

General Education Artifact Sampling: An Overview of the History for Faculty and Deans

Overview Updated 5-9-2018 Laura Blasi, Director, Institutional Assessment

Why We Are Sampling Artifacts

We sample papers in order to obtain a representative cross-section of student artifacts to assess using the faculty-developed General Education Assessment checklist. This essay focuses on the primary type of assessment that has required sampling at the college¹ which has been in the General Education program for artifact assessment. We briefly outline the history of the related decisions made since 2007.

While the history can be traced back to the 2007 academic year, in 2014-2015 we developed a theoretical model that was then redefined to account for contextual factors. The model is outlined in Graphic 1: Sampling Guidelines for General Education. Our Institutional Research office also refers to the model when responding to faculty requests for samples, and stratifies the samples according to campus, mode of instruction, and several other factors. In 2015-2016, faculty members in the Humanities and Comp I & II courses collected papers using the model.

When faced with a large number of papers (300) and inconsistent responses from faculty called to participate in the assessment, both groups pared down the number of papers they would assess. Then they assigned two reviewers to each paper and so cut the numbers further in order to manage faculty workloads related to paper assignments. In that year Comp I & II had 98 papers read twice (without consensus scores) and Humanities had 132 papers read twice (they also did not have consensus scores). Current discussions are underway to use the Gordon Rule assessment as an opportunity to improve the artifact review and sampling practices.

¹ While it is possible to use online calculators or a Table of Populations in order to calculate the required samples needed, the consideration of how we sample in General Education also takes into account other factors – such as the number of faculty members available to assess. This context requires more than the application of a frequently-used sampling formulae. When sampling is used for a survey or a poll the concern is focused on getting a randomized, representative group of respondents to answer a set of questions to document their knowledge or opinions. In General Education assessment, when focused on artifacts, we are not trying to document that the results to questions we are asking are reflective of the larger population. We are sampling student work to document patterns in student writing or speeches in order to understand and to improve the General Education program.

Next Steps in the Sampling of Artifacts

Rather than concentrating on an increase of the number of papers sampled, when we step back and look at the assessment process the areas in most need of attention include: (a) increased faculty participation paired with technical accuracy; (b) reduction of challenges faced when leading assessment activities; (c) the use of consensus scores by a third readers when two readers disagree; (d) the meaningful use of data to make decisions; (e) the implementation and follow-through with decisions focused on teaching and learning. Moving forward we aim to use a college-wide panel approach to sampling and assessing student artifacts.

An Overview of the History (2007-2015) of Sampling in General Education for Learning Outcomes Assessment at Valencia College for Faculty and Administrators

Laura Blasi, Ph.D., Institutional Assessment
lblasi@valenciacollege.edu

Prior to the fall of 2007, all Freshman Composition I students had to complete and pass a written exit essay exam². The Freshman Composition I assessment project was then created and it required that student work be sampled. The sampling formula was defined by Institutional Research using the SAS software program. Each year the actual return rates were much lower than expected. Subsequent reports noted the need for an improved response rate and the desire for higher rates of faculty participation in Assessment Day activities.

How We First Sampled in 2007

Using a computer script written in SAS, a statistical software package, a simple random sample of students was chosen from the population of students currently enrolled in a particular course or courses. The sample of students, although chosen at random, was selected to intentionally mimic the college enrollment in terms of particular variables. The sample size is determined based upon the total number of students in the target population and the assessment tool that is being used (rubric, etc.). Specifics may vary and the results will look different based upon the assessment but will follow the ENC 1101 model. An example is provided below from Fall 2007:

- There were 4,529 students enrolled in ENC 1101, ENC 1101H, and IDH 1110 and a sample size of 115 students was used. This sample size was determined using a margin of error of $\pm .20$ (5% of the scale – a 4 point rubric) and a confidence level of 95%. The sample size for a population size of 4,529 should have been 98. However, from 200310 to 200630, the college-wide withdrawal rate for all courses (W only, not including WP or WF) ranged anywhere from 10.33% to 15.44%.
- So in preparation for a potential loss of selected students (due to withdrawal), the population was oversampled (115 as opposed to the suggested 98). The sample of students intentionally echoed the college enrollment in terms of the following: campus of enrollment, day/evening (time of day course is offered), Full-time/Part-time (employment status of instructor), Online/Traditional (delivery method) (this history was excerpted from historical records regarding methods used at the college, 2007.)

² Hardin, C. (2011). *English Composition I (ENC1101) student-learning outcome assessment*. Orlando, FL: Valencia College.

