Connecticut Alternate Science Assessment Standard Setting

Science
in Grades 5, 8, and 11

CSDE
CONNECTICUT STATE
DEPARTMENT OF EDUCATION

Submitted to
Connecticut State Department of Education
by American Institutes for Research

Table of Contents

List of Tables

List of Figures

1. Executive Summary

The Individuals with Disabilities Education Act (IDEA) of 1997 established a legal requirement for all students to participate in statewide content-area assessments. The goal of this requirement was to ensure that every child—including special-education students with the most significant cognitive disabilities—would have access to rigorous curriculum, effective instruction, and reasonable and high expectations for achievement of academic content. While students with the most significant cognitive disabilities do not always participate in the same grade-level academic classroom instruction with general education students, they are nevertheless expected to receive grade-level instruction with appropriate academic content and skills with simplifications in the breadth, depth, or complexity of the content standards.

The Connecticut Alternate Science Assessments (CTAS) is an alternate assessment based on alternate achievement standards for students with significant cognitive disabilities. The CTAS has been developed to ensure that all students with significant cognitive disabilities can participate in an assessment that is a measure of what they know and can do in relation to the Next Generation Science Standards (NGSS). The CTAS includes six performance tasks that are intended to be administered throughout the year as teachers work with eligible students to rate student performance on the CTAS Core Extensions. Teachers administer various activities to the students and submit performance ratings into the Data Entry Interface (DEI). The CTAS must be administered to eligible students with significant cognitive disabilities in grades 5, 8, and 11.

In 2019 AIR calibrated the CTAS using the 1PL item response theory (IRT) model. After the spring 2019 administration, a Bookmark standard-setting workshop was held to determine cut scores for three performance standards for each test.

On July 29–30, 2019, the American Institutes for Research (AIR), under contract to the CSDE, invited a panel of 25 teachers and administrators to recommend performance standards (cut scores) for the test. The CSDE recruited a broadly representative panel ensuring that a diverse range of perspectives informed the standard-setting process. Panelists included special-education teachers, curriculum specialists, education administrators, and other stakeholders. The panel was also broadly representative of Connecticut's special education teacher population in terms of gender, race/ethnicity, and regional composition. The CSDE designated the most knowledgeable and experienced panelists at the workshop as table leaders.

For each test, the panelists recommended three cut scores, or performance standards: Approaches Proficiency, Meets Proficiency, and Exceeds Proficiency.

1.1. The Standard-Setting Process

Connecticut used the Bookmark procedure (Mitzel, Lewis, Patz & Green, 2001), which is the most common procedure used throughout the country. In this process, the panelists review items ordered by difficulty in an ordered-item booklet (OIB) for each test. Each OIB contains a set of items that meet the test blueprint. The panelists also reviewed the corresponding Connecticut Essence Statements and Core Extensions and performance-level descriptors (PLDs) for each test. With this information in mind, the panelists selected pages in the OIB that best represent the cut scores on the test. The Bookmark standard-setting process was described in a standard-setting plan submitted to the CSDE. The CSDE approved the plan before the workshop.

The standard-setting workshop was held over two days. The first day was devoted to training and review of materials, and the second day was devoted to two rounds of standard setting. At the end of the activity, the panelists completed a survey that evaluated the workshop.

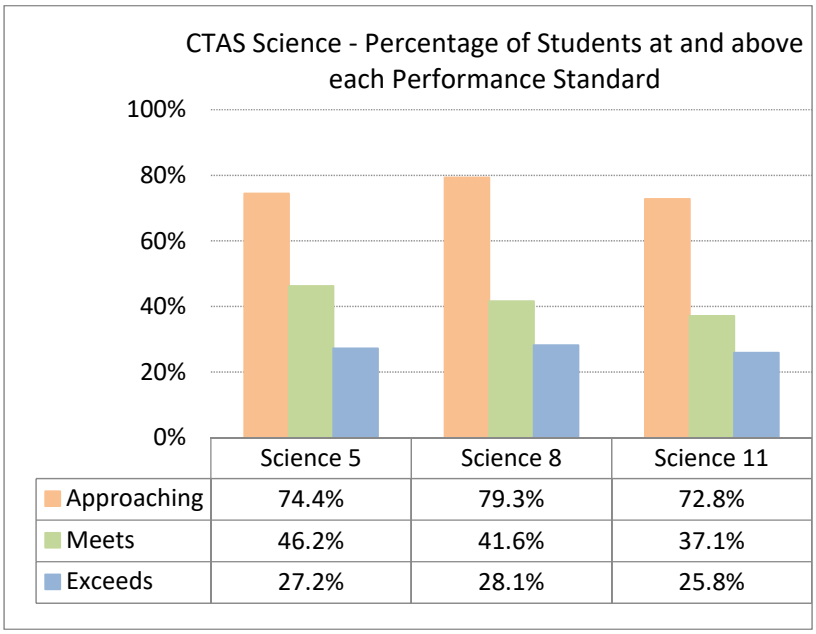
1.2. Performance Standards Recommended by the Panel

The recommended performance standards are presented in Table 1. The percentage of students reaching each standard and percentage of students within each standard are summarized in Figure 1 and Figure 2. Note that each test has an independent scale, so the cut scores should not be compared across grades. The percentages are based on students who took the 2019 operational field tests.

Table 1. CTAS Performance Standards

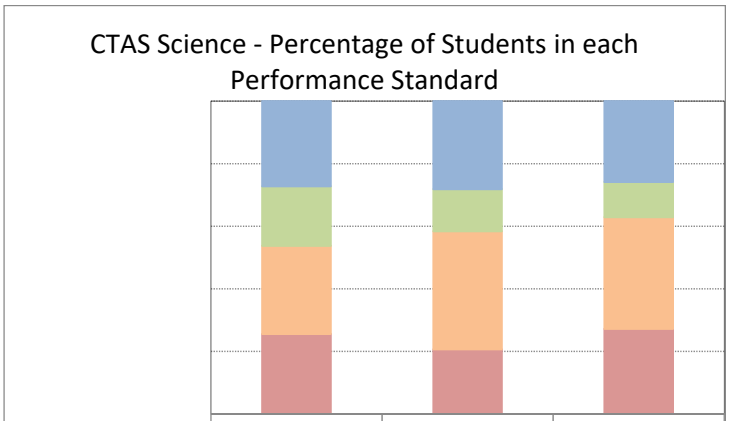
Grade	Performance Standard		
	Approaching	Meets	Exceeds
5	32	57	65
8	26	57	64
11	32	57	65

Figure 1. Percentage of Students Reaching Performance Standard



	Science 5	Science 8	Science 11
Approaching	74.4%	79.3%	72.8%
Meets	46.2%	41.6%	37.1%
Exceeds	27.2%	28.1%	25.8%

Figure 2. Percentage of Students Within Each Performance Standard



	Science 5	Science 8	Science 11
Exceeds	27.2%	28.1%	25.8%
Meets	19.0%	13.4%	11.2%
Approaching	28.2%	37.7%	35.7%
Not Meet	25.6%	20.7%	27.2%

2. Introduction

Connecticut developed an innovative new performance-task-based science assessment system for students with disabilities in grades 5, 8, and 11, the Connecticut Alternate Science Assessment (CTAS). The CTAS tests are for students with significant cognitive disabilities who participate in a school curriculum that includes both academic and life skills. The CSDE provides a summary of the new tests at https://portal.ct.gov/SDE/Student-Assessment/CTAA-Skills-Checklist/Connecticut-Alternate-Assessments.

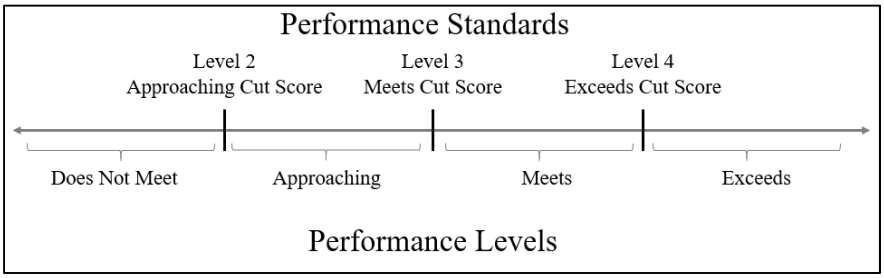
The Connecticut State Department of Education (CSDE) and its assessment vendor, the American Institutes for Research (AIR) developed and field tested the CTAS in the 2017–2018 school year and administered the first operational assessment in the 2018–2019 school year. New tests require new performance standards to link performance on the test to the content standards. Establishing performance standards on the new tests uses student responses from the first live administration combined with educator expertise via a process known as *standard setting*. The CSDE and AIR implemented a defensible, valid, and technically sound method; provided training on standard setting to all participants; oversaw the process; computed real-time feedback data to inform the process; and produced a technical report documenting the method, approach, process, and outcomes.

3. Standard Setting

Twenty-five out of thirty recruited educators from Connecticut (approximately ten for each grade-level test) convened at the Red Lion in Cromwell, Connecticut, from July 29 through July 30, 2019, to complete two rounds of standard setting and recommend three performance standards for the CTAS.

Standard setting is the process used to define achievement on the CTAS. Performance standards, or cut scores, define how many of the content standards students must know and be able to do to meet the minimum for each performance level. As shown in Figure 3, three performance standards define Connecticut's four performance levels.

Figure 3. Three Performance Standards Defining Connecticut's Four Performance Levels



The cut scores are derived from the knowledge and skills measured by the test items that students at each performance level are expected to be able to answer correctly.

3.1. The Bookmark Method

The Bookmark method of standard setting is appropriate for setting cut scores on tests, like the CTAS, that are scored using item response theory (IRT) and that use mixed-type items. This approach is appropriate for these types of tests and simplifies the decision process for panelists by allowing them to make the same judgment task for all items, regardless of item type. Because the Bookmark method relies on judgments made by experts, the panelists and stakeholders report high confidence in the outcomes. It has proven to be technically sound in litigation, and more than 30 states have selected and implemented it, making it the most frequently used method of setting performance standards on high-stakes state accountability assessments (Karantonis and Sireci [2006]; Lewis, Mitzel, Mercado, and Schulz [2012]; Mitzel, Lewis, Patz, and Greene [2001]). For these reasons, the CSDE chose to apply the Bookmark standard-setting method to establish new performance standards.

The Bookmark method derives its name from the primary task required of panelists: the placement of a bookmark in an OIB to represent a cut score recommendation. Over multiple rounds of judgments, panelists consider feedback, reference, and benchmark data provided for each round to recommended content-based cut scores using the policy descriptors for content, target student performance descriptors, test content viewed in the OIBs, and panelist discussions.

3.2. Workshop Structure

One large meeting room served as an all-participant training and meeting room. Each grade level was designed to have two tables with one table leader and four panelists at each table. Of the 30 panelists the CSDE recruited, only 25 panelists were able to participate. The recruited panelists who were unable to participate informed the CSDE on the day of the meeting, citing personal or other last-minute scheduling conflicts. Panelists sat at separate tables based on their grade and table assignment, as shown in Table 2.

Table 2. Room and Panel Structure

Panel	Table	Table Leader Panelists	Panelists	Facilitator	Facilitator Assistants
Grade 5	1	1	3	Peter Pluckebaum	Vanessa Brayman and Kevin Cleary
	2	1	3		
Grade 8	1	1	3		
	2	1	4		
Grade 11	1	1	2		
	2	1	4		

The standard-setting participants brought a variety of experience and expertise in the content areas for which standards were being set.

3.3. Participants and Roles

3.3.1. Connecticut State Department of Education Staff

Staff from the CSDE were present throughout the process and provided overall leadership and policy background and answered any policy-related questions that arose. They included the following staff:

- Abe Krisst, Performance Office, Bureau Chief
- Janet Stuck, Special Populations
- Jeff Greig, Science
- Michelle Rosado, Connecticut SAT School Day
- Pei-Hsuan Chiu, Psychometrics Team
- Mohamed Dirir, Psychometrics Team
- Michael Sabados, Data Team

3.3.2. AIR Staff

AIR facilitated the workshop and each of the content-area rooms, provided psychometric and statistical support, and oversaw technical setup and logistics. AIR team members included the following:

- Dr. Gary Phillips, AIR Vice President and Institute Fellow, facilitated the workshop. Dr. Phillips provided training to all participants, including the facilitators, the table leaders, and all participants; supervised the psychometric analyses conducted during and after the workshop; and presented impact and benchmark data to panelists after each round.
- Jennifer Chou, Program Director, oversaw the project and managed processes and logistics throughout the meeting.
- Dr. Tzu-Chun Kuo, Senior Psychometrician, provided additional psychometric support.
- Nicholas Kalich, Psychometric Support Manager, oversaw analytics technology and psychometrics; Alana Hutchinson provided support.
- Drew Azar and Dotun Adebayo set up, tested, and troubleshot technology during the workshop.

An AIR room facilitator and two assistant facilitators guided the process for all three grades. Facilitators had science content expertise and experience in leading standard-setting processes and could answer any questions about the process, the tests, or what the performance tasks measure. They also monitored time

and motivated panelists to complete tasks within the scheduled time. Peter Pluckebaum facilitated the room and Vanessa Brayman and Kevin Cleary provided facilitation support. All facilitators and assistant facilitators participated in a full-day process training and a separate technology training before each workshop. This training covered six important functions:

1. Operating and following the steps in the online standard-setting tool
2. Taking the online assessment
3. Placing bookmarks online
4. Practicing leading discussions and getting feedback on information from rounds 1 and 2
5. Reviewing all workshop materials
6. Conducting an online evaluation

3.3.3. Educator Panelists

To establish performance standards, the CSDE recruited a diverse set of participants from across the state. Panelists included educators with various areas of expertise in science, special education, or both including students with significant cognitive disabilities across grade levels. All participants needed to be familiar with the NGSS content as well as with the CTAS.

Table 3 summarizes the characteristics of each grade-level panel. Panelists included special education and general education teachers, specialists, and representatives from other stakeholder groups (e.g., parents, instructional coaches) to ensure that a diverse range of perspectives contributed to the standard-setting process and product. In recruiting panelists, the CSDE targeted the recruitment of participants to be representative of the geographic representation of the teacher population found in Connecticut. The CSDE also recruited panelists to bring a varied set of specific skills and expertise (e.g., experience with specific disabilities, grade levels, or subject matter). Panelists held both school and district positions and represented a range of district sizes and urbanicity (rural, suburban, and urban).

Table 3. Panelist Characteristics

	Percentage of Panelists by Panel		
	Grade 5	Grade 8	Grade 11
Characteristics			
Male	0%	25%	25%
Non-White	11%	0%	38%
Stakeholder Group*			
Teacher	44%	75%	50%
Science Specialist	11%	0%	13%
Administrator	0%	0%	0%
Instructional Coach	22%	0%	0%
Special Education Teacher	33%	13%	38%
ESL Teacher	0%	0%	0%
Higher Education	0%	0%	0%
Other	33%	38%	13%
Current Position			
School	67%	88%	88%
District	11%	0%	0%
Other	22%	13%	13%
District Size			
Large	44%	50%	50%
Medium	22%	13%	25%
Small	33%	38%	25%
District Urbanicity			
Urban	56%	38%	50%
Suburban	33%	63%	25%
Rural	11%	0%	25%
Number of Schools Represented	8	7	7
Number of Counties Represented (6 in CT)	4	4	6
Primary Grades Taught			
ES (grades 1–5)	33%	0%	0%
MS (grades 6–8)	0%	38%	0%
HS (grades 9–12)	0%	0%	63%
ES and MS	33%	25%	0%
MS and HS	0%	13%	13%
ES, MS, and HS	22%	13%	13%
N/A (non-educators)	11%	13%	13%
Subjects Taught			
Science	78%	88%	75%
Other (including N/A)	22%	13%	25%

* Each panelist could have multiple roles in this group

Table 4 summarizes the qualifications of each panel. Panelists were well educated, with the majority holding at least a master's degree. They were also highly experienced, with most having more than 10 years' experience in the classroom teaching students and many having additional professional experience outside the classroom. Nearly all had experience working with diverse student groups, including students from low socioeconomic backgrounds, English language learners (ELL), and students with disabilities. Appendix A provides additional information about the individuals participating in the standard-setting workshop.

Table 4. Panelist Qualifications

	Percentage of Panelists by Grade		
	Grade 5	Grade 8	Grade 11
Highest Degree			
Bachelor's	11%	13%	0%
Master's	44%	50%	63%
Sixth-year degree	33%	13%	25%
Doctorate	11%	25%	13%
Years of Teaching Experience			
0 years	11%	0%	0%
1–5 years	22%	0%	0%
6–10 years	11%	25%	13%
11–15 years	22%	13%	13%
16–20 years	11%	50%	25%
21+ years	22%	13%	50%
Years of Teaching Experience in Assigned Grade/Subject			
0 years	11%	0%	13%
1–5 years	44%	0%	13%
6–10 years	22%	38%	38%
11–15 years	0%	25%	13%
16–20 years	0%	25%	13%
21+ years	22%	13%	13%
Years of Professional Experience in Education			
0 years	33%	25%	63%
1–5 years	33%	0%	0%
6–10 years	0%	25%	13%
11–15 years	22%	13%	13%
16–20 years	0%	13%	13%
21+ years	11%	25%	0%
Experience Teaching Special Student Populations			
Students receiving free/reduced price lunch	89%	100%	88%
English language learners	100%	100%	100%
Students on an IEP	100%	100%	100%

Note: "Years of professional experience in education" refers to experience outside the classroom, such as being an administration or specialist.

3.3.4. Educator Table Leaders

For each table in a panel, the CSDE pre-selected table leaders from the participant pool for their leadership or specialized knowledge or experience working with the assessment or the alternate standards. All were familiar with and knowledgeable of the assessment and the alternate standards because they had previously served on the committee that developed the CTAS. Table leaders also served as panelists, and as such participated in the recommendation of the cut scores.

As with room facilitators, it was necessary to ensure that each table leader was knowledgeable of the constructs, processes, and technologies used in standard setting and able to adhere to a standardized process across the grade/subject committees.

Table leaders trained as a group early in the morning of the first day. Abe Krisst from the CSDE welcomed table leaders and introduced the workshop. Dr. Gary Phillips from AIR provided the training. Training consisted of an overview of their responsibilities and some process guidance.

Table leaders were given 45 minutes of initial training before the full group of participants convened on the first day of the workshop. Table leader training focused on the table leaders' unique roles and responsibilities. Following this, whole group training was conducted, and all participants in the workshop received training on all the crucial standard-setting concepts and procedures they would be using throughout the workshop.

