

Introduction to National Assessment of Educational Progress

Emmanuel Sikali, Ph.D.
National Center of Education Statistics

Transcript

Transcription is provided in order to facilitate communication accessibility and may not be a totally verbatim record of the proceedings.

Introduction to National Assessment of Educational Progress

(Slide 1)

Greetings, I am Emmanuel Sikali with the Assessment Division of the National Center of Education Statistics (NCES). This webinar is provided as a joint task between NCES, NCSER, and AIR to provide an introduction to NAEP.

(Slide 2)

During this webinar, I will give a general introduction to the National Assessment of Educational Progress (NAEP).

Then the assessment design, the sampling design, and how assessment results are reported.

(Slide 3)

Let's start with the introduction to NAEP.

(Slide 4)

The National Assessment of Educational Progress is made up of two major assessments, the Main NAEP and the Long-Term Trends.

The Main NAEP is grade-based while the Long Term Trends is age-based. Let's start with the main NAEP.

The Main NAEP assesses students in public and non-public schools at grades 4, 8, and 12. The public school students at grade four and eight for states and trial urban districts. The trial urban district assessment is conducted for students in selected school districts. The National Assessment Governing Board selects participating districts.

The Long-Term Trend assessment assesses students in mathematics and reading for public and non-public school students at ages 9, 13, and 17.

This webinar will focus only on the Main NAEP assessment.

(Slide 5)

Here is the list of all the subjects that NAEP used to assess either at the national, state or urban district level.

You can see all 12, together with their respective icons. Until 2017, all the subjects on this slide appeared on the schedule of assessments. Some of these subjects were administered only at the 12th grade level like economics, while others such as Technology and Engineering Literacy (TEL) at grade 8.

Introduction to National Assessment of Educational Progress

The TEL assessment assesses skills that might have been learned out of school. Moving forward, all the grayed out subjects will no longer be assessed due to budgetary constraints.

(Slide 6)

NAEP reports results at a national level for public and non-public students for all subjects that it assesses.

(Slide 7)

Mathematics and reading are the only two subjects that have results for all 50 states, the District of Columbia, the Department of Defense schools and Puerto Rico. We should note here that in Puerto Rico, students are assessed only in mathematics at grade four and eight but not in reading.

(Slide 8)

Here is a map showing the 27th Trial Urban District Assessment (TUDA) districts. Again, these are large urban districts that have volunteered to administer NAEP to an oversample in their districts so that they can obtain their own results.

(Slide 9)

Now, let us focus on Assessment Design.

(Slide 10)

During a NAEP administration, students receive cognitive and non-cognitive items. Cognitive items are developed using specifications from the specific subject framework.

These frameworks, developed by the National Assessment Governing Board (NAGB), are the test blueprints. Each framework specifies the content and sub-content area to be assessed, the complexity, the item type to be used in this assessment. Non-cognitive items are used to collect some individual characteristics of students, their teachers, and the school that they attend.

They are developed using a survey information framework. Before these items are used in an assessment, they undergo intensive review by experts before conducting cognitive labs, pilot, or field testing.

(Slide 11)

A NAEP framework specifies many aspects of the content targeted for assessment.

The framework identifies the number of sub-components, sometimes referred to as subscales and the weights of each subscale to the overall scale for an assessed content.

Introduction to National Assessment of Educational Progress

For example, the framework identifies five sub-components for mathematics and two subcomponents for reading. Similarly, there are sub-components for other subjects. The number of sub-components can vary between subject and between grades within a subject.

The number of sub-components can change between years. In addition to the number of sub-components, the framework specifies distributions of items by subscales, item complexity, and item format.

(Slide 12)

Instrument development encompasses three principles.

The goal for developing the content of a NAEP instrument is to have the full coverage of what the Governing Board has specified in the assessment framework.

Approach is to develop a large pool of items within the available resources and present a smaller portion of those items to an individual student to minimize student burden.

Procedure involves multiple committees in the development of the framework, content of the instrument, and administration of the instrument.

(Slide 13)

The cognitive part of the NAEP assessment includes discrete items, scenario-based tasks, and hybrid hands-on tasks. There are two response types.

Selected response types include, multiple-choice, multiple-select multiple-choice, drag and drop, zone, drop down. Constructed response type includes short and extended constructed responses.

For some assessments, ancillary materials are provided to students. For example, a periodic table and mathematics formula are provided at grade 12, depending on the subject, while a writing handout is provided at grades 4, 8, or 12.

Starting from 2017, NAEP transitioned from paper and pencil assessment to digital assessment in mathematics and reading at grades 4 and 8. For these digital assessments, the testing environment or in it include digital tools such as rulers, bar graphs, interactive item components, equation editors, calculators, just to name a few.

(Slide 14)

After completing the two 30-minute content sections, students are prompted to answer questions about their educational experiences inside and outside the classroom.

Similar surveys are also given to teachers and school administrators. As part of most NAEP assessments, three types of survey questionnaires are used to collect information. Student

Introduction to National Assessment of Educational Progress

questionnaires collect information on students' demographic characteristics, opportunity to learn in and outside of the classroom and educational experiences.

They are completed by students during the 15 minutes allowed for questionnaires following the cognitive component of the assessment. Students are encouraged to answer as many questions as they feel comfortable with and they can skip any question they do not want to answer.

All responses are kept confidential. Students' names are never reported with their responses or with the other information collected by NAEP. Teacher questionnaires gather information on teacher training and instructional practices.

They are completed by teachers mainly at grade 4 and 8, rarely at grade 12. School questionnaires gather information on school policies and characteristics and are completed by the principal or assistant principal. To ensure that necessary accommodations are provided, schools also provide information about students with disabilities (SD) or English language learners, also known as ELL.

Here are some example questions from the school questionnaire.

Does your school or district offer tenure to teachers?

In the last school year, how many full-time teachers were new to your school?

In this school year, which of the following types of computers or digital devices are available in your school for student use? Select all that apply.

In your school, is there a wireless internet connection that students can use for schoolwork?

Is there a literacy coach available full-time or part-time to every teacher at your school?

(Slide 15)

Because NAEP is a large group assessment, each student takes only a small part of the overall assessment. Each sampled student takes two cognitive item blocks that contain content area questions. For digitally-based assessments (DBA), two 30-minute blocks. For paper and pencil assessment, two 25-minute blocks.

Due to the NAEP's design, data at the school or individual to student level is not reliable for reporting. This is because students take too few items for the data to be reliable for individuals. We should note that only when the student scores are aggregated at levels such as the urban district, state or national that the data are considered reliable and valid estimate of what students know and can do in the content area. This is why school or student level results are never reported.

(Slide 16)

NAEP is mandated by public law 107-279. Section 303 of this law specifies, among other things, the type of sampling method and lists subgroups of the population that NAEP must report results.

To accomplish that, the design, operations, and reporting of NAEP must function efficiently and with rigor. There are many techniques and methods applied to the operations of NAEP and some innovations in the field of large-scaled assessment came out of the needs and opportunities NAEP created.

(Slide 17)

One method used to construct each NAEP assessment is matrix sampling. Matrix sampling is a sampling plan in which different samples of respondents take different samples of items. The number of cognitive items developed vary by subject and grade.

For example, the number of items in any one assessment for reading is between 100 and 170. And for mathematics is between 180 and 250. Each student does not take each item. In fact, every item is exposed to about one fourth of the student sample.

The figure shows how cognitive items are assembled into blocks and then blocks into booklets.

As you can see, in step one, items are selected. In step two, items are paired to develop a block of items, that is, item one paired with item two in block one, item three paired with item four in block two and item five paired with item six in block three.

Such pairing can be more complex as the number of items increases. Blocks are created such that they are similar in average item difficulty, content distribution, and complexity level.

At step three, blocks are paired to develop booklets. Block one is paired with block two, in booklet A, and again, block one is paired with block three in booklet B, while block two is paired with block three in booklet C.

Systematic pairing is necessary to ensure proper distribution of items across the population.

(Slide 18)

Further, NAEP test booklets are created with interlocking blocks of items. Each block appears once in every position of a test form administered to students. Booklet design can vary by subjects and grade.

(Slide 19)

In distributing booklets during an assessment administration, booklets are organized according to sequences designed to ensure equal distribution of them across all sessions.

Booklets are alternated in a sequence within each session when multiple subjects are assessed in a year. We also make sure that one single booklet should not be distributed to too many students or should not be placed at the top of all sets of booklets.

Introduction to National Assessment of Educational Progress

Rotating booklets through different positions within a bundle, the chance of not using any one type of booklet is equalized.

(Slide 20)

Each sampled student takes two cognitive item blocks in a booklet. For DBA, the time limit is 30 minutes for each block.

