



The slide features a dark blue background with a lighter blue curved graphic element on the left side. In the top left corner, there is a circular logo for the National Assessment Governing Board (NAGB), which includes a stylized star and the text 'NATIONAL ASSESSMENT GOVERNING BOARD'. To the right of the logo, the title 'NAEP Achievement Levels' is displayed in yellow text. Below the title, the names and titles of the speakers are listed in white text.

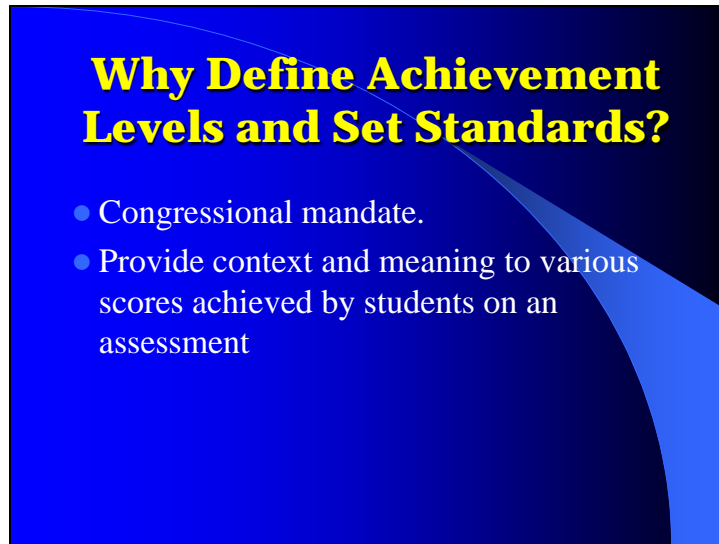
**NAEP
Achievement Levels**

Michael Ward, Chair of COSDAM
Susan Loomis, Assistant Director
NAGB
Christina Peterson, Project Director
ACT

This November 2006 presentation begins with sharing some general observations about setting achievement levels for the National Assessment of Educational Progress (NAEP) from Michael Ward, the Chair of the National Assessment Governing Board's Committee on Standards, Design and Methodology (COSDAM). Then more specifics on the standard-setting process will be presented by Susan Loomis of the Governing Board Staff. Finally, Christina Peterson of ACT will provide an update on the economics standard-setting project.



There are four levels of performance on NAEP based on cut scores set for Basic, Proficient, and Advanced. The fourth level, below basic, is by default and denotes performance that falls below our lowest performance level. The goal of standard setting is to translate the statements of what students should know and be able to do to the score scale and to represent those statements by cut scores on that scale. These cut scores are used to report how students performed on NAEP. It is important to mention that these lines on the scale are neither evenly nor consistently spaced across different assessments.

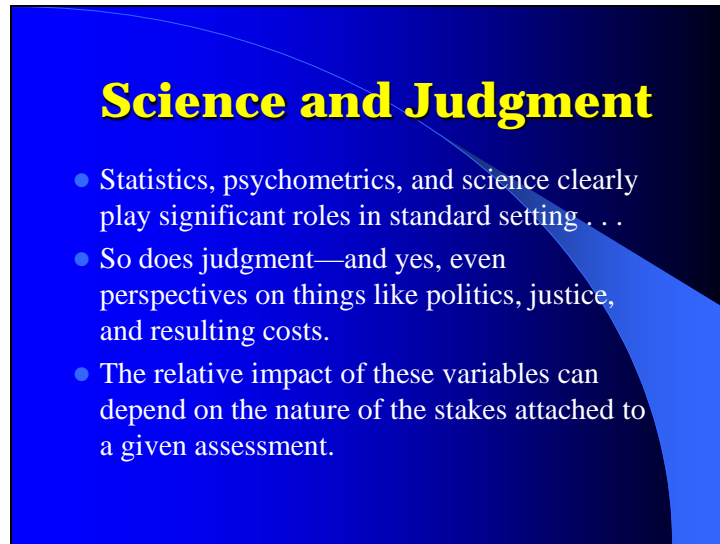


Why Define Achievement Levels and Set Standards?

- Congressional mandate.
- Provide context and meaning to various scores achieved by students on an assessment

Congress has mandated that the Board define achievement levels and standards for performance on NAEP.

Because a score on an assessment scale does not, in itself, have substantial meaning, the levels and standards defined by the Governing Board give context to the various scores achieved by students.



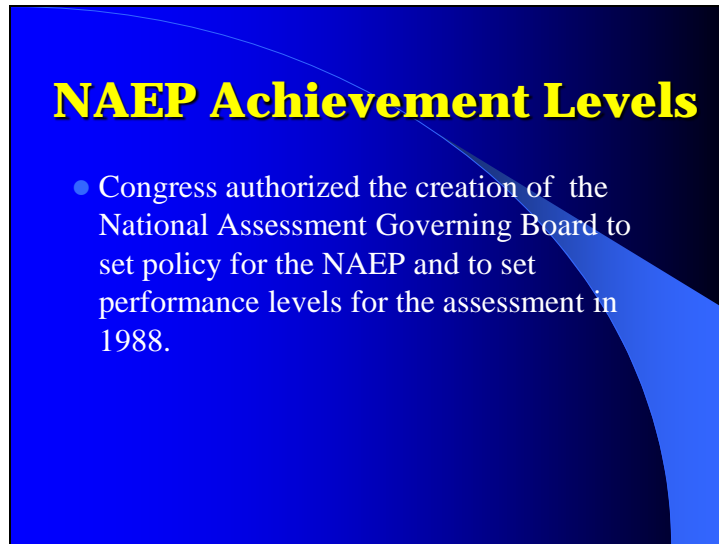
Science and Judgment

- Statistics, psychometrics, and science clearly play significant roles in standard setting . . .
- So does judgment—and yes, even perspectives on things like politics, justice, and resulting costs.
- The relative impact of these variables can depend on the nature of the stakes attached to a given assessment.

NAEP has invested a lot of time, money, and expertise to ensure that sound processes are used for setting achievement levels.

When the stakes associated with test results increase, additional variables enter into the decision-making process regarding standards. NAEP has relatively low stakes for the student, the school, and the district, because NAEP results are not tied directly to student-, school-, or district-level decisions. Alternatively, a state-level high school graduation test has high stakes.

