

Program Evaluation Toolkit

Module 4, Chapter 1: Evaluation Samples

Regional Educational
Laboratory
Central

From the National Center for Education Evaluation at IES

Speaker 1:

Welcome to the first chapter of the fourth module in the program evaluation toolkit. Now that you understand logic models, evaluation questions, and evaluation design categories, let's consider sampling for data collection. To ensure that your findings align with your evaluation needs, the participants included in your dataset (also known as the "sample") must represent the individuals and contexts to which you hope to apply the results of your evaluation. Therefore, you need to consider the characteristics of the sample from which you will collect data as well as the characteristics of the population to which you will apply the results.

At the end of this module, you will have a greater understanding of what sampling is, how you can use sampling procedures to obtain a desirable sample, and what the limitations of sampling procedures are.

This module includes three chapters, each focusing on a different component of sampling. Chapter 1 answers the question "What is the purpose of sampling?" and defines key terms. Chapter 2 answers the question "What sampling techniques should I consider?" and provides resources and activities for you to practice applying the learnings. Finally, chapter 3 answers the question "How do I make a sampling plan?" and will support you in drafting a sampling plan.

Please refer to the resources page on the website, which includes worksheets, templates, and other resources to help you develop your own sampling plan.

Let's get started with our first chapter, in which you will learn the purpose of sampling and key terms used in sampling, including *census*, *population*, *representativeness*, and *generalizability*.

Sampling is necessary because the answer to the question "Can data be collected from everyone?" is usually "no." Each decade, the U.S. government attempts to collect information about the entire population of the country. The *population* includes all possible participants. When it is possible to collect information from everyone in a population, the process is called a *census*. You may have heard or read about how difficult this process is.

When you conduct an evaluation, it is often more feasible to identify a subset of the entire population for data collection. This smaller subset is called a *sample*. If the sample does not represent the population in important ways, however, the results may not be useful. When you decide who to include in a sample, there are important considerations and established techniques to help you ensure the sample yields useful results.

You do not have to eat the entire bowl of ice cream to know how it tastes. This is true of the population of an evaluation as well. You don't have to sample the entire population in order to describe that population. Instead, you can take a representative sample of the population you are interested in describing.

However, it is important that you don't take a haphazard approach, or you will not get the full picture of the population. For example, if you take only one bite from a banana split, you might taste only whipped cream, banana, or one flavor of the ice cream. That won't allow you to describe the entire dish. You need to be sure you've taken enough bites to know all the characteristics. In program evaluation, this is called sampling.

Scientists have devised systematic sampling approaches for gathering information from part of a population in order to understand the entire population. This module includes guidance on how to sample a population.

Not all evaluations involve making inferences to a larger population. For example, if you are implementing a program at one school, and you only need to understand how well the program is working at that school, then you can consider the individuals at that school as your population. In this case, you may be able to collect data from the entire population.

However, if you are not able to collect data from the entire population in an evaluation, you need to consider three things: *representativeness*, *generalizability*, and *weighting*. Representativeness refers to how well the subset from which you can collect data represents the entire population of interest. Generalizability is the extent to which your results will apply to different types of individuals and contexts. Weighting refers to making statistical adjustments to your data to help improve representativeness.

Participants in your evaluation may have characteristics that will affect the results. These characteristics are often outside of your control as the evaluator. If you ensure that your sample is representative of the entire population, you can have increased confidence that the results will be similar to what you would have found if you collected data from everyone. When generalizations are made back to the population but the sample does not represent the population, the results may not be valid due to sampling error.

In the figure to the right, each icon of a particular color represents an individual in the population with a certain characteristic. Let's say that each color represents attendance at a different middle school. The middle group is larger, so you need to make sure that there are more participants from that group in your sample so that it represents the entire population accurately. Thus, the sample in the figure includes two individuals from the middle group and one individual from each of the other, smaller groups in the population.

Attendance at different schools is only one characteristic you might consider. You might also consider locale, gender, age, race/ethnicity, and other characteristics specific to the context of your evaluation. Even though you may not be able to ensure that the sample is representative with respect to all these characteristics, you should consider the characteristics in data collection and again in the interpretation and limitations of the results. Choosing which characteristic or

characteristics to prioritize when sampling depends on what evaluation questions you want to answer. If you want to ensure representativeness with respect to many characteristics, you may need a larger sample size, or you may need to prioritize certain characteristics over others.