Table leaders fulfilled the following functions throughout the workshop:

- Helping panelists see the big picture
- Leading table discussions
- Supporting panelists with their multiple tasks
- Monitoring security of materials
- Reporting issues or misunderstandings to room facilitators
- Maintaining a supportive atmosphere of professionalism and respect

3.4. Materials

3.4.1. Alternate Content Standards (Essence Statements) and Core Extensions

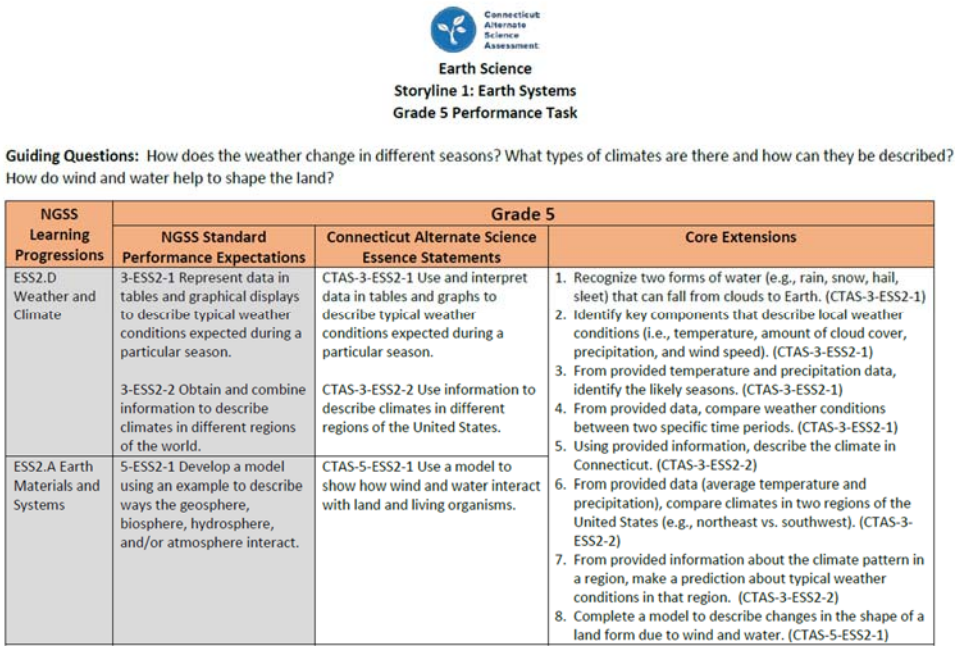
Connecticut's Alternate Content Standards describe what students with the most significant cognitive disabilities should know and be able to do. They identify clear expectations for students, parents, and teachers and improve teaching and learning.

To create the standards, Connecticut educators extracted Performance Expectations from the NGSS Learning Progressions that were appropriate for Connecticut students with significant cognitive disabilities. They simplified some of the more complex Performance Expectations into Essence Statements and then broke the Essence Statements into Core Extensions that describe specific student performances that would demonstrate the knowledge and skills described in each Essence Statement. The Essence Statements are uniquely numbered and comprise Connecticut's Alternate Content Standards. The Core Extensions break each into specific tasks that demonstrate the knowledge and skills required by the

Essence Statements and are the basis for the performance tasks. Some Essence Statements include more than one Core Extension.

Figure 4 shows the relationship between elements of the NGSS and the CTAS Essence Statements and Core Extensions.

Figure 4. Relationship between NGSS and CTAS Structure



Earth Science
Storyline 1: Earth Systems
Grade 5 Performance Task

Guiding Questions: How does the weather change in different seasons? What types of climates are there and how can they be described? How do wind and water help to shape the land?

NGSS Learning Progressions	Grade 5		
	NGSS Standard Performance Expectations	Connecticut Alternate Science Essence Statements	Core Extensions
ESS2.D Weather and Climate	3-ESS2-1 Represent data in tables and graphical displays to describe typical weather conditions expected during a particular season. 3-ESS2-2 Obtain and combine information to describe climates in different regions of the world.	CTAS-3-ESS2-1 Use and interpret data in tables and graphs to describe typical weather conditions expected during a particular season. CTAS-3-ESS2-2 Use information to describe climates in different regions of the United States.	1. Recognize two forms of water (e.g., rain, snow, hail, sleet) that can fall from clouds to Earth. (CTAS-3-ESS2-1) 2. Identify key components that describe local weather conditions (i.e., temperature, amount of cloud cover, precipitation, and wind speed). (CTAS-3-ESS2-1) 3. From provided temperature and precipitation data, identify the likely seasons. (CTAS-3-ESS2-1) 4. From provided data, compare weather conditions between two specific time periods. (CTAS-3-ESS2-1) 5. Using provided information, describe the climate in Connecticut. (CTAS-3-ESS2-2) 6. From provided data (average temperature and precipitation), compare climates in two regions of the United States (e.g., northeast vs. southwest). (CTAS-3-ESS2-2) 7. From provided information about the climate pattern in a region, make a prediction about typical weather conditions in that region. (CTAS-3-ESS2-2) 8. Complete a model to describe changes in the shape of a land form due to wind and water. (CTAS-5-ESS2-1)
ESS2.A Earth Materials and Systems	5-ESS2-1 Develop a model using an example to describe ways the geosphere, biosphere, hydrosphere, and/or atmosphere interact.	CTAS-5-ESS2-1 Use a model to show how wind and water interact with land and living organisms.	

3.4.2. CTAS Performance Tasks

Performance tasks measure the Alternate Content Standards. To create the tasks, educators created Storylines capturing the NGSS Performance Expectations, Essence Statements, and Core Extensions within a specific content area. Together, there are six storylines for each grade level (see Table 5).

Table 5. CTAS Storylines

Connecticut Alternate Science Assessment (CTAS)	
Content Area	Storyline (Performance Task)
Earth Science	Earth Systems
	Natural Resources
Life Science	Living Organisms
	Health Ecosystems
Physical Science	Forces and Motion
	Using Energy Every Day

Accompanying each storyline is a guiding question that introduces the topic and identifies the Core Extensions that the performance task assesses. One performance task assesses each storyline, and each performance task consists of multiple activities. Each activity includes a script for instructing students during the activity and another for asking test questions, teacher notes, instructions for scaffolding, and guidelines for rating and recording student responses. Resource packets are specific to each performance task and include materials such as posters, graphs, and sentence strips. A test includes a performance task (comprised of multiple activities) from each storyline.

Teachers score each activity on a 0–2 scale, where "0" indicates the student does not demonstrate understanding, "1" indicates the student demonstrates limited understanding typically requiring additional support through scaffolding, and "2" indicates the student demonstrates understanding independently without scaffolding.

The CSDE makes the performance tasks and associated materials available at: https://ct.portal.airast.org/ctas-required-materials/.

3.4.3. Ordered-Item Booklets

The Bookmark method uses OIBs as the key tool for setting standards. Each OIB contains all performance task activities from the 2019 operational CTAS test. Activities in the OIB are the same as those on actual student tests. Because activities can be worth up to two score points, they occur twice in the OIB, with one page for each possible point. OIBs are ordered by difficulty, so easier activities are in the front of the OIB and more difficult activities are in the back of the OIB. The OIBs consist of 84–88 pages each.

The OIB presents items (activities ratings) sorted in ascending order by IRT item (rating) difficulty, calculated by Winsteps and indicated by a response probability of .50 (RP50). For a one-point activity, RP50 is the item difficulty point where 50% of students correctly performed the activity. For two-point items, two RP50 values represent the difficulty level where 50% of the students earned each of the two score points.

Panelists place electronic bookmarks where 50% of students would be able to earn the score point of the item. Each page of the OIB can correspond to a cut score; thus, when panelists place their bookmark for a performance level, they are, in fact, selecting the performance standard, indicated by the RP50 value of the item, for that performance level.

3.4.4. Performance-Level Descriptors

With the adoption of new standards and the development of new assessments to assess achievement of those standards, the CSDE must also establish a new system of performance standards to determine

whether students have met the learning goals defined by the content standards. Determining the categories in which to classify students is a prerequisite to standard setting.

These categories, or performance levels, are associated with PLDs, that define the content area knowledge, skills, and processes that students at each level can demonstrate. Two types of PLDs link the content standards to the performance standards.

1. **Range PLDs.** Provided to panelists to review during the workshop, these detailed grade- and content-area-specific descriptions communicate exactly what students performing throughout the range of each performance level know and can do.
2. **Just Barely PLDs.** Created during and used for standard setting only, these describe what a student just barely scoring at the bottom of each performance level knows and can do. Sometimes these are also called "Target PLDs."

Connecticut uses four performance levels to describe student performance: "Level 1: Does Not Meet the Alternate Achievement Standard," "Level 2: Approaching the Alternate Achievement Standard," "Level 3: Meets the Alternate Achievement Standard," and "Level 4: Exceeds the Alternate Achievement Standard."

3.5. Workshop Technology

Panelists used AIR's online application for standard setting. From this application, panelists placed multiple rounds of bookmarks, reviewed the content alignment and score points for each item, and evaluated the impact that proposed cuts will have on students. Panelists also saw their own bookmarks, their table's bookmarks, the other tables' bookmarks, and the overall bookmarks for all tables. They could add notes and comments on the items as they reviewed each item. Impact data were also presented for each item onscreen after being introduced at round 2.

Each panelist used an AIR laptop or Chromebook to review items and place bookmarks. The panelists experienced the test in its paper format along with the ancillary materials that accompany the assessment. The laptops were loaded with the AIR online standard-setting tool that has been used for multiple state standard-setting activities.

Two full-time AIR IT specialists oversaw laptop setup and quality control testing, answered questions, and ensured that technological processes ran smoothly and without interruption throughout the meeting. No technological issues arose during the workshop that in anyway disrupted the smooth operation of the standard-setting task.

3.6. Workshop Events

The standard-setting workshop occurred over two days. Table 6 summarizes each day's events, and this section describes each event listed in greater detail. Appendix B**: Workshop Agenda** provides the full workshop agenda.

Table 6. Standard Setting Agenda Summary

Day 1: Monday, July 29, 2019
• Table leader training • Large-group orientation and training • Test administration • Performance task, Essence Statement, Core Extensions, and resource review • Range PLD review • Discussion of skills at each performance level and creation of "Just Barely" PLDs • OIB review
Day 2: Tuesday, July 30, 2019
• OIB review (continued) • Bookmarking and RP50 training • Bookmark placement practice • Standard-setting quiz • Readiness assertion • Round 1 bookmark placement, feedback, impact data, benchmark data, and articulation: review and discussion • Round 2 bookmark placement, feedback, impact data, benchmark data, and articulation: review and discussion across grades • Workshop evaluation

3.6.1. Table Leader Training

Table leaders trained as a group early in the morning of the first day to ensure that each leader was knowledgeable of the constructs, processes, and technologies used in standard setting and was able to adhere to a standardized process across the grade and subject panels.

Table leaders provided the following support throughout the workshop; they

- helped panelists see the big picture;
- led table discussions;
- supported panelists with tasks;
- monitored security of materials;
- monitored panelist understanding and reported issues or misunderstandings to room facilitators; and
- maintained a supportive atmosphere of professionalism and respect.

Training consisted of an overview of their responsibilities and some process guidance. Table leaders also served as panelists and set individual cut scores.

3.6.2. Large-Group Orientation

Abe Krisst from the CSDE welcomed panelists to the workshop and provided context and background for the workshop. He reviewed the development of the CTAS and explained the need for standard setting. He outlined the roles and responsibilities of the three groups at the workshop: panelists, AIR staff, and the CSDE personnel.

Next, Janet Stuck from the CSDE reviewed the CTAS including eligibility and individualized education program (IEP) planning and described characteristics of Connecticut's special education student population. She summarized the development process for the CTAS, highlighting the contribution of Connecticut educators, and described the structure of the alternate standards and assessments.

Dr. Gary Phillips then oriented participants to the workshop by explaining the process that would unfold over the next two days and outlining the events that would take place each day. He explained that the CSDE selected panelists because they were experts and how, by design, the process elicits and applies their expertise to recommend new cut scores. Finally, he described how standard setting works and what would happen once the panelists had finalized their recommendations. Appendix C provides the large-group training presentation.

3.6.3. Confidentiality and Security

Workshop leaders and room facilitators addressed confidentiality and security during orientation and again in each room. Standard setting uses live test items from the operational CTAS tests. Although the CTAS performance tasks are not secure, panelists adhered to security and confidentiality requirements to maintain their integrity. Workshop leaders asked participants not to do any of the following during or after the workshop:

- Discuss the test items outside of the meeting
- Remove any materials from the room on breaks or at the end of the day
- Discuss judgments or cut scores with anyone outside the meeting
- Discuss items with non-participants
- Use cell phones in the meeting rooms
- Take notes on anything other than provided materials
- Bring any other materials to the workshop

Participants could have general conversations about the process and days' events, but workshop leaders warned them against discussing details, particularly those involving test items, cut scores, and any other confidential information.

3.6.4. Workshop Technology

The panelists used AIR's online application for standard setting. Each panelist used an AIR laptop or Chromebook on which he or she took the test; reviewed test questions, OIBs, and ancillary materials; and recommended cut scores for each performance level. Through this application, panelists could review each activity in the OIB, examine the content alignment and score points for each item, and evaluate the percentage of students who would fall into each performance level given their proposed cuts. Panelists also saw their own bookmarks, their table's bookmarks, and the overall bookmarks for both tables. They could add notes and comments on the items as they reviewed them and examine impact and benchmark data onscreen following each round. On the last day of the workshop, panelists completed online evaluation forms.

Two full-time AIR IT specialists oversaw laptop setup and testing, answered questions, and ensured that technological processes ran smoothly and without interruption throughout the meeting.

3.6.5. Experience the Test

The first standard-setting task was for panelists to see the test and performance tasks and activities on which they would be setting standards. Reviewing the tests provides the opportunity to interact with and become familiar with the items and test administration. Doing so also allowed panelists to see the connection between Essence Statements and Core Extensions and to understand how the embedded scaffolding provides additional access and support for students who need it. Panelists reviewed a form of the test that students took in 2019 in the grade level for which they would be setting performance standards.

3.6.6. Review Alternate Content Standards and Range PLDs

After reviewing the test, panelists reviewed the Essence Statements, Core Extensions, and Resources for their assigned grade. Tables discussed separately at first and then joined for an all-grade discussion. Reviewing and discussing these materials ensured that participants understood the expectations for what students in Connecticut should know and be able to do and how much knowledge and skill students should be able to demonstrate at each level of performance.

3.6.7. Draft "Just Barely" Student Descriptions

After reviewing and discussing the alternate content standards, performance tasks and activities, and range PLDs, panelists worked in their table groups to draft target PLDs that described the skills that students scoring "Just Barely" in one performance level have but that students scoring just below the performance level do not have.

Each Essence Statement is associated with a set of range PLDs describing what performance looks like for each of the four performance levels. Looking at each Essence Statement, panelists identified the skills needed to just barely perform at the "Meets" level and noted this in a worksheet. The point of this exercise is to think of the student who is just barely at each of the levels and how he or she differs from the student who is well into each level.

Target PLDs describe students who are not typically at a performance level, although, at "Just Barely," they do reach the standard. Panelists, working across tables, drafted descriptions for "Meets Standard," "Exceeds Standard," and "Emerging."

3.6.8. Review OIBs

After completing the "Just Barely" PLDs (target PLDs), the panelists reviewed the OIB. The facilitators explained that the objective of standard setting is aspirational, to identify what all students should know and be able to do and not to describe what they currently know and can do. To accomplish this, as panelists review the items in the OIB, they think about the target PLDs that describe students "Just Barely" meeting each performance level. "Just Barely" students are more likely to be able to answer items at the beginning of the OIB correctly and less likely to be able to answer items towards the end of the OIB correctly. For each item, the panelists think of what students need to know and be able to do to answer each item correctly and what makes each item more difficult than the preceding item. They could note these characteristics for each item for reference as they placed their bookmarks later.

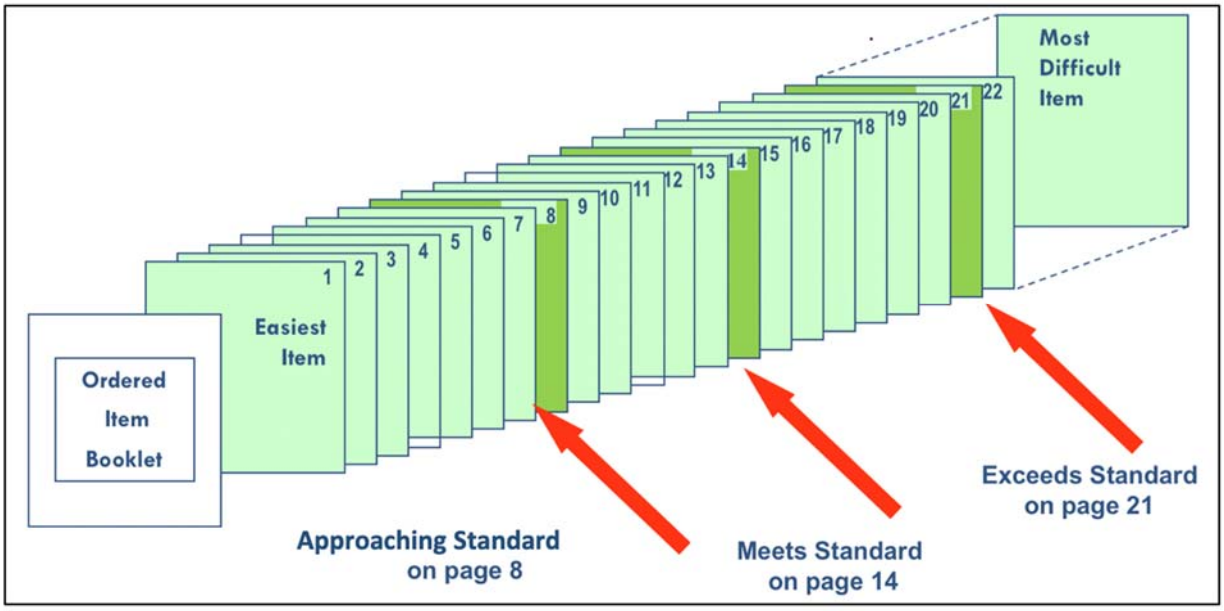
The facilitators advised panelists that while some activities may seem out of order, the order is determined by difficulty, which represents actual student performance on the activity, not content or cognitive processes. The ordering of activities in the OIB does not follow the sequence of instruction or the order of presentation on the test. To keep panelists focused on the standard-setting task and not on item critique, panelists could refer item-related questions or comments to workshop facilitators and the CSDE staff to investigate. Panelists independently reviewed and annotated each score point for all activities in the online OIB.

