations. As these numbers indicate, some instructors taught more than one Cluster 4 course and completed the survey separately for each course (because pedagogy may differ); consequently each course-instructor combination is represented by a separate case, so the maximum number of cases is 40.

Pedagogies Used

Percentage of Cluster 4 Courses Using:	
Multiple-choice test or quiz questions	75%
Open-ended test or quiz questions	68%
Papers	75%
Other writing assignments	68%
Oral presentations	35%
Lectures	85%
Videos/films/movies	68%
Simulations	15%
Full-class discussions	83%
Small-group discussions	43%
Small-group projects that happened only in the classroom	23%
Small-group projects that required group work outside of class	18%
Service learning	0%
Online activities	70%
Classroom response systems	15%

Types of Writing Assigned

Percentage of Cluster 4 Courses Assigning:	
Research paper (requiring the use and documentation of sources)	50%
Reflective writing (requiring students to think about their learning or their educational experiences)	48%
Writing about readings (writing assignments, small or large, that require student to discuss course readings)	68%
Argumentative writing (requiring student to take a position and defend it using reasoning and evidence)	63%
Low-stakes writing (i.e., writing students did either inside or outside of class that was either not graded or graded only for completion, not for quality)	35%
Journaling	3%
Discussion boards	40%
Blogging	3%
Wikis	5%

In addition to these statistics, the survey also revealed that faculty assign an average of 16 pages of writing in their Cluster 4 courses.

…

As described above, the purpose of collecting these data is to assess whether certain teaching practices were more effective in promoting student competency on Cluster 4 outcomes. This assessment requires some attention to the nature of the data and the appropriate statistical methodologies required. First, and most importantly, it is important to utilize a multivariate model, rather than simple correlations, to draw conclusions about causality; because courses involve multiple pedagogies and multiple types of writing, multivariate models help rule out spurious relationships and help to identify those specific practices that are really driving student learning.

Generally, some type of regression model would be well-suited for this task, but there are characteristics of the data that require a relatively complicated approach. First, the data are "hierarchical," in that some variables (the student learning outcomes) are measured at the student-level while other variables (teaching practices) are measured at the course-level, and cases at the first level fall into groups on the second level. Second, the dependent variable—an individual student's score on a particular learning outcome—is an ordinal-level variable that takes on discrete values (0, 1, 2, 3). To manage these facets of the data and produce valid estimates and standard errors, a multilevel, mixed-effects ordered logistic regression procedures was utilized.

Below are tables presenting the results of these multivariate analyses. The main entries are unstandardized ordered logit coefficients, which convey the effect of each teaching practice on that particular learning outcome. Because the coefficients are derived from a logistic regression model, the values do not have a straightforward interpretation in terms of the strength of the effect, but one can conclude that positive coefficients indicate the teaching practice was associated with better student performance on that learning outcome (and negative coefficients with worse performance).

Because, as discussed above, the learning outcomes are substantially parallel, it is possible to pool students' scores across 4A, 4B and 4C when assessing the effects of pedagogical practices. The major advantage of doing so is that it greatly increases the number of cases available for analysis and provides much greater variation on the key variables. Thus, we have pooled the cases in this manner in the analyses reported below.

The Impact of Pedagogies on Student Learning Outcomes

	Knowledge	Information Literacy	Critical Thinking
Multiple-choice test or quiz questions	**.72 (.33)****	-.48 (.56)	-.11 (.47)
Open-ended test or quiz questions	**-.84 (.25)*****	.66 (.60)	.25 (.34)
Papers	-.20 (.38)	.57 (.85)	.67 (.45)
Other writing assignments	-.22 (.30)	.19 (.57)	.16 (.38)
Oral presentations	-.54 (.40)	-1.14 (.74)	-.44 (.56)
Lectures	**-.92 (.54)***	**-2.36 (1.00)****	-1.01 (.65)
Videos/films/movies	**1.12 (.38)*****	-.55 (.79)	.10 (.50)
Simulations	.20 (.36)	-.26 (.59)	-.43 (.48)
Full-class discussions	-.26 (.43)	.35 (.85)	.05 (.60)
Small-group discussions	**.88 (.28)*****	**1.02 (.52)***	**1.11 (.33)*****
Small-group projects in the classroom	**-1.81 (.29)*****	**-.99 (.52)***	**-1.21 (.37)*****
Small-group projects outside of class	**1.63 (.43)*****	1.02 (.70)	.83 (.57)
Online activities	-.42 (.37)	-.74 (.66)	-.08 (.50)
Classroom response systems	.30 (.38)	.34 (.71)	.39 (.48)
Level-1 Observations (students)	578	354	342
Level-2 Groups (courses)	35	30	28