2011

In the 2011 academic year the approach to sampling and the overall assessment process was redefined. ENC1101 instructors were asked by their campus English Coordinators to submit artifacts of student learning and the sample size was determined by the Learning Evidence Team at the college based on the total number of students enrolled in all ENC1101 classes at the college. The random sample was determined to be 100 and 54 essays were returned and assessed by forty-four faculty members³. On Assessment Day thirty-three faculty participants voted to work on improving instruction specific to Information Literacy.

2012

In 2012 again from a sample of 100 and 51 essays were submitted for review and thirty-three faculty members evaluated the artifacts on Assessment Day. The end-of-year improvement plan stated that faculty members “agreed to integrate more focus on teaching students how to properly document sources within an essay in MLA format.” In 2013 the coordinator received 39 essays from a sample of 100. As a result of the assessment when faculty members were asked if the students properly documented the sources within the essay, “of the 48 essays assessed, 60% received an answer of ‘yes’ to the assessment question” (end-of-year plan, 2013.)

2013

Since 2010 the English faculty members had the largest samples and the highest response rates and still struggled to obtain the number of artifacts they desired. The other General Education disciplines also aspired to higher response rates and faculty participation with the ultimate goal of meaningful decisions to be implemented and re-evaluated for impact. In 2013 for the first time we worked to define sample sizes using a consistent method for all disciplines. For some of the disciplines this was anticipated and warmly received and for others this change was met with dissatisfaction, but in the end all adopted the 5% sample size. We used similar sampling parameters to those that has been consistently used by Institutional Research since 2007. The samples were stratified by campus, mode, and faculty status. The focus was on the Full Term standardizing the communications cycle and timeline for the coordinators.

³ Hardin, C. (2011). *English Composition I (ENC1101) student-learning outcome assessment*. Orlando, FL: Valencia College.

For Comp I and Comp II with 9,803 students enrolled a 5% sample called for 490 papers. Thanks to the focused teamwork of the faculty members they had a return rate of 3.4% (336) which was six times the number of papers that had ever been collected and required a more complex approach to norming and managing the assessment activities. The sample of 490 was much higher than the sample size identified in the table of populations, which would have been 380⁴. Using a sample size calculator online a simple random the sample would have been 370 (using a confidence level 95% and a margin of error + or - 5%) given the number of students taking Comp I and II in the Full Term (see figure 1.)

Figure 1 Online Calculator

Determine Sample Size

Confidence Level: 95% 99%

Confidence Interval:

Population:

Sample size needed:

We requested a 5% sample for English, Humanities, Political Science, and Speech and Student Success courses were sampled by section (which resulted in a 9% sample of students.) We were able to report the use of a consistent method in our ten-year report to SACS and show results for the assessment conducted in each discipline (see figure 2.)

Figure 2 Table of Select Samples at Valencia, 2013

<p>Student Success (SLS): N=693 observed of 809 students / 33 sections (9% of the total population.) Speech rubric used for assessment.</p>	<p>English: N=336 of 480 students sampled at random (5% of the total population for the full term)</p>
<p>Speech: N=799 of 1,031 students / 45 sections sampled at random (5% of the total population for the full term.) Speech rubric used for assessment.</p>	<p>Humanities: N=151 of 210 students sampled at random (5% of the total population for the full term)</p>
	<p>U.S. Government: N=139 of 150 students sampled at random (5% total population for the full term)</p>

⁴ See attachment: *Table for Determining Sample Size Given a Population* using a confidence level 95% and a margin of error + or - 5%.

We were also able to report targets and document the scores across disciplines using a shared checklist to evaluate the work by General Education outcome. For example – for the Communications outcome we were able to report that expectations were partially met (see figure 3.) The results featured targets that aimed to have 70% complete at a 3 or higher on a 4 point scale. The bar was not set very high; it was set without college-wide faculty discussion; and for the most part this target was only partially met or not at all met for the General Education outcomes.

These targets were set via E-mail exchanges with deans and learning outcomes assessment lead faculty members and Assessment Day 2013 discussions were again held by discipline rather than by outcome or with cross-disciplinary conversations. While the faculty leaders expressed surprise at the patterns that emerged when they saw the targets that were not met and when they were able to compare across disciplines, the college still needs to establish a forum for faculty to discuss and make decisions based on this kind of cross-disciplinary, outcomes-based data related to teaching and learning in General Education.

