

SAMPLING GUIDELINES FOR ASSESSMENT

This guide, developed by **Academic Planning and Accountability**, provides practical guidance for **FIU faculty, staff, assessment coordinators, and other personnel involved in academic or administrative assessment**. It explains when sampling may be appropriate, how to select a sampling strategy, how to consider representation across populations and locations, and how to interpret and document results. The guide also addresses the use of sampling for direct and indirect measures, pilot assessments, preliminary benchmarking, and scaling assessment efforts over time.

For questions about these guidelines or assistance with sampling for assessment, contact **Katherine Perez** at kathpere@fiu.edu.

CONTENTS

Purpose.....	3
Population, Census, and Sample.....	3
Summary Chart: Census or Sample?	4
Example of Using a Census	4
Example of Using a Sample	4
When Sampling May Be Appropriate	5
Summary Chart: Common Uses of Sampling in Assessment	5
Assessing Large Populations	6
Ensuring Representation Across Relevant Groups	6
Assessing Student Learning	6
Conducting a Pilot Assessment.....	6
Establishing a Preliminary Benchmark.....	7
Scaling an Assessment Process	7
Addressing Feasibility.....	7
Representation Across Populations, Methods of Delivery, and Locations.....	7
Sampling Methods	8
Summary Chart: Choosing a Sampling Method	9
Simple Random Sampling.....	9
Systematic Random Sampling	10
Stratified Random Sampling	10
Cluster Sampling.....	10

Convenience Sampling.....	11
Purposive Sampling.....	11
Choosing an Appropriate Sampling Strategy.....	11
Sample Size Considerations.....	12
Summary Chart: Working With Small Samples	12
Strategies When Available Evidence Is Small.....	13
Include All Available Evidence.....	13
Use Multiple Sources or Methods of Evidence When Appropriate	13
Gather Data at Multiple Points	13
Use Well-Designed Assessment Instruments	14
Sampling for Piloting, Benchmarking, and Scaling	14
Summary Chart: From Pilot to Broader Implementation	14
Documenting Sampling in Assessment Reports	15
Example	15
Interpreting Results From a Sample	15
Sampling at a Glance	16

PURPOSE

Sampling is a data-collection strategy that allows faculty and staff to gather evidence from a subset of a larger population when assessing the entire population is unnecessary, impractical, or not appropriate for the purpose of the assessment. When thoughtfully designed, sampling can provide useful and sufficiently representative evidence to support assessment, planning, decision-making, and continuous improvement.

These guidelines apply to **academic and administrative assessment** and may be used with both **direct and indirect measures**. Sampling may be appropriate when assessing student learning, evaluating programs or services, examining stakeholder experiences, conducting pilot assessments, establishing preliminary benchmarks, or testing an assessment process before expanding data collection to a larger population.

Sampling should be purposeful and appropriate to the assessment question. The sampling strategy should consider:

- the population being assessed;
- the type of evidence being collected;
- how the results will be used;
- the conclusions that will be drawn from the results; and
- whether relevant groups, methods of delivery, instructional locations, service locations, or other characteristics of the population need to be represented.

The purpose of sampling is not simply to reduce the amount of data collected. A well-designed sampling strategy should provide evidence that is appropriate for answering the assessment question and sufficiently representative to support meaningful interpretation and improvement decisions.

POPULATION, CENSUS, AND SAMPLE

A **population** is the entire group about which the assessment is intended to provide information. Depending on the assessment, a population might include all students enrolled in an academic program, all students completing a particular course, all participants in a university program or activity, all employees served by an administrative unit, all individuals using a particular service, or all student artifacts eligible for assessment.

A **census** includes the entire defined population in the assessment.

A **sample** includes a subset of the defined population selected for assessment.

For smaller populations, or when data can be collected and analyzed efficiently for all eligible participants or artifacts, a census may provide the most complete evidence. For larger populations, or

when assessment requires substantial time or resources, an appropriately designed sample may provide sufficient evidence for meaningful interpretation and decision-making.

SUMMARY CHART: CENSUS OR SAMPLE?

Consideration	Consider a Census When...	Consider Sampling When...
Population Size	The population is relatively small.	The population is large.
Availability of Evidence	All eligible participants, records, or artifacts are readily available.	Reviewing all eligible evidence would be difficult or unnecessary.
Assessment Method	Data can be collected or analyzed efficiently for everyone.	Assessment requires substantial manual review, scoring, interviewing, or analysis.
Representation	Including everyone provides the clearest picture of the population.	A well-designed sample can appropriately represent the population.
Resources	Assessing the population is feasible within available resources.	A census would require disproportionate time or resources.
Intended Use	Decisions require information about the entire population.	The sample can provide sufficient evidence for the conclusions and decisions being made.