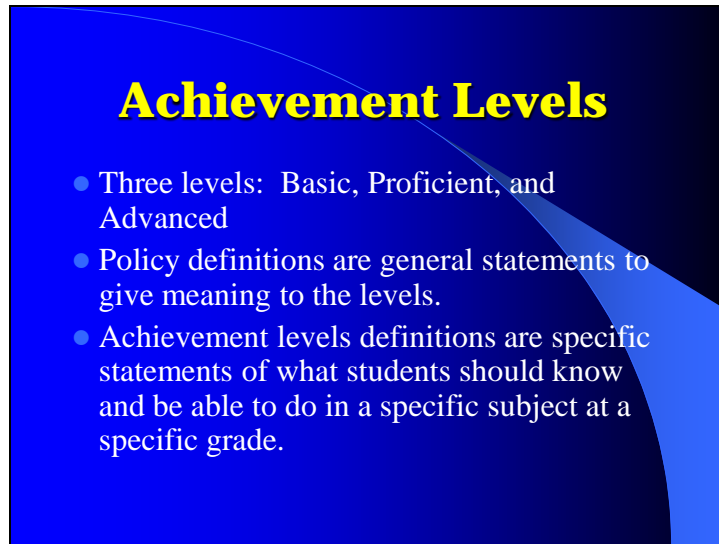
A cut score is a location on a score scale that delineates the boundary between achievement levels. In setting the cut score for a state assessment to delineate pass and fail, a state board of education typically considers variables such as costs, remediation, and other support systems, particularly, as the cut score increases. Other considerations include demographic issues, teacher incentives, school accreditation, and proficiency-based promotion policies.

A presentation slide with a blue background and a dark blue curved graphic on the right side. The title 'NAEP Achievement Levels' is in bold yellow text. A single bullet point in white text describes the creation of the National Assessment Governing Board in 1988.

NAEP Achievement Levels

- Congress authorized the creation of the National Assessment Governing Board to set policy for the NAEP and to set performance levels for the assessment in 1988.

On this slide, two core objectives of the Board are listed.



Achievement Levels

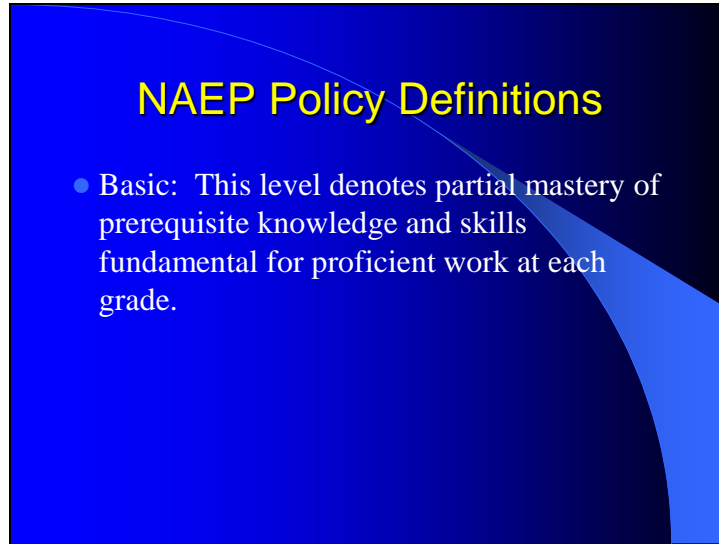
- Three levels: Basic, Proficient, and Advanced
- Policy definitions are general statements to give meaning to the levels.
- Achievement levels definitions are specific statements of what students should know and be able to do in a specific subject at a specific grade.

Policy definitions of the three achievement levels are general statements that give meaning to the levels, but there are also specific statements of what students should know and be able to do at each achievement level in a specific subject at a specific grade. These specific statements are referred to as achievement level definitions. Policy definitions guide the development of achievement level definitions.

NAEP Policy Definitions

- Proficient: This level represents *solid academic performance* for each grade assessed. Students reaching this level have demonstrated *competency over challenging* subject matter, including subject-matter knowledge, applications of such knowledge to real world situations, and analytical skills appropriate to the subject matter.

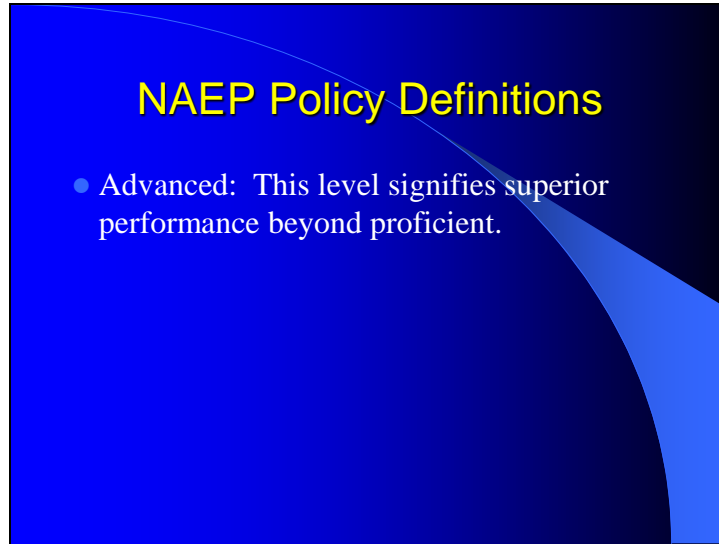
The policy definition for the Proficient level is more specific, compared to the policy definitions for the Basic and Advanced levels. Intentionally, the Proficient achievement level has a high standard, and reflects a high level of expectation.

A slide with a blue background and a dark blue curved graphic on the right side. The title "NAEP Policy Definitions" is in yellow. A bullet point in white text describes the "Basic" level.

NAEP Policy Definitions

- Basic: This level denotes partial mastery of prerequisite knowledge and skills fundamental for proficient work at each grade.

Basic demonstrates partial mastery of Proficient level knowledge and skills. Both the Basic and Advanced achievement levels reference the Proficient level. Basic performance is defined as “partial mastery”—below Proficient, and Advanced is defined as superior performance—beyond Proficient.

A blue rectangular slide with a dark blue gradient background. The title "NAEP Policy Definitions" is written in yellow text at the top. Below the title, there is a single bullet point in white text: "Advanced: This level signifies superior performance beyond proficient."

NAEP Policy Definitions

- Advanced: This level signifies superior performance beyond proficient.

The policy definitions do not change from subject to subject, from grade to grade, or from framework to framework. Basic, Proficient, and Advanced levels always have the same policy meaning—partial mastery, solid academic performance, and superior performance, respectively. The policy definition for the Advanced level is presented here.

NAEP Achievement Level Definition for Grade 8 Proficient Performance in Mathematics

- *Eighth-grade students performing at the Proficient level should apply mathematical concepts and procedures consistently to complex problems in the five NAEP content strands.*

As part of the Governing Board's work, achievement levels are also detailed in specific terms appropriate to each subject and grade assessed. This is NAEP's current achievement level definition for grade 8 Proficient performance in mathematics. The Board developed this summary paragraph to state what eighth-grade students need to know and be able to do in order to perform at the Proficient level in mathematics. This definition applies to all the assessments that have been developed since 1990 for grade 8 mathematics.

8th grade Proficient continued

- Eighth graders performing at the Proficient level should be able to conjecture, defend their ideas, and give supporting examples. They should understand the connections among fractions, percents, decimals, and other mathematical topics such as algebra and functions. Students at this level are expected to have a thorough understanding of basic level arithmetic operations—an understanding sufficient for problem solving in practical situations.