Figure 3 Example of Cross-disciplinary Findings, 2013

Student Learning Outcomes and Indicator(s)	Expectations Met?	Expected Performance (%) and Method of Assessment	Actual Performance (%) Comparative When Possible
Communication 2 of 3: Students will be able to demonstrate effective written communication skills.	<i>Expectations partially met</i>	English (Comp I & II) Humanities and US Gov't: 70% or higher satisfactory / acceptable on an assignment evaluated using a shared checklist.	Skill Assessed: Writing: Logic (English – 70.2%/ Humanities – 76%/ US Gov't – 66%) Skill Assessed: Writing: Accuracy (English – 70.1%/ Humanities – 68%/ US Gov't – 79%)

2015-2016

In 2015 Valencia College moved away from an approach to sampling that used an across-the-board percentage. The across-the-board percentage was originally implemented for consistency. It was held steady while Comp I and II faculty, who led the way, were joined by others in the Humanities and in the Social Sciences. As faculty members in each discipline were trying to collect papers from their colleagues, they were sometimes disappointed by the numbers of papers returned. After a meeting of the Learning Outcomes Leaders (LOLs) in Feb. 2015, we agreed upon a new approach to sampling that worked on a sliding scale instead of an overall percentage (see “Graphic 1”). This sliding scale took into account the average number of sections being offered for each course in each discipline. We also took into account the number of full-time faculty members who were expected to participate in the assessment for each discipline based on the voters list used for decisions about the programs. Ideally the sample would be manageable (given the number of faculty available) and representative (given the number of sections that could be included in the sample).

Graphic 1: Sampling Guidelines for General Education

Course Short Title	COURSE	Enrolled at End of Term Fall 2014	Approx count of students enrolled at end-of-term - ranges	Proposed artifact sample size for this range by count of student ONLY	Number of Full-time from Faculty-Voter Lists	Exception Rule: If 10 or fewer faculty projected to participate then this number of artifacts
Freshman Comp I	ENC1101	6750	6000+	300	43	
Intermediate Algebra	MAT1033C	4670	4000-5999	250	70	
College Algebra	MAC1105	4627	4000-5999	250	70	
Fundamentals of Speech	SPC1608	4107	4000-5999	250	20	100
Freshman Comp II	ENC1102	3559	2000-3999	200	43	
U.S. Government	POS2041	3531	2000-3999	200	11	100
General Psychology	PSY2012	3282	2000-3999	200	13	100
Intro To Humanities	HUM1020	2766	2000-3999	200	24	
Fundamentals Of Biology	BSC1010C	2112	2000-3999	200	32	
Statistical Methods	STA2023	2078	2000-3999	200	66	
Humanities 20th Century	HUM2250	1420	800-1999	150	24	
Developmental Math (C)	MAT0022C	1384	800-1999	150	70	
Hum Greek/Roman	HUM2220	1284	800-1999	150	24	
Precalculus Algebra	MAC1140	1087	800-1999	150	70	
College Trigonometry	MAC1114	1062	800-1999	150	70	
Principles of Economics	ECO2013	1058	800-1999	150	7	100
Interpersonal Communication	SPC1017	993	800-1999	150	20	100
Computer Fundamentals	CGS2100C	989	800-1999	150	13	100
Introductory Sociology	SYG2000	936	800-1999	150	5	100
Human Anatomy and Physiology	BSC2093C	935	800-1999	150	32	
College Mathematics	MGF1106	923	800-1999	150	70	
Mythology	HUM2310	887	800-1999	150	24	
			<800	100		
New Student Experience	SLS1122	4906	4000-5999	250		

This approach was amended in November 2015, in response to discussions held with Learning Outcomes Leaders facilitating the work for disciplines collecting papers that year. They planned on having each paper read by two or more faculty members. As a result, the expectation for 800 papers to be collected was adjusted to 600 with the understanding that efforts to increase the response rate would be amplified. If the response rate was low, there could be an additional collection in the spring. We also clarified that no more than 15 papers would be sampled from each faculty member.

2015-2019

Sampling for General Education varied according to the number of student artifacts obtained and the number of faculty available to score. Faculty members in the Humanities have consistently collected papers using the Gen Ed Assessment checklist to assess over this span of time. Moving forward we aim to use a college-wide panel approach to sampling and assessing student artifacts.

Additional Resources:

When communicating with Ashley Finley at the Association of American Colleges and Universities (AAC&U) in 2015 about our sampling challenges she offered an additional resource, as she explained: “We have devised sampling guidelines for a large multi-state initiative that involves 9 states and 69 two and four year institutions. All of the institutions are drawing samples of student work across three outcomes (written communications, quantitative reasoning, and critical thinking).” They offered tiered options which ranged from a sample of 50 artifacts to 300, depending on several factors outlined in their related guidelines. For more information see: <https://www.aacu.org/value/msc-resources> (May 9, 2018).

VanVoorhis, C. W., & Morgan, B. L. (2007). Understanding power and rules of thumb for determining sample sizes. *Tutorials in Quantitative Methods for Psychology*, 3(2), 43- 50.