EXAMPLE OF USING A CENSUS

A graduate program has eight students completing a capstone experience during the assessment period. Because the population is small and all eight artifacts are available, faculty assess all eight student artifacts rather than selecting a sample.

EXAMPLE OF USING A SAMPLE

An undergraduate program has 250 eligible capstone papers, each of which requires substantial faculty time to assess using a common rubric. Faculty select an appropriately designed sample of the papers for assessment rather than scoring all 250 artifacts.

WHEN SAMPLING MAY BE APPROPRIATE

Sampling may support several different assessment purposes.

SUMMARY CHART: COMMON USES OF SAMPLING IN ASSESSMENT

Purpose	How Sampling Can Be Used	Example
Assess a Large Population	Select sufficient evidence from a larger population rather than assessing every eligible case.	Faculty assess a randomly selected group of capstone papers using a common rubric.
Ensure Representation	Structure the sample to represent relevant groups, methods of delivery, locations, or populations.	Student artifacts are sampled from the applicable methods of delivery in which a program is offered.
Assess Student Learning	Sample evidence for direct or indirect measures of learning.	Faculty sample student projects for direct assessment or students for an indirect learning survey.
Conduct a Pilot	Test an instrument or assessment process with a smaller group before wider implementation.	A new rubric is piloted in selected course sections.
Establish a Preliminary Benchmark	Use initial assessment data to develop a starting point for understanding performance.	First-year results from a pilot provide baseline information used to inform future criteria for success.
Scale an Assessment	Begin with a manageable group, evaluate the process, and expand after refinement.	A service assessment begins at one location and later expands to additional locations.
Address Feasibility	Reduce the volume of evidence when a census would require disproportionate resources.	A sample of lengthy research papers is assessed rather than all eligible papers.

Assessing Large Populations

Sampling can reduce the resources required to review large numbers of student artifacts, survey responses, records, or other forms of evidence while still providing useful information about the population.

For example, manually scoring 50 randomly selected student research papers using a rubric may provide a more feasible assessment process than scoring several hundred papers when the sample is appropriate for the assessment purpose.

Ensuring Representation Across Relevant Groups

Sampling can be designed to include students, participants, or other stakeholders from different methods of delivery, instructional or service locations, programs, sections, campuses, or other groups relevant to the assessment.

For example, a program offered through multiple methods of delivery may structure its sample so that eligible students or artifacts from each applicable method of delivery are appropriately represented.

Assessing Student Learning

Sampling may be used with both **direct and indirect measures of student learning**.

Direct measures may include examinations, projects, research papers, presentations, performances, portfolios, or other student work assessed using established criteria or rubrics.

Indirect measures may include surveys, interviews, focus groups, self-assessments, or other measures that capture students' perceptions, experiences, or reflections related to learning.

The sampling approach should be appropriate for both the measure being used and the conclusions that will be drawn from the results.

Conducting a Pilot Assessment

A smaller sample may be used to test an assessment instrument, rubric, survey, process, or data-collection strategy before implementing the assessment more broadly.

Pilot assessment can help faculty and staff determine whether the instrument produces useful information, items or rubric criteria are clearly understood, the assessment process is feasible, data can be collected and analyzed as intended, and modifications are needed before broader implementation.

Results from a pilot should be interpreted within the context of the population and sample used.

Establishing a Preliminary Benchmark

Data from an appropriately selected sample may be used to establish a **preliminary benchmark** when no prior assessment data are available.

For example, a unit implementing a new measure may collect baseline data from an initial sample and use the results to better understand current performance before establishing or refining future criteria for success.

Benchmarks established through a pilot or small initial sample should be identified as preliminary and reconsidered as additional evidence becomes available.

Scaling an Assessment Process

Units may begin with a smaller population or selected group to determine whether an assessment process is feasible and useful before expanding it.

This approach allows units to **pilot, learn, refine, and scale** assessment processes while maintaining attention to the representativeness and limitations of the initial data.

Addressing Feasibility

Sampling may also be appropriate when assessing an entire population would require substantial resources or more time than reasonably available, or when complete population data cannot be obtained.

