Sample Sizes in Assessment: How much is enough?

A common refrain in assessment of student learning is, "How much student work do I need to collect?" As a reminder, the overarching goal of assessment is to establish a process that is meaningful, manageable, and produces useful results at the local level. In order to do this faculty members need to balance the time and energy available to thoughtfully examine artifacts of student learning with the effort required to produce reasonably accurate and useful representations of student ability. The effort required to analyze student work varies with the type of work collected and the complexity of the learning outcome being examined. In cases where the most appropriate student products for assessing a learning outcome are multiple choice or short answer, it often makes sense to use the entire population, especially when the work is machine graded. However, when the student work products that are best suited for demonstrating the learning outcome(s) require in-depth analysis (e.g. exploring lines of computer code, examining student essays, or inspecting the steps shown to solve a mathematical problem) it is generally not reasonable to expect faculty members to give such close scrutiny to the entire population of artifacts beyond what is already occurring within the courses themselves. Therefore, Suskie (2009) argues that the answer to "How much evidence is enough?" is to apply common sense:

> Collect enough evidence to feel reasonably confident that you have a representative sample of what your students have learned and can do. The sample should be large enough and representative enough that you can use the results with confidence to make decisions about a course or program. And take careful steps to ensure the accuracy and truthfulness of your assessment findings.
>
> Suskie, L. (2009). Assessing student learning: A common sense guide (2nd ed.). San Francisco, CA: Jossey-Bass. p. 47.

Additionally, the amount of work must be manageable for the members of the assessment committee to undertake. According the the Middle States Commission on Higher Education, "Effective assessment processes are useful, cost-effective, reasonably accurate and truthful, carefully planned, and organized, systematic, and sustained."[1] This is to say, trying to apply the assessment process to a number of student artifacts that is large enough to conduct a statistical analysis at the expense of providing a thoughtful examination of student work and student learning does not add value to the assessment process. Yet, the question persists, just how much is enough?

---

[1] Middles States Commission on Higher Education. (2007). Student learning assessment: options and resources (2nd ed.) p.55.

If it is decided that statistical analysis is necessary for the results to be useful, then an appropriate sample size needs to be calculated. In order to determine the sample size, it is necessary to know the size of the population and decide on a satisfactory confidence level and an acceptable margin of error. The first step to conducting a statistical analysis is determining the true size of the population that the sample will be drawn from (all students taking Calculus I, the entire sophomore class, every student within a major, and so on). The next step is deciding on an appropriate confidence level. The confidence level refers to the amount of uncertainty that is acceptable. Confidence levels are typically 90%, 95%, or 99% and represent the degree of confidence one has in terms of the result actually being within the margin of error for the sample. To increase the confidence level it is necessary to increase the sample size. Related to the confidence level is the last step, determining the margin of error or range around the answer in which the true result can be found. The margin of error is expressed as the result for the sample plus/minus a percentage often 5% or less. To decrease the margin of error it is necessary to increase the sample size.

For example, for the admitted class of 2021 at the Naval Academy, the population was 1,215 midshipmen. To determine the percentage of males admitted (assuming we didn't already know that it was actually 73% for the population), with a confidence level of 95% and a margin of error +/-5%, it would be necessary to randomly sample a minimum of 293 midshipmen. The result would likely read something like, "It is with 95% confidence that between 68% and 78% of the incoming class is male."

Sample size calculators are commonly found on line:

http://www.raosoft.com/samplesize.html

Or, the following table provides general advice for times when statistical precision is required to obtain usable information.

Population	90% Confidence Level		95% Confidence Level		99% Confidence Level	
	+/-3%	+/- 5%	+/-3%	+/- 5%	+/-3%	+/- 5%
10	10	10	10	10	10	10
20	20	19	20	20	20	20
30	29	28	30	28	30	29
40	39	35	39	37	40	38
50	47	43	48	45	49	47
75	69	59	71	63	73	68
100	89	74	92	80	95	88
150	126	97	132	109	139	123
200	159	116	169	132	181	154
250	188	131	203	152	221	182
500	301	176	341	218	394	286
1000	430	214	517	278	649	400

The above calculations assume a valid assessment instrument (one that accurately corresponds to the learning outcome it purports to assess) and a sample selected using a simple or stratified random sample method. Non-random samples likely contain systematic biases with the result being that the confidence level and margin of error cannot be assumed to be accurate.