Here is a continuation of the example of grade 8 mathematics at the Proficient level. This illustrates the level of detail in the achievement level definitions. As noted earlier, the more general policy definitions are used to guide development of these detailed achievement level definitions for each subject and grade assessed.

8th grade Proficient continued

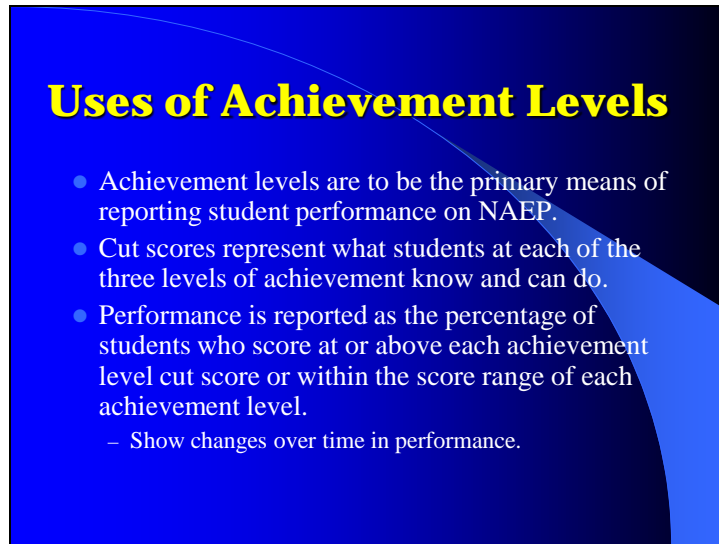
- Quantity and spatial relationships in problem solving and reasoning should be familiar to them, and they should be able to convey underlying reasoning skills beyond the level of arithmetic. They should be able to compare and contrast mathematical ideas and generate their own examples. These students should make inferences from data and graphs; apply properties of informal geometry; and accurately use the tools of

The same example continues. As seen here, achievement level definitions can be lengthy.

8th grade Proficient continued

- technology. Students at this level should understand the process of gathering and organizing data and be able to calculate, evaluate, and communicate results within the domain of statistics and probability.

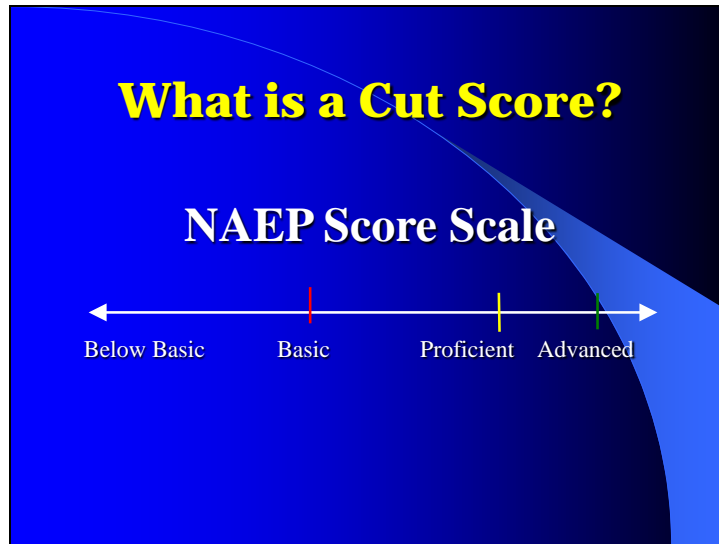
The example continues here.



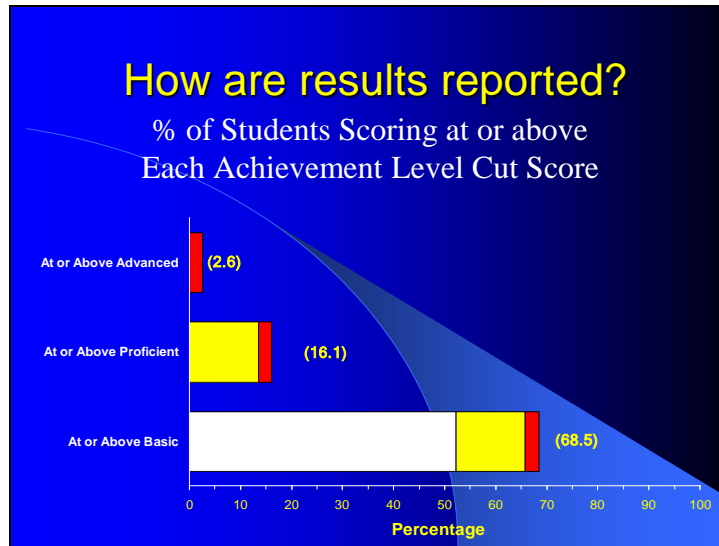
Uses of Achievement Levels

- Achievement levels are to be the primary means of reporting student performance on NAEP.
- Cut scores represent what students at each of the three levels of achievement know and can do.
- Performance is reported as the percentage of students who score at or above each achievement level cut score or within the score range of each achievement level.
 - Show changes over time in performance.

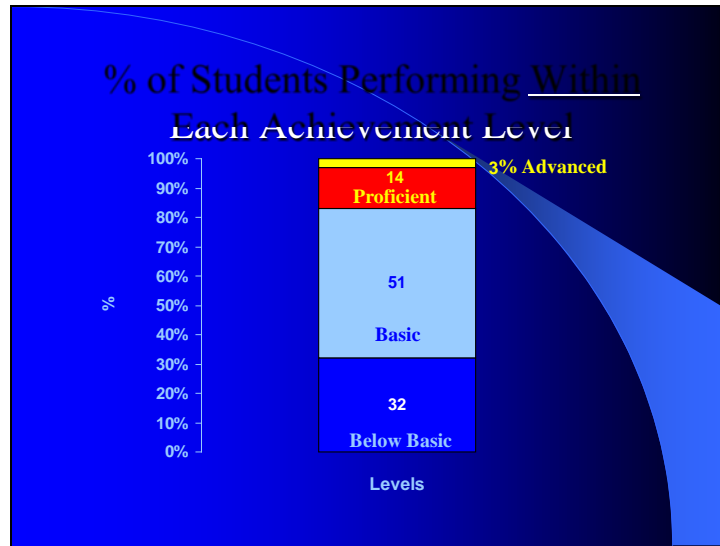
Achievement levels are used for several purposes. First, they are meant to be the primary means of reporting student performance on NAEP. The cut scores that are associated with achievement levels represent what students know and are able to do. Student performance is reported, typically, as the percentage of students who perform at or above each of the achievement levels.