Now it's time to try your hand at selecting a representative sample. The *Representative Sample Activity*, available on the resources page of the website, includes a list of 20 teachers who volunteered to be interviewed about the usefulness of AMMP!. Both the activity and this slide include district demographics: gender, race/ethnicity, age, academic degree, and experience. The AMMP! evaluation team can select only 10 teachers to interview because of resource constraints. Which 10 would you select if you were on the evaluation team? How does your chosen sample compare to the district demographics? Be sure to answer the reflective questions at the end of the activity. This activity will take you approximately 20 minutes.

You may have noticed when completing the activity that it was challenging to construct a sample that was representative of the population with respect to all the characteristics. Random sampling is a solution to this problem. Random sampling can ensure that, at least on average, a sample is representative of the entire population. However, to effectively represent the population, a random sample must be relatively large. Chapter 2 discusses random sampling and shows how to select a random sample using Microsoft Excel. Chapter 3 discusses sample size requirements.

Another key consideration when sampling is *generalizability*, which concerns the extent to which your results will apply to different types of individuals and contexts. For example, in the previous activity, all the teachers volunteered to be interviewed. It is important that the AMMP! evaluation team is aware of this limitation when the team decide to which populations the team might be able to generalize the results obtained from the sample. The 20 teachers who volunteered may be fundamentally different in some way from teachers who did not volunteer. The results might generalize to a population of teachers who were willing to be interviewed, but they would not generalize to all possible teachers. The evaluation team should explain this limitation when the team report the results so that the audience clearly understands the constraints of the evaluation.

Sometimes it may be impossible to ensure that your sample is representative of the entire population with respect to particular characteristics. In these cases, you may allow some responses to count more than other responses so that the results more accurately reflect the target population. You can do this by weighting.

For example, the AMMP! evaluation team wants to survey middle school teachers about their approval of the program. The composition of all teachers in the middle school is 50 percent grade 8 teachers, 25 percent 7, and 25 percent 6 grade teachers. However, the team's sample includes five teachers from each grade. In other words, the composition of the sample is even for all grades, but grade 8 teachers make up a larger proportion of all math teachers. The team may want to statistically adjust its data so that the data more accurately represent the grade distribution in the middle school. Weighting allows the team to adjust the data so that the statistics the team computes from the data more accurately represent what the team would likely obtain from the entire population.

To weight the data, the AMMP! evaluation team needs to multiply the grade 8 surveys by 1.5. The team also needs to multiply each grade 6 and 7 data point by 0.75. This adjustment allows the total weights for the grade 8 and grade 6 and 7 teachers in the sample to count the same. Using these weights will ensure that the responses from grade 6 and 7 teachers are not influencing the results more than they would in the entire population. When the team reports the results, it is important that the team acknowledges that this adjustment assumes that grade 8 teachers who respond to the survey answer in the same way as other grade 8 teachers would.

For more information on generalizability, you can follow the link [here](#) or on the resource page to the Generalizer. The Generalizer is a tool to guide you in selecting a sample of schools that is representative of a particular population of public schools. It can also help you determine how well the results of a completed evaluation might generalize to a particular population of interest.

By using the Generalizer, you can define a broadly or narrowly representative sample of public schools by specifying geographic and demographic criteria. The Generalizer will use your criteria to identify an appropriate sample by querying its database of schools, derived from the U.S. Department of Education's Common Core of Data and the U.S. Census Bureau's American Community Survey. Take a moment to explore the Generalizer if you need to identify a particular sample of schools to meet your evaluation needs. It includes step-by-step directions that will guide you in using the tool.

The Generalizer can be very useful for targeting schools that meet certain criteria for an evaluation. For instance, it can help you identify schools that have large enrollments, are located in urban areas, or enroll large percentages of students from racial/ethnic minority groups. However, the tool is limited in that the information provided is specific to public schools. It will not help you choose a sample of school districts with certain characteristics, nor will it help you choose a sample of students with certain characteristics in a particular school.

Next, chapter 2 explores different sampling techniques and shows how each technique might be used in the context of the AMMP! example, the after-school middle-grades math program introduced in module 1.

This handout was prepared under Contract ED-IES-17-C-0005 by Regional Educational Laboratory Central, administered by Marzano Research. The content does not necessarily reflect the views or policies of IES or the U.S. Department of Education, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.