3.6.9. Bookmark Placement Training

Panelists placed the bookmark at the point where students scoring at and above that level are described by the PLD. They applied a 50/50 response probability rule when placing bookmarks. This rule required panelists to identify the page in the OIB at which 50% of students who "Just Barely" meet the standard (those at the lowest end of the target PLDs) should be able to correctly complete the activity. As panelists work through the OIB, they come across an activity, or small group of activities, that they think about half of the "Just Barely Meets Standard" students (for example) would answer correctly. Activities before that point in the OIB are those that more than half of the "Just Barely Meets Standard" students would correctly complete. Activities beyond that point in the OIB are those that less than half of the "Just Barely" students would answer correctly.

As they progress through the OIB, for each activity and score-point, panelists ask "Out of 100 students who are 'just barely' at the Meets level, what percent would likely get this Activity Rating?" As shown in Figure 5, panelists placed their bookmarks on the first page in the OIB where they believe the "Just Barely Meets" student would not have at least a 50% chance of answering correctly. Panelists repeated this process for the "Just Barely" Approaching student and the "Just Barely" Exceeds student.

Figure 5. Example Bookmark Placement



Panelists practiced placing bookmarks in the OIB. The practice round ensures panelist comfort with the technology, performance tasks, and bookmark placement procedures prior to determining any consequential cut scores. Panelists asked questions, and the room facilitators provided clarifications and further instructions until everyone had successfully completed the practice round.

Panelists were not to place bookmarks on any activity or rating (score-point) that they disagreed with or felt might be incorrect or unfair. Finally, panelists were not to set standards for individual students who they knew, or for students in their classrooms, but to set performance standards for all students across the state.

3.6.10. Bookmarking Quiz

Following the practice round, panelists completed a short quiz assessing their understanding of the bookmark-setting process. The quiz assessed panelists' understanding of the standard-setting task in multiple ways. For example, they must be able to

- answer questions about the bookmarking process and online application;
- identify more and less difficult items and demonstrate placing bookmarks in the OIB;
- describe where "Just Barely" students fall on a performance scale; and
- indicate on a diagram how performance standards define performance levels.

Room facilitators reviewed the quizzes with the panelists and provided additional training for any incorrect responses on the quiz. Appendix D**: Bookmark Placement Quiz** provides the quiz panelists completed.

3.6.11. Readiness Assertions

After completing the practice round and standard-setting quiz, and prior to placing the round 1 bookmarks, panelists completed a readiness assertion form. On this form, panelists asserted that trained sufficiently prepared them to understand the following concepts and tasks:

- The concept of a student who just barely meets the criteria described in the PLDs;
- The structure, use, and importance of the OIB; and
- The process to determine and place bookmarks in the standard setting tool.

The readiness form for round 2 focused on affirming understanding of the impact and benchmark data supplied after round 1. On this form, panelists affirmed the meeting training had fully prepared them for the following:

- Understanding the impact and benchmark data;
- Understanding the round 2 task; and
- Readiness to complete the round 2 task.

Room facilitators reviewed the assertions and were prepared to provide additional training to anyone not confident in their readiness. Following additional training, these panelists would reaffirm their readiness.

However, every panelist affirmed readiness before beginning to bookmark in both rounds of the workshop. Appendix E**: Readiness Form** provides the form panelists completed.

3.6.12. Bookmark Placement and Feedback

Panelists independently placed each of three cut scores demarcating the four performance levels using the content standards, the PLDs, their notes from reviewing the OIB, and their knowledge and experience with students. They first determined the "Meets" standard, then the "Approaching" standard, and finally the "Exceeds" standard.

AIR psychometricians then computed and summarized new data for consideration in round 2 based on the round 1 cut scores. This new information included feedback data, impact data, benchmark data, and articulation.

3.6.13. Feedback Data

Feedback for each round included the scaled scores corresponding to the bookmarks placed by each panelist and the median bookmarks placed by each table and for the room overall (across both tables). This information allows panelists to compare their marks to other panelists' marks to see how their expectations compare.

3.6.14. Impact Data

Applying the round 1 scaled scores to student data from the 2019 administration of the CTAS provided impact data. Impact data describe the projected percentage of Connecticut students who would fall into each of the performance levels, given the proposed cut scores. This information provides panelists with an idea of how the proposed cut scores will impact students and teachers throughout the state.

3.6.15. Benchmark Data

Benchmarking provides panelists with an external reference so that they can see how their recommendations compare with the standards on other similar assessments or with similar populations. For Connecticut, benchmark data describes the percentage at or above the Approaches and Meets levels using data from the 2019 Mathematics and ELA Connecticut Alternate Assessment (CTAA). To make it easier for panelists to apply the benchmark data, AIR psychometricians mapped these percentages onto the page numbers in the science OIBs corresponding to the same percentages. Comparing the results of round 1 on the CTAS against the CTAA results, panelists could see how the proposed standards for the new alternate science assessment compare to those for the existing alternate mathematics and ELA assessments and judge the reasonableness and rigor of the proposed performance standards for the new CTAS.

3.6.16. Articulation

Performance standards for a statewide system must be coherent across grades and subjects. No irregular peaks and valleys should appear, and standards should be similar across subjects and grades with no dramatic differences in expectations. AIR psychometricians described the need for articulated performance standards and presented example cut scores based on the average percentage of students approaching and meeting the standard on the mathematics and ELA CTAA. This information shows panelists how they could maximize articulation in Performance Expectations across grade and subject, ensuring similar expectations for students with disabilities on the math, ELA, and science alternate assessments.

After reviewing this new information, workshop facilitators provided panelists with additional instruction for completing round 2. First, they described the goal of round 2 as one of convergence, but not consensus, on a common performance standard. A second goal was articulation across grade levels.

Round 2 bookmark placement began with panelist discussion of this new information, beginning at each table and then progressing across tables. Each table spent time reviewing and discussing feedback data, impact data, benchmark data, and articulation data and considered this information in placing round 2 bookmarks. After completing these discussions, panelists again worked independently through the OIB, placing their round 2 cut scores for all three performance levels. They again determined the "Meets" standard first, followed by the "Approaching" standard, and then the "Exceeds" standard.

3.7. Workshop Results

3.7.1. Round 1

The AIR online standard setting tool automatically computes and verifies the results and impact data for each round. AIR psychometricians conduct an additional round of QC and then present the round 1 results for each grade.

For the grade-level median OIB pages, the facilitator showed the percentage of students who would fall into each performance level for each grade and pointed out any inconsistencies across grades. They explained that approximately 30% of students meet the standard on the benchmark assessment in mathematics and ELA and that dramatically lower (or higher) percentages of students meeting the standard on the CTAS would mean that students with disabilities were held to higher (or lower) standards in science than they were in mathematics or ELA. They provided an example of articulated standards to panelists to consider but emphasized that the decision was theirs and should ultimately be based on content. The psychometricians also stressed that panelists could consider articulation, impact data, and benchmark data as they deemed them appropriate. The CSDE repeatedly communicated that they would

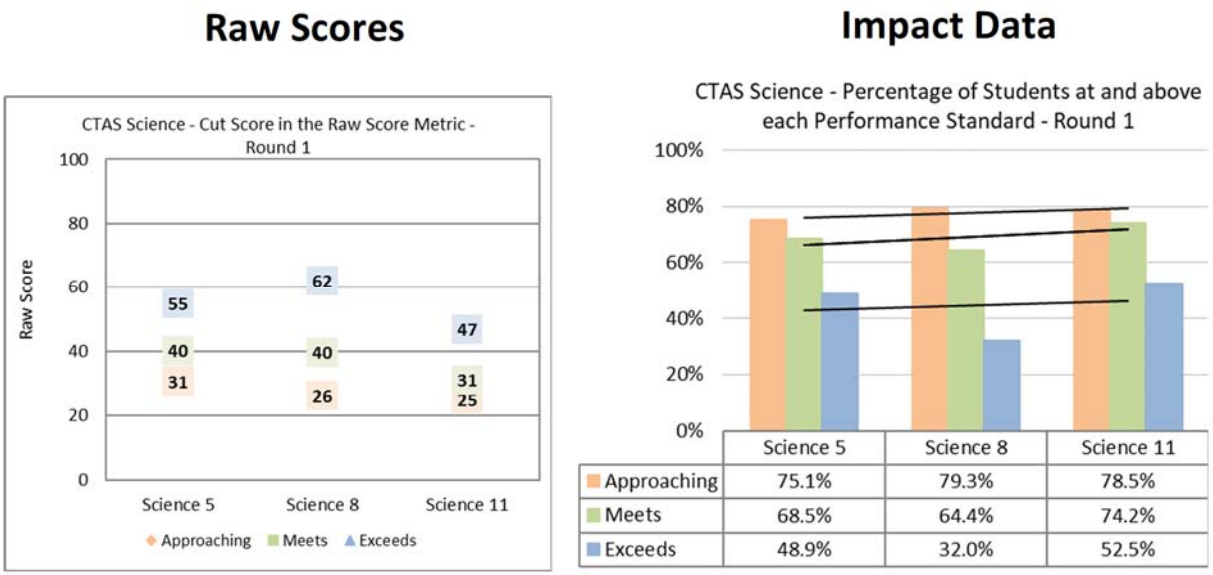
advocate for a more (or less) rigorous test for students with disabilities than for general education students, if that was the recommendation of the panel. Table 7 presents the round 1 median recommended page number for each grade, associated impact data, and benchmark data. Figure 6 displays the cut scores and impact data from round 1.

Table 7. Round 1 Results

Grade	Median Round 1 Bookmark (Page No.)			Impact Data			Benchmark Data	
	A	M	E	A	M	E	M (ELA)	M (Math)
5	23	38	59	75.1%	68.5%	48.9%	35.9%	38.1%
Table 1	25	50	69	74.4%	62.6%	32.9%		
Table 2	17	28	56	81.4%	73.2%	56.0%		
8	25	41	68	79.3%	64.4%	32.0%	28.3%	48.5%
Table 1	26	44	66	79.3%	57.6%	32.0%		
Table 2	16	37	69	83.2%	66.5%	30.7%		
11	20	31	49	78.5%	74.2%	52.5%	41.9%	36.3%
Table 1	18	30	49	79.1%	74.2%	52.5%		
Table 2	20	31	56	78.5%	74.2%	40.2%		

Note: The "Grade" row summarizes the room data (the median across both tables). Benchmark data describe the percentage at or above the proficient performance level on the general education tests. Performance-Level Abbreviation Key: A = Approaching, M = Meets, E = Exceeds.

Figure 6. Round 1 Raw Scores and Impact Data



Round 1 performance standards were low, half as low as those for ELA and mathematics. Higher percentages of students would Meet or Exceed the Standard than expected. At grades 5 and 11, approximately half of students would Exceed the Standard and at grades 5 and across all grades, two-thirds or more would Meet the Standard. At grade 11, the Approaches Standard and Exceed Standard cut scores were quite close. Because the science standards are brand new content in the state, it may be unreasonable for such high percentages of students to already meet the standard. Contributing factors for the low initial standards may include the following:

- Round 1 standards frequently differ from final standards because they are based solely on content judgments. Once educators consider feedback, impact, benchmark, and articulation data in addition to content, results often naturally articulate.
- The test was an easier test than teachers anticipated. The items on the CTAS tended to be less difficult than students' ability. (There were more easy items on the test than students at the lower end of the ability distribution.)
- Anecdotal evidence suggests a possible psychological aversion to going deep into longer OIBs.
- Time spent understanding and defining "Just Barely" students may have predisposed panelists to initially set low standards.

Because the standards were lower than expected, AIR and the CSDE leadership briefed table leaders on the round 1 results before reviewing them with all panelists. Workshop leaders wanted table leaders to understand the new data and the implications of the cut scores so they could help panelists incorporate the information into their round 2 decisions. AIR and the CSDE staff reviewed the outcomes and consequences of round 1 and previewed the information (feedback data, impact data, benchmark data, and articulation data) that would factor into round 2 decisions. Table leaders reviewed and discussed the bookmarks and examined them relative to the CTAA benchmark data before returning to their panels.

Workshop leaders then presented the results of round 1 to all panelists and reviewed the feedback, impact, benchmark, and articulation data. Panelists discussed and then made their round 2 decisions.

3.7.2. Round 2

Round 2 performance standards were higher and well-articulated. Given the recommended cut scores, for all grades, between 37% and 46% of students would meet the recommended standard and between 26% and 28% would exceed the standard. Table 8 presents the round 2 median recommended page number for each grade, associated impact data, and benchmark data for round 2. Figure 7 presents the raw scores and impact data graphically. **Error! Reference source not found.** describes the percentage of students falling into each performance level.

Table 8. Round 2 Results

Grade	Median Round 2 Bookmark (Page No.)			Impact Data			Benchmark Data	
	A	M	E	A	M	E	M (ELA)	M (Math)
5	25	62	71	74.4%	46.2%	27.2%	35.9%	38.1%
Table 1	25	59	73	74.4%	48.9%	25.2%		
Table 2	25	62	71	74.4%	46.2%	27.2%		
8	26	61	72	79.3%	41.6%	28.1%	28.3%	48.5%
Table 1	25	59	72	79.3%	41.6%	28.1%		
Table 2	26	62	72	79.3%	36.7%	28.1%		
11	34	60	76	72.8%	37.1%	25.8%	41.9%	36.3%
Table 1	38	60	75	68.6%	37.1%	27.0%		
Table 2	30	59	77	74.2%	37.1%	23.7%		

Note: The "Grade" row summarizes the room data (the median across both tables). Benchmark data describe the percentage at or above the proficient performance level on the general education tests. Performance-Level Abbreviation Key: A = Approaching, M = Meets, E = Exceeds.

Figure 7. Round 2 Raw Scores and Impact Data

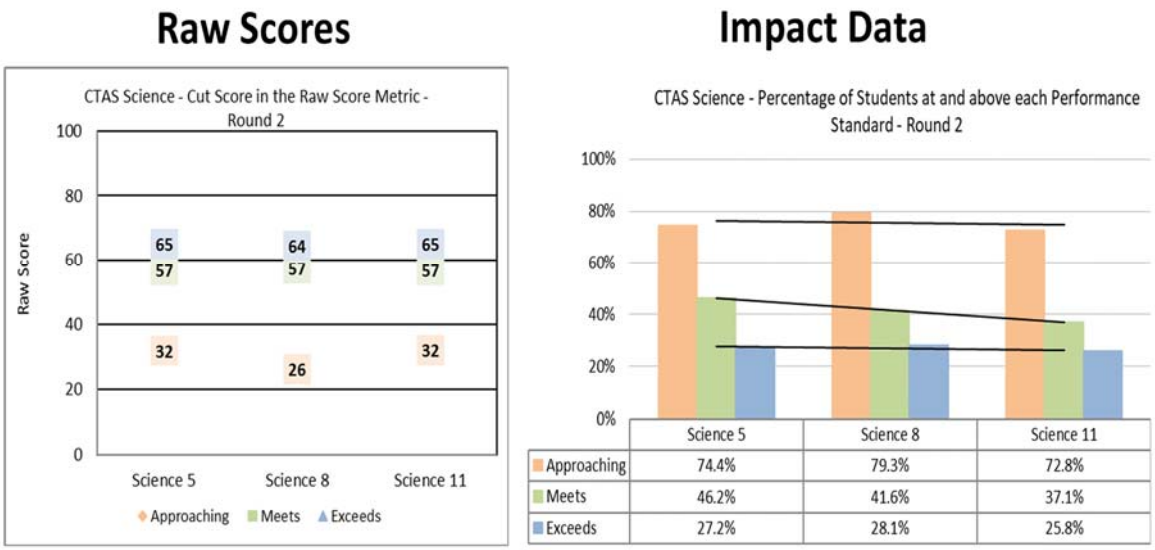
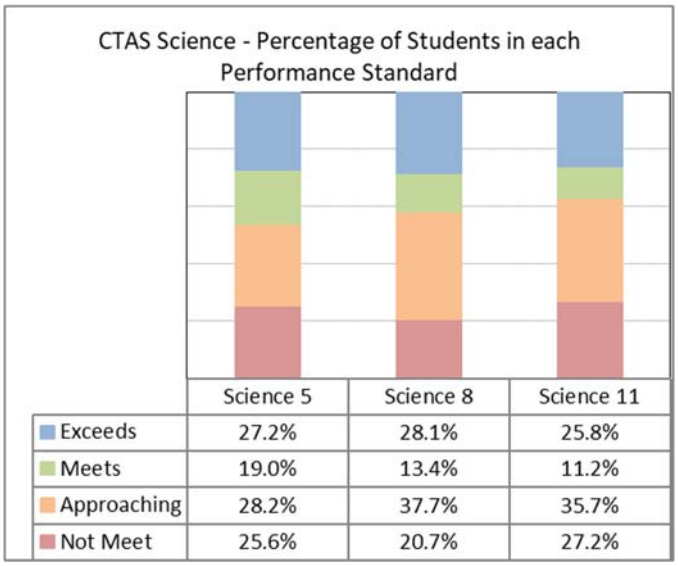


Figure 8. Percentage of Students Within Each Performance Standard



3.8. Workshop Evaluations

3.8.1. Evaluation Ratings

After finishing workshop activities, all panelists independently completed online meeting evaluations in which they described and evaluated their experience taking part in the standard setting. Table 9 through Table 13 summarize the results of the evaluations. Table 9 summarizes panelist ratings of clarity in instructions of materials and the standard-setting process. Panelists indicated that workshop materials and processes were clear, although some reported that the workshop instructions provided during orientation were not as clear as they could have been. The grade 8 panel found the PLDs and impact data to be less clear than did the grade 5 and 11 panels.

Table 9. Evaluation Results: Clarity of Materials and Process

Please rate the clarity of the following components of the workshop.	Percentage Responding "3" or "4"		
	Grade 5	Grade 8	Grade 11
Instructions provided by the Workshop Leader	78%	75%	100%
PLDs	100%	88%	100%
OIB	100%	100%	100%
Panelist agreement data	100%	100%	100%
Impact data (percentage of students who would achieve at the level indicated by the OIB page)	100%	88%	100%

Note: Number of responses = 24. Participants responded on a 4-point scale, where "1" = Not Clear" and "4" = Very Clear.

Table 10 summarizes panelist ratings of the appropriateness of the time allocated to the processes used to set standards. Most panelists indicated having enough time to complete processes, although the grade 8

panel indicated the least satisfaction with the timing allocated to discussing just barely students, reviewing the OIB, placing bookmarks, and discussing round 1 results.

Table 10. Evaluation Results: Appropriateness of Process

How appropriate was the amount of time you were given to complete the following components of the standard-setting process?	Percentage Responding "3" or "4"		
	Grade 5	Grade 8	Grade 11
Large group orientation	100%	100%	100%
Experiencing the online assessment	89%	100%	100%
Review of the performance-level descriptors (PLDs)	89%	88%	86%
Discussion of skills demonstrated by students who are "Just Barely" described by each PLD	100%	75%	86%
Review of the ordered-item booklet (OIB)	89%	63%	86%
Placement of your bookmarks in each round	100%	75%	100%
Round 1 discussion	100%	88%	100%

Note: Number of responses = 24. Participants responded on a 4-point scale, where 1 = "Too Little Time" and 4 = "Enough Time."

