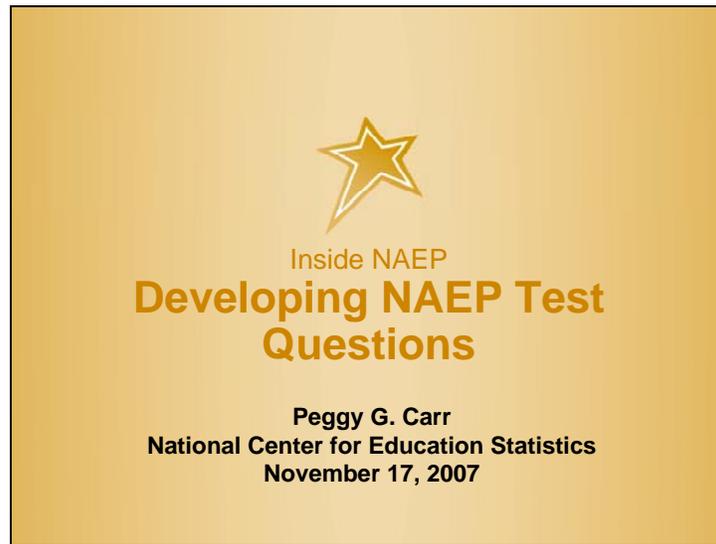


Slide 1



The major goal of this presentation is to explain the NAEP item development process and the schedule of activities involved.

Slide 2



Inside NAEP
Developing NAEP Test Questions

- Explain the item development process and schedule
 - Complexity
 - Inclusiveness of the process
 - Attention to quality

2

This presentation will hopefully provide you with a sense of the complexity and inclusiveness of the item development process and the attention to quality. The quality of this process is what makes NAEP the gold standard of assessments.

Slide 3



Inside NAEP
Developing NAEP Test Questions

New vs. Existing Frameworks

- For new frameworks
 - Create all new items
 - More than 4 years from development to implementation
- For existing frameworks
 - Ongoing development for replacement blocks (3 per assessment per grade)
 - Nearly 3 years from development to implementation
- For both types of frameworks
 - More than 300 people involved in each assessment
 - Twice the number of items needed are created for paper and pencil items

3

What will be described to you is the “replacement items process” for existing assessment frameworks. There also is a protocol for developing items associated with new frameworks. For both new and existing frameworks there are numerous people involved with development and review work.

Slide 4

Inside NAEP
Developing NAEP Test Questions

How It Works

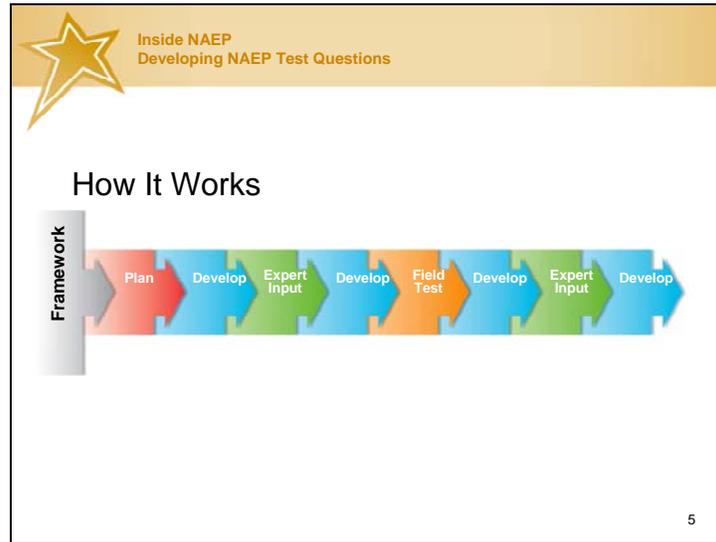
Framework →

4

The slide features a gold star icon in the top left corner. The title 'How It Works' is centered in the upper half. Below it, a grey arrow points to the right, with the word 'Framework' written vertically inside the arrow's shaft. The slide number '4' is located in the bottom right corner.