Note: Entries are unstandardized ordered logit coefficients; standard errors are in parentheses; *p<.10; **p<.05; ***p<.01

The Impact of Writing on Student Learning Outcomes

	Knowledge	Information Literacy	Critical Thinking
Research paper	**.51 (.30)***	.15 (.40)	.40 (.34)
Reflective writing	-.42 (.32)	**-1.07 (.43)****	**-.62 (.27)****
Writing about readings	-.15 (.41)	-.56 (.47)	-.50 (.34)
Argumentative writing	.44 (.37)	.35 (.49)	.47 (.34)
Low-stakes writing	-.28 (.33)	-.30 (.44)	.40 (.31)
Discussion boards	-.04 (.34)	-.11 (.45)	**-.58 (.30)***
Total pages of writing	.02 (.01)	**.04 (.02)****	**.02 (.01)****
Level-1 Observations (students)	578	354	342
Level-2 Groups (courses)	35	30	28

Note: Entries are unstandardized ordered logit coefficients; standard errors are in parentheses; *p<.10; **p<.05; ***p<.01

The analyses suggest a number of practices have positive and statistically significant relationships with student performance: multiple-choice exams, videos, small group discussions, small group projects outside of class, and research papers. Interestingly, a number of practices are associated with *worse* student outcomes, including open-ended exams, lectures, and small group projects in the classroom. It is worth noting that the conventional approach to classroom practice (lectures and multiple-choice exams) was a mixed bag in terms of subject knowledge development. Lectures actually appeared to be counter-productive, while multiple choice exams had a positive relationship with knowledge outcomes.

Information literacy outcomes were higher when instructors used small group discussions and when they assigned more pages of writing overall. The connections between these practices and information literacy are difficult to explain, as is the negative effect of reflective writing assignments. It is also notable that research papers were not associated with better information literacy outcomes.

The results regarding critical thinking are also somewhat difficult to explain. Students did better when their courses involved small-group discussions, and they did worse when there were small-group projects in the classroom and when they engaged in reflective writing or discussion board writing. It does make some sense that students do better at critical thinking when their courses contain more writing, insofar as writing assignments are more likely to involve critical evaluation than other types of pedagogies.

We can look at the results from another angle by considering whether certain pedagogical practices tend to have relationships with student learning in general. From this perspective, small group discussions in class look like a high-impact practice, displaying positive relationships with all three outcomes. However, when these groups engage in projects in class, instead of just having discussions, the results are consistently negative. The reason for this relationship is unclear. In addition, lectures were associated with worse scores on both knowledge and information literacy outcomes.

Finally, though simply assigning papers was not associated with higher student performance, the data do indicate that more pages of writing overall were related to higher achievement, at least on the two "skill" outcomes (information literacy and critical thinking). Interestingly, reflective writing (requiring students to think about their learning or their educational experiences), had a negative relationship with both of these skill outcomes. Beyond that, no specific type of writing had a particularly strong relationship with information literacy and critical thinking; more writing of any type seemed to be key variable.

Overall, the results confirm what could be viewed as a number of reasonable hypotheses about teaching practice. For example, it is not entirely surprising that students do better when they write more or that the much-maligned lecture is ineffective (or counterproductive). However, many of the results are puzzling. It is difficult to understand why many practices have *negative* relationships with student learning. It is understandable that some practices would be wholly unrelated to student outcomes, but why might a practice make students do worse? It is difficult to see how reflective writing assignments or open-ended exam questions, for example, make students do *worse* on the intended learning outcomes.

One might postulate that time spent on one kind of pedagogy comes at the expense of other, perhaps more effective pedagogies, but the multivariate modeling used here controls for these

effects. In the end, one has to consider whether there are limitations of the research design that proscribe the usefulness of these results to inform best practices. There are, of course, a number of uncontrolled variables left out of the model—most importantly, those measuring the capabilities and motivations of the individual students. But also missing are other aspects of the classroom experience related to content and context. For example, it might be important to control for the time of day the class met or the inherent interest students have in the course topics. Some of these factors may interact in ways that are important to model properly.

These considerations suggest faculty interested in improving student learning on particular outcomes should examine the research literature on the scholarship of teaching and learning. The studies in this literature are more likely to reflect research designs with higher levels of internal validity and therefore they can tell us much more about the causal connections between teaching practices and student learning.