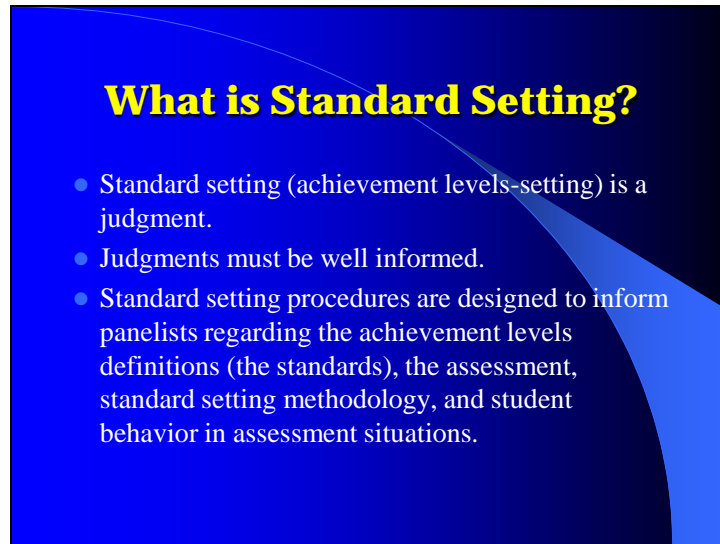
Cut scores are set through this achievement level setting process. Cut scores represent one of three component parts of the achievement levels—the definitions, the cut scores, and the items to represent that performance. The process of setting achievement levels is to determine where cut scores fall and how close or far apart they are on the score scale.



For illustrative purposes, the slide shows that 68.5 percent of students perform at or above the Basic level. That means that 68.5 percent of students score from the cut score for Basic and higher. At the same time, 16.1 percent of students score at or above the Proficient cut score, and this includes Proficient and Advanced levels of performance. Finally, 2.6 percent of students score at the Advanced level.



Reports also include the percentages of students who perform within each of the achievement level score ranges. This chart shows 32 percent of students performing below the Basic cut score, 51 percent within the Basic score range, 14 percent within the Proficient score range, and about 3 percent within the Advanced score range.

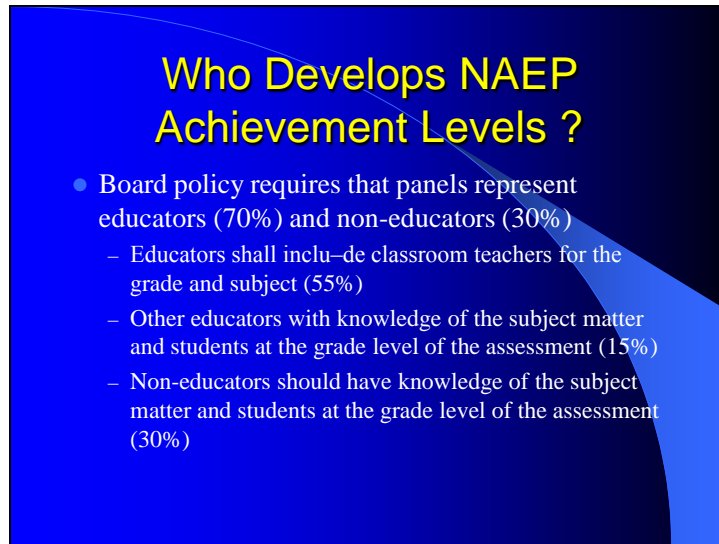


What is Standard Setting?

- Standard setting (achievement levels-setting) is a judgment.
- Judgments must be well informed.
- Standard setting procedures are designed to inform panelists regarding the achievement levels definitions (the standards), the assessment, standard setting methodology, and student behavior in assessment situations.

Standard-setting procedures establish the criteria to apply in determining cut scores. The criteria in the case of standard setting for NAEP are the achievement level definitions, which are used for making decisions and judgments about where to set the cut scores.

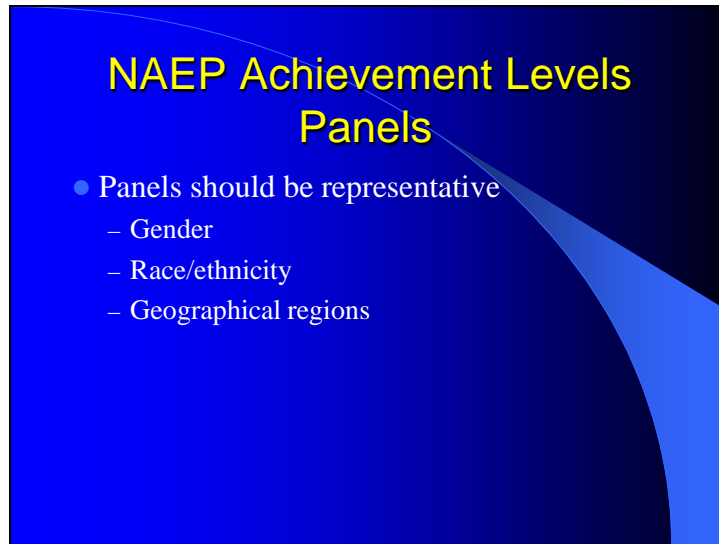
Standard setting is a judgmental process that must be well informed by facts and information relevant to setting cut scores. Some of this information includes the difficulty of an item, how likely it is that students will answer certain kinds of questions on tests, and how the judgments of the panelists relate to those of other panelists and to the performance of students. An array of information is provided about items, about overall performance on the test, and about the agreement among judges regarding how to represent the achievement levels definitions on the score scale.



Who Develops NAEP Achievement Levels ?

- Board policy requires that panels represent educators (70%) and non-educators (30%)
 - Educators shall include classroom teachers for the grade and subject (55%)
 - Other educators with knowledge of the subject matter and students at the grade level of the assessment (15%)
 - Non-educators should have knowledge of the subject matter and students at the grade level of the assessment (30%)

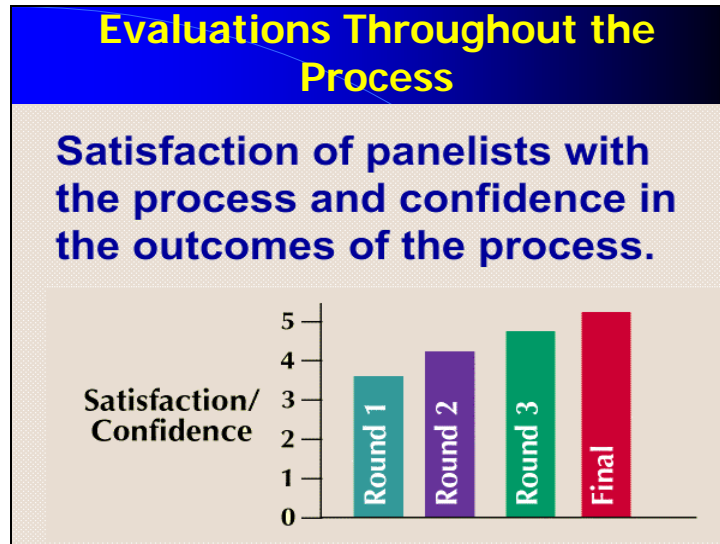
The NAEP achievement levels are developed by panels in the standard-setting process. The Board has set policy regarding the composition of these panels. That policy is fairly unusual for standard setting as it requires a mix of educators and non-educators. The educators are subdivided as classroom teachers who teach the subject matter at the grade level for which they are a standard-setting panelist and non-teacher educators who are typically curriculum directors or specialists in the districts or at the state level. They may be faculty members in colleges of education who teach that subject area for teacher education programs in the appropriate grade levels or college faculty who teach freshman courses in that subject area. Non-educators are individuals who have training in the subject matter and familiarity with students at the grade level, but no significant career experience in education. It is necessary that all people serving on the panels be familiar with and have a background in the subject matter of the assessment and that they know about students at the grade levels for which they are serving as panelists.

A presentation slide with a dark blue background and a lighter blue curved graphic on the right side. The title 'NAEP Achievement Levels Panels' is written in yellow. Below the title is a bulleted list in white text.

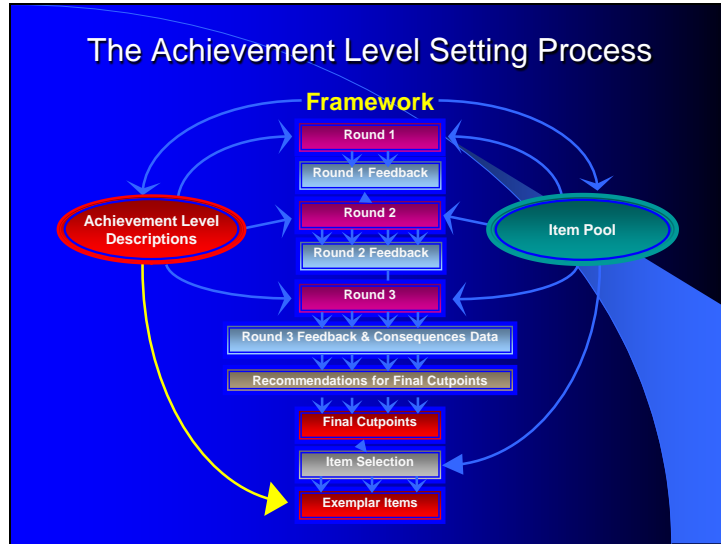
NAEP Achievement Levels Panels

- Panels should be representative
 - Gender
 - Race/ethnicity
 - Geographical regions

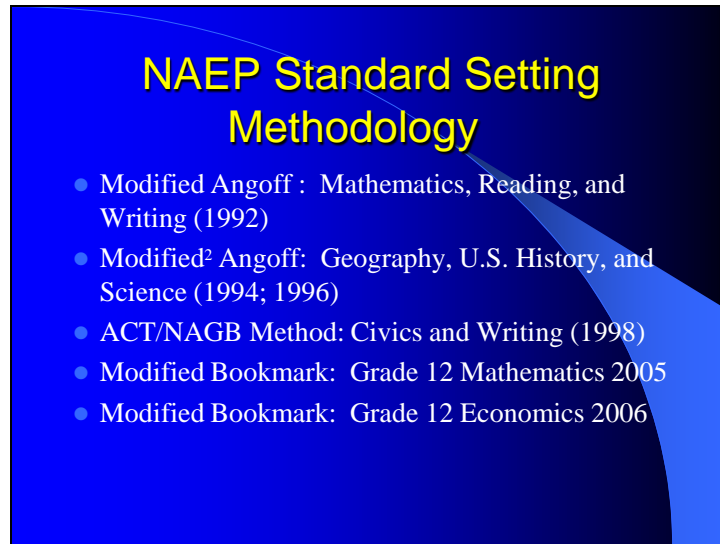
In addition, the panels are selected to be representative of gender, race/ethnicity, and geographic regions, approximately in proportion to the population of the nation. It should be noted that recruiting panelists is an intricate process.



Evidence of procedural validity is one important aspect of a standard-setting process. Thus, it is imperative that panelists' evaluations be collected throughout the process to determine their level of knowledge and confidence about the process and their work as panelists, and their level of satisfaction with the outcomes of the process. It is desirable to see the panelists' confidence increase across the stages of the process, and this tends to be associated with greater satisfaction in the outcomes.



Shown here is what is considered a “Gold Standard” for a standard–setting process. The iterative process involves many steps with feedback following each of several opportunities for making judgments. Key inputs to the process are the achievement level definitions and the assessment item pool. The panelists then begin their work of making decisions about performance on the item pool based on the achievement level definitions. For each round of judgments, panelists receive feedback regarding their judgments and the implications of the cut scores. The process continues with panelists revisiting the item pool using the achievement level definitions as the criteria that guide the process. At the end of the process, there are cut scores or cut points, denoted as numerical representations on the score scale. As part of the process, the panelists select a subset of items as exemplars to represent the achievement level definitions in reports of NAEP results.

A blue rectangular slide with a dark blue gradient background. The title "NAEP Standard Setting Methodology" is centered at the top in yellow text. Below the title is a bulleted list of five items, each starting with a yellow circle. The items describe different NAEP standard setting methodologies and their application years.

NAEP Standard Setting Methodology

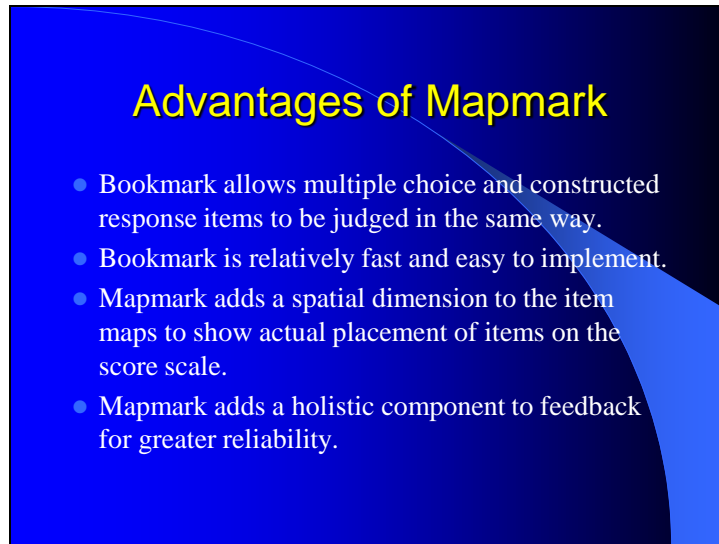
- Modified Angoff : Mathematics, Reading, and Writing (1992)
- Modified² Angoff: Geography, U.S. History, and Science (1994; 1996)
- ACT/NAGB Method: Civics and Writing (1998)
- Modified Bookmark: Grade 12 Mathematics 2005
- Modified Bookmark: Grade 12 Economics 2006

The methodologies that have been used for NAEP standard setting have evolved over time, moving from an item-by-item methodology to a more holistic approach. The first contract that was awarded for NAEP achievement level setting required that a modified Angoff process be used for setting the achievement levels cut scores for mathematics, reading, and writing in 1992. In the modified Angoff process, panelists make judgments on how students who meet the minimal criteria defining achievement at each level—Basic, Proficient, and Advanced—would perform on each item. A modified Angoff rating methodology was used again in 1994 for geography and U.S. history, in 1996 for science, and in 1998 for civics and writing.

With the 1994 and 1996 standard-setting projects, there was also experimentation to investigate how item maps could be used to facilitate panelists' judgments. Item maps show locations of items at score points on the NAEP scale and they were used to help panelists understand how the cut scores delineate what students at each achievement level can do, cannot do, or would find challenging.

In the 1998 achievement level setting, a process was used that further modified the Angoff methodology. This procedure involved the use of Reckase charts. Mark Reckase designed the charts and methodology used to provide feedback to panelists. It was a successful process, still based on a modified Angoff process. By this point, however, many considered the method sufficiently different from the Angoff procedure to warrant calling it the ACT/NAGB method.

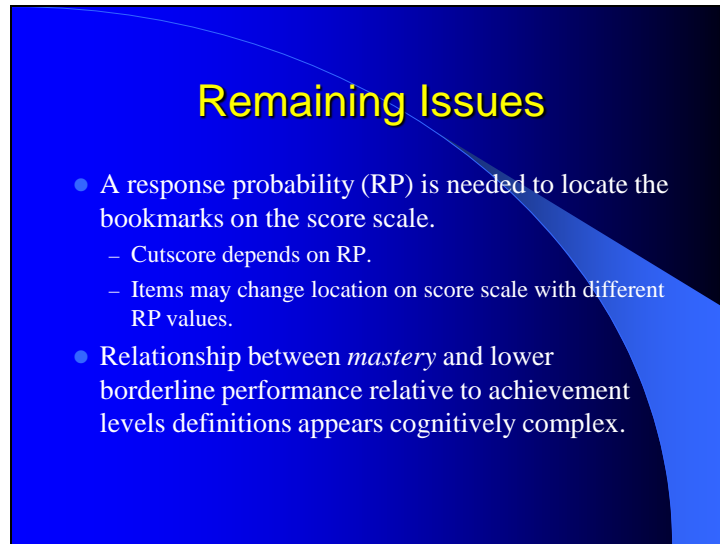
For the 2005 grade 12 mathematics achievement level setting, a Bookmark modification called Mapmark, now called Mapmark with Domains was used. This is one of the methods that is to be piloted in the standard-setting work for the 2006 NAEP Economics assessment at grade 12.



Advantages of Mapmark

- Bookmark allows multiple choice and constructed response items to be judged in the same way.
- Bookmark is relatively fast and easy to implement.
- Mapmark adds a spatial dimension to the item maps to show actual placement of items on the score scale.
- Mapmark adds a holistic component to feedback for greater reliability.

The Bookmark and Mapmark methodologies have noted advantages over the Angoff method. First, Bookmark was designed to incorporate both constructed response and multiple choice items and allows the same judgment process for both types of items. Bookmark is also relatively fast and easy to implement, and the computation of cut scores is simple. The Mapmark modification to the Bookmark procedure has added a spatial dimension to the item maps, so that panelists can actually see how different two items are in terms of where they are located on the score scale. This feature assists panelists in making judgments about the placement of bookmarks, a marker that indicates the lower score boundary of an achievement level. Finally, the Mapmark procedure designed by ACT is an improvement over the Bookmark method in that it adds an holistic aspect, rather than limiting the judgments solely to individual items.



Remaining Issues

- A response probability (RP) is needed to locate the bookmarks on the score scale.
 - Cutscore depends on RP.
 - Items may change location on score scale with different RP values.
- Relationship between *mastery* and lower borderline performance relative to achievement levels definitions appears cognitively complex.

There are some remaining issues. An important one is that the Bookmark and Mapmark procedures require a response probability criterion or a method of locating the items on the item score scale in the item map. The response probability is defined as the probability of a student answering an item correctly, if the item difficulty matches his or her ability. This is an assumption, and so any response probability may be selected as the criterion. This criterion is used to map items to scale scores. The issue, here, is that the cut score depends on the response probability criterion, and this will be demonstrated in the next slide. For a given item (bookmark placement) the lower the response probability, the lower the score at which the item will map. As a result, the bookmark, although placed in the same location (i.e., at the same item in the set or ordered items), could move up or down the score scale depending on the response probability associated with that item. Thus, individual items could change their ordering or their relative location on the scale score depending on the response probability criterion.

Another issue is the term “mastery” and how it is used to designate performance at the lower borderline of an achievement level. Mastery may pose problems for some panelists when making their judgments.

When thinking about what students can do, the translation of achievement level definitions to the score scale is different depending upon the response probability

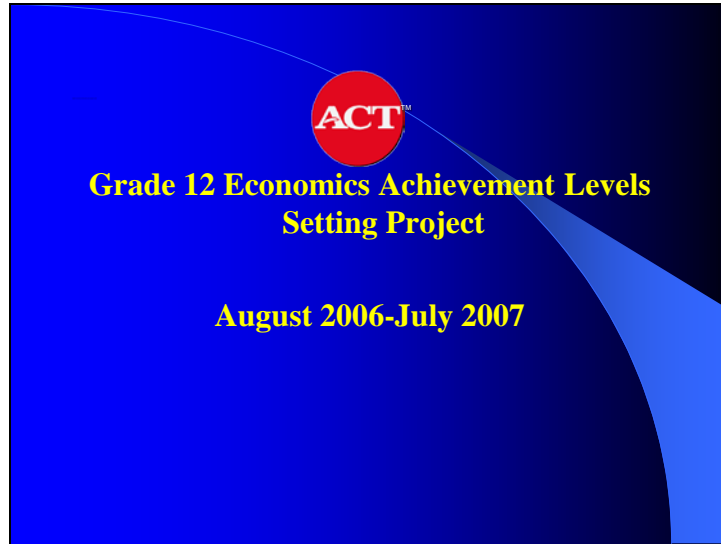
criterion used. When the term mastery is considered, a higher response probability may be required as evidence of mastery.

Item		RP Criterion		
Order	Item Description	80%	65%	50%
1	Newspapers delivered in 5 days	256	237	216
2	Marble from bag most likely red	260	249	238
3	Understand concept of a variable	265	255	246
4	Runner passed after 14 minutes	286	266	248
5	Find angle of N-gon	295	277	262
6	Extent a given pattern	299	277	258
7	Apply concept of minimum	316	290	263
8	Find cost using percent	329	309	292
9	Find intersection of a rectangle	340	323	308
10	Estimate using a circle graph	383	358	335
11	Find area of trapezoid	416	389	368
	BASIC = 262			
	PROFICIENT = 299			
	ADVANCED = 333			

This slide illustrates another issue with response probabilities and scale scores. At a given scale score, items having a lower probability of correct response are relatively more difficult than those having a higher probability of correct response.

A response probability of 80 is one possible mastery criterion, 65 is another, and 50 is a third. These are all illustrated on this slide. One can see that the scale scores vary for these items and that it is possible for items to be classified into different achievement levels depending upon the response probability criterion that was used to map the item. The relationship between achievement levels, scale scores, and the response probability criterion is a large area of research associated with the Bookmark methodology, and will be for Mapmark in the future.

Slide 27



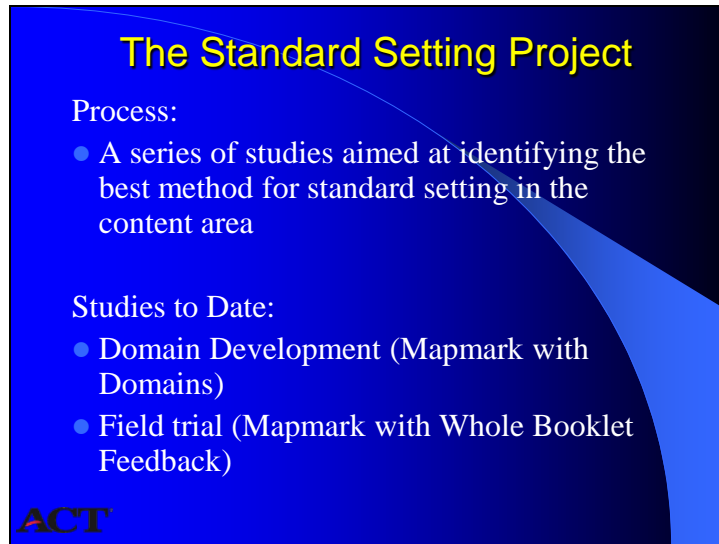
Christina Peterson will now discuss more specifically the Mapmark procedure that ACT has designed for the 2006 NAEP Economics assessment.

Proposed Methods

<u>Mapmark with Domains</u>	<u>Mapmark with Whole Booklet Feedback</u>
<ul style="list-style-type: none">● Knowledge, Skills and Abilities (KSA) Review● Feedback and Domain Information● Feedback and Consequences Data	<ul style="list-style-type: none">● Knowledge, Skills and Abilities (KSA) Review● Feedback and Student Performance Information● Feedback and Consequences Data

ACT

For the economics assessment, there are two proposed methods for standard setting, Mapmark with Domains and Mapmark with Whole Booklet Feedback. Both of these methods are Bookmark-based, a procedure commonly used at the state level. Like Bookmark, they begin the first round with what is called a knowledge, skills, and abilities review, or a KSA review, where the panelists look at the items ordered from easiest to hardest, and identify what is needed to answer each item correctly relative to the other items in that ordering. Also, like Bookmark, both methods end with a panelist review of the consequences data, which includes the proportion of students scoring at or above each of the three achievement levels based on the cut scores that emerged from the previous round. The difference between Mapmark and Bookmark is that Mapmark uses a spatially representative item map, which allows panelists to see the distinction and difference between two consecutive items. The other difference is in the nature of the holistic feedback that is provided in the second round. In the feedback for Mapmark with Domains, panelists are shown expected percent correct scores on all areas of content in economics for each of the three achievement levels. In Mapmark with Whole Booklet Feedback, domain-specific information is not shown. Instead, panelists are shown examples of test booklets of actual student performance at each of the three achievement levels.



The Standard Setting Project

Process:

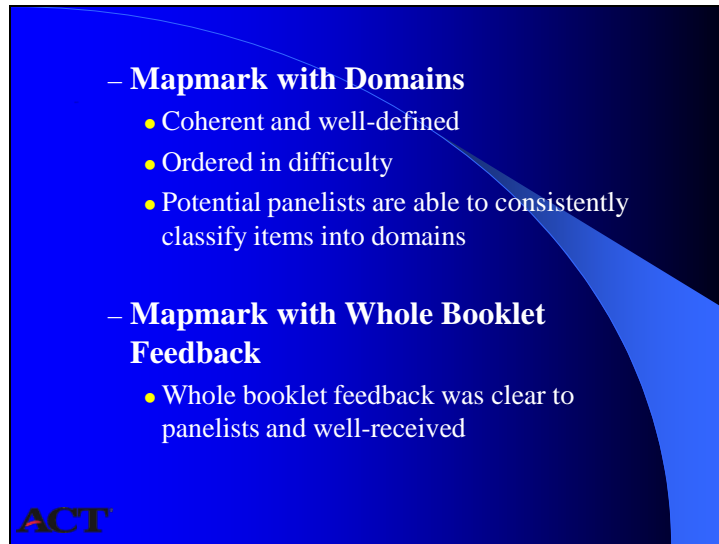
- A series of studies aimed at identifying the best method for standard setting in the content area

Studies to Date:

- Domain Development (Mapmark with Domains)
- Field trial (Mapmark with Whole Booklet Feedback)

ACT

The process for standard setting is to conduct a series of studies aimed at identifying which of the proposed methods would be most effective for use in establishing the achievement levels or the cut scores in a given content area. ACT has identified four studies to precede the actual achievement level setting meeting. The first two were designed to examine the viability of the two Mapmark methods with economics content. The first study was a domain development to establish the domains or sub-areas of the content within economics to ascertain if those domains were meaningful and coherent, and to determine if they varied in difficulty. The second study was a field trial of the Mapmark with Whole Booklet Feedback where panelists were asked to evaluate and provide feedback on the method, specifying what was useful and what was not useful.



- **Mapmark with Domains**
 - Coherent and well-defined
 - Ordered in difficulty
 - Potential panelists are able to consistently classify items into domains
- **Mapmark with Whole Booklet Feedback**
 - Whole booklet feedback was clear to panelists and well-received

ACT

Findings revealed the domains did emerge as coherent and well defined, such that panelists not involved in the domain development process were able to consistently classify items into the domains using solely the domain definitions. Also, in the Mapmark with Whole Booklet Feedback very positive results from the panelists were received. They responded quite well to the booklets and felt that the ability to actually see student performance at the cut scores made the process more tangible, and enabled panelists to feel more confident about subsequent decisions. Based on this positive feedback, both Mapmark methods were recommended for implementation in the pilot.

Project Schedule to Date

Event	Date	Purpose
Domain Development Meeting	September 17-19, 2006	Develop domains for use in Mapmark with domains standard setting method.
Domain Item Classification	October 13, 2006	Evaluate domains.
Field Trial	October 20-21, 2006	Evaluate and refine Mapmark with whole booklet feedback.
COSDAM Meeting	November 16-18, 2006	Review results of domain development and field trial. Determine two methods for use in the pilot study.

ACT

This is a schedule of the activities and meetings for the two methodology studies discussed in the previous slide.

Project Schedule: Tasks Remaining

Event	Date	Purpose
Pilot Study	December 6-9, 2006	Compare two methods.
Validity Studies	January 11-13, 2007	Assess the validity of the ALS cut scores.
COSDAM Meeting	February 2, 2007 (Tentative)	Review results of pilot and decide on ALS method.
NAGB Meeting	March 1-3, 2007	Review results of pilot study and ALS method.