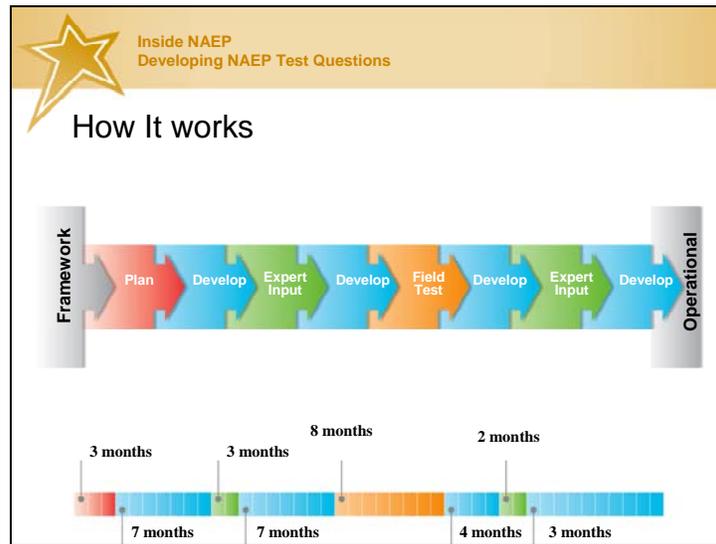
The first step in the process of item development is for the National Center for Education Statistics (NCES) to interpret the assessment framework and its associated specifications document after they have been approved by the National Assessment Governing Board. The framework is the conceptual definition of the construct which needs to be operationalized in the instrument—the test—which the child actually sees. The specifications are the detailed blueprint for constructing the assessment.

Slide 5



There are four major stages involved in item development: planning, development, receiving expert input and field testing. As can be seen in the graph, this is an iterative process, where planning leads to a first round of development, followed by multiple rounds of expert input on the items. In the middle of the process, there is an opportunity to field test assessment items.

Slide 6



The iterative nature of this process means that the time involved may be as long as three years.

If NAEP were developing items associated with a new framework, the process could be even longer. There would be two major data collections, pilot testing in addition to field testing. These are important steps in the process as they are opportunities to collect large amounts of information from students regarding how they are performing and how they engage and interact with the items.

This is a time intensive process that keeps NAEP very busy. For example, in 2008: NAEP administered the arts assessment; field tested items for the 2009 reading assessment; reviewed and revised items for the reading 2011 assessment; and developed the planning document for the writing assessment.

Slide 7



Inside NAEP
Developing NAEP Test Questions

Contractors

- Educational Testing Service (ETS)
- NAEP Education Statistics Services Institute (NESSI)
- Pearson
- Westat
- Fulcrum
- Logistics

7

There are multiple contractors involved in the item development process:

Educational Testing Service (ETS) is the primary contractor responsible for the development of the items. They have between 12 and 14 full time personnel and access to other specialists on an as-needed basis.

The NAEP Education Statistics Services Institute (NESSI) is a group of individuals who provide assessment services to NCES. Within this entity, there is a team of Ph.D. level item development and Master's specialists who work within the NAEP Division.

Pearson performs the scoring and the training of the scorers for the NAEP assessments.

Westat conducts the data collection in the field.

Fulcrum performs technology tasks related to the NAEP assessments.

Finally, a logistics contractor arranges meetings and travel for the various activities related to assessment development.

Slide 8



Inside NAEP
Developing NAEP Test Questions

Outside Experts

- Planning Committee (12)
 - Representatives from framework and standing committees
- Standing Committee (12-18)
 - Members from framework committee
 - Curriculum specialists, university education faculty, etc.
- State and District Review Panel (100-150)
 - Representatives from each state or district
 - Assessment directors, content specialists
- Academic Review Panel (5-9)
 - Experts in the subject area (mathematicians, scientists)

8

NAEP vets and reviews the NAEP assessment items at various stages of development with four groups of outside experts.

The Planning Committee consists of approximately twelve individuals. We include some of the framework development committee members in the Planning and Standing Committees as required by Board policy. This is an important overlap between the Board's framework development process and the item writing process.

There are Standing Committees for every subject assessed. These individuals are at the top of their fields, whether in education or industry.

The State and District Review Panels are required by law for every major development process. Whether NAEP is developing items for new frameworks or replacement items for existing frameworks, input and review are required legislatively from the states and the local jurisdictions. These state and district panel members are composed of assessment directors, content specialists, and curriculum specialists.

The Academic Review Panel varies in number and is fairly new to the NAEP vetting process. These highly specialized individuals review the items to ensure quality and appropriateness of the items for the assessments.



Inside NAEP
Developing NAEP Test Questions

Ensuring Quality

- The number of reviews
- The qualifications of our reviewers
- Examination of empirical data collected during both pilot and operational data collections
- Quality control measures

9

Quality control is the overriding component of NAEP assessment development. The foundation of our quality is in the number of reviews and the stellar experience of the reviewers.