Table 11 summarizes how important panelists found various factors in making their bookmark placements. All panelists rated the PLDs, panel discussions, and impact data as important. A few panelists rated their perception of item difficulty, the external benchmark data, their experience with students, and feedback data as being less important to their judgements. Fewer grade 11 panelists indicated that their perceptions of item difficulty and experience with students were helpful in placing their bookmarks.

Table 11. Evaluation Results: Importance of Materials

How important were each of the following factors in your placement of your bookmarking decisions?	Percentage Responding "3" or "4"		
	Grade 5	Grade 8	Grade 11
Performance-level descriptors (PLDs)	100%	100%	100%
Your perception of the difficulty of the items	89%	88%	57%
Your experience with students	100%	88%	57%
Discussions with other panelists	100%	100%	100%
External benchmark data	89%	75%	100%
Room agreement data (room medians and individual bookmark placements)	89%	100%	100%
Impact data (percentage of students who would perform at the level indicated by the OIB page)	100%	100%	100%

Note: Number of responses = 24. Participants responded on a 4-point scale, where 1 = "Not Important" and 4 = "Very Important."

Panelists tended to agree with the statements shown in Table 12. They understood the purpose of the workshop, felt well trained for the task, appreciated taking the test, and found workshop materials helpful. Slightly fewer grade 8 panelists endorsed some statements than did other grade-level panelists.

Table 12. Evaluation Results: Understanding Processes and Tasks

At the end of the workshop, please rate your agreement with the following statements.	Percentage Responding "3" or "4"		
	Grade 5	Grade 8	Grade 11
I understood the purpose of this standard-setting workshop.	100%	100%	100%
The procedures used to recommend performance standards were fair and unbiased.	89%	88%	100%
The training provided me with the information I needed to recommend performance standards.	100%	100%	100%
Taking the online assessment helped me to better understand what students need to know and be able to do to answer each question.	100%	100%	100%
The performance-level descriptors (descriptions of what students within each performance level are expected to know and be able to do) provided a clear picture of expectations for student performance at each level.	100%	63%	100%
I was able to develop an understanding of the knowledge and skills demonstrated by students who are "Just Barely" described by the performance-level descriptors.	100%	75%	100%
I understood how to review each page in the OIB to determine what students must know and be able to do to answer each item correctly.	100%	100%	100%
I was able to interpret having approximately a 50% chance of answering an item correctly as indicating mastery.	100%	75%	71%
I understood how to place my bookmarks.	100%	100%	100%
I found the benchmark data and discussions helpful in my decisions about where to place my bookmarks.	100%	100%	100%
I found the panelist agreement data (room medians and individual bookmark placements) and discussion helpful in my decisions about where to place my bookmarks.	100%	100%	100%
I found the impact data (percentage of students who would achieve at the level indicated by the OIB page) and discussions helpful in my decisions about where to place my bookmarks.	100%	100%	100%
I felt comfortable expressing my opinions throughout the workshop.	100%	88%	100%
Everyone was given the opportunity to express his or her opinions throughout the workshop.	100%	100%	100%

Note: Number of responses = 24. Participants responded on a 4-point scale where 1 = "Strongly Disagree" and 4 = "Strongly Agree."

Finally, panelists indicated that students performing at each performance level meet the expectations for that level. Slightly fewer grade 8 panelists endorsed these statements than did other grade-level panelists (see Table 13).

Table 13. Evaluation Results: Student Expectations

Please read the following statement carefully and indicate your response.	Percentage Indicating "Yes"		
	Grade 5	Grade 8	Grade 11
A student performing at Level 3 meets the expectations for the grade level.	100%	88%	100%
A student performing at Level 2 is approaching expectations for the grade level.	100%	88%	100%
A student performing at Level 4 exceeds expectations for the grade level.	100%	88%	100%

Note: Number of responses = 24. Evaluation options included "Yes" and "No."

3.8.2. Participant Feedback

Finally, panelists responded to two open-ended questions: "What suggestions do you have to improve the training or standard-setting process?" and "Do you have any additional comments? Please be specific." Sixteen participants responded to the first question, and twelve responded to the second question.

Most suggestions concerned workshop event timing, such as allowing more (and less) time for PLD review and providing more time for the "Just Barely" conversations. One panelist suggested providing materials to panelists to read to prepare themselves prior to the workshop. Additional comments included having the facilitators project the same screen as the panelists during technology training (rather than training slides) and appreciation for being a part of the process.

Individual comments included the following:

"The experience of seeing through the creation of the test, implementation and assessment analysis was rewarding professionally in so many ways."

"Collaboration between SpEd and regular ed teachers is a HUGE benefit. We learned so much from each other. We shared classroom tips of course, and that was great, but the best part was the different approaches we brought to the table. All teachers should have this experience, regularly!"

"This was an invaluable process that I have learned many things from. The clear understanding that we were working on setting standards for the widely variable 1% was a key concept for the work."

REFERENCES

American Educational Research Association, American Psychological Association, National Council on Measurement in Education. (2014). Joint Committee on Standards for Educational and Psychological Testing. *Standards for educational and psychological testing*. Washington, DC: AERA.

Cizek, G. J., & Bunch, M. B. (2007). *Standard setting: A guide to establishing and evaluating performance standards on tests*. Thousand Oaks, CA: Sage.

Huynh, H. (2006), A Clarification on the Response Probability Criterion RP67 for Standard Settings Based on Bookmark and Item Mapping. *Educational Measurement: Issues and Practice 25*, 19–20.

Karantonis, A. & Sireci, S. (2006). The Bookmark Standard-Setting Method: A Literature Review. *Educational Measurement: Issues and Practice 25*. 4–12.

Lewis, D. M., Mitzel, H. C., Mercado, R. L., & Schulz, E. M. (2012). The bookmark standard setting procedure. In G. J. Cizek (Ed.), *Setting performance standards: Foundations, methods, and innovations* (2d ed.) (pp. 225–253). New York: Routledge.

Mitzel, H. C., Lewis, D. M., Patz, R. J., & Greene, D. R. (2001). "The Bookmark procedure: Psychological perspectives." In G. J. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives*. Mahwah, NJ: Erlbaum.

Appendix A: Standard-Setting Panelists

Panel	First Name	Last Name	Position	Level	District	Education	Years Teaching	Years Teaching in Assigned Grade	Gender	Ethnicity
5	Mae	Dalton	Special Education Teacher	School	Norwich	Master's	11–15 years	1–5 years	Female	Asian
5	Christina	Zucaro	Teacher	School	New Haven County	Master's	1–5 years	1–5 years	Female	White
5	Rebecca	Gaetano	Special Education Teacher; Early Interventionist B23	Families Homes	New Haven and Fairfield	Bachelor's	1–5 years	1–5 years	Female	White
5	Heidi	Gold	Teacher; Scientist, science curriculum developer	Little Scientists	New Haven	Doctorate	21+ years	21+ years	Female	White
5	Jennifer	Miller	Special Education Teacher	School	Hartford	Master's	11–15 years	0 years	Female	White
5	Tracey	Purcell	Instructional Coach	School	Hartford	Sixth-Year Degree	6–10 years	6–10 years	Female	White
5	Valerie	Saltzman	Teacher; Specialist	School	Fairfield	Master's; Sixth-Year Degree	21+ years	6–10 years	Female	White
5	Roseanne	Haughton	Instructional Coach; Elementary Science Curriculum Coordinator	District	Fairfield	Sixth-Year Professional Degree	16–20 years	1–5 years	Female	White
5	Maura	Graham-Vecellio	Teacher	School	New Haven	Master's	21+ years	21+ years	Female	White
8	Sarah	Seals	Special Education Teacher	School	Hartford	Master's	16–20 years	11–15 years	Female	White
8	Meghan	Pogonelski	Teacher	School	Fairfield	Doctorate	6–10 years	6–10 years	Female	White
8	Chris	Bombara	Teacher	School	Hartford	Master's	16–20 years	16–20 years	Male	White
8	Kathleen	Foley	Teacher	School	Fairfield	Master's	16–20 years	6–10 years	Female	White
8	Jennifer	Reilly	Teacher	School	Middlesex	Master's	11–15 years	11–15 years	Female	Caucasian

Panel	First Name	Last Name	Position	Level	District	Education	Years Teaching	Years Teaching in Assigned Grade	Gender	Ethnicity
8	Valerie	LeBlanc	Teacher; Science Curriculum Coordinator	School	Fairfield	Master's	16–20 years	16–20 years	Female	White
8	Jonathan	Jaekle	Teacher	School	Fairfield County	Bachelor's	6–10 years	6–10 years	Male	White
8	Terry	Contant	Science Education Consultant	Home	New Haven	Doctorate	21+ years	21+ years	Female	White
11	Joan	Donlon	Teacher	School	New Haven	Master's	21+ years	21+ years	Female	White
11	Charles	Detelich	Teacher	School	Fairfield	Master's	16–20 years	16–20 years	Male	White
11	Tara	Bellefleur	Specialist; Special Education Teacher	School	Hartford	Master's + 30	21+ years	6–10 years	Female	White
11	Andrea	LaRosa	Teacher	School	Fairfield	Sixth-Year Degree	11–15 years	6–10 years	Female	Hispanic; Asian
11	Sarah	Parsons	Special Education Teacher	School	Avon	Master's	21+ years	0 years	Female	White
11	Smita	Worah	Consultant	SERC	Hartford	Doctorate	21+ years	11–15 years	Female	Asian
11	Ann-Marie	Stevenson	Special Education Teacher	School	Tolland	Master's	6–10 years	6–10 years	Female	White; Native American
11	Michael	Gomola	Teacher	School	Waterbury	Master's	16–20 years	1–5 years	Male	White

Appendix B: Workshop Agenda

CSDE
CONNECTICUT STATE
DEPARTMENT OF EDUCATION

2019 Standard Setting for the Connecticut Alternate Science Assessment

SCIENCE EDUCATOR PANEL AGENDA

July 29 – July 30, 2019

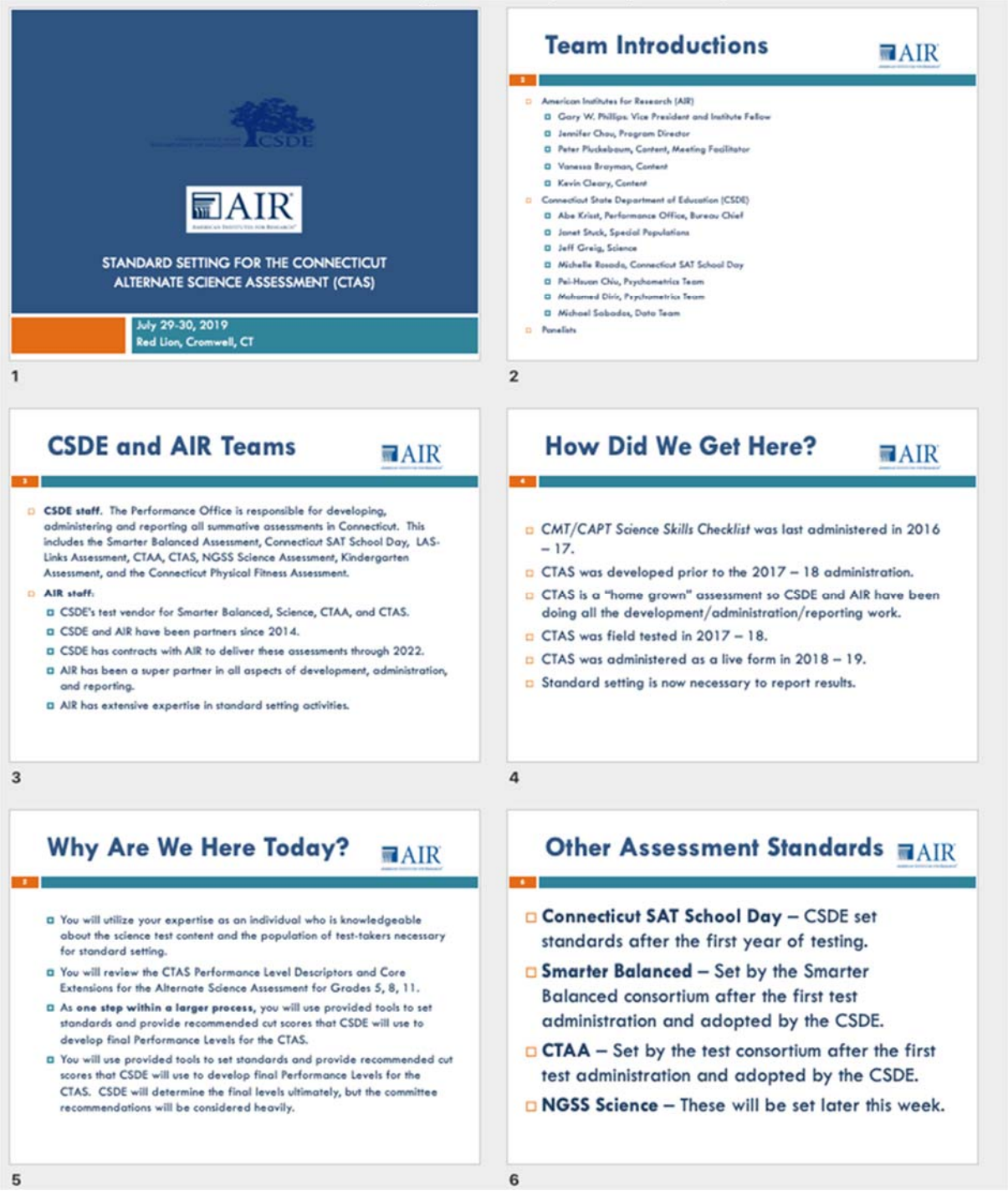
Day 1 – Monday, July 29, 2019	
8:00 – 8:30 a.m.	Registration, Table Leader Orientation, Breakfast
8:30 – 9:30 a.m.	Welcome, Introductions, General Orientation
9:30 – 10:30 a.m.	Overview of Connecticut Alternate Science (CTAS) Assessment Development, Overview of Activities for Standard Setting
10:30 – 10:45 a.m.	Break
10:45 – 12:00 p.m.	Take the CTAS Review Performance Tasks, Essence Statements, and Core Extensions and Resources
12:00 – 1:00 p.m.	Lunch
1:00 – 2:30 p.m.	Panelists Review Range Performance Level Descriptors
2:30 – 3:30 p.m.	Panelists Summarize Skills of Students for Each Performance Level and Create an Understanding of Student Characteristics for "Just Barely"
3:30 – 3:45 p.m.	Break
3:45 – 5:00 p.m.	Panelists Review Ordered Item Booklet in Grade Level Groups
5:00 p.m.	Adjourn

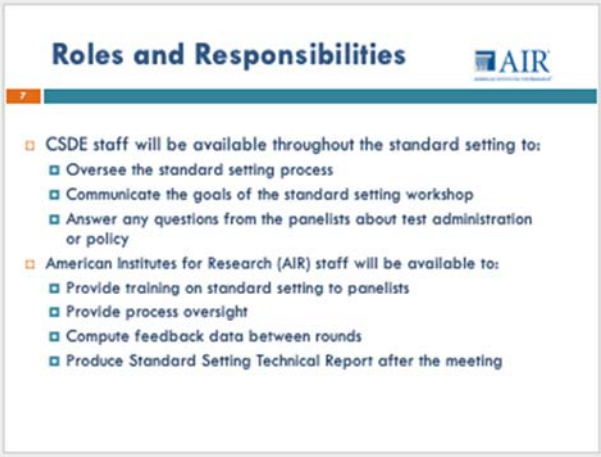
AIR ASSESSMENT | AIR.ORG 1

Day 2 – Tuesday, July 30, 2019	
8:00 – 8:30 a.m.	**Breakfast, Sign In, Obtain Materials**
8:30 – 9:30 a.m.	**Panelists Continue to Review Ordered Item Booklet in Grade Level Groups**
9:30 – 9:45 a.m.	**Panelists Review 50/50 Chance Criterion**
9:45 – 10:15 a.m.	**Panelists Practice Bookmarking Method and Complete Standard Setting Quiz**
10:15 – 10:30 a.m.	**Break**
10:30 – 12:00 p.m.	**Round 1 Bookmark Placement in Grade Level Groups (5, 8, 11)** Review of Bookmark Procedures and Key Concepts Sign Readiness Form Round 1 Bookmark Placement
12:00 – 1:00 p.m.	**Lunch**
1:00 – 1:30 p.m.	**Panelists Review and Discuss Round 1 Results**
1:30 – 2:15 p.m.	**Round 2 Bookmark Placement**
2:15 – 2:45 p.m.	**Break**
2:45 – 3:15 p.m.	**Panelists Review Round 2 Results**
3:15 – 3:30 p.m.	**Panelists Complete Online Workshop Evaluation**
3:30 – 5:00 p.m.	**Panelists Share Feedback on Recommendations Across Grades in Large Group**
5:00 p.m.	**Adjourn**

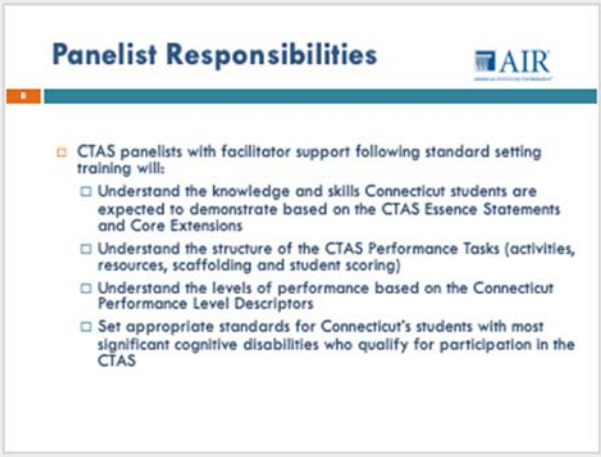
Appendix C: Training Slides

Figure C1. Large Group Training




Roles and Responsibilities
AIR
7
CSDE staff will be available throughout the standard setting to:
Oversee the standard setting process
Communicate the goals of the standard setting workshop
Answer any questions from the panelists about test administration or policy
American Institutes for Research (AIR) staff will be available to:
Provide training on standard setting to panelists
Provide process oversight
Compute feedback data between rounds
Produce Standard Setting Technical Report after the meeting

7


Panelist Responsibilities
AIR
8
CTAS panelists with facilitator support following standard setting training will:
Understand the knowledge and skills Connecticut students are expected to demonstrate based on the CTAS Essence Statements and Core Extensions
Understand the structure of the CTAS Performance Tasks (activities, resources, scaffolding and student scoring)
Understand the levels of performance based on the Connecticut Performance Level Descriptors
Set appropriate standards for Connecticut's students with most significant cognitive disabilities who qualify for participation in the CTAS