ACT

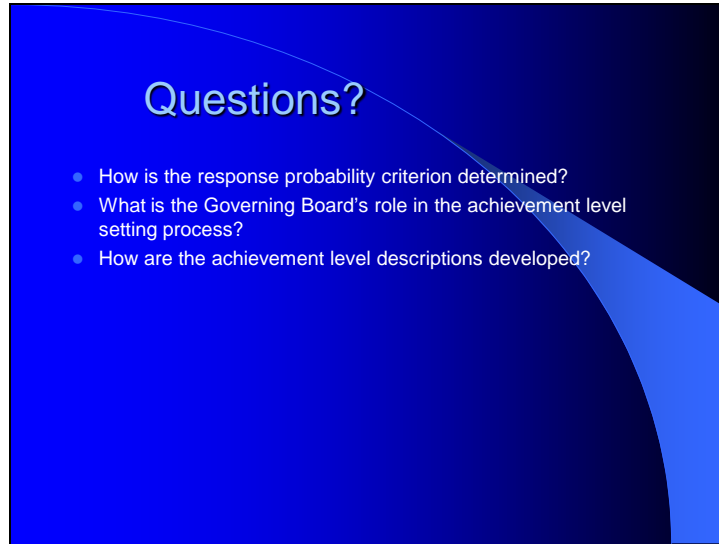
In early December 2006, both methods will be piloted, and the panelists will be asked to evaluate both. Data on the similarity of the cut scores that emerge will be collected, so that comparisons of the two methods both on the similarity of the cut scores and on the evaluative data can be made. In January 2007, additional data for validity will be collected, but analysis of these data will not be conducted until the cut scores from the achievement level setting meeting are obtained in mid-March 2007. At this January 2007 meeting, an independent group of economists, educators, and non-educators will provide the data for two validity studies. First, they will be provided with booklets of student performance across a range of the achievement scales and asked to classify the booklets into the achievement levels based on the achievement level descriptions alone. Second, they will be provided with all of the achievement items and asked to classify these items into the achievement levels based on the achievement level descriptions alone. Analysis will determine if their classifications match the empirical classifications that come out of the achievement level setting meeting in mid-March 2007. In February 2007, a special meeting with COSDAM will be held to review the results of the pilot study, and the comparative data, as well as to present recommendations for the method to be used for the achievement level setting meeting in mid-March 2007. At the March Governing Board Meeting, the pilot results will also be presented to the entire Board, as well as the results from the discussion with COSDAM.

Project Schedule: Tasks Remaining

Event	Date	Purpose
ALS Meeting	March 7-10, 2007	Set new Grade 12 cut scores and select exemplar items.
COSDAM Meeting	May 8, 2007	Review ALS and Validity Studies results. Prepare recommendation of cut scores for the full board.
NAGB Meeting	May 17-19, 2007	Review ALS and Validity Studies results. Action on cut scores.

ACT

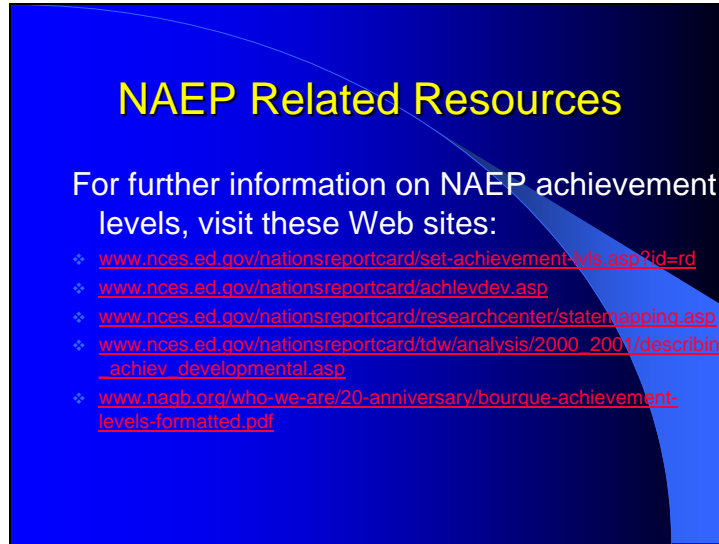
Later in March 2007, the actual achievement level setting meeting will occur utilizing the standard setting method that was decided upon. The panelists at this meeting will set new grade 12 cut scores for economics and recommend exemplar items for each of the three achievement levels. In May 2007, a presentation of the results of the achievement level setting meeting will be presented to COSDAM, along with the results of the validity studies. COSDAM will use this information to prepare a recommendation of cut scores for the Board. At the next Board meeting in May 2007, the results of the achievement level setting meeting and the validity studies will again be reviewed by COSDAM. At this time, the Governing Board will take action on the final achievement levels.



Questions?

- How is the response probability criterion determined?
- What is the Governing Board's role in the achievement level setting process?
- How are the achievement level descriptions developed?

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

A blue rectangular slide with a dark blue gradient background. The title "NAEP Related Resources" is centered at the top in yellow. Below it, the text "For further information on NAEP achievement levels, visit these Web sites:" is in white. A list of five red hyperlinks follows, each preceded by a small white diamond symbol.

NAEP Related Resources

For further information on NAEP achievement levels, visit these Web sites:

- ❖ www.nces.ed.gov/nationsreportcard/set-achievement-lev.asp?id=rd
- ❖ www.nces.ed.gov/nationsreportcard/achlevdev.asp
- ❖ www.nces.ed.gov/nationsreportcard/researchcenter/statemapping.asp
- ❖ www.nces.ed.gov/nationsreportcard/tdw/analysis/2000_2001/describing_achiev_developmental.asp
- ❖ www.nagb.org/who-we-are/20-anniversary/bourque-achievement-levels-formatted.pdf

These NAEP and Governing Board Web sites will keep you abreast of any further developments in the achievement levels and standard-setting processes.

How is the response probability criterion determined?

The response probability criterion is determined by policy decision. The response probability is selected as a criterion to use for locating items on the item map and to represent the probability that students are likely to get an item correct. Field tests are informative, but they are not necessary to set the criterion.

What is the Governing Board's role in the achievement level setting process?

When the panels are convened to set the standards and cut scores, Board members are invited to observe. Board members do not enter into policy discussion at this point in time. The Board's policy discussion proceeds from what is produced in the process. The contract officer oversees the achievement level setting projects and attends the meetings to explain what NAEP and the Governing Board are about and what the goals and purposes are for the standard setting process.

How are the achievement level descriptions developed?

The achievement level descriptions are written as part of the framework development process. The framework development panels are requested to draft preliminary achievement level definitions. These are based on the policy definitions that the Board has adopted, and they focus on the parts of the framework that are appropriate for the Basic, Proficient, and Advanced levels of performance for each of the three grade levels for the subject area described in the framework. Panels are then convened consisting of content experts to review and finalize those definitions so that the finalized definitions can be used to guide the item development process and the standard setting process. Revisions of achievement level definitions are produced from an iterative process. They are sent out multiple times to the broader community and are modified based on the recommendations until there is general agreement that the achievement level definitions are statements that reflect what needs to be addressed for each grade level in the subject area of the assessment.