Empirically, there is also a great emphasis placed on feedback. Items are revised based upon the empirical information obtained from both the field tests and the operational assessment.

The evaluation of NAEP is mandated by law to occur on an annual basis. In 2007, the Buros Institute for Assessment Consultation and Outreach was our external evaluator for NAEP. And they, like their predecessors, evaluated NAEP items and assessed whether the items fit the framework. Recently, we requested an external evaluation of the math framework and the items by the NAEP Validity Studies Panel.

Inside NAEP
Developing NAEP Test Questions

Plan

Ensure that we are following both the letter and intent of the framework

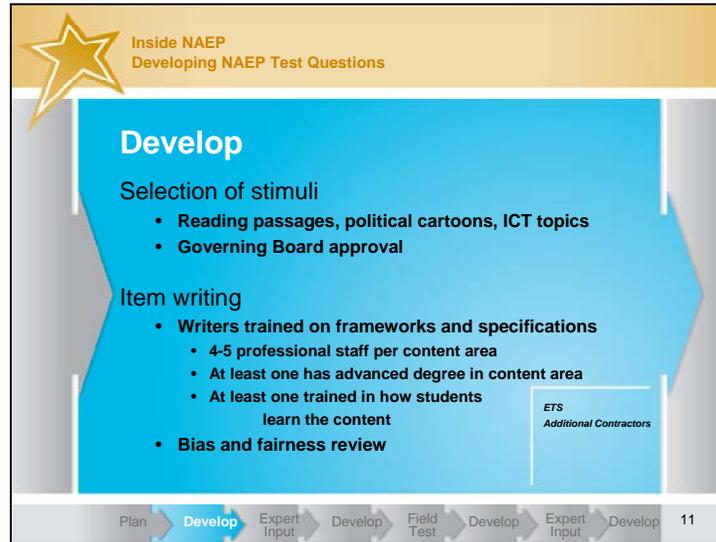
- An implementation plan for the entire process
- Guidelines/blueprint for the item development contractor
- Inventory and evaluation of items in current pools

NESI
ETS

Plan Develop Expert Input Develop Field Test Develop Expert Input Develop 10

In the planning stage—the first step in the item development process—an implementation plan is developed based on the framework. The implementation plan provides guidelines for item development and specifies any necessary additional work, research, or special studies. There are common block structures that we try to achieve with our assessments to allow us to spiral different subjects in the same assessment session. In the planning stage, we must determine how many items and blocks of items are needed to operationalize the framework objectives, how long the assessment will take, and how various combinations of these blocks will be spiraled.

Inventory and evaluation of the current item pool determine the development of new items.



In the early development process, before we start writing items associated with any given stimulus, such as reading passages, political cartoons, or integrated computer task (ICT) topics, Board approval is required.

The item developers primarily at ETS consist of professionals that are concentrated in their content area (e.g., reading and writing curricula).

ETS, like any other large entity in the assessment industry, is required to follow the joint standards from the American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME) in the development of items, which include attention to bias and fairness as well as reliability and validity.

Inside NAEP
Developing NAEP Test Questions

Develop

Small Scale Study

- About 50 responses per item

For special items (ICTs, HOTs, etc.)

- Cutting-edge
- Additional time and research
- Platform and administrative processes and procedures
- Special working groups of experts

ETS

Plan → **Develop** → Expert Input → Develop → Field Test → Develop → Expert Input → Develop → 12

Small scale studies are another important component of the early item development process. Small scale studies can be classroom tryouts of items or think-alouds where students are asked what they were thinking when they responded to a particular test item. Focus groups are also used for this purpose. Responses from 50 students are typical for any given item at this phase.

Special studies are conducted for cutting edge item types. For these studies, special working groups are organized and special administrative procedures and processes are often established.

Slide 13

Inside NAEP
Developing NAEP Test Questions

Expert Input

Ensure diversity of opinions and input

- **Standing Committee**
- **State and District Reviewers**

NESI
ETS
Logistics

Plan → Develop → **Expert Input** → Develop → Field Test → Develop → Expert Input → Develop 13

As mentioned earlier, expert input is an important part of the process. There are two major expert groups: the Standing Committee and the State and District Reviewers. Both groups are essential to our review process.