Feasibility alone, however, should not determine the sampling strategy. The resulting evidence must still be appropriate for the assessment purpose.

REPRESENTATION ACROSS POPULATIONS, METHODS OF DELIVERY, AND LOCATIONS

An important consideration in sampling is whether the sample adequately represents the population about which conclusions will be made.

When a population includes meaningful subgroups, faculty and staff should consider whether those groups should be represented in the assessment.

Depending on the purpose of the assessment, relevant groups may include different:

- methods of delivery;
- instructional locations or campuses;
- service locations;

- academic programs or levels;
- course sections;
- participant or stakeholder groups; or
- other characteristics directly related to the assessment question.

For academic assessment, if a program is offered through multiple **methods of delivery or instructional locations**, the sampling strategy should provide appropriate representation of the populations included within the scope of the assessment.

For administrative assessment, units that provide services across multiple FIU locations or to distinct populations should similarly consider whether their sample adequately represents the individuals or groups served.

Representation does not necessarily require an equal number of participants from every group. Depending on the purpose of the assessment, samples may be selected proportionally to the size of each group or intentionally structured to provide sufficient information about smaller groups.

The key question is whether the sample provides appropriate evidence for the population and conclusions being examined.

SAMPLING METHODS

The sampling method should be selected based on the assessment purpose, characteristics of the population, type of evidence being collected, and conclusions that will be drawn from the results.

Sampling methods generally fall into two categories: **probability sampling** and **non-probability sampling**.

Probability sampling uses random selection and is generally preferable when the goal is to obtain a sample that supports conclusions about a larger population.

Non-probability sampling does not use random selection and may be appropriate for exploratory assessment, qualitative inquiry, pilot testing, or situations in which random selection is not feasible. Greater caution should be used when generalizing results from a non-probability sample to the larger population.

SUMMARY CHART: CHOOSING A SAMPLING METHOD

Sampling Method	Type	Best Used When...	Example
Simple Random	Probability	Eligible cases can be identified and each should have an equal opportunity for selection.	Randomly select 50 of 200 eligible capstone papers.
Systematic Random	Probability	A complete ordered list is available and a systematic selection process is efficient.	Select a random starting point and then every 10th eligible participant.
Stratified Random	Probability	Important subgroups should be appropriately represented.	Randomly sample artifacts from each applicable method of delivery.
Cluster	Probability	The population naturally occurs in intact groups such as course sections.	Randomly select three course sections and assess all eligible artifacts within them.
Convenience	Non-probability	Readily available participants can provide useful information and broader generalization is not required.	Survey students who attend an advising workshop.
Purposive	Non-probability	Specific participants or cases possess experiences directly relevant to the assessment question.	Conduct a focus group with students who used a newly implemented service.

Simple Random Sampling

Definition: Each eligible person, case, or artifact in the population has an equal opportunity to be selected through a random process.

Example: A program has 200 eligible capstone papers. Faculty use a randomized selection process to select 50 papers for assessment using a common rubric.

Systematic Random Sampling

Definition: Members of a population are placed on a list, a random starting point is selected, and every *n*th person, case, or artifact is selected thereafter.

Example: A unit has an ordered list of 300 eligible participants. A random starting point is selected among the first 10 individuals, and every 10th individual thereafter is included in the sample.

Care should be taken to ensure that the ordering of the list does not contain a pattern related to the characteristic being assessed that could systematically influence who is selected.

Stratified Random Sampling

Definition: The population is divided into relevant subgroups, or **strata**, and participants or artifacts are randomly selected from each subgroup.

Stratified sampling is particularly useful when representation of specific groups is important to the assessment.

Example: An academic program is offered through multiple methods of delivery. Eligible student artifacts are grouped by method of delivery, and artifacts are randomly selected from each group in proportion to enrollment.

A unit may also intentionally select a larger proportion from a smaller subgroup when sufficient evidence from that group is necessary for the assessment purpose. When this occurs, the sampling approach should be documented.

Cluster Sampling

Definition: Naturally occurring or intact groups, or **clusters**, are randomly selected, and all eligible participants or artifacts within the selected clusters are assessed.

Example: A course has 10 sections. Three sections are randomly selected, and all eligible student artifacts within those sections are assessed.

Because individuals within a cluster may share experiences or characteristics, faculty and staff should consider whether the selected clusters adequately reflect the broader population before drawing conclusions about that population.

Convenience Sampling

Definition: Participants, cases, or evidence are selected primarily because they are readily available or accessible.