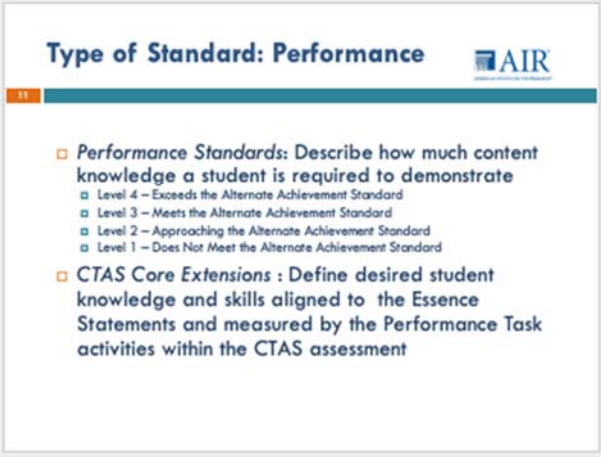
8


Agenda
AIR
9
Day 1 – Monday, July 29, 2019
8:00 – 8:30 a.m. Registration, Table Leader Orientation, Breakfast
8:30 – 9:30 a.m. Welcome, Introductions, General Orientation
9:30 – 10:30 a.m. Overview of Connecticut Alternate Science (CTAS) Assessment Development
Overview of Activities for Standard Setting
10:30 – 10:45 a.m. Break
10:45 – 12:00 p.m. Take the CTAS
Review Performance Tasks, Essence Statements, and Core Extensions and Resources
12:00 – 1:00 p.m. Lunch
1:00 – 2:30 p.m. Panelists Review Range Performance Level Descriptors
2:30 – 3:30 p.m. Panelists Summarize Skills of Students for Each Performance Level and Create an Understanding of Student Characteristics for "Just Barely"
3:30 – 3:45 p.m. Break
3:45 – 5:00 p.m. Panelists Review Ordered Item Booklet in Grade Level Groups
5:00 p.m. Adjourn

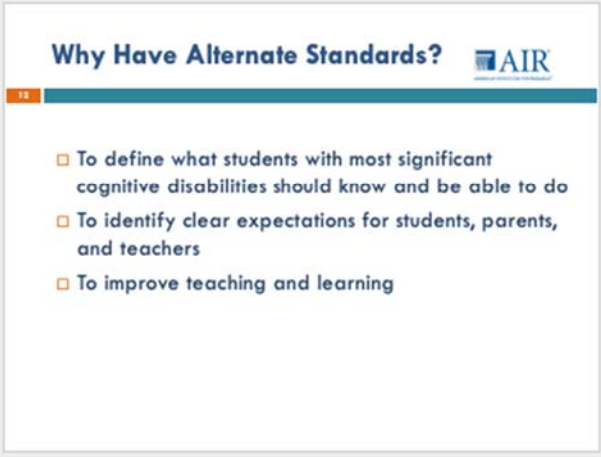
9


Agenda
AIR
10
Day 2 – Tuesday, July 30, 2019
8:00 – 8:30 a.m. Breakfast, Sign In, Obtain Materials
8:30 – 9:30 a.m. Panelists Continue to Review Ordered Item Booklet in Grade Level Groups
9:30 – 9:45 a.m. Panelists Review 50/50 Chance Criterion
9:45 – 10:15 a.m. Panelists Practice Bookmarking Method and Complete Standard Setting Quiz
10:15 – 10:30 a.m. Break
10:30 – 12:00 p.m. Round 1 Bookmark Placement in Grade Level Groups (5, 8, 11)
Review of Bookmark Procedures and Key Concepts
Sign Readiness Form
Round 1 Bookmark Placement
12:00 – 1:00 p.m. Lunch
1:00 – 1:30 p.m. Panelists Review and Discuss Round 1 Results
1:30 – 2:15 p.m. Round 2 Bookmark Placement
2:15 – 2:45 p.m. Break
2:45 – 3:15 p.m. Panelists Review Round 2 Results
3:15 – 3:30 p.m. Panelists Complete Online Workshop Evaluation
3:30 – 5:00 p.m. Panelists Share Feedback on Recommendations Across Grades in Large Group

10


Type of Standard: Performance
AIR
11
Performance Standards: Describe how much content knowledge a student is required to demonstrate
Level 4 – Exceeds the Alternate Achievement Standard
Level 3 – Meets the Alternate Achievement Standard
Level 2 – Approaching the Alternate Achievement Standard
Level 1 – Does Not Meet the Alternate Achievement Standard
CTAS Core Extensions : Define desired student knowledge and skills aligned to the Essence Statements and measured by the Performance Task activities within the CTAS assessment

11


Why Have Alternate Standards?
AIR
12
To define what students with most significant cognitive disabilities should know and be able to do
To identify clear expectations for students, parents, and teachers
To improve teaching and learning

12

When Is Standard Setting Necessary?

13

- Standard setting becomes necessary whenever any of the following occur:
  - New tests with new content standards are developed
  - Test Blueprints change
  - Performance Level Descriptors (PLDs) change

13

What Is Standard Setting?

14

- A process of deriving levels of performance on educational assessments, by which decisions or classifications of persons will be made. (Cizek, 2006)
- Test scores can be used to group students into meaningful performance levels.
- Standard setting is the process whereby we draw the lines that separate the test scores into various performance levels.

14

Confidentiality

15

- As we will describe shortly, the CTAS is a non-secure assessment. Therefore the items are publicly available on the CSDE Portal. However, as a panelist we ask that you:
  - Do not share the content of the standard setting process.
  - Do not remove any materials from the room.
  - Do not discuss judgments or cut scores (yours or others) with anyone outside of the meeting.
- General conversations about the process and days' events are acceptable, but participants should avoid discussing details, particularly those involving items, cut scores, and any other confidential information.
- Notes should be taken using provided materials only.
- The only materials allowed on the table are standard setting materials.
- Please turn off cell phones and do not use cell phones during the workshop.

15

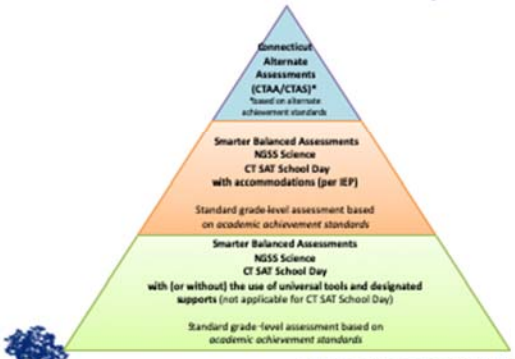
CONNECTICUT STATE DEPARTMENT OF EDUCATION

Overview of the CTAS

CONNECTICUT STATE DEPARTMENT OF EDUCATION

16

Connecticut's Summative Assessment Options



CONNECTICUT STATE DEPARTMENT OF EDUCATION

17

CTAS Eligibility

- Criteria reflect the **pervasive nature of a significant cognitive disability**
- PPT makes assessment decisions
- IEP includes **detailed evidence** of
  - significant cognitive disability
  - content learned based on grade-level standards; and
  - extensive direct individualized support needed

CONNECTICUT STATE DEPARTMENT OF EDUCATION

18

1:13

Considerations for PPT Decision Making

What type of evidence is included throughout the IEP (i.e. Present Levels of Performance, Goals/Objectives, Support and Accommodations, Assistive Technology, Augmentative and Alternative Communication) that address access to instruction and assessment?

CSDE CONNECTICUT STATE DEPARTMENT OF EDUCATION

19 1:13

Characteristics of Alternate Assessment Students

Students participating in the Alternate Assessment System are a relatively small population who:

(1) are identified with one or more of the existing categories of disability under the IDEA (intellectually disabled, autism, multiple disabilities, and traumatic brain injury, are the most common); and

(2) have cognitive impairments that may prevent them from attaining grade-level achievement standards, even with systematic instruction and accommodations.

Student Individualized Education Program (IEP) records indicate a pervasive disability or multiple disabilities that significantly impact intellectual functioning and adaptive behavior defined as essential for someone to live independently and to function safely in daily life.

CSDE CONNECTICUT STATE DEPARTMENT OF EDUCATION

20 1:13

Development of the CT Alternate Science Assessment

Prior to beginning the design and development of the CT Alternate Science Assessment, the CSDE sought informal and formal feedback from educators across the state on the science assessment format that would be most relevant and appropriate for students with the most significant cognitive disabilities who were eligible for the alternate assessment. Based on that feedback, the following guiding principles were established for the Alternate Science Field Test

This assessment should:

- be meaningful and accessible to participating students;
- guide science curriculum and instruction throughout the year by providing a coherent sequence of assessment activities;
- allow for administration throughout the year;
- include an appropriate balance of the breadth and depth of NGSS Learning Progressions across grade bands;
- assess the three-dimensions of NGSS (i.e., science and engineering practices, disciplinary core ideas, and crosscutting concepts);
- incorporate scientific phenomena that students make sense of or use to solve a problem; and
- expect consistent demonstration of the performance expectations by students statewide.

CSDE CONNECTICUT STATE DEPARTMENT OF EDUCATION

21 1:13

Initial Development of the CTAS

CTAS Committee work:

- Committee consisted of 25 educators with various areas of expertise in Science and/or special education including students with significant cognitive disabilities across grade levels
- Rated NGSS Learning Progressions and selection of Performance Expectations for students with significant cognitive disabilities
- Developed Essence Statements and Core Extensions
- Finalized Essence Statements and Core Extensions
- Reviewed/Developed initial Performance Task ideas

CSDE CONNECTICUT STATE DEPARTMENT OF EDUCATION

22 1:13

How was the CTAS designed?

The CTAS is compromised of Performance Tasks consisting of a Storyline capturing the NGSS Performance Expectations, Essence Statements, and Core Extensions within a specific content area (Earth Science, Life Science, and Physical Science).

CSDE CONNECTICUT STATE DEPARTMENT OF EDUCATION

23 1:13

CTAS Structure

CTAS is organized into six Storylines (two per content area) for each assessed grade level – grades 5, 8, and 11.

Connecticut Alternate Science Assessment	
Content Area	**Storyline and Performance Task**
Earth Science	1. Earth Systems
	2. Natural Resources
Life Science	3. Living Organisms
	4. Healthy Ecosystems
Physical Science	5. Forces and Motion
	6. Using Energy Every Day

CSDE CONNECTICUT STATE DEPARTMENT OF EDUCATION

24 1:13

CTAS Test Design

- Each Connecticut Alternate Science Essence Statement is associated with 2–4 Core Extensions.
- Core Extensions describe specific student performances and are aligned to provided activities administered by TEA.

CONNECTICUT STATE DEPARTMENT OF EDUCATION

25 1:13

CTAS Test Design

Connecticut Alternate Science Essence Statements

CTAS-3-ESS2-1 Use and interpret data in tables and graphs to describe typical weather conditions expected during a particular season.

CTAS-3-ESS2-2 Use information to describe climates in different regions of the United States.

CONNECTICUT STATE DEPARTMENT OF EDUCATION

26 1:13

CTAS Test Design

Core Extensions

1. Recognize two forms of water (e.g., rain, snow, hail, sleet) that can fall from clouds to Earth. (CTAS-3-ESS2-1)
2. Identify key components that describe local weather conditions (i.e., temperature, amount of cloud cover, precipitation, and wind speed). (CTAS-3-ESS2-1)
3. From provided temperature and precipitation data, identify the likely seasons. (CTAS-3-ESS2-1)
4. From provided data, compare weather conditions between two specific time periods. (CTAS-3-ESS2-1)
5. Using provided information, describe the climate in Connecticut. (CTAS-3-ESS2-2)
6. From provided data (average temperature and precipitation), compare climates in two regions of the United States (e.g., northeast vs. southwest). (CTAS-3-ESS2-2)
7. From provided information about the climate pattern in a region, make a prediction about typical weather conditions in that region. (CTAS-3-ESS2-2)
8. Complete a model to describe changes in the shape of a land form due to wind and water. (CTAS-5-ESS2-1)

CONNECTICUT STATE DEPARTMENT OF EDUCATION

27 1:13

Guiding Questions

Guiding Questions accompany each Storyline as an introduction to a science inquiry topic and capture the Core Extensions in which the Performance Task Activities are based.

CONNECTICUT STATE DEPARTMENT OF EDUCATION

28 1:13

Components of the CTAS

Each grade-specific CTAS set contains:

- **Performance Tasks**, which include:
  - a guiding question and a general overview of the task
  - a list of materials needed
  - instructions for preparing materials
  - step-by-step activities with built-in script and scaffolding for TEAs
  - scoring guidance
- **Resource Packets**, which are specific to each Performance Task, and include materials such as posters, graphs, sentence strips
- **Student Score Worksheet**
  - grade specific
  - completed by trained teachers administering the CTAS
  - completed through out the school year and then is submitted through the Data Entry Interface (DEI) during the testing window

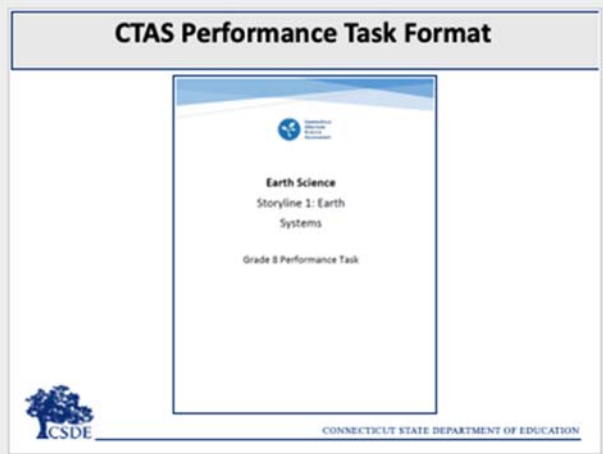
CONNECTICUT STATE DEPARTMENT OF EDUCATION

29 1:13

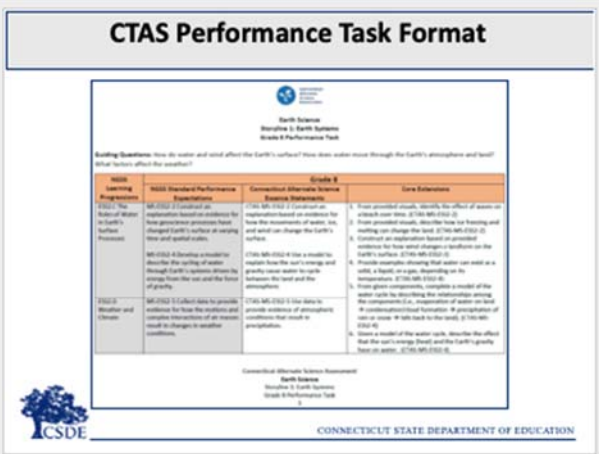
CTAS Performance Task Format

CONNECTICUT STATE DEPARTMENT OF EDUCATION

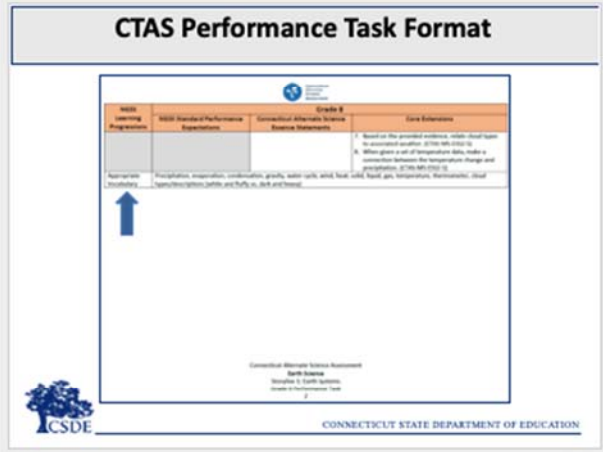
30 1:13



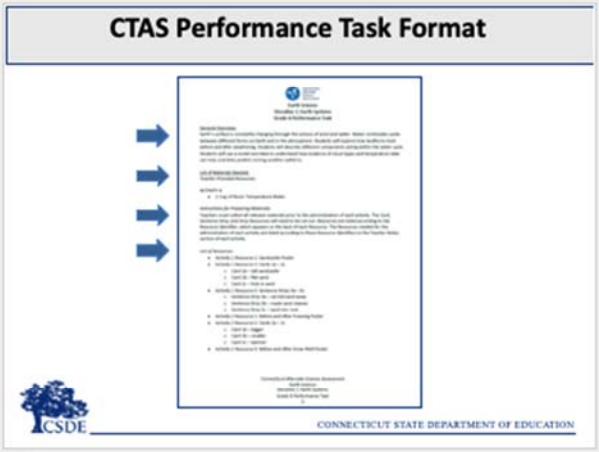
31 1:13



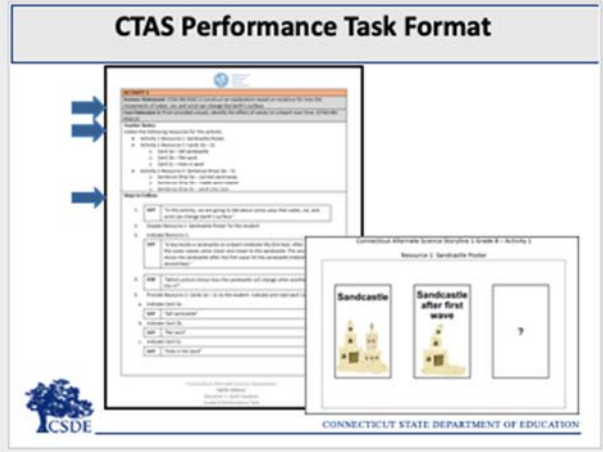
32 1:13



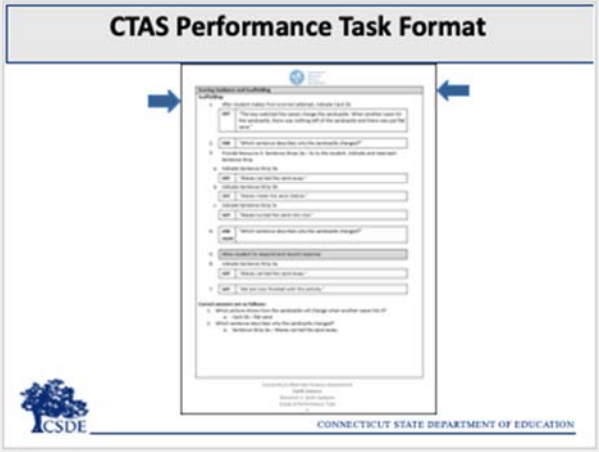
33 1:13



34

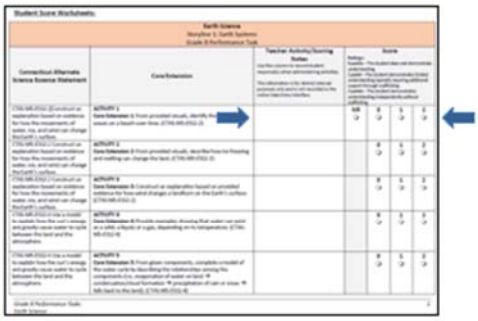


35



36

Student Score Worksheet



CONNECTICUT STATE DEPARTMENT OF EDUCATION

37

Alternate Assessment System Teacher Training

Connecticut requires that **all** teachers who will be administering the alternate assessments participate in annual online training and complete the associated quiz with a score of 80% or better.

The designated trained teacher administering the alternate (TEA) is then provided the permissions to register the student for participation in the alternate assessment system, access to the Connecticut Alternate Assessment (CTAA) materials and administration interface for ELA and Math, and the ability to submit the CTAS scores.

CONNECTICUT STATE DEPARTMENT OF EDUCATION

38

1:13

Overall Structure of the Workshop Panel

39

Connecticut, CTAS, Red Lion, Cromwell, Hartford, date - July 29-30, 2019
CTAS, Online-Bookmark

Panel	Room	Grade	Table Leader Panelists	Panelists	Facilitator	Facilitator Assistant	Other AIR Staff
Science	1	5	2	8	Peter Pluckebaum	Vanessa Brayman	Gary Phillips
		8	2	8			Nik Kalich
		11	2	8		Kevin Cleary	Drew Azar (IT)
							Dotun Adebayo (IT)
							Jennifer Chou
Totals	1		6	24	1	2	5

39

Main Activities of the Workshop Panel

40

- Receive Table Leader training
- Receive large group training
- Receive grade and subject-specific training
- Panelists will
  - take the test.
  - review the content standards
  - review Performance Level Descriptors
  - create "Just Barely" summary PLDs
  - review the ordered item booklet
- Recommend three performance standards in two rounds
- Provide a workshop evaluation

40

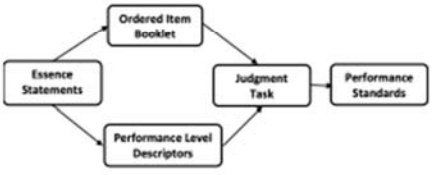
What Will Be Available Online for the Panelists?

41

- Tests for all grades
- Ordered item booklets
- Bookmark placements (2 rounds)
- Feedback results
- Impact data
- Evaluation

41

From Content Standards to Performance Standards



42

What method are we using for standard setting?

- Bookmark Method*
  - Research-based procedure
  - Used in many state assessment programs
  - Proven to be technically sound in litigation

*Mitzel, H. C., Lewis, D. M., Patz, R. J. & Green, D. R. (2001). "The Bookmark procedure: Psychological perspectives." In G. Cizek (ed.), *Setting performance standards: Concepts, methods and perspectives*. Mahwah, NJ: Erlbaum.

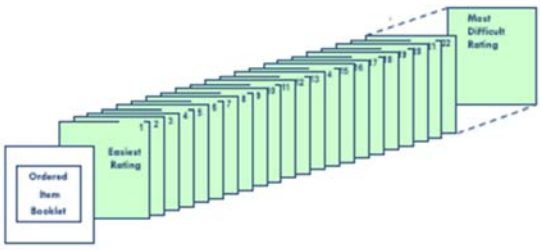
43

Bookmark Method using Ordered Item Booklet (OIB)

- Items (Activity Ratings) are ordered by difficulty
- Each page is a rating point on an activity
- Each activity appears twice in the OIB (once for a rating of 1 and again for a rating of 2)

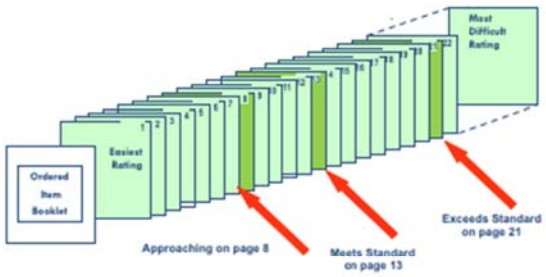
44

Ordered Item Booklet (OIB)



45

Bookmarking Pages in the Ordered Item Booklet



46

Bookmarking Pages in the Ordered Item Booklet

- The terms "bookmarking pages" and "ordered item booklet" historically come from a paper-pencil testing environment.
- For the CTAS, the ordered item booklet will be online; the pages are selected from a drop-down menu.

47

Day 1 Overview: What Are Content Standards?

- CTAS Core Extensions
- Specify what students know and can do

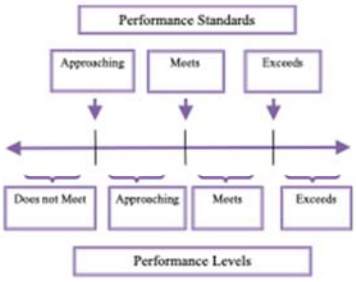
48

Day 1 Overview: What Are Performance Standards?

- Specify how much of the content standards students must know and be able to do in order to meet each performance level
  - Three performance standards (cut scores)
  - Four performance levels

49

Day 1 Overview: What Are Performance Standards?



50

Day 1 Become Familiar with the CTAS Assessment

Benefits:

- Understand the connection between the Essence Statements and Core extensions
- Have an opportunity to interact with items
- Understand the student experience with the assessment
- Understand how the embedded scaffolding provides additional access and support for students who need it.

51

Day 1 Performance Level Descriptors (PLDs)

- Specify what students in each performance level are expected to know and be able to do
- PLDs are the link between content and performance standards
- PLDs are used to develop a mental representation of students at each performance level
- Place the bookmark at the point where students scoring at and above that level can be accurately described by the PLD

52

PLDs

- ***Range*** PLDs: Describe what students should know and be able to do at different proficiency Description of what students know and are able to do throughout the range of each performance level. For example, the range PLD for Level 3 describes what students know and can do at that level all the way up to just below the Level 4.
- ***"Just Barely"*** PLDs: Created during the standard setting workshop and are used for standard setting only. The Just Barely PLDs describe what a student just barely scoring at the bottom of each performance level knows and can do.

53

Day 1 Just Barely PLDs

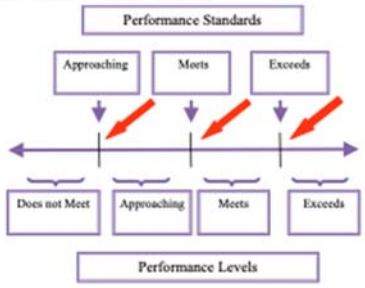
- Not typical of students in performance level; although just barely, they reach the standard.

54

Day 1 Just Barely PLDs

AIR

- Not typical of students in performance level; although just barely, they reach the standard.



55

Day 1 Ordered Item Booklet

AIR

- Consider each activity and answer two questions:
  1. What do students need to know and be able to do to receive this rating on this activity?
  2. Why is this activity more difficult than the previous one?

56

Day 1 Ordered Item Booklet
What If a rating Seems Out of Order?

AIR

- The order of the ratings in the OIB is based on student performance
- Activities may seem out of order because they are ordered by difficulty, not by content or cognitive process
- The sequence of activities in the OIB will not match the sequence of instruction taught throughout the school year
- The ordering of activities in the OIB will not match the ordering of activities on the test

57

Day 1 Ordered Item Booklet

AIR

- Remember, Standard Setting Is Aspirational
- Standard setting is all about what students **should** know and be able to do, not about what they actually know and are able to do in your classroom.
- **Do not set standards for your classroom.** You are setting standards for all students across the state.

58

Chance of Getting an Activity Rating

AIR

- Imagine that you have 100 students who are "just barely" at the Meets level.
- For each page in the OIB, you should ask yourself the question, "Out of 100 students who are 'just barely' at the Meets level, what percent would likely get this Activity Rating?"
  - Toward the beginning of the OIB, you would expect almost all the "just barely" Meets students would likely get 1 or 2 points on the Activity because the Activities are easy for the Meets student.
  - Toward the end of the OIB, you would expect very few of the "just barely" Meets students would likely get a 1 or 2 points on the Activity because the Activities are difficult for the Meets student.

59

Chance of Getting an Activity Rating

AIR

- What the concept in the previous slide shows is that a student who is "just barely" Meets will not have an equal likelihood of getting 1 or 2 points on each Activity in the OIB.
- In the Bookmark procedure, the panelist is trying to locate the Activity Rating, or small group of Activity Ratings, where the "just barely" Meets student has about a 50/50 chance of receiving the rating.

60

Chance of Getting an Activity Rating

61

- What this shows is that a student who is "just barely" at Meets Proficiency will not have an equal likelihood of correctly answering each item in the OIB.
- In the Bookmark procedure, the panelist is trying to locate the item, or small group of items, where the "just barely" Meets Proficiency student has about a 50/50 chance of getting this rating.

61

Making your Bookmark

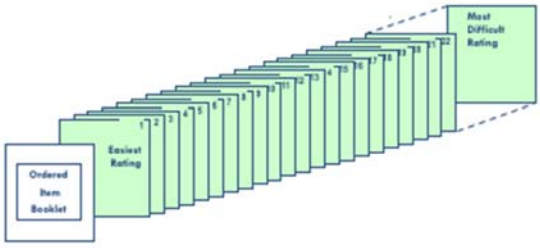
63

- As you work through the OIB, you will come across an Activity Rating, or small group of ratings, where you think about 50% of the "just barely" Meets Proficiency students would likely get that rating.
- Activity Ratings before that point in the OIB are ratings that you feel more than 50% of the just barely students would receive.
- Items beyond that point in the OIB are items that you feel less than 50% of the just barely students would receive.
- For each page, ask yourself the question - would a student who just barely Meets the Proficiency standard have about a 50/50 chance of earning this point?
- Panelists place a bookmark on the last page in the OIB where they believe the "just barely" student for that level would have about a 50/50 chance of receiving that rating.

62

Illustration of Selecting Cut Scores

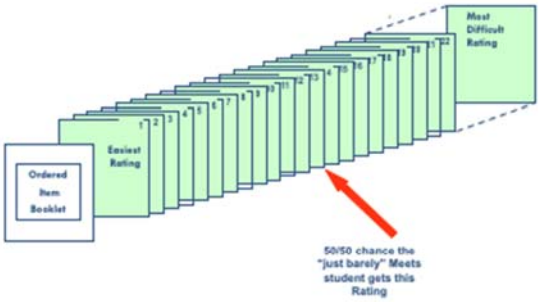
63



63

Illustration of Selecting Cut Scores

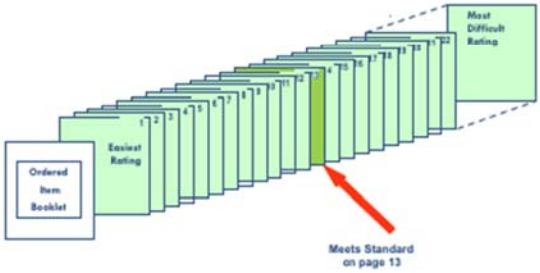
64



64

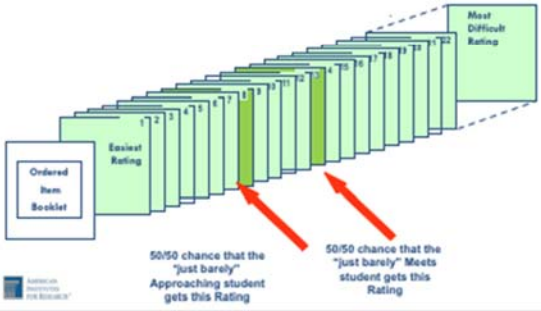
Illustration of Selecting Cut Scores

65

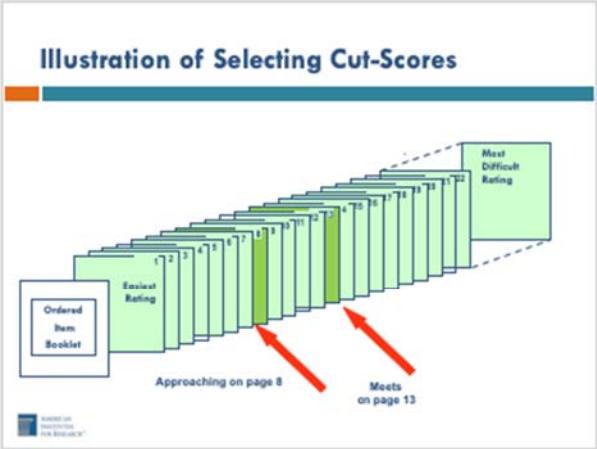


65

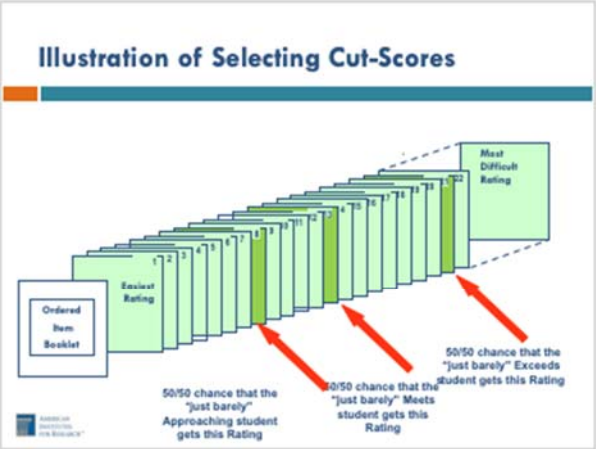
Illustration of Selecting Cut-Scores



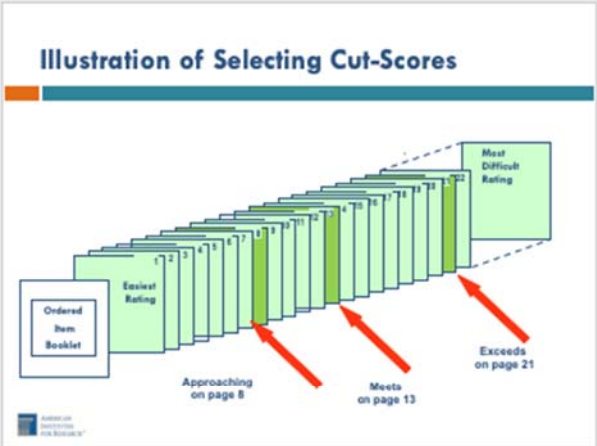
66



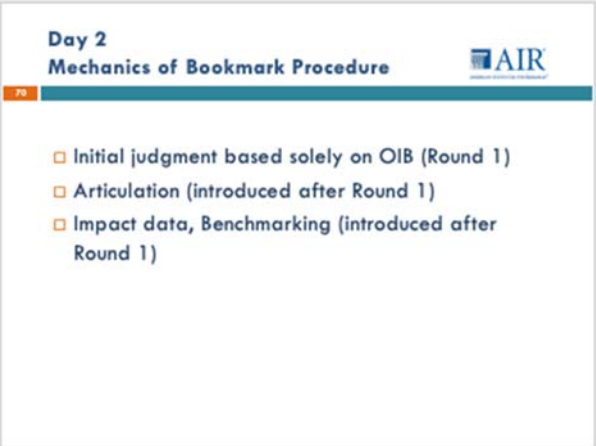
67



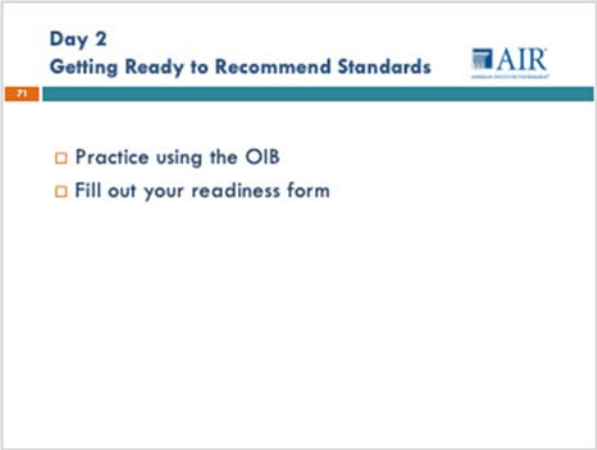
68



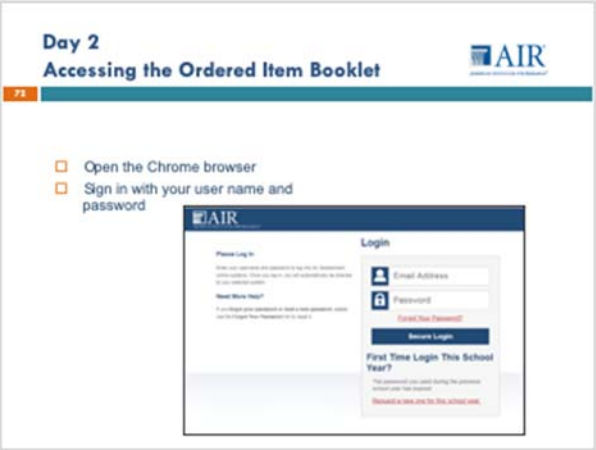
69



70



71



72

Day 2

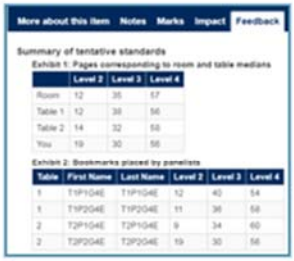
73

- Submit recommendations for round 1

73

Feedback

74



74

Articulation

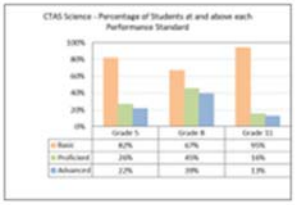
75

- Each group is recommending standards for different grades.
- Performance standards for a statewide system must be coherent across grades:
  - Articulated across grades with no anomalous peaks and valleys

75

Illustration of Disarticulated Standards

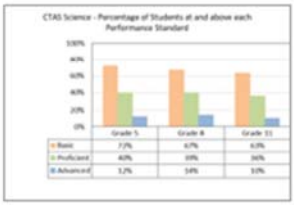
76



76

Illustration of Articulated Standards

77



77

Articulation

78

- AIR will estimate page numbers that will represent articulated standards.
- The articulated standards would be communicated to the panelists at the beginning of round 2.
- With articulated cut scores in hand, you are now asked to judge whether it makes sense from a content point of view to place your bookmark at or near the OIB page associated with each articulated cut score.

78

Impact Data

79

- Impact data show the percentage of students who would reach any standard that you select.
- Impact data are introduced at round 2, after round 1 recommendations were made based solely on content considerations.
- Impact data are used as context to inform the panelists' recommendations but should not determine their recommendations.
- In the end, the panelists' recommendations should have a content justification (OIB, PLDs, Essence Statements).

79

How Do We Display the Impact Information?

80

- As the panelists scroll through the online OIB, they will be shown the impact percentages associated with each page.
- After round 2 recommendations are made, a graph will show the percent of students that would score at and above the Performance Standard.
  - Your table selected
  - You selected

80

Impact Data

81

81

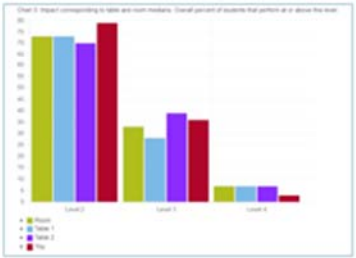
Impact Data

82

82

Impact Data

83



83

Benchmark Data

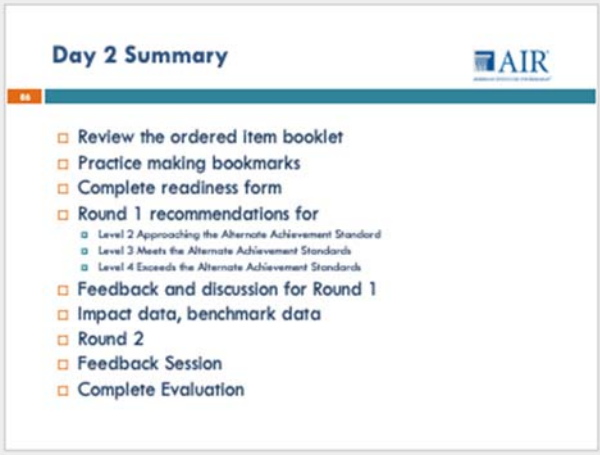
84

- Benchmarking provides the panelists with external references so that they can see how their recommendations compare with other assessments
- The idea of benchmarking is that we would like the performance standards for students with most significant cognitive disabilities to be as rigorous for the 1% population as the general education standards are for the general education population.
- For benchmarking we will use data from Connecticut Alternate Assessment (CTAA) in Round 2.

84


Day 1 Summary
AIR
85
Table Leader Training
Large Group Training
Small Group Training
Take the Test
Review Essence Statements
Review Performance Level Descriptors
Create Just Barely Performance Level Descriptors

85

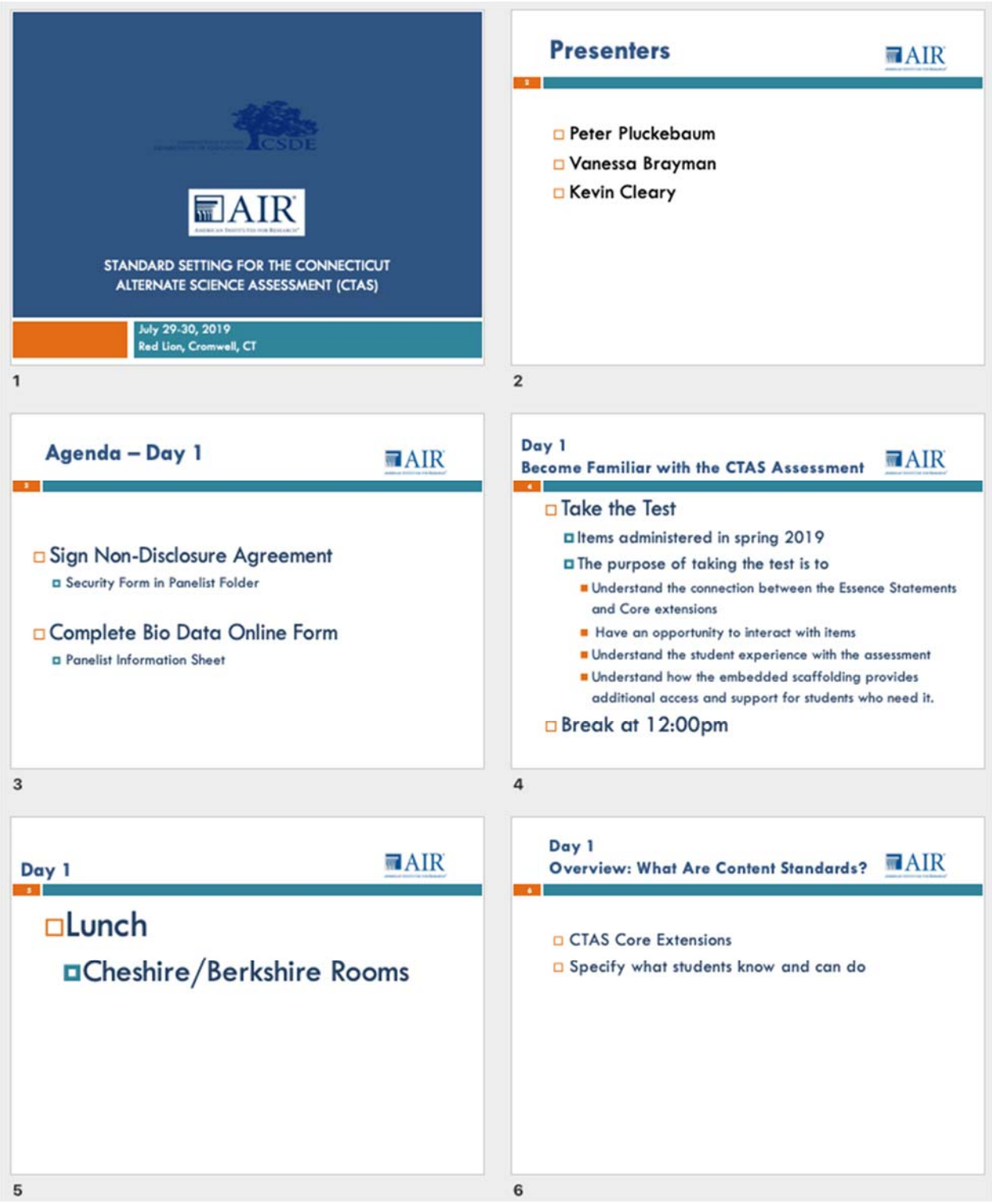

Day 2 Summary
AIR
86
Review the ordered item booklet
Practice making bookmarks
Complete readiness form
Round 1 recommendations for
Level 2 Approaching the Alternate Achievement Standard
Level 3 Meets the Alternate Achievement Standards
Level 4 Exceeds the Alternate Achievement Standards
Feedback and discussion for Round 1
Impact data, benchmark data
Round 2
Feedback Session
Complete Evaluation

86


AIR
87
Questions?

87

Figure C2. Breakout Room Training Slides (Day 1)



Day 1
Content Standards

AIR

7

Grade 5 Content Standard Example

CT Standards Code	Essence Statements
CTAS-3-ESS2-1	CTAS-3-ESS2-1 Use and interpret data in tables and graphs to describe typical weather conditions expected during a particular season.
CTAS-3-ESS2-2	Use information to describe climates in different regions of the United States.
CTAS-5-ESS2-1	Use a model to show how wind and water interact with land and living organisms.

7

Type of Standard: Performance

AIR

8

- *Performance Standards*: Describe how much content knowledge a student is required to demonstrate
  - Level 4 – Exceeds the Alternate Achievement Standard
  - Level 3 – Meets the Alternate Achievement Standard
  - Level 2 – Approaching the Alternate Achievement Standard
  - Level 1 – Does Not Meet the Alternate Achievement Standard
- *CTAS Core Extensions* : Define desired student knowledge and skills aligned to the Essence Statements and measured by the Performance Task activities within the CTAS assessment

8

Day 1 Overview: What Are Performance Standards?

AIR

9

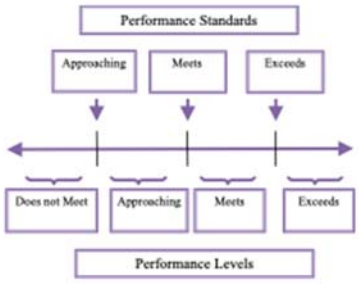
- Specify how much of the content standards students must know and be able to do in order to meet each performance level
  - Three performance standards (cut scores)
  - Four performance levels

9

Day 1 Overview: What Are Performance Standards?

AIR

10



10

Day 1
Performance Level Descriptors (PLDs)

AIR

11

- Specify what students in each performance level are expected to know and be able to do
- PLDs are the link between content and performance standards
- PLDs are used to develop a mental representation of students at each performance level
- Place the bookmark at the point where students scoring at and above that level can be accurately described by the PLD

11

Day 1
Performance Level Descriptors (PLDs)

AIR

12

Grade 5 Science (Example)

Domain	NGSS Standard Performance Expectation	CT Standards Code	Essence Statements	Level 1 - Does Not Meet	Level 2 - Approaching	Level 3 - Meets	Level 4 - Exceeds
Earth Science	5-ESS2-1 Represent data in tables and graphical displays to describe typical weather conditions expected during a particular season.	CTAS-5-ESS2-1	CTAS-5-ESS2-1 Use and interpret data in tables and graphs to describe typical weather conditions expected during a particular season.	Identify two forms of water (e.g., rain, snow, hail, sleet) that can fall from clouds to Earth.	Identify key components that describe local weather conditions (e.g., temperature, amount of cloud cover, precipitation, and wind speed).	From provided temperature and precipitation data, identify the likely season.	From provided data, compare weather conditions between two specific time periods.
Earth Science	3-ESS2-2 Obtain and combine information to describe climates in different regions of the world.	CTAS-3-ESS2-2	Use information to describe climates in different regions of the United States.	Identify a climate (e.g., cold and wet, hot and dry).	Using provided information, describe the climate in Connecticut.	From provided data (average temperature and precipitation), compare climates in two regions of the United States (e.g., northeast vs. southwest).	From provided information about the climate pattern in a region, make a prediction about typical weather conditions in that region.

12

Day 1

13

- Review PLDs

13

PLDs

14

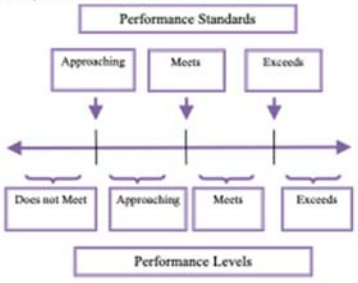
- ***Range*** PLDs: Describe what students should know and be able to do at different proficiency Description of what students know and are able to do throughout the range of each performance level. For example, the range PLD for Level 3 describes what students know and can do at that level all the way up to just below the Level 4.
- ***"Just Barely"*** PLDs: Created during the standard setting workshop and are used for standard setting only. The Just Barely PLDs describe what a student just barely scoring at the bottom of each performance level knows and can do.

14

Just Barely PLDs

15

- Not typical of students in performance level; although just barely, they reach the standard.



15

Just Barely PLDs

16

- Not typical of students in performance level; although just barely, they reach the standard.

16

Developing Just Barely Summary Statements

17

- Think about what skills, concepts, or knowledge a just barely student would need to have to enter into each level.
- As a group we will summarize the skills that a just barely student needs to have to gain entry into each of the three levels for a grade as an example.
- We will start as a large group and work through a couple of bullets together at each performance level.
- We will then break into our tables and write Just Barely Summary Statements for remaining bullets within our tables.
- Then we will review the statements as small groups.

17

Developing Just Barely Summary Statements

18

Grade: Subject:	Level 2: Approaching	Level 3: Meets	Level 4: Exceeds
CT Standard Code: Essence Statement:	*	*	*
CT Standard Code: Essence Statement:	*	*	*

18

Developing Just Barely Summary Statements

19

For each performance level think about:

- What is important for a just barely student to be able to demonstrate?
- How does this differ from the upper range of the adjacent performance level?

19

Day 1

20

Break

20

Ordered Item Booklet (OIB)

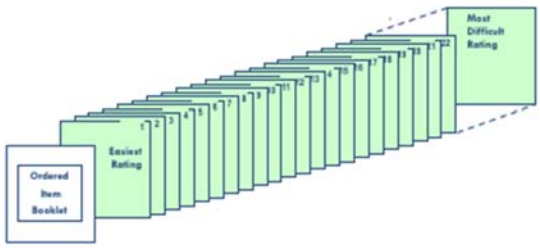
21

- Items (Activity Ratings) are ordered by difficulty
- Each page is a rating point on an activity
- Each activity appears twice in the OIB (once for a rating of 1 and again for a rating of 2)

21

Ordered Item Booklet (OIB)

22



22

Day 1
Accessing the Ordered Item Booklet

23

- Open the Chrome browser
- Sign in with your user name and password

23

Day 1
Navigating the OIB

24

- Description label indicates test and step being worked on
- Page forward in the OIB, or select a page from the drop-down menu
- View more about the item

24


Day 1
Navigating the OIB
AIR
Review Panel
More about this item tab
Content Alignment
Item Content
Answer Key
Scoring Guidelines
Notes tab
Other tabs are not relevant at this stage
25
Day 1
Evaluating Partial Credit
AIR
First time you review this item:
What does a student have to know and be able to do to achieve 1 of 2 points on this item?
Score points
Second time you review this item:
What does a student have to know and be able to do to achieve 2 of 2 points on this item?
Score points
26
Day 1 Ordered Item Booklet
What If a rating Seems Out of Order?
AIR
The order of the ratings in the OIB is based on student performance
Activities may seem out of order because they are ordered by difficulty, not by content or cognitive process
The sequence of activities in the OIB will not match the sequence of instruction taught throughout the school year
The ordering of activities in the OIB will not match the ordering of activities on the test
27
Day 1 Ordered Item Booklet
AIR
Remember, Standard Setting Is Aspirational
Standard setting is all about what students **should** know and be able to do, not about what they actually know and are able to do in your classroom.
Do not set standards for your classroom. You are setting standards for all students across the state.
28
Day 1 Ordered Item Booklet
AIR
Consider each activity and answer two questions:
1. What do students need to know and be able to do to receive this rating on this activity?
2. Why is this activity more difficult than the previous one?
Adjourn at 5:00PM
29

Figure C3. Breakout Room Training Slides (Day 2)

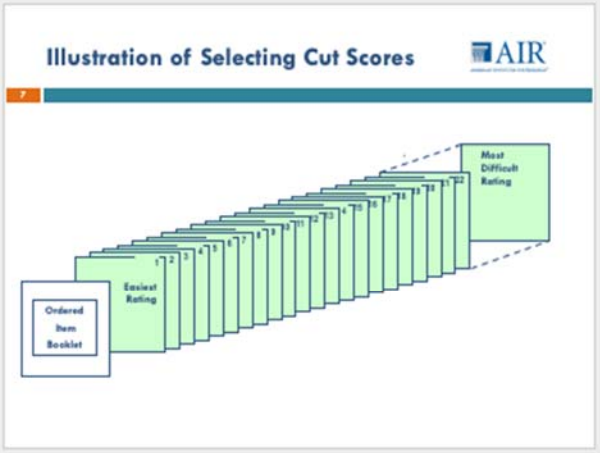


Illustration of Selecting Cut Scores
AIR
Most Difficult Rating
Easiest Rating
Ordered Item Booklet

7

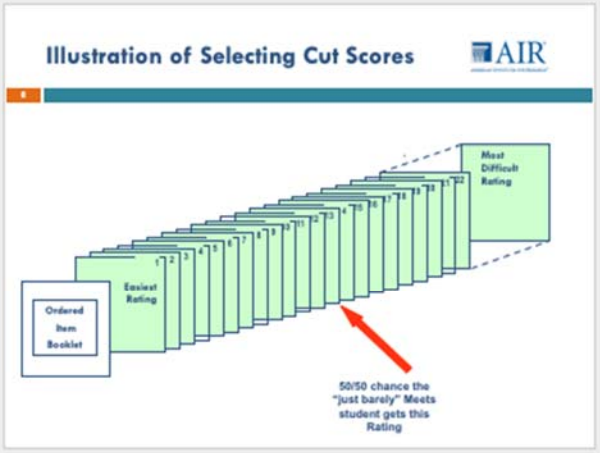


Illustration of Selecting Cut Scores
AIR
Most Difficult Rating
Easiest Rating
Ordered Item Booklet
50/50 chance the "just barely" Meets student gets this Rating

8

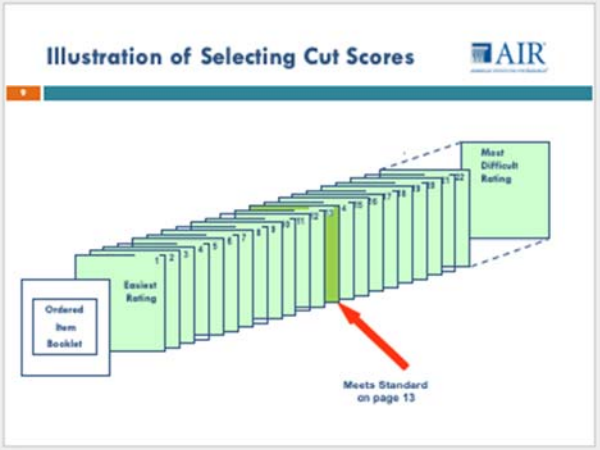


Illustration of Selecting Cut Scores
AIR
Most Difficult Rating
Easiest Rating
Ordered Item Booklet
Meets Standard on page 13

9

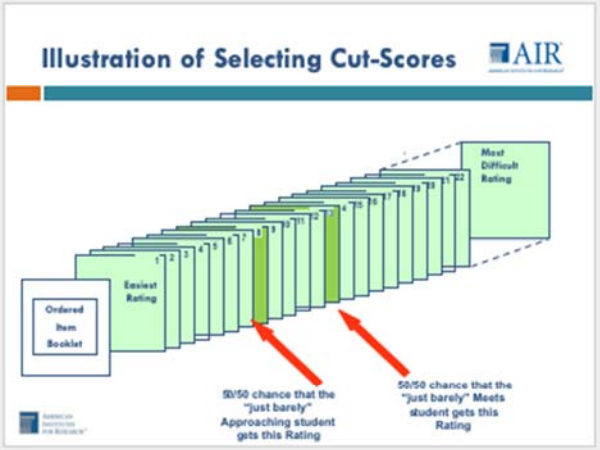


Illustration of Selecting Cut-Scores
AIR
Most Difficult Rating
Easiest Rating
Ordered Item Booklet
50/50 chance that the "just barely" Approaching student gets this Rating
50/50 chance that the "just barely" Meets student gets this Rating

10

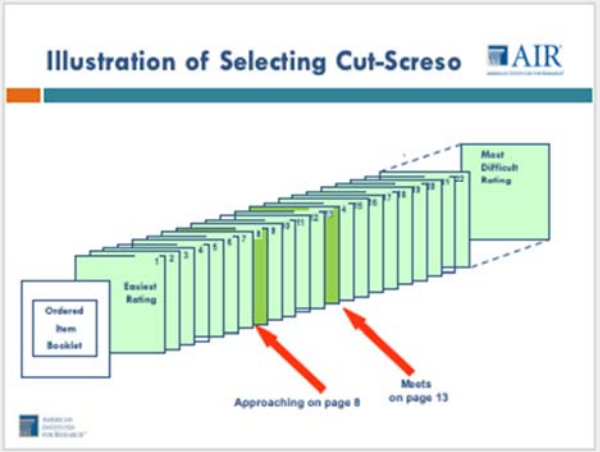


Illustration of Selecting Cut-Screso
AIR
Most Difficult Rating
Easiest Rating
Ordered Item Booklet
Approaching on page 8
Meets on page 13

11

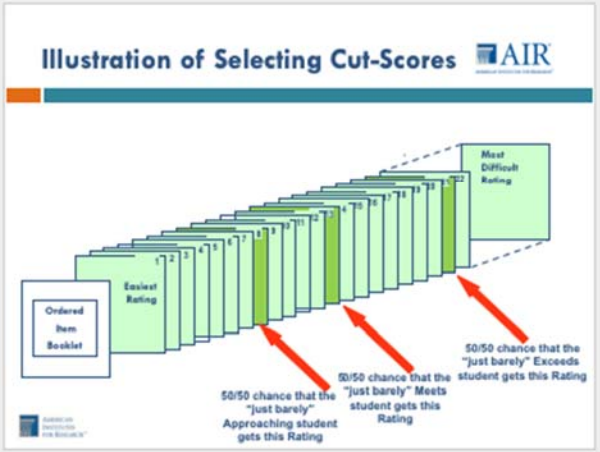
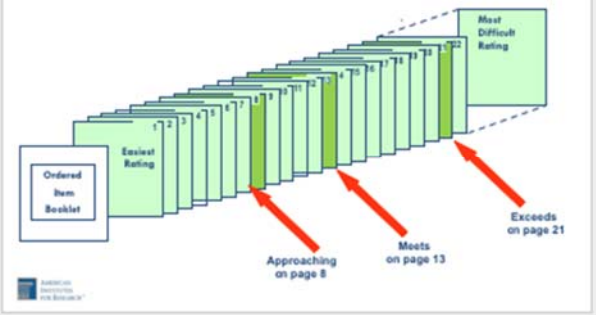


Illustration of Selecting Cut-Scores
AIR
Most Difficult Rating
Easiest Rating
Ordered Item Booklet
50/50 chance that the "just barely" Approaching student gets this Rating
50/50 chance that the "just barely" Meets student gets this Rating
50/50 chance that the "just barely" Exceeds student gets this Rating

12

Illustration of Selecting Cut-Scores



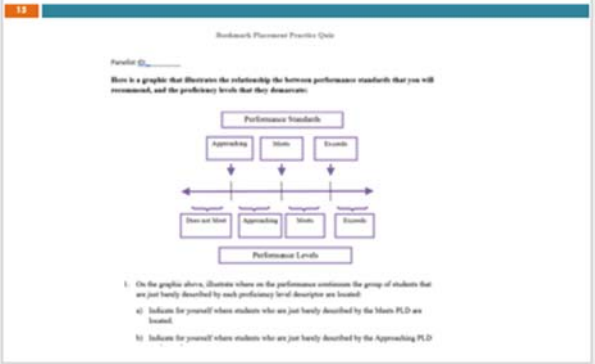
13

Practice Task

- Review the practice OIB
- Address any areas of confusion
- Remember
  - What does a student have to know and be able to do to achieve full credit on this item?
  - Why is this page more difficult than previous pages?

14

Standard Setting Quiz



15

Day 2

- Break
- 10:15 to 10:30 AM
  - Meet back at 10:30 AM

16

Day 2
Mechanics of Bookmark Procedure

- Initial judgment based solely on OIB (Round 1)
- Keep in mind:
  - "Just Barely" described by the PLD
  - Set your bookmark on the last page where 50% of students who just barely meet the performance level will answer correctly.
    - Fewer than 50% of just barely students would be expected to respond successfully to the next item in the OIB.
  - Three different bookmarks
    - Level 3 Meets
    - Level 2 Approaching
    - Level 4 Exceeds

17

Day 2
Are you ready? Completing the Readiness Form

2019 CTAS Standard Setting Educator Panel – Readiness Form

Preparation for Round 1 Bookmark Method

Panelist ID _____

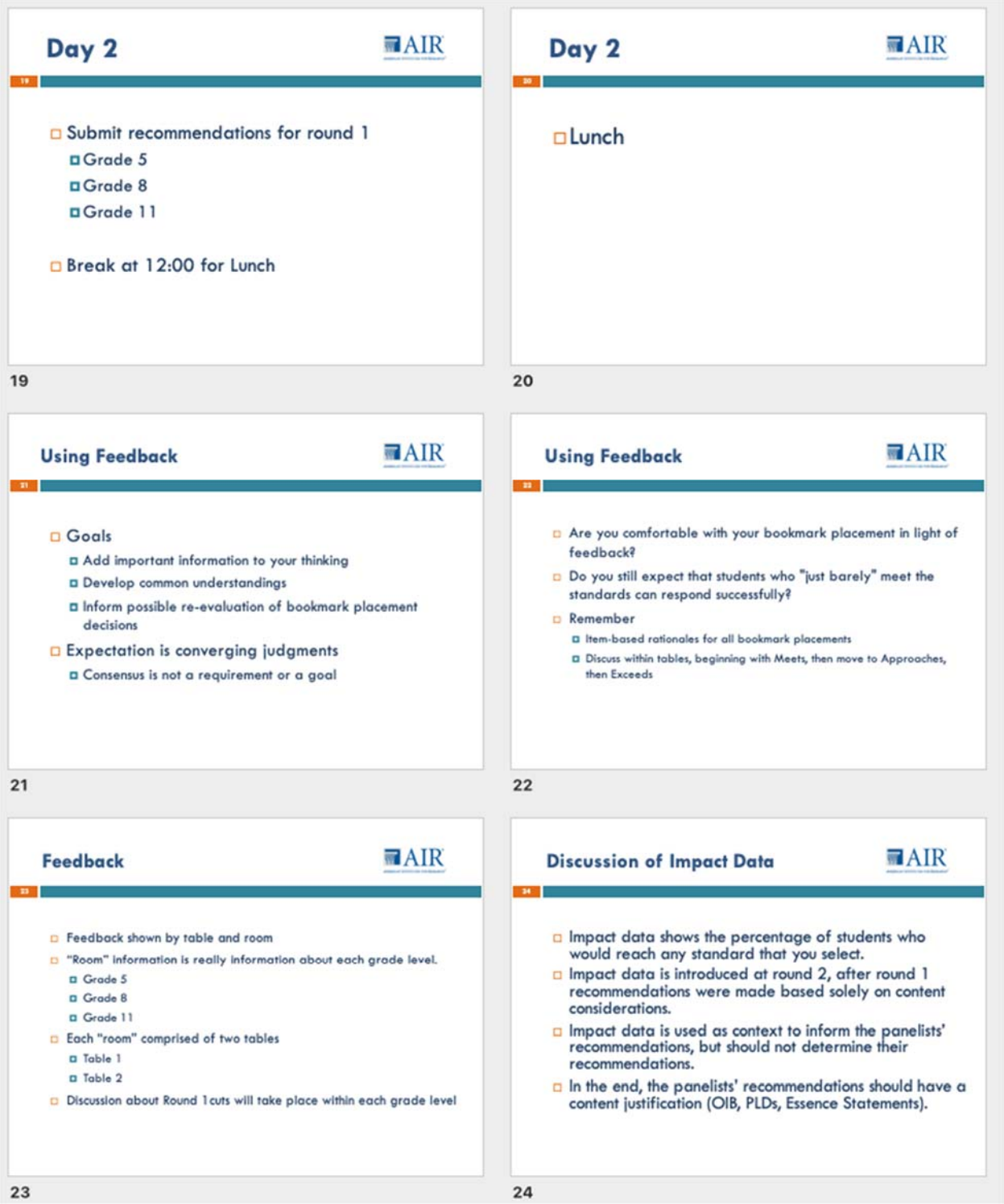
Committee (e.g., Grades 5 Science) _____

	Yes	No
a. The training fully explained the concept of a student who just barely meets the criteria described in the Performance Level Descriptors.	☐	☐
b. The meeting training has prepared me to review the Ordered Item Book (OIB).	☐	☐
c. The meeting training has prepared me to set bookmarks in the Standard Setting Tool.	☐	☐

I have answered, "Yes" to the above questions and I understand what I need to do to place my bookmarks.

Yes ____ No ____ Initials _____

18


Day 2
AIR
19
Submit recommendations for round 1
Grade 5
Grade 8
Grade 11
Break at 12:00 for Lunch
19
Day 2
AIR
20
Lunch
20
Using Feedback
AIR
21
Goals
Add important information to your thinking
Develop common understandings
Inform possible re-evaluation of bookmark placement decisions
Expectation is converging judgments
Consensus is not a requirement or a goal
21
Using Feedback
AIR
22
Are you comfortable with your bookmark placement in light of feedback?
Do you still expect that students who "just barely" meet the standards can respond successfully?
Remember
Item-based rationales for all bookmark placements
Discuss within tables, beginning with Meets, then move to Approaches, then Exceeds
22
Feedback
AIR
23
Feedback shown by table and room
"Room" information is really information about each grade level.
Grade 5
Grade 8
Grade 11
Each "room" comprised of two tables
Table 1
Table 2
Discussion about Round 1 cuts will take place within each grade level
23
Discussion of Impact Data
AIR
24
Impact data shows the percentage of students who would reach any standard that you select.
Impact data is introduced at round 2, after round 1 recommendations were made based solely on content considerations.
Impact data is used as context to inform the panelists' recommendations, but should not determine their recommendations.
In the end, the panelists' recommendations should have a content justification (OIB, PLDs, Essence Statements).
24

How Do We Display the Impact Information?

- As the panelists scroll through the online OIB, the impact percentages associated with each page will display in the "Impact" tab
- After round 2 recommendations are made, a graph will show the percent of students that would score at and above the Performance Standard.
  - The entire room selected
  - Your table selected
  - You selected

25

How Do We Display the Impact Information?

26

Benchmark Data

- Benchmarking provides the panelists with external references so that they can see how their recommendations compare with other assessments
- The idea of benchmarking is that we would like the performance standards for students with the most significant cognitive disabilities to be as rigorous for the 1% population as the general education standards are for the general education population.
- For benchmarking we will use data from Connecticut Alternate Assessment (CTAA) in Round 2.

27

Benchmarks

Benchmarks (CTAA ELA)

	ELA 5	ELA 8	ELA 11
	% at and Above	% at and Above	% at and Above
Meets	36%	28%	42%
Approaching	65%	60%	63%
	ELA 5	ELA 8	ELA 11
	OIB Page	OIB Page	OIB Page
Meets	61	70	53
Approaching	42	43	42

28

Benchmarks

Benchmarks (CTAA Math)

	Math 5	Math 8	Math 11
	% at and Above	% at and Above	% at and Above
Meets	38%	49%	36%
Approaching	73%	69%	69%
	Math 5	Math 8	Math 11
	OIB Page	OIB Page	OIB Page
Meets	58	50	61
Approaching	31	36	39

29

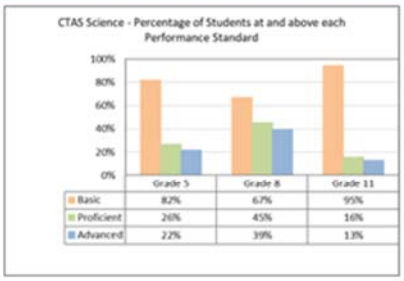
Articulation

- Each group is recommending standards for different grades.
- Performance standards for a statewide system must be coherent across grades:
  - Articulated across grades with no anomalous peaks and valleys

30

Illustration of Disarticulated Standards

31

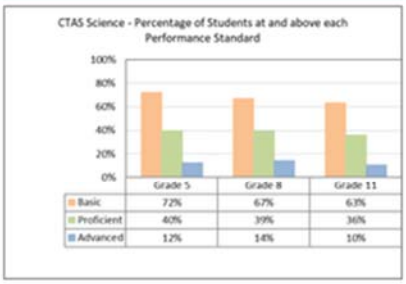


	Grade 5	Grade 8	Grade 11
Basic	82%	67%	95%
Proficient	26%	45%	16%
Advanced	22%	39%	13%

31

Illustration of Articulated Standards

32



	Grade 5	Grade 8	Grade 11
Basic	72%	67%	63%
Proficient	40%	39%	36%
Advanced	12%	14%	10%

32

Articulation

33

- AIR estimated the page numbers that will represent articulated standards.
- With articulated cut scores in hand, you are now asked to judge whether it makes sense from a content point of view to place your bookmark at or near the OIB page associated with each articulated cut score.

33

Articulation

34

(INSERT SLIDE FROM GARY HERE)

34

Day 2: Round 2 Remember to

35

- Set your bookmark on the last page where **50%** of students who "just barely" meet the performance level will answer correctly
  - Fewer than 50% of "just barely" students would be expected to respond successfully to the next item in the OIB
- Record the bookmarked page number by clicking "Set Here" under the marks tab
  - Begin with Meets, then move to Approaches, and finally Exceeds
  - Confirm your bookmarks

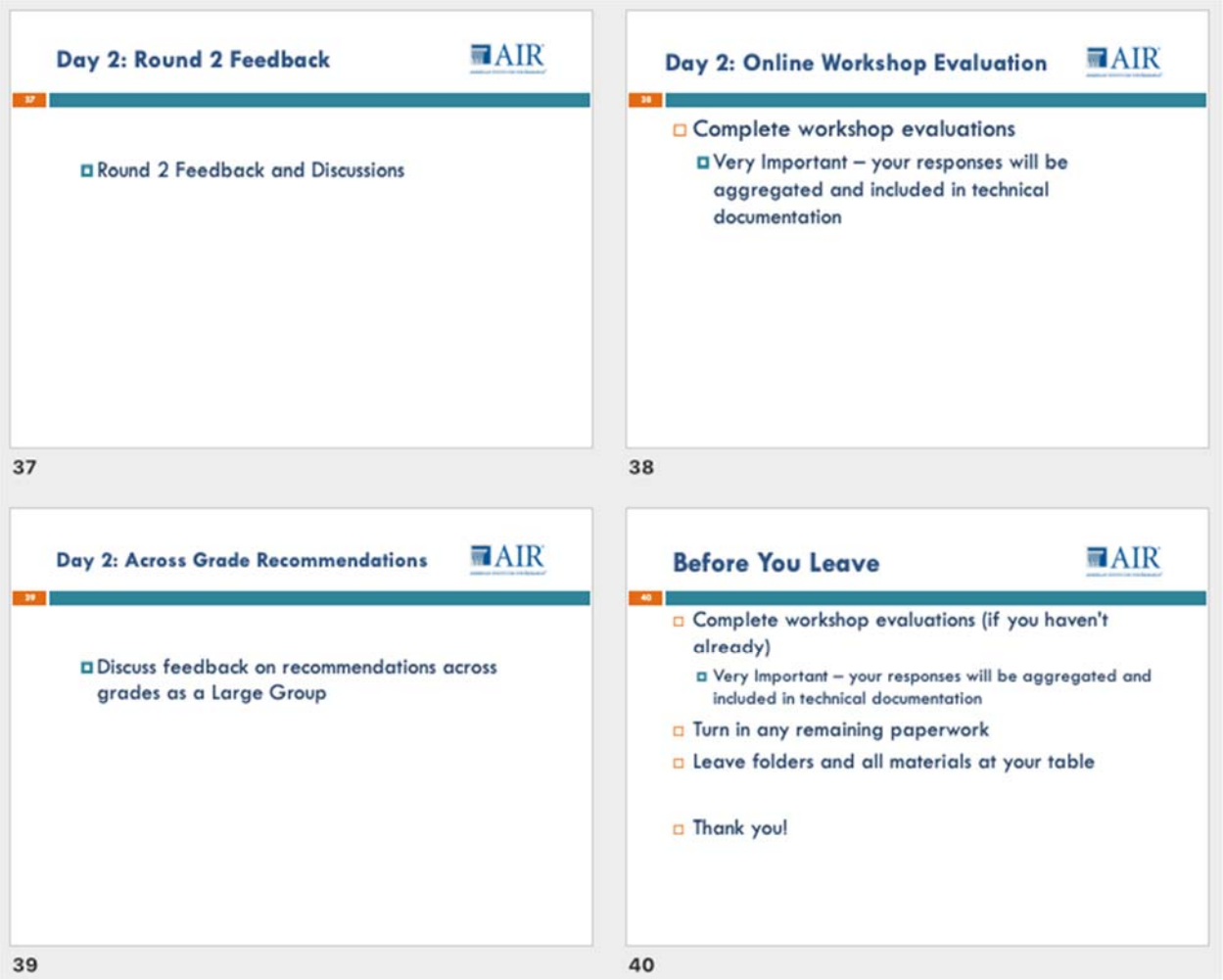
35

Day 2: Round 2

36

- Submit recommendations for Round 2
  - Grade 5
  - Grade 8
  - Grade 11
- Break from 2:15 to 2:45
- Return to room at 2:45

36

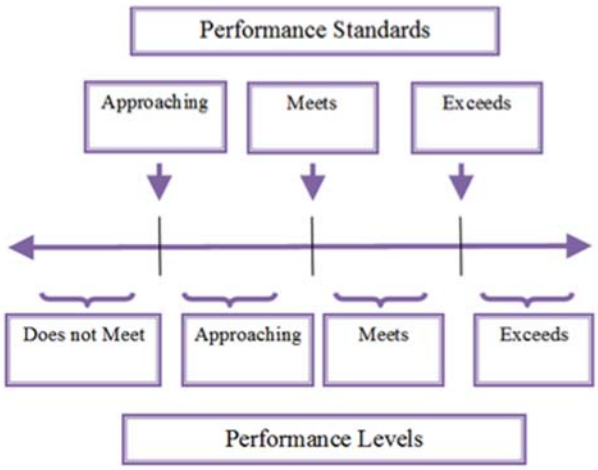

Day 2: Round 2 Feedback
AIR
37
Round 2 Feedback and Discussions
37
Day 2: Online Workshop Evaluation
AIR
38
Complete workshop evaluations
Very Important – your responses will be aggregated and included in technical documentation
38
Day 2: Across Grade Recommendations
AIR
39
Discuss feedback on recommendations across grades as a Large Group
39
Before You Leave
AIR
40
Complete workshop evaluations (if you haven't already)
Very Important – your responses will be aggregated and included in technical documentation
Turn in any remaining paperwork
Leave folders and all materials at your table
Thank you!
40

Appendix D: Bookmark Placement Quiz

Bookmark Placement Practice Quiz

Panelist ID:_____

Here is a graphic that illustrates the relationship the between performance standards that you will recommend, and the proficiency levels that they demarcate:



1. On the graphic above, illustrate where on the performance continuum the group of students that are just barely described by each proficiency level descriptor are located:

