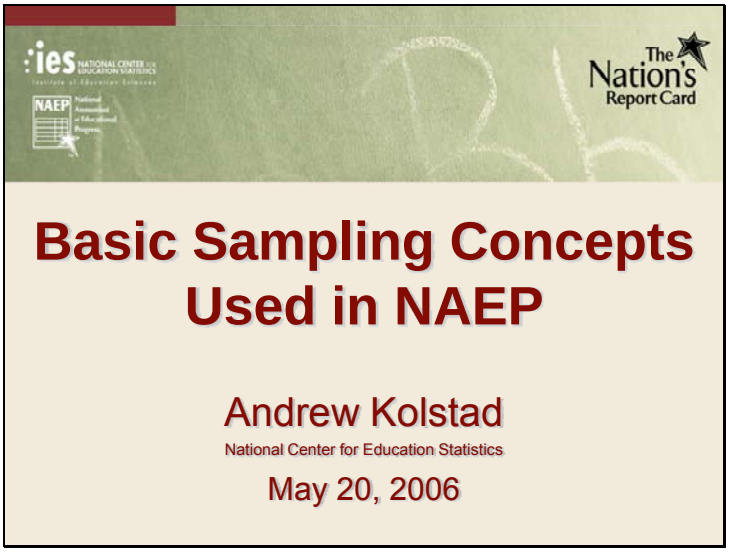
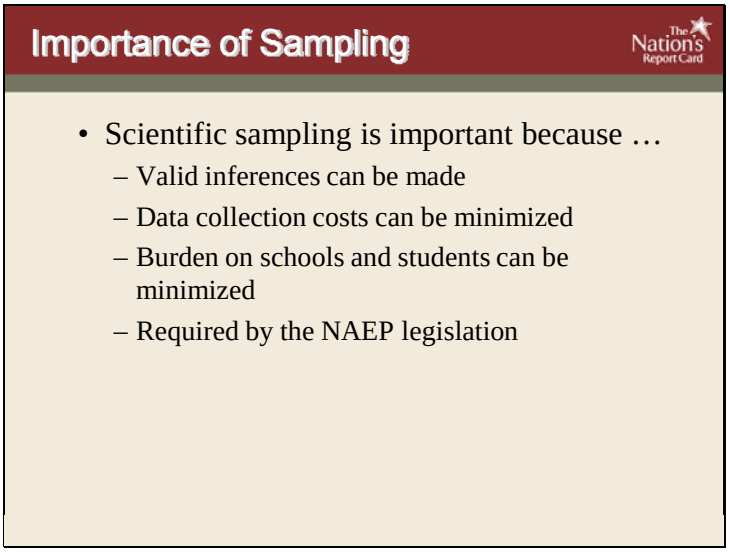
Slide 1



This presentation provides an introduction to the basic concepts of sampling and its importance to the National Assessment of Educational Progress (NAEP).

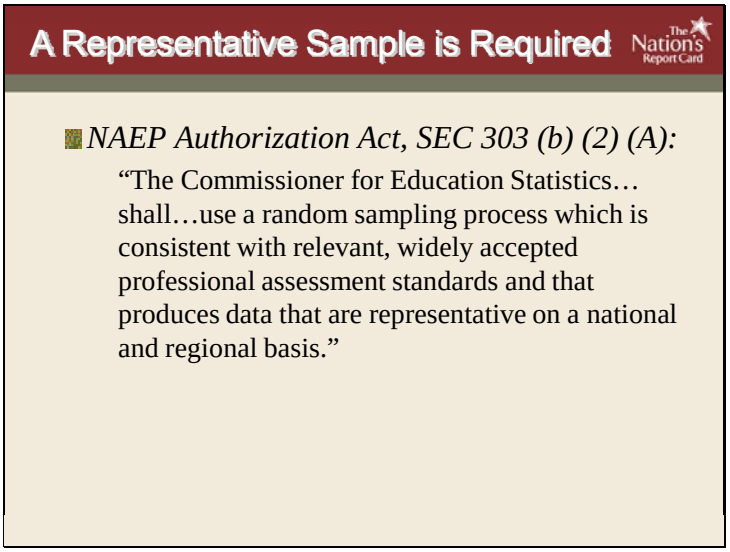
Slide 2



Sampling is a branch of statistics that is critical to conducting the NAEP assessments. NAEP is required by legislation to use scientific sampling procedures. By using scientific sampling, NAEP is able to:

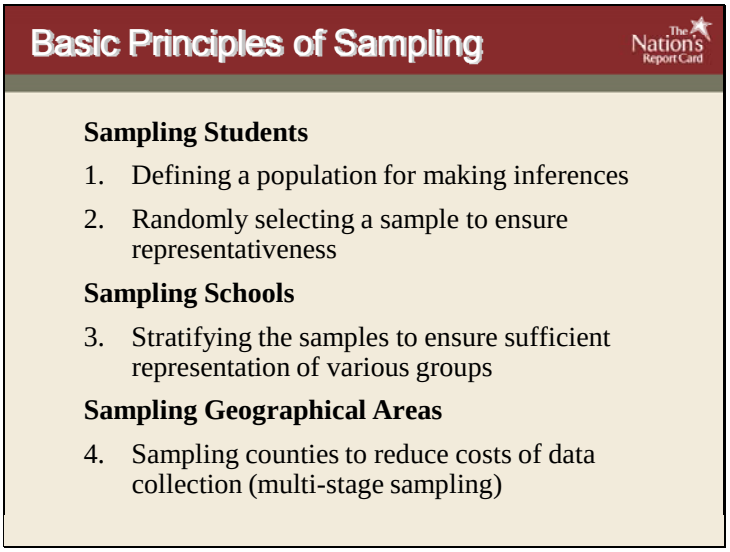
- Make valid statements about the performance of student groups across the nation (12 million students total), even though only one–tenth of these students are assessed.
- Contain costs and concentrate data collection efforts.
- Minimize the burden to the nation's school systems by only assessing a representative sample of their students.

Slide 3



The NAEP legislation requires "a random sampling process" that produces representative data on a national and regional basis. The sampling methods used by NAEP must have a strong scientific basis consistent with widely accepted professional testing standards.

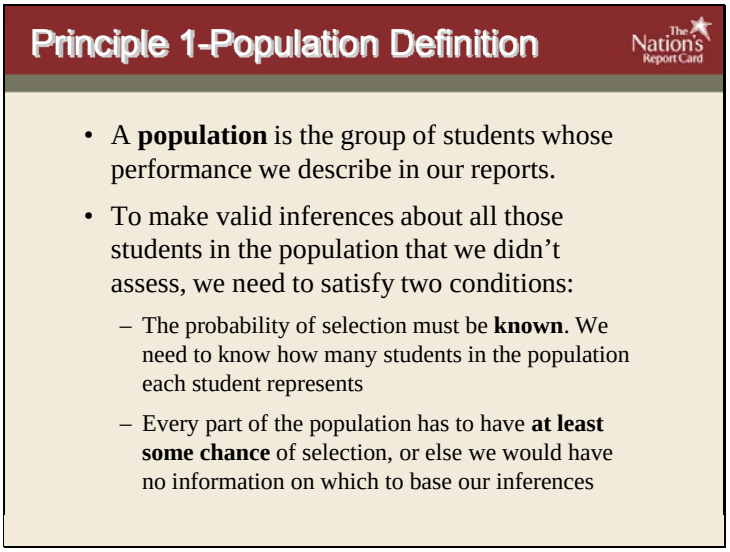
Slide 4



There are three stages of sampling. The first stage is sampling students. To sample students, NAEP defines the population for making inferences and randomly selects a sample to ensure representativeness. The second step, sampling schools, is a little more complicated and sometimes requires sample stratification by school type or various other groupings of schools. The third level of sampling applies to sampling geographical areas.

The following slides address each of the basic principles in sampling students, schools, and geographical areas.

Slide 5



The first principle of sampling is defining a population for making inferences. In this case, a population is the group of students for whom NAEP test performance is reported, even those not assessed. In this country the total population of students is:

- 4.1 million fourth–grade students; but NAEP only tests 180,000
- 3.8 million eighth–grade students; but NAEP only tests 180,000
- 2.6 million twelfth–grade students; but NAEP only tests 16,000
  (The number of students NAEP tests in grades 4 and 8 is larger because NAEP produces results at the state level and at the large urban district level for several subject areas. At grade 12, NAEP reports at the national level only, so a smaller number of students is needed.)

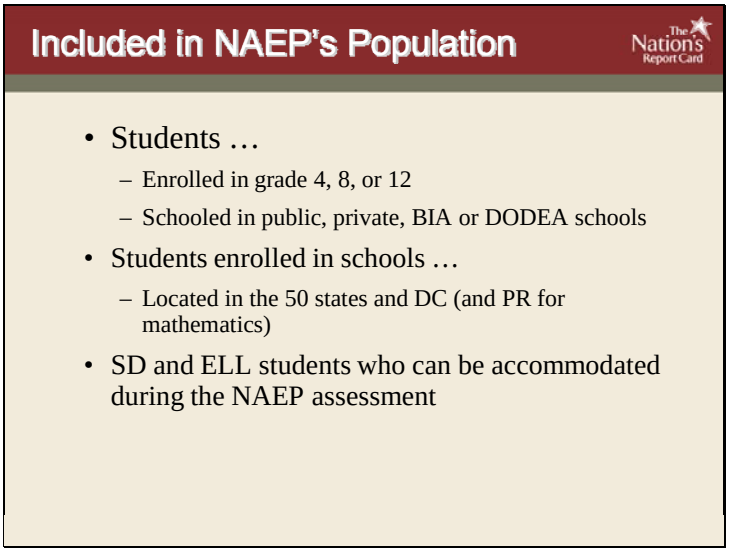
Because of scientific sampling, it is possible to generalize test performance to the total population of students from only the fraction of students who are assessed. To generalize—or to make valid inferences—about the total population of students, even those not assessed, two conditions must be satisfied:

The first condition is that the probability of selection must be known. For example, with college admissions tests, the assessment programs do not know the probability of selection because the test takers are self-selected. So, those assessment programs cannot generalize to anyone who does not take their tests. However, NAEP knows the

probability of selection of all students because they are selected by NAEP to take the assessment.

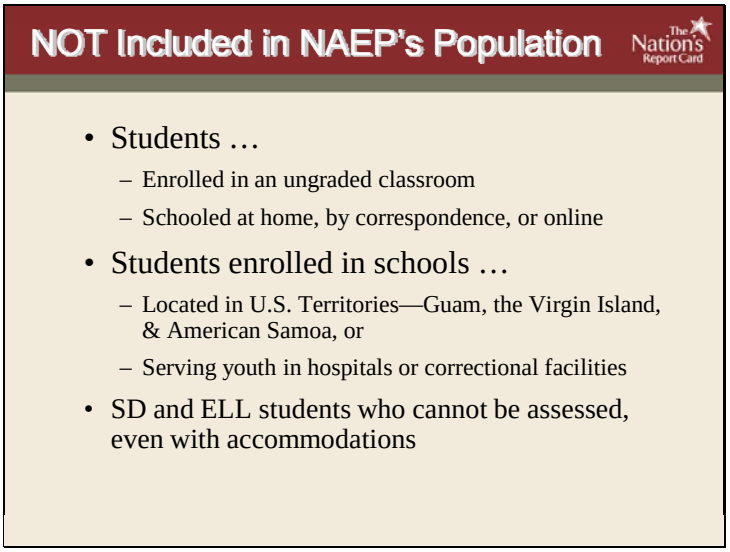
The second condition is that every group must have at least some chance of being selected. If some part of our student population has no chance of selection, NAEP would not have the information to generalize about the performance part of that group.

Slide 6



NAEP's target population for the national assessments consists of 4th , 8th , and 12th grade U.S. students who are enrolled in public and private elementary and secondary schools. This includes the Bureau of Indian Affairs (BIA) and the Department of Defense Education Activity (DODEA). Students with disabilities (SDs) and English language learners (ELLs) who can be accommodated sufficiently during testing are also included in the sample.

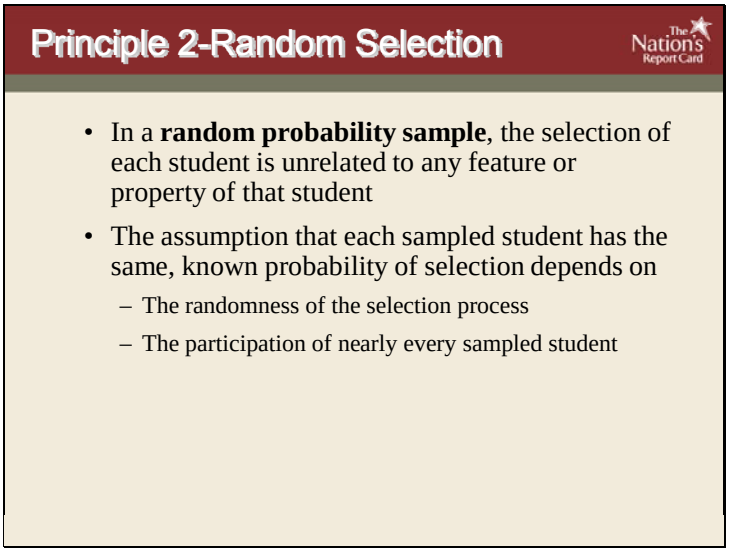
Slide 7



Students who are not part of NAEP's population, because they cannot be sampled are:

- Ungraded school children – If NAEP does not know whether students are in 4th, 8th, or 12th grade, they are not assessed. These students tend to be SDs, ELLs, or students who are otherwise involved in an alternative curriculum.
- Home-schooled children – NAEP does not know how to locate and assess these students and the resources required would entail an enormous financial outlay.
- Students living in U.S. Territories – Changes in NAEP's legislation when the No Child Left Behind (NCLB) law was enacted explicitly excluded some of the Territories from NAEP's target population. Other students not included in the population are those located in schools that serve youth in hospitals or correctional facilities.
- Students with special needs who cannot be accommodated appropriately are not assessed – NAEP knows how many students are in this group and teachers' ratings of their capabilities.

Slide 8

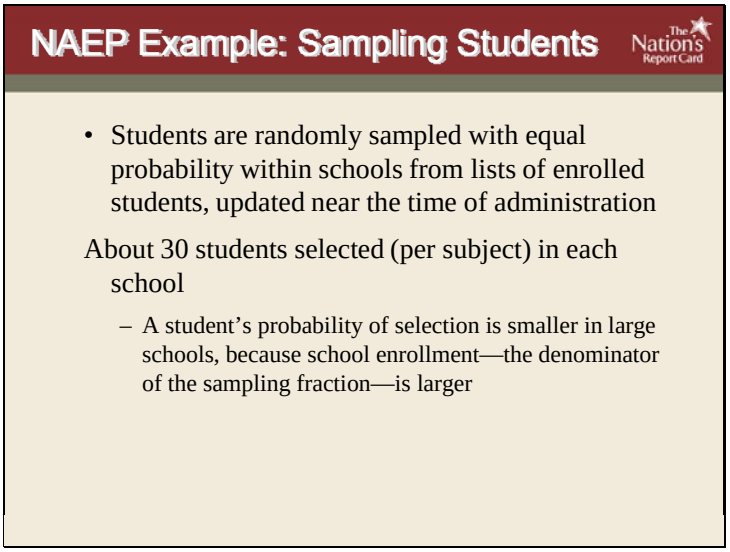


The second principle of sampling is randomly selecting a sample to ensure representativeness. In a random probability sample, the selection of each student is unrelated to anything about that student. There is an assumption that each sampled student has the same known probability of selection. This assumption depends on:

- The randomness of the process – random selection can be controlled through random number generators and good operational procedures in the field.

- The participation of all the students sampled – participation is more difficult to control. If the proportion of sampled students or schools that participates is low, the probability of selection is not very accurate, and unknown biases may occur. High participation rates are important to ensure valid inferences. That is why there are NAEP policies about response rates to ensure valid generalizations.

Slide 9



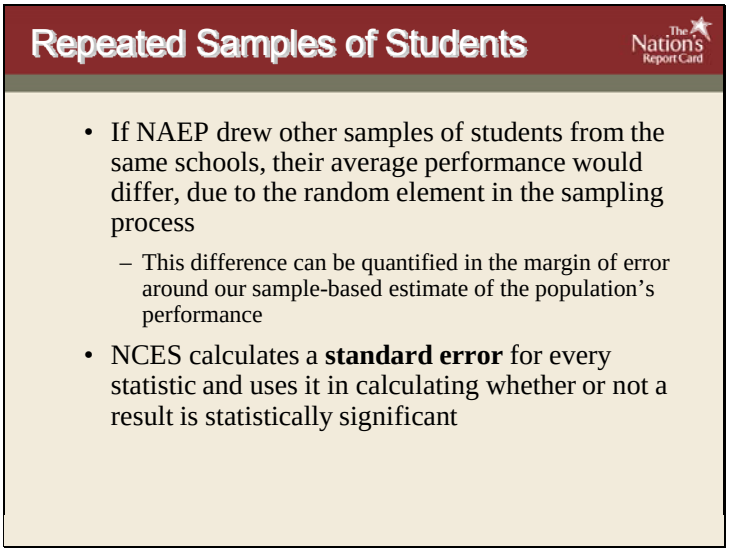
Students are sampled from lists provided by the states or the schools. These lists of students are updated not long before data collection to ensure all students have a chance of being selected. The selection process is strictly random so that all students have an equal probability of selection within any given school.

Thirty students are selected (per subject) in each school. Because this number is constant, a student's probability of selection is smaller in large schools, because school enrollment is larger.

To summarize some key points:

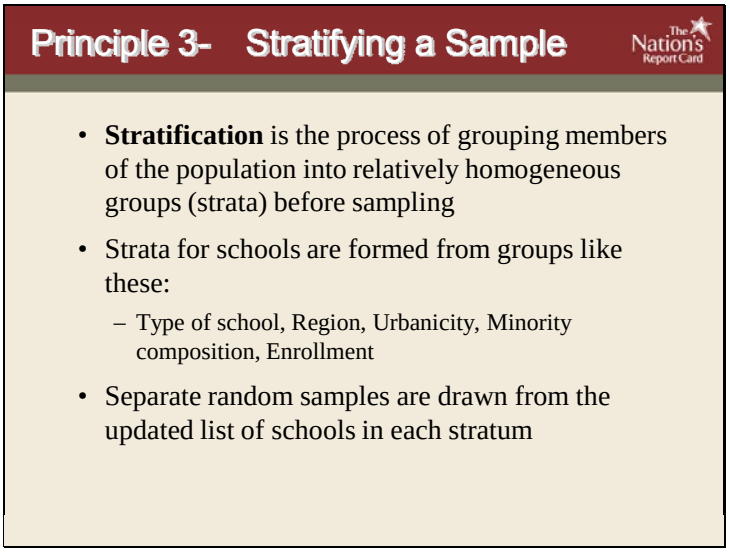
- Students and schools are selected by NAEP via a scientific sampling process; they do not choose whether to be in the NAEP sample—if they are not in the sample, they cannot participate. (NAEP is voluntary to students. So, if a student is selected to be in the sample, the student may choose whether or not to participate.)
- The selection process is strictly random so that all students have an equal probability of selection within any given school.
- NAEP selects students without identifying whether they are taking a class in the subject area of the assessment. For example, a student may not currently be in a science class, but he or she can still be selected to take the NAEP science assessment.

Slide 10



If repeated samples of students from the same schools are drawn, their average performance would differ somewhat, due to the random element in the sampling process. Statistical theory provides a quantitative measure of the deviation of the sample from the true value that is expected due to random error. The term "error" does not mean mistakes were made, but only that there is always a margin of error. As long as the error is randomly distributed, the National Center for Education Statistics (NCES) can feel comfortable with the "error." NCES reports this deviation in the form of standard errors and uses them in calculating whether or not a finding is statistically significant. If a finding is statistically significant, there is a high chance that if another sample were drawn, the same finding would result.

Slide 11

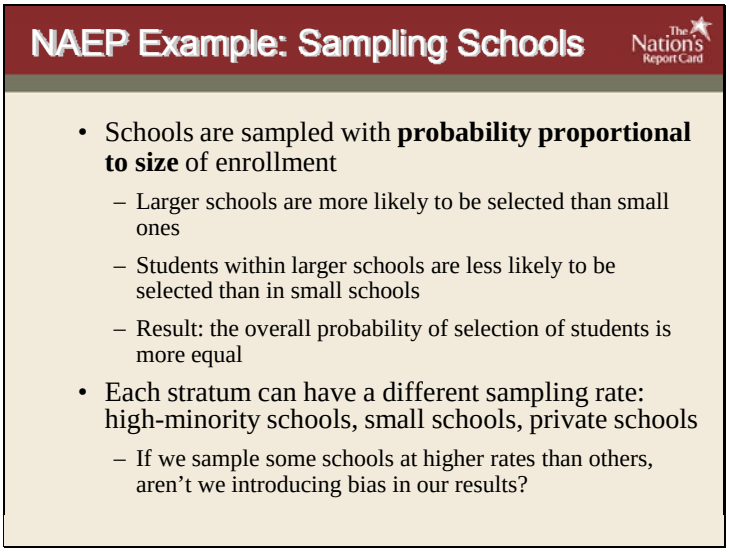


The third principle of sampling is stratification. This is a process of grouping members of the population into relatively homogeneous groups before sampling. Each group is referred to as a stratum. There are a number of different groups of schools that NAEP wants to involve in the assessment. Characteristics of these groups include:

- Type of school, such as whether it is public, private, managed by the DODEA, or the BIA;
- Regions in which the schools are located, such as whether it is urban or metropolitan; and
- Size of the school.

Random samples are drawn from the list of schools in each stratum. Lists are updated to ensure newly formed schools have a chance of selection. NAEP does not sample at the same rate within each stratum. Over–sampling is usually performed on high–minority and private schools, and under–sampling is performed on very small schools, because the cost per student is high in smaller schools.

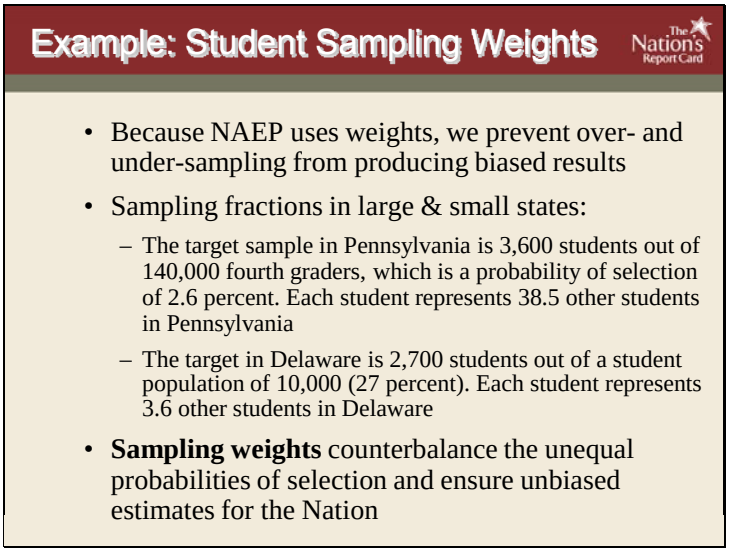
Slide 12



Schools are sampled with probability proportional to the size of school; thus, larger schools are more likely to be selected than smaller schools. However, when the probability of selection of schools and students are multiplied together, the overall probability of selection of students becomes equalized, because as mentioned before, students are less likely to be selected in larger schools.

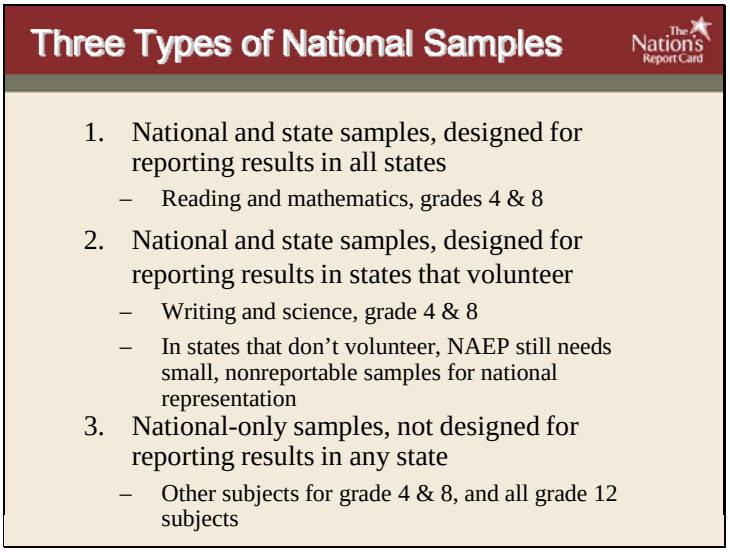
The question arises whether bias is introduced by having too many of some kinds of schools and too few of others. NAEP uses sampling weights to counterbalance the unequal probability of selection to obtain unbiased representative sample estimates for the nation as a whole.

Slide 13



Sampling weights work to prevent over– and under–sampling from producing biased results. NAEP targets about 3,000 students per state at grades 4 and 8 to produce state–level results for several NAEP assessments. Because states differ in population size, this causes a difference in the fractions of students sampled in each state. A weight is the inverse of the probability of selection. If the probability of a student's selection in Pennsylvania is .026, the student would represent 38.5 other students. If the probability of selection is .270 as in Delaware, the student's performance would represent that of 3.7 other students. The combination of Delaware and Pennsylvania, and all the other states with their different rates add up to the 4.1 million fourth graders being represented in the national assessment. If the weights were not used, the results would be biased.

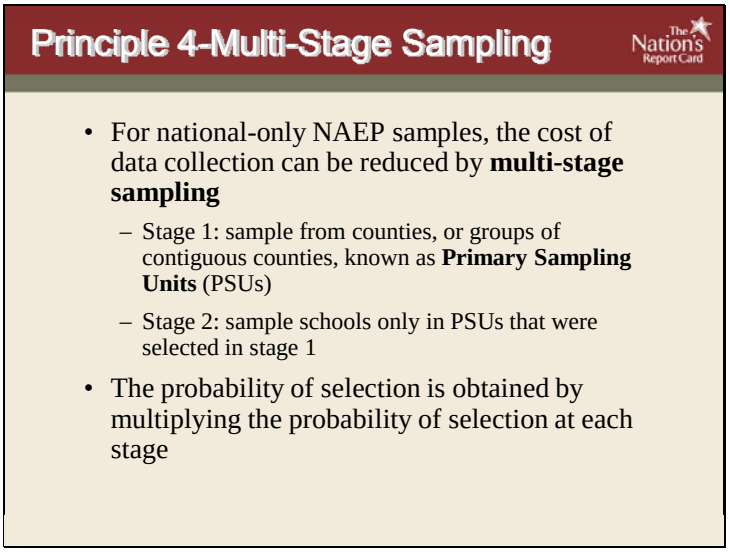
Slide 14



NAEP produces national samples in three ways. The first type is a national and state sample that is designed for reporting results for every state. These are the NCLB subjects, reading and math in grades 4 and 8. In these NCLB subjects, performance is reported in every state, and the national results are based on the state results. The second type is a national and state sample, such as in science and in writing, for which states are not required to participate. But, in states that do not volunteer, NAEP still needs small, state samples for national representation, though these small samples are not representative of the state. There is a third type of national sample for subjects such as history and civics that is not designed for reporting results in any state.

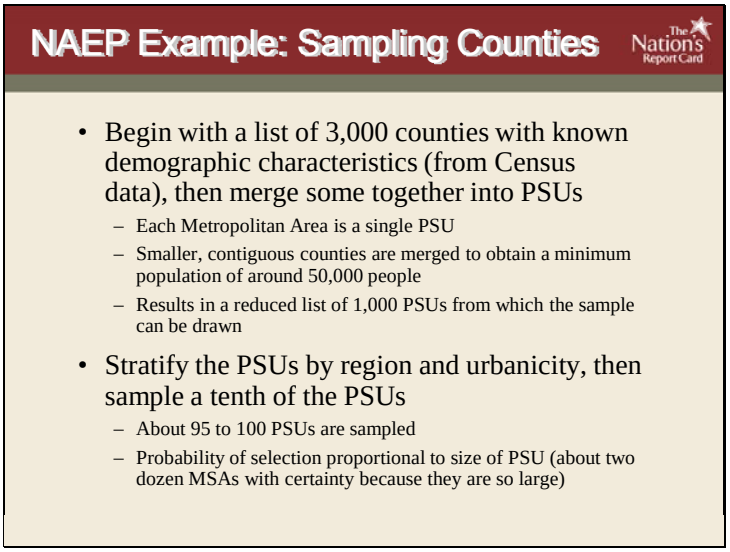
In the first two types, where both states and the nation are sampled, lists of all schools in each state are used. Any school in each of the states could be selected. In the third type, NAEP reduces the cost of data collection by first sampling counties, and then only sampling schools from those counties. This process is known as multi–stage sampling.

Slide 15



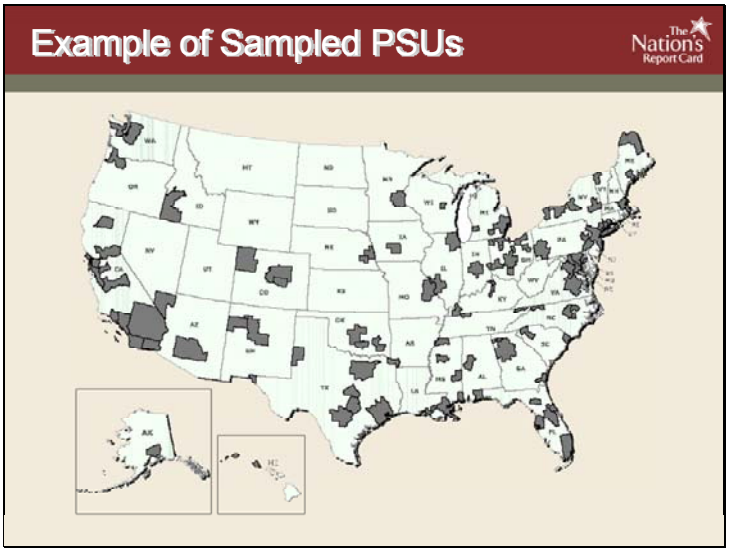
The fourth principle is multi–stage sampling. For national–only NAEP samples, the cost of data collection can be greatly reduced by multi–stage sampling. In the first stage, NAEP samples from counties, or from groups of contiguous counties known as primary sampling units (PSUs), because the unit is not a single county. In the second stage, NAEP samples only those schools in PSUs that were selected in stage 1. The probability of selection is obtained by multiplying the sampling fraction at each stage. When the PSUs are sampled, the probability of selection of the county group is obtained. Then NAEP multiplies the PSU probability, the school probability, and the student probability to obtain the overall chance of selection for a student, and the result is how the weight is obtained. The probability of selection is needed in order to get the correct sampling weight.

Slide 16



Multi–stage sampling works in the following manner: NAEP begins with a list of the 3,000 counties in the USA. The counties' demographic characteristics are obtained from U.S. Census data. The counties are merged into PSUs. Smaller, contiguous counties are merged to obtain a minimum population of around 50,000 people. Metropolitan areas are considered single PSUs. The sample frame—the list from which samples are drawn—is now reduced from 3,000 to 1,000. This list is then stratified by regions and urbanicity and one–tenth of these PSUs are sampled. The probability of selection is proportional to the size of the PSU in order to minimize the difference in the overall probability of selection of students across the Nation. This sampling stage results in about 95 to 100 PSUs. Once the selection of PSUs is completed, the sampling of schools and students proceeds in the same manner.

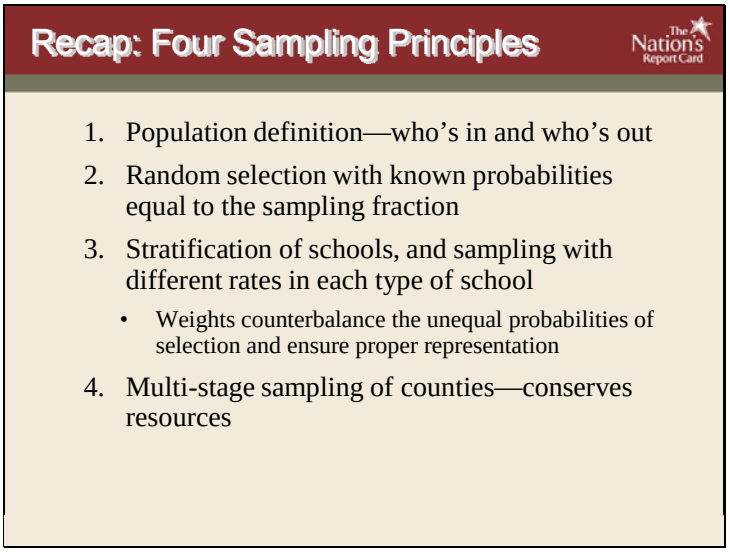
Slide 17



This map shows a sample of PSUs chosen for a NAEP, national–only sample. As seen, large metropolitan areas were selected, but few large, unpopulated areas. The crucial point here is that all counties in the nation had a chance of being selected and were part of the population that was sampled.

The overall probability of selection of a student is then the product of three factors: the probability that the PSU was sampled, the probability that the school was sampled within a PSU, and the probability that the student was sampled within a school. Sampling weights are then created to counterbalance the overall probability of selection.

Slide 18



To summarize this presentation, the first key point about scientific sampling for NAEP is identifying the population and being clear about who is and who is not in the population on which NAEP will report. In other words, from the population of several million students, the first task is to determine whose performance is being measured with the sample.

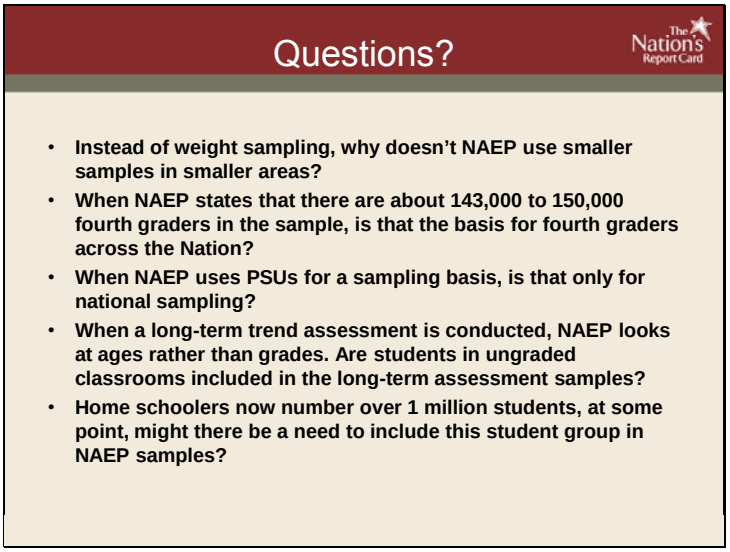
The second key point is that random sampling assures representation of the population. Random selection is necessary to determine the probability of selection and to enable computing the sampling fraction, that is, the ratio between the target sample that is selected and the population list or number of students the selection represents. If the sample is not random, it is not certain that a sample is representative of the population.

The third key point is that stratification can be used to assure that specific attributes of the population are represented. Stratification allows for over–sampling in some sectors and under–sampling in others to ensure enough of each kind of school is represented in the performance results. Any bias effect resulting from unequal probabilities of selection can be counterbalanced, or removed by using weights to ensure proper representation for overall results.

The fourth key point is that multi–stage sampling is used to increase the efficiency of the sampling process. The overall probability of selection of any student is obtained by

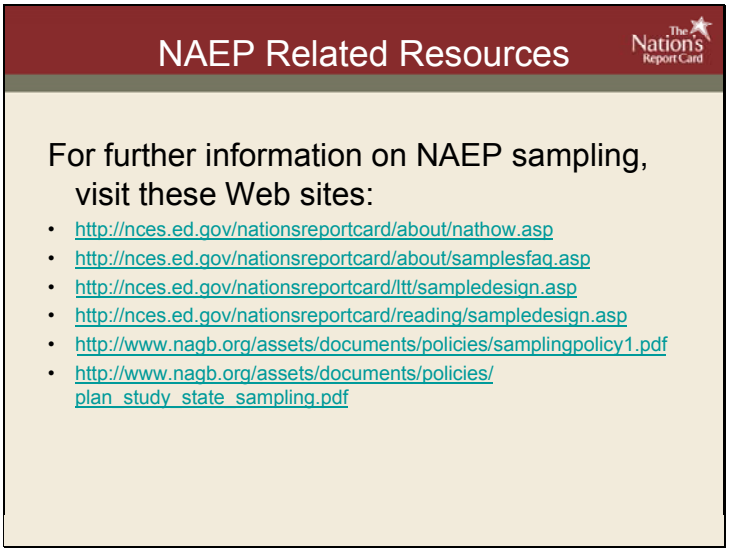
multiplying the selection probability at each stage—the PSU, the school, and the student.

Slide 19



These are questions that were elicited from the presentation. To obtain the response, please click on the question.

Slide 20



These Web sites provide more information on NAEP sampling procedures and policies.

Instead of weight sampling, why doesn't NAEP use smaller samples in smaller areas?

It is because of state-level reporting. In order to get enough cases to report at the state level larger samples are needed in some cases. For example, if NAEP chose a sampling rate of four percent in Delaware, there might be only 400 cases, and that would not be enough to report on the performance of the total student population in Delaware.

When NAEP states that there are about 143,000 to 150,000 fourth graders in the sample, is that the basis for fourth graders across the Nation?

Yes, and then NAEP filters that number down to 3,000 per state.

When NAEP uses PSUs for a sampling basis, is that only for national sampling?

Yes, for national-only sampling, no states are reported. Regions are sometimes reported using PSU sampling.

When a long-term trend assessment is conducted, NAEP looks at ages rather than grades. Are students in ungraded classrooms included in the long-term assessment samples?

If it is a regular school, not a special education school, and the whole school is ungraded, then they can be included in the sample.

Home schoolers now number over 1 million students. At some point, might there be a need to include this student group in NAEP samples?

There is an increasing portion of this population across the nation. By law, however, NAEP is not applicable to home schooled students. Even if some home school students are enrolled or accounted for by a school, NAEP would not be able to talk about home schoolers as a sector because these students would only be a portion of home schoolers, and there would be many who had no chance of selection, so it would be a biased sample.