

Sampling for the Assessment of Student Learning Outcomes

Rationale: Why sample?

For Student Learning Outcomes (SLO's), we must assess artifacts that reflect the course's desired outcomes. Sampling facilitates the assessment process when it is not feasible to assess all students—for example when programs/courses have large numbers of students or when artifacts take a long time to review. The portion evaluated is the *sample* of the entire population.



Best practice. A subjective artifact using a rubric (e.g. a research project in a capstone course or a paper) may be used **only if it is scored by an evaluation group, not the individual instructor.**



Best practice. When scoring subjective artifacts with a rubric, the evaluation group must norm before scoring. This is especially important for rubrics assessing complicated critical-thinking outcomes. In addition, it is a best practice that each artifact be scored independently by two different evaluators—that is, scored twice by two scorers who don't know that the other gave it.

Census vs. Sampling

For programs that are small, assessing the entire population may yield a more accurate measure of student learning. Assessing the entire population is called a **census** whereas assessing only part of the population is called a **sample**.

Example of Using a Census:

- An Honors section of Music Appreciation ends the course with four students, each of whom is required to write a 10-15 page paper. All four of the course's outcomes are to be assessed by the paper using a rubric. An evaluation group reads all four student papers.
- The Math Department runs eight sections of Intermediate Algebra involving 163 students. One of the outcomes is to be assessed by three questions on an exam. The exam is given in common to all 163 students on a shared platform like Course Compass which permits instructors to see the aggregate results of individual questions. The aggregate results of all eight sections are gathered, comprising responses from all 163 students.

Example of Using a Sample:

- The English Department runs five sections of Critical Thinking Through Argument involving 98 students. Two of the course's four outcomes are to be assessed by a 8-10 page paper scored by a rubric. The English department selects 20 papers randomly from the five sections.

Sampling Procedures

Before evaluating artifacts or data for the SLO, you must:

1. Decide whether you will use a sample or the whole population.
2. Choose an appropriate sample size based on percentage, artifact size and complexity.
3. Choose an appropriate sampling method.

Determining Sample Size

If you have a large program (over 100 students), you may not have the people and time to evaluate 100 artifacts. Therefore, you would choose a specific percentage of students or artifacts.



Best practice. A common standard for sampling is 10% or 10 artifacts, whichever is greater. So for populations less than 100, choose 10; for populations over 100% choose 10%.

Whether or not to sample and the size of the sample depend on three factors, all of which must be kept in mind when making sampling decisions:

1. The length and complexity of the assignments/artifacts.
2. The number of students enrolled in the course or program.
3. The number of faculty members serving as the artifact evaluators.

Length and complexity of the assignments. If the assignment or artifact is of a capstone level (e.g. research project), then a smaller percentage of students might be chosen.

The number of students enrolled in the course or program. If your course or program has less than 100 students, then you should consider using a larger percentage or the entire population. Remember that the acceptable minimum is 10 students.

The number of faculty members serving on the faculty committee. If the program has only three faculty members on the faculty committee, then a smaller sample size would be more appropriate depending on the complexity of the assignment. However, programs with many faculty members and short assignments could have a much larger sample size since there are many more people available to evaluate the artifacts.

Examples

- The Social Sciences Department runs 12 sections of Introduction to Psychology involving 250 students. Three of the course's four outcomes are assessed by a term-ending 8-10 page research paper scored by a rubric. The department has 5 full-time faculty members on staff and 1 adjunct who is willing to participate. This 6-person evaluation committee agrees to score 10% of papers. The committee selects 25 papers randomly from all 12 sections, meets, norms, reads

two papers apiece for best practice, and aggregates the results. The fourth outcome is assessed by three questions on the final exam. The decision is made that it is feasible to assess all students: percentages are gathered independently by each instructor for each section and sent to the chair for compiling.

- The Humanities Department runs 12 sections of Western Civilization involving 250 students. It also has three of the course's four outcomes assessed by a term-ending paper scored with a rubric. However, only 2 members are on the committee, so they assess only 5% of papers. They select 12 papers randomly, meet, norm, read two papers each for best practice, and aggregate the results. The fourth outcome is assessed by a short-essay answer on the final exam scored by a rubric. The 2 members agree that due to the lesser degree of complexity a 10% sampling size is feasible. They select 25 answers randomly, norm, read two apiece, and aggregate the results.

Common Types of Sampling

There are a variety of sampling methods. Simple random, stratified, systemic, and cluster sampling are examples of four common and appropriate sampling methods for institutional assessment activities.

Simple Random Sampling: You randomly select a certain number of students or artifacts. Random sampling can be done easily enough by compiling a list of all students completing the artifact and then using a random number generator, referring to a random number table, or picking out of a hat.

Example: The Business Department runs 3 sections of Introduction to Business involving 112 students. All four outcomes are to be assessed by the final exam, and the department has decided on a 20% sample size. The faculty chair gathers the final exams from the instructors, creates a computerized list in Excel of students who completed the exam, and uses the program's random number generator to identify 22 students for each outcome (different 22 students for each outcome). The results are aggregated.

Stratified Sampling. Students are sorted into homogenous groups and then a random sample is selected from each group. This is useful when there are groups that may be underrepresented.

Example: Child Development 101 has traditionally had few male students. In compiling the data for outcome assessment with a sample size of 20%, the faculty chair makes sure to get 20% of male students by breaking students into gender first before randomly selecting 20% from each group.

Systematic Sampling. You select the nth (e.g. 7th, 9th, 20th) student or artifact from a list.

Example: An Introduction to Art course has been delivered in three sections with a total of 83 students finishing the course. One of the outcomes involves scoring a student art

project with a rubric. The faculty chair has sampled 10% of the student population in the past, but this semester because a section got cancelled, 10% of completing students (8) falls below the minimum threshold of 10. So the faculty chair generates an alphabetical listing of all 83 students who completed the assessment artifact and then selects every 8th student for a minimum of 10.

Cluster Sampling. You randomly select clusters or groups (e.g. classes or sections), and you evaluate the assignments of all the students in those randomly selected clusters or groups.

Example: The English Department offers Freshman Composition in eight sections across four sites including two sections online, involving a total of 143 students. There is a single assessment instrument: a final paper scored by a rubric that contains all SLO's. Two of these sections are randomly chosen and all papers from those two sections assessed.

A Note about Stratified Sampling:

For the purposes of Cerro Coso, this method should be preferred when courses are offered in some combination online and onsite sections, or when offered at multiple sites.

A Note about Cluster Sampling:

Given the amount of variation that can happen between different instructors in different locations in different delivery modes, reservations have been raised about the viability of this choice for giving the department and the college a clear picture of any particular course assessed by it.

Nevertheless, the consensus has been to move ahead with this option for this year until it can be assessed in turn as serving our needs or not. However, clear ground rules need to be established in order to avoid the perception of bias that could result from such a selection. In particular, it must be understood that the selection of the section(s) be *random*.

To that end, the following process will be observed: sections to be identified for cluster sampling will be chosen randomly according to the best practices in this document by the Institutional Effectiveness Committee (faculty chairs are welcome to be present at the time of selection) and will be chosen after the mid-point of the semester in which the course is to be assessed. This assures that 1) no perception of bias is present, and 2) that all sections are deploying the assessment instruments as intended.