In sum, while the assessment process provides some opportunity to examine connections between pedagogy and learning, it is not an ideal context. In the end, the teaching development process must be the place where faculty improve their pedagogy using evidence about best practices. It is possible, and perhaps ideal, that the assessment arc feed into the teaching development arena. For example, if assessment data reveal that students are doing particularly poorly on certain outcomes, professional development programming might be directed at these outcomes.

In that light, it is worth pointing out that the lowest (or nearly lowest) score in each of the three Cluster requirements was the critical thinking outcome. This pattern would suggest a focus on critical thinking in the Cluster 4 course reapproval process and a concomitant emphasis on helping faculty understand what the research literature says about best practices in teaching critical thinking.

Cluster 5B Student Assessment, 2016-2017

The total number of student respondents included in these data is 574. The data set combines 508 responses collected by the Leduc Center from service learning enrollments in 2016-2017 with a survey by the General Education Director of 66 students who were enrolled in non-service learning 5B courses in 2016-2017.

Relative Frequencies and Means

Please rate how well you think you were able to do each of the following before you started the course.

	(1) Not at all	(2)	(3) Moderately well	(4)	(5) Very well	Mean	N
Identify needs and resources of the community.	6.4	23.8	44.7	14.5	10.7	3.0	551
Apply knowledge and skills gained to real problems/opportunities in my community.	3.8	17.8	44.3	21.0	13.3	3.2	549
Make connections between learning and issues/needs of the community.	5.1	20.2	40.6	22.4	11.7	3.1	549
Articulate the value of engagement to other members of the community.	8.2	19.8	42.1	19.1	10.8	3.0	546
Communicate effectively orally and in writing.	3.6	13.6	34.9	26.7	21.1	3.5	550
Evaluate and integrate information from multiple sources.	4.0	10.8	38.1	28.5	18.6	3.5	548

Please rate how well you think you are able to do each of the following now.

	(1) Not at all	(2)	(3) Moderately well	(4)	(5) Very well	Mean	Increase in Mean	N
Identify needs and resources of the community.	2.0	3.1	28.5	43.9	22.5	3.8	.8*	547
Apply knowledge and skills gained to real problems/opportunities in my community.	2.0	3.1	24.4	41.2	29.3	3.9	.7*	546
Make connections between learning and issues/needs of the community.	2.4	4.1	24.7	40.3	28.6	3.9	.8*	543
Articulate the value of engagement to other members of the community.	3.5	6.3	29.7	34.6	26.0	3.7	.7*	543
Communicate effectively orally and in writing.	2.4	2.2	18.8	39.5	37.1	4.1	.6*	544
Evaluate and integrate information from multiple sources.	2.0	1.3	20.0	42.4	34.3	4.1	.6*	540

*difference is statistically significant at the .0001 level

Please indicate whether you agree or disagree with each statement.

This Learning Through Engagement/Service Learning course:

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree	Mean	N
Showed me the impact that I can have on solving problems that face my local community.	5.0	5.7	30.2	46.6	12.5	3.6	560
Deepened my concern about community issues.	5.7	5.4	32.9	44.1	12.0	3.5	560
Showed me that it's personally important to influence the political process.	8.5	6.3	37.2	35.8	12.2	3.4	556
Showed me that contributing to the solutions of social problems is my responsibility.	7.0	6.1	32.6	44.3	10.0	3.5	558
Made me aware that I am uncomfortable working with people who are different from me in such things as race, wealth, and life experiences.	24.6	40.1	17.7	13.3	4.3	2.3	558
Made me aware that I am able to see a situation from someone else's point of view.	3.2	3.8	24.9	52.8	15.4	3.8	559
Made me want to invest time in learning about social issues and problems (for example, check the web, read the paper or magazines, attend community meetings).	7.3	6.6	32.7	39.7	13.6	3.5	559
Showed me that action by groups can solve social problems.	3.8	5.4	25.7	49.7	15.4	3.7	557
Made me plan to improve my community by volunteering in the near future.	9.6	6.9	38.3	36.6	8.7	3.3	554

Helped me identify needs and resources of the community.	4.2	5.1	34.1	43.1	13.5	3.6	554
Challenged me to apply knowledge and skills gained to real problems/opportunities in my community.	4.2	4.2	28.0	47.7	16.6	3.7	554
Helped me make connections between learning and issues/needs of the community.	4.7	4.9	30.0	47.1	13.4	3.6	554
Helped me articulate the value of engagement to other members of the community.	4.0	4.9	34.2	46.0	11.0	3.6	555
Challenged me to communicate effectively orally and in writing.	4.5	3.4	27.3	48.2	16.6	3.7	556
Challenged me to evaluate and integrate information from multiple sources.	2.9	2.9	27.8	49.4	17.1	3.8	555