

## Frequently Asked Questions About NAEP Sampling

*Developed in support of reporting NAEP 2005 reading, mathematics, and science results*

### Who is assessed by NAEP?

- NAEP assesses representative samples of students in certain grades or at certain ages in public and nonpublic schools in the United States.
- For the national assessments, NAEP samples students from grades 4, 8, and 12 in public and nonpublic schools.
- For long-term trend assessments, NAEP samples students at ages 9, 13, and 17.
- For the state assessments, NAEP samples students from grades 4 and 8 and assesses students in reading, mathematics, writing, and science. As part of the *No Child Left Behind* legislation, NAEP assesses fourth- and eighth-grade students every 2 years in reading and mathematics (beginning in 2003).
- NAEP does not provide scores for individual students; instead it offers results regarding subject-matter achievement, instructional experiences, and school environment for populations of students (e.g., grade 4) and subgroups of those populations (e.g., female students, Hispanic students). NAEP results are based on samples of these student populations of interest.

### How many students are selected for the state and national assessments?

- For the state assessments, NAEP samples approximately 2,500 to 3,000 students in the state in each grade for each subject area. To do this, NAEP samples about 100 to 200 schools in the state at each grade (4 and 8), and students are sampled within those schools.
- From each school, NAEP samples 30 students for each subject. In 2005, 60 students were sampled from each school in states not participating in the science assessment, and 90 students were assessed from each school in states participating in all three state subject assessments—reading, mathematics and science. At grade 4, schools could elect to assess all students (up to 120).
- In 2005, sample sizes for reading, mathematics, and science assessments were increased in four states. California received a triple sample (three times the normal sample size); Texas a double sample; and New York and Florida received a 50 percent increase. This was due to recognition that these states have diverse populations, and increasing sample sizes will allow meaningful breakdowns of the results at finer levels than have been possible in the past. Increasing sample sizes in these large states will also improve the precision of national estimates, overall as well as by demographic subgroups.
- Ten urban districts participated in the Trial Urban District Assessment (TUDA) in 2005. The sampling design for TUDA districts provides for oversampling. For the four largest TUDA districts—New York City, Los Angeles, Chicago, and Houston—the target student sample sizes are three-quarters the normal size of the state sample. For the other six districts (Atlanta, Austin, Boston, Charlotte, Cleveland, San Diego), the target student sample sizes are half the normal size of the state sample. The larger samples allow reliable reporting about subgroups in these districts.
- Students in the TUDA samples are considered part of the state and national samples. For example, the data for students tested in the Chicago sample will be used to report results for Chicago, and will also contribute to Illinois' estimates and to the national estimates, with appropriate weights. Chicago has approximately 20

5/24/2006

This is not an NCES publication

percent of the students in Illinois, therefore Chicago will contribute 20 percent and the rest of the state will contribute 80 percent to Illinois' results.

- The national samples for the 2005 reading and mathematics assessments contain the combined sample of students assessed in reading and mathematics in each state. Beginning with the 2002 assessments, NAEP has selected a combined sample of public schools for state and national NAEP rather than selecting separate state and national samples. This approach reduces the burden on states and schools by decreasing the total number of schools participating in state and national NAEP. The full data set is analyzed together, allowing all data to contribute to the final results, and setting a single scale for the assessment and improving the reliability of the national estimates.
- The national sample for 2005 science consists of the combined sample of students assessed in each state that participated in the state science assessment, plus a sample of students from the states that did not participate in the state assessment. The following states did not participate in the science assessment: Alaska, Iowa, Kansas, Nebraska, New York (except New York City), Pennsylvania, and the District of Columbia. Including the state-specific samples of students in the national samples improves the reliability of the national estimate.
- The design of the 2005 national sample for science is similar to the design for 2002 (when only some states participated in the state-by-state assessment), whereas the 2005 national sample design for reading and mathematics is similar to the 2003 design (when all states participated in the state-by-state assessments).
- A separate, national sample of nonpublic schools is also selected for grades 4, 8, and 12. This sample is designed to produce national estimates and estimates for the major types of nonpublic schools (Catholic, Conservative Christian, and Lutheran).

#### **How does NAEP select the schools and the students for the assessments?**

- As the Nation's Report Card, NAEP must report accurate results for populations of students and subgroups of these students (e.g., minority students, students from low-income families). To ensure accurate results, the relatively small samples of students must be truly representative of the entire student population in the nation (for the national assessments) or the state (for the state assessments) or the district (for the TUDA assessments).
- NAEP uses a multistage sampling design that relies on stratification (i.e., classification into groups having similar characteristics) to choose samples of schools and student populations. To ensure an accurate representation of public schools, the samples are randomly selected from groups of schools that have been stratified by variables such as the extent of urbanization, percentage of minority enrollment, and school level results on state achievement tests. For national assessments not involving state-by-state samples, region of the country and median household income of the area where the school is located are also used. A similar approach is used for nonpublic schools, but using somewhat different characteristics.
- NAEP selects a representative sample of students by first randomly selecting schools and then selecting the students within those schools who will participate in a given NAEP assessment. Every school has some known chance of being selected for the sample. Within a selected school, all students within a participating grade have an equal chance of being selected. The probability of students and schools being selected into the sample varies based on factors such as grade, subject, public and nonpublic school status, and so on. Those probabilities are important in producing NAEP results, and NAEP takes them into account in the calculation of results through the process of applying sampling weights.

**Why are some schools always selected?**

- NAEP usually selects 100 public schools for each subject at each grade for the state's sample—each school would then represent about 1 percent of the students in public schools in the grade being assessed in that state. If a school is chosen repeatedly, typically it is because they have more than about 1 percent of the state's enrollment in the grade. Other schools, with about 0.5–1 percent of the enrollment, are selected frequently though not *always*—however, it probably seems like *always* to those schools.

**Why does NAEP use sampling? What are the benefits of sampling for NAEP?**

- Sampling minimizes the assessment time required per student while allowing complete coverage of the subject being assessed. In order to have a valid and reliable assessment of the NAEP content, several hundred assessment questions are needed. Testing a student on the entire collection of assessment questions that make up each NAEP assessment is too time consuming and impractical. Hence, no single student takes the entire assessment.
- Sampling produces accurate estimates of student achievement while reducing the amount of time and cost to administer and score the assessment. Administering NAEP to all students in a state or the nation would be very expensive—there are many constructed-response questions, which are very expensive to score.
- NAEP does not report data for individual students, schools, or districts—except for a few districts participating in the TUDA assessment—and, therefore, it is not necessary to assess and report results for every student in every school.

**What is matrix sampling and what are its advantages for NAEP?**

- In matrix sampling, different portions from the entire pool of assessment questions are printed in separate booklets and administered to different but equivalent samples of students. Matrix sampling allows NAEP to assess the entire subject area within a reasonable amount of testing time.