Example: Students attending an advising workshop are invited to complete a post-session survey.

Convenience samples can provide useful assessment information, particularly about participants who experienced a specific program or service. However, the results may not represent individuals who did not participate or were not available for assessment.

Purposive Sampling

Definition: Participants, artifacts, or cases are intentionally selected because they possess characteristics, experiences, or perspectives that are relevant to the assessment question.

Example: An administrative unit conducts a focus group with students who used a newly implemented service to gather detailed feedback about their experiences.

Purposive sampling is frequently appropriate for qualitative assessment when the objective is to obtain detailed information from individuals with specific knowledge or experience rather than to statistically represent an entire population.

CHOOSING AN APPROPRIATE SAMPLING STRATEGY

There is no single sampling strategy that is appropriate for every assessment. The strength of a sample is determined not simply by its size but by whether the sampling strategy is appropriate for the assessment purpose and supports the intended interpretation of the results.

Faculty and staff should consider:

1. **What is the assessment question?**
2. **Who or what constitutes the population?**
3. **How large is the population?**
4. **What type of evidence will be collected?**
5. **Is the objective to describe the larger population, understand a particular group, pilot a process, or explore an issue in greater depth?**
6. **Are there important groups, methods of delivery, instructional locations, or service locations that should be represented?**
7. **How will the results be used?**
8. **What conclusions will be drawn from the data?**
9. **What resources are available for data collection and analysis?**

SAMPLE SIZE CONSIDERATIONS

There is no universal minimum sample size that is appropriate for all assessment activities.

An appropriate sample size depends on factors such as the size of the population, sampling method, assessment instrument, type and complexity of the evidence collected, variation within the population, number of relevant subgroups, intended use of the results, and conclusions the unit intends to draw.

For assessment purposes, the central question is whether the available evidence is sufficient to support **reasonable and justifiable conclusions and improvement decisions**.

When the population is small and all eligible evidence can reasonably be assessed, using a census is generally preferable to selecting an even smaller sample.

SUMMARY CHART: WORKING WITH SMALL SAMPLES

Situation	Recommended Approach	Interpretation Consideration
Small Population	Assess the entire population when feasible.	A census provides the most complete information available for that population.
Few Available Artifacts	Include all eligible artifacts and consider additional evidence when appropriate.	Distinguish the number of students from the number of independent artifacts.
Low Survey Response	Consider additional outreach, another collection period, or complementary evidence.	A low response rate may introduce nonresponse bias.
Small Qualitative Sample	Use structured interviews or focus groups to obtain detailed evidence.	Findings may provide valuable insight but should not automatically be generalized to the entire population.
Pilot Data	Use results to evaluate the instrument or process and establish preliminary information.	Treat initial benchmarks and conclusions as preliminary.
Insufficient Evidence at One Point	Collect additional evidence across appropriate sections, locations, or time periods.	Document what was combined and consider contextual differences.

Examples of situations requiring additional consideration include:

- A capstone course enrolls 40 students, but students complete the assessed project in four teams. Although 40 students participated, only four independent project artifacts are available for assessment.
- A survey is distributed to 100 individuals, but only 10 respond. The number of invitations does not compensate for the limited number of responses available for analysis.
- Six students participate in a focus group. Although the number is small, the qualitative data may provide detailed information useful for identifying patterns, experiences, or opportunities for improvement. Findings should not automatically be generalized to the entire population.
- A graduating cohort includes only five students, and all five complete the dissertation or capstone artifact being assessed. Assessing all available artifacts provides a census of the cohort.

Small numbers do not automatically make assessment evidence unusable. They do, however, affect the conclusions that can reasonably be drawn from the results.

STRATEGIES WHEN AVAILABLE EVIDENCE IS SMALL

Include All Available Evidence

When the number of eligible participants or artifacts is small, assess the entire available population when feasible rather than selecting an even smaller sample.

Use Multiple Sources or Methods of Evidence When Appropriate

When different measures address the same outcome or assessment question, multiple sources of evidence may provide a more complete understanding of performance.

Converging evidence from multiple sources or methods can strengthen confidence in the interpretation of results and provide additional information for improvement. Different measures should be selected because they contribute meaningful evidence—not simply to increase the volume of available data.

Gather Data at Multiple Points

If one period of data collection produces insufficient evidence, additional data may be gathered at another appropriate point in time.

This may include another semester, additional course sections, expanded participation opportunities, additional applicable service locations, or a subsequent assessment cycle.

When combining data across multiple points in time, units should document the periods included and consider whether changes in the population, instrument, or context affect interpretation.

Use Well-Designed Assessment Instruments

High-quality instruments can provide more meaningful information even when the number of participants or artifacts is limited.

Examples include:

- **Surveys:** Use well-designed or validated items clearly aligned with the assessment question or outcome.
- **Tests or examinations:** Use appropriately designed items that provide sufficient evidence of the knowledge or skills addressed by the outcome.
- **Projects, essays, performances, or presentations:** Use clearly defined rubric criteria aligned with the outcome.
- **Interviews and focus groups:** Use structured or semi-structured protocols aligned with the assessment question and systematically analyze responses for relevant patterns and themes.

The objective is to maximize the **quality and relevance of the information collected**, rather than simply increasing the number of questions, criteria, or data points.

SAMPLING FOR PILOTING, BENCHMARKING, AND SCALING

Sampling can be especially useful when developing or expanding an assessment process.

SUMMARY CHART: FROM PILOT TO BROADER IMPLEMENTATION

Stage	Purpose	Key Question
1. Pilot	Test the assessment with a manageable sample or population.	Does the instrument and process work as intended?
2. Evaluate	Examine the quality and usefulness of the evidence generated.	Did the assessment provide information that can support decisions?
3. Refine	Modify the instrument, procedures, or analysis based on the pilot.	What should be improved before broader implementation?
4. Benchmark	Use initial findings to develop a preliminary understanding of current performance.	What does the initial evidence suggest about current performance?
5. Scale	Expand assessment to additional groups, locations, methods of delivery, or populations.	How can the assessment be implemented more broadly while maintaining quality and appropriate representation?

Any benchmark developed from limited pilot data should be identified as **preliminary**. As additional evidence becomes available, units should revisit preliminary benchmarks, criteria for success, instruments, and sampling approaches to determine whether revisions are warranted.

DOCUMENTING SAMPLING IN ASSESSMENT REPORTS

When a sample is used, the assessment report should provide enough information for a reader to understand **who or what was eligible for assessment, what was actually assessed, and how the sample was selected**.

As applicable, document:

- the population being assessed;
- the number of eligible participants, cases, or artifacts;
- the number included in the sample;
- the sampling method used;
- how the sample was selected;
- representation of relevant methods of delivery, instructional locations, service locations, or other meaningful groups; and
- known limitations that may affect interpretation of the results.

EXAMPLE

The eligible population consisted of 286 graduating students across applicable methods of delivery. A stratified random sample of 90 student artifacts was selected, with artifacts sampled proportionally from each method of delivery represented in the program. Faculty assessed all selected artifacts using the program's common rubric.

Documentation does not need to be unnecessarily complex. It should provide sufficient information to demonstrate that the sampling approach was intentional and appropriate for the assessment.

INTERPRETING RESULTS FROM A SAMPLE

Assessment results should be interpreted within the context of the sampling strategy used.

Faculty and staff should avoid drawing conclusions that extend beyond what the available evidence can reasonably support. In particular:

- results from a convenience or purposive sample should not automatically be generalized to the entire population;
- very small samples should be interpreted cautiously;
- low survey response rates may introduce nonresponse bias;

- samples that omit relevant groups or locations may not adequately represent the full population;
- pilot results should generally be treated as preliminary; and
- unexpected findings may warrant additional data collection before major conclusions or decisions are made.

Limitations do not necessarily invalidate assessment results. Identifying limitations helps faculty and staff determine the appropriate scope of interpretation and whether additional evidence is needed.

SAMPLING AT A GLANCE

Principle	What It Means for Assessment
Purposeful	Begin with the assessment question and intended use of the results.
Appropriate	Match the sampling strategy to the population, measure, and type of evidence needed.
Representative When Needed	Consider relevant groups, methods of delivery, instructional locations, service locations, and populations.
Sufficient	Collect enough appropriate evidence to support reasonable conclusions and improvement decisions.
Feasible	Balance methodological quality with available time and resources.
Transparent	Document the population, sample size, selection process, and relevant limitations.
Interpreted Appropriately	Draw conclusions that are supported by the sampling method and available evidence.

The goal of sampling is not to assess the fewest possible participants or artifacts. The goal is to collect sufficient, appropriate, and useful evidence to support meaningful assessment and continuous improvement.