Slide 14

Inside NAEP
Developing NAEP Test Questions

Expert Input

Standing Committee Goal

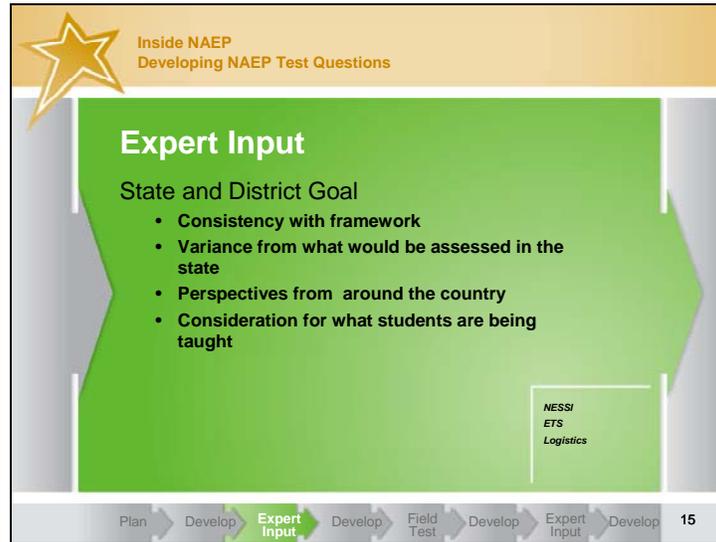
- Consistency with framework
- Grade or age appropriate
- Alignment with the scoring rubric
- Pool of items match with blueprint
- Appropriate and adequate coverage of:
 - Content
 - Difficulty
 - Dimensions
 - Item type

NESI
ETS
Logistics

Plan → Develop → **Expert Input** → Develop → Field Test → Develop → Expert Input → Develop

14

The Standing Committee has the goal of ensuring that the letter and intent of the framework is met. The specifications describe item and assessment development parameters in detail. The committee pays attention to the grade and age appropriateness of the items and ensures that they are aligned with the scoring rubrics. These content specialists review item difficulty, distracters, and the proportional representation of different item types.



Inside NAEP
Developing NAEP Test Questions

Expert Input

State and District Goal

- Consistency with framework
- Variance from what would be assessed in the state
- Perspectives from around the country
- Consideration for what students are being taught

NESI
ETS
Logistics

Plan → Develop → **Expert Input** → Develop → Field Test → Develop → Expert Input → Develop 15

The expert input from the states and the districts is valuable because the learning that takes place is two-way. Many of the states are ahead of NAEP in the assessment process for some of the content areas – on-line writing is an example. States and districts are also important in ensuring that the NAEP assessments are not too far ahead of the students.

Slide 16

Inside NAEP
Developing NAEP Test Questions

Develop

Further Item Development

- Resolution of comments from reviews
- Items revised
- Quality control
- Governing Board comments

ETS

Plan → Develop → Expert Input → **Develop** → Field Test → Develop → Expert Input → Develop → 16

After we have received expert input, there is another development phase to review the comments, revise test items, address quality control issues and prepare the items for Governing Board review and approval.

By the time the items reach this stage, there is a high degree of confidence in them due to the thoroughness of the process.

Inside NAEP
Developing NAEP Test Questions

Pilot Test

- Includes production time, administration, and data analysis
- At least 1500 responses per item
- Administered at same time as operational assessments

NESI
ETS
Pearson
Westat

Plan → Develop → Expert Input → Develop → **Field Test** → Develop → Expert Input → Develop 17

The field test stage includes the development of the block structures (groups of items included in an assessment booklet), production time (including the printing of the booklets), and the actual data collection.

The data collection for pilot items generally takes place during an operational assessment. Personnel are already in the field, and they simply spiral these items in with the operational assessment. For example, one student may be taking the reading assessment and another may be taking some pilot items for mathematics. Between 1,000 and 1,500 responses per item are collected during this phase of the project. These responses serve as valuable input for further revisions.

Inside NAEP
Developing NAEP Test Questions

Develop

Further Item Development

- Proposed selection of items for the operational assessment
- Revisions based on empirical data from the pilot assessment

ETS

Plan → Develop → Expert Input → Develop → Field Test → **Develop** → Expert Input → Develop → 18

Sound, informed item selection and revision decisions are made based on information collected during the field tests. One process conducted during this phase is a differential item functioning analysis for various subgroups, such as race, ethnicity, and gender. This is done in order to determine if the items are behaving differentially for students that have comparable abilities. If a reason is found for the differential functioning, the item is eliminated. It is rare, however, that an item is excluded for this reason.

Slide 19

Inside NAEP
Developing NAEP Test Questions

Expert Input

Standing Committee

- Review of the proposed item pool in its entirety
- Review of individual items for quality

Academic Panel

- Mathematicians, scientists, etc.
- Accurate representation of the subject matter
- Balance of accuracy and grade- level appropriateness

NESI
ETS
Logistics

Plan → Develop → Expert Input → Develop → Field Test → Develop → Expert Input → Develop 19

At this stage in the process, the Standing Committee is consulted again reviewing the entire proposed item pool for quality. The Academic Review Panel, consisting of practitioners and subject matter experts, is also engaged to determine whether the item pool is representative of the subject matter and to ensure the items are accurate and grade-level appropriate.

Slide 20

Inside NAEP
Developing NAEP Test Questions

Develop

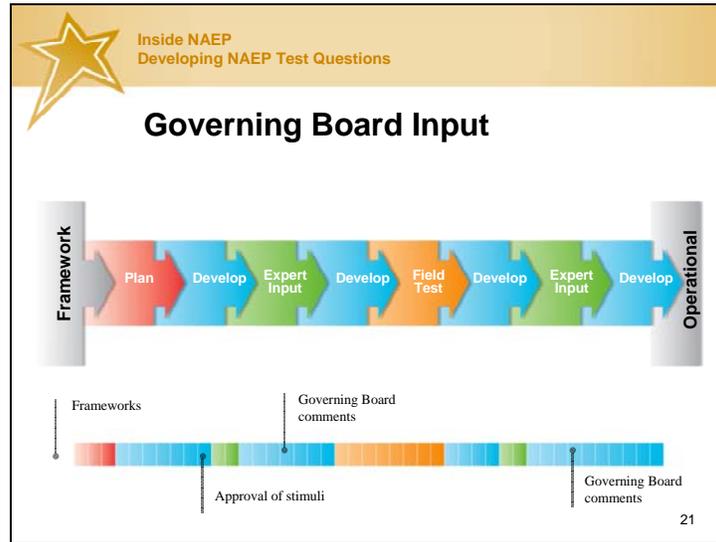
Further Item Development

- Resolution of comments from reviews
- Items revised
- Quality control
- Governing Board comments

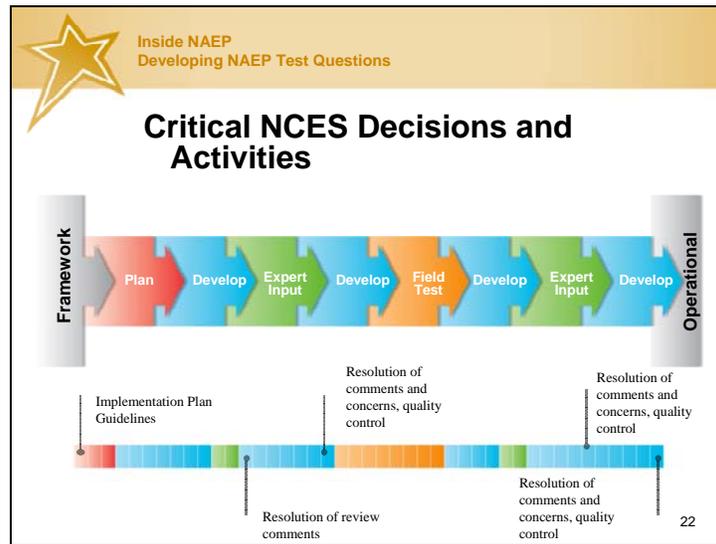
ETS

Plan → Develop → Expert Input → Develop → Field Test → Develop → Expert Input → **Develop** 20

Finally, the items are revised based on the feedback from the Standing Committee and Academic Review Panel and prepared for the final Board review and approval.



Again, item development is a very iterative process. This graphic is simply a representation of how the process works. Often the items are reviewed many more times than is indicated here. It should also be mentioned, that if items were being developed for a new framework, rather than for a revised or existing one, there would be many more stages involved.



To restate our mission, NCES is committed to developing quality items that have been thoroughly reviewed and vetted. This presentation covered the item development process for revised frameworks. Due to time constraints, there was much that could not be covered. For example, this presentation did not address: the development of different item types – multiple choice items, constructed response items, hands on tasks, and computer-based tasks. All of these item types would require a different level of thinking and advancement through these stages.



Inside NAEP
Developing NAEP Test Questions

NAEP Related Resources

For further information regarding NAEP item development, visit these NAEP and NAGB websites:

http://nces.ed.gov/nationsreportcard/tdw/item_development/

<http://nagb.org/policies/pl-index.htm>

23

The NAEP and the Governing Board websites offer ongoing information on NAEP development issues. These websites address item development specifically and the Governing Board policies applicable to the process.