

Sampling and the Texas Assessment Program

When did Texas start holding schools and districts accountable for student performance?

In 1993, the 73rd Legislative Session passed Senate Bill 7, overhauling the Texas school finance system around the principles of equalized per student expenditures and mandating the creation of the nation's first comprehensive accountability system.

Why does Texas require testing?

As part of Senate Bill 7, goals for the public education system were established that would be monitored by the new accountability system. These goals included:

- 1) *Closing the achievement gap*
- 2) *Raising student performance in Texas compared to national and international communities*
- 3) *Delivering a well-balanced and standardized curriculum to all students*
- 4) *Attracting and retaining qualified and effective personnel*
- 5) *Improving instruction and administration across all schools*

To monitor these goals, the Texas Assessment of Academic Skills (TAAS) program was expanded in the 1993–1994 school year to assess all students in grades 3–8 and 10 in reading and mathematics, and all students in grades 4, 8, and 10 in writing. Assessments in science and social studies were added in subsequent years according to statute. In 2001, the Texas model of accountability and assessment was adopted nationwide following the passage of the No Child Left Behind Act of 2001 (PL 107-110).

What is sampling?

Sampling is the process used to select a representative subset of a population. The characteristics of a sample should match the characteristics of the population so that the sample can be used to make inferences about the population. The sample size needed depends on whether the sample information needs to be aggregated into smaller subgroups and the desired level of confidence for making inferences.

Why is sampling not currently used for the STAAR program?

The United States Department of Education (USDE) requires annual testing of all students in reading and mathematics in grades 3–8 and once in grades 10–12. All students are required to test in science at least once in grades 3–5, 6–9, and 10–12. Sampling would not meet these requirements.

What would be lost by using sampling instead of administering STAAR assessments to all students?

Individual student results would not be available if sampling were used. Without individual student scores students may not be as motivated to show what they know on the assessment, year-to-year progress information for each student would not be available, and teachers would not be able to use test results to tailor instruction toward individual student needs. Additionally, the ability to disaggregate student subgroup information at the campus level would be lost unless repeated sampling of the same students was conducted. This means that for minority subgroups to be represented in the sample, the same students would have to be tested year after year.

What would be gained by using sampling instead of administering STAAR assessments to all students?

Sampling would reduce the testing burden for some students. However, districts would still need to participate in testing and would have to negotiate the logistics surrounding testing some students but not others. Sampling could reduce costs related to test materials production, test administration, and score reporting. However, the cost savings may not be directly proportional to the reduction in the number of students tested because expenses for test development would be the same regardless of the number of students tested.

How would sampling affect the campus-based accountability system used in Texas?

If campus-level inferences are needed for accountability, all campuses must be part of the sample. Sampling would benefit larger campuses more than smaller campuses. The larger the campus is, the smaller the proportion of tested students needs to be to have confidence in the sample results. Smaller campuses must test a larger proportion of students to have the same level of confidence in the sample results. Additionally, if results need to be disaggregated (for example, by race/ethnicity or socio-economic status) these student subgroups need to be included in the sample. Sampling to reduce student burden is therefore less beneficial for smaller campuses and minority groups.

How would sampling affect Texas' ability to monitor progress towards the Senate Bill 7 goals?

Expanded testing was put in place so that the Texas accountability system could monitor the goals laid out in Senate Bill 7. The use of sampling would limit the ability to monitor some goals:

1) *Closing the achievement gap*

The achievement gap could not be evaluated unless sampling was done at the subgroup level so that results could be disaggregated. Using a sample to make subgroup comparisons requires students in these subgroups to test and minimizes the benefit of sampling to reduce testing (see attached).

2) *Raising student performance in Texas compared to national and international communities*

This goal could be evaluated by using a sample.

3) *Delivering a well-balanced and standardized curriculum to all students*

This goal could be evaluated by using a sample.

4) *Attracting and retaining qualified and effective personnel*

Sampling at the classroom level is needed if test results are to be linked to teachers. This requires some students from each classroom to test and minimizes the benefit of sampling to reduce testing.

5) *Improving instruction and administration across all schools*

To evaluate instruction and administration across schools requires sampling at the campus level. This level of sampling could provide some reduction in testing but still requires all campuses to participate (see attached).

In order to monitor these goals the type of sampling that must be done is restricted and does not allow Texas to experience the full benefits of sampling. The Senate Bill 7 goals do not align well with a sample-based approach to assessment.

Where does the Texas assessment program use sampling?

Although sampling is not used for STAAR administrations, sampling is used within the Texas assessment program for other purposes. Embedded field-testing is done through sampling so that information can be gathered on the quality of items without all students having to take all field-test items. Prompt studies use sampling as a way to obtain information about student responses to a variety of writing prompts without all students in Texas needing to participate. Audits for TELPAS and STAAR Alternate involve sampling so that documentation can be collected from a representative portion of the state rather than all of the state.