(Slide 21)

Starting from 2017, the mathematics and reading assessments at grade four and eight were administered digitally. Under this assessment delivery model, each sampled school is assessed in one day. The field staff drops into the school and sets up a control environment that relies entirely on NAEP-owned device.

As you can see on the slide in the control environment, there are student tablets and a monitor tablet. During the testing, the field staff monitors students' progression. At the end of this session, each student's data are saved and transmitted to the administrator's tablet.

And the administrator's tablet transmits this data to the NAEP central server.

(Slide 22)

The data captured from each student are sequences of individual actions taken by the examinee.

This action could be mouse clicks, the change of the default background, the use of calculators, the selection of a response, the use of text-to-speech, just to name a few. For more information about this data, refer to the presentation on the introduction to NAEP Process Data.

(Slide 23)

The next section will give a brief introduction to NAEP sampling.

(Slide 24)

NAEP uses the common core of data (CCD) to sample public schools. The schools on this database are classified by school location and the racial-ethnic composition of the school, as well as the school's performance on the state test.

In each stratum, schools are ordered, then sampled with probability proportional to the school size. The final list is sent to the State Department of Education. They are tasked to confirm that the information is accurate and up-to-date.

(Slide 25)

NAEP sampling ensures approximately equal sample sizes for each state public school. In each state, approximately 100 to 150 schools were selected for each grade and subject.

In 2017, there were approximately 2700 students assessed for each grade and subject and each student in a school has an equal chance of being selected for NAEP.

Individual student demographics do not play a role in which students are selected to participate within a school. In each school, about an equal number of students are assessed in reading and mathematics. The preset number of students is about 30 students per subject per school.

Within each school and grade to be assessed, students are chosen at random to participate in NAEP.

(Slide 26)

The mixed sampling used by NAEP to create the shortest form used to assess students effectively creates missing data by design and NAEP samples are stratified and clustered, which means that special training and software is needed to analyze the NAEP data.

(Slide 27)

Let us transition to NAEP Reporting.

(Slide 28)

NAEP reports scale scores, percentile scores, and achievement levels. Scale scores range from 0 to 500 for mathematics and reading at grade 4 and 8, or 0 to 300 for other subjects.

In NAEP, the percentile score is defined as the percentage of students scoring lower than a particular scale score. For example, NAEP reports on scores added to either 10th, 25th, 50th, 75th, and 90th percentile.

For example, if the 25th percentile score is 281, then 25% of assessed students score below 281.

There are three achievement levels for NAEP. They are NAEP Basic, which denotes partial mastery of prerequisite knowledge and skills that are fundamental for proficiency work at each grade.

NAEP Proficient is solely academic performance for each grade assessed. Student reaching this level has demonstrated competency over challenging subject matters, including subject matter knowledge, application of such knowledge to real-world situations, and analytical skills appropriate to the subject matter.

NAEP Advance is superior performance.

Introduction to National Assessment of Educational Progress

(Slide 29)

This example is taken from the 2019 mathematics grade A report card. It shows the trend line from 1990 to 2019. For example, we can see that in 2019, students' skill score was 282 and in 1990 was 263.

We can also see that compared to 2017, students' scores were lower. This is denoted by the star on 283.

(Slide 30)

In this example, taken from this same NAEP report card, we can see that in comparison to 2017, the 2019 mathematics scores were statistically not different for eighth-graders performing at the 75th and 90th percentiles, but they were statistically lower for students at the 10th and 25th percentiles.

(Slide 31)

If we focus only on the top bar, we can see that in 2019 there were 31% of students scoring below NAEP Basic and 10% scoring at NAEP Proficient.

NAEP will also report that in 2019, 69% of eighth-graders scored at or above NAEP Basic. Compared to 2017, the percentage of students scoring below Basic was statistically larger in 2019. The asterisks indicate statistical significance.

(Slide 32)

We learn that NAEP has two main components. The main NAEP that is a grade-based assessment and a long term trend an age-based assessment.

NAEP uses complex sample design to select the students. Each student is assessed using only short-form made of sample of items and survey questionnaires. Starting from 2017, NAEP mathematics and reading assessment has been administered digitally.

During these digital assessments, NAEP collects process data. Analysts must account for the NAEP psychometrics and complex sample design while analyzing the NAEP data.

(Slide 33)

Please submit your questions until July 30th, at the email address provided on the slide. We will make sure to frequently post responses to all the questions that you might have had while going through this presentation.

(Slide 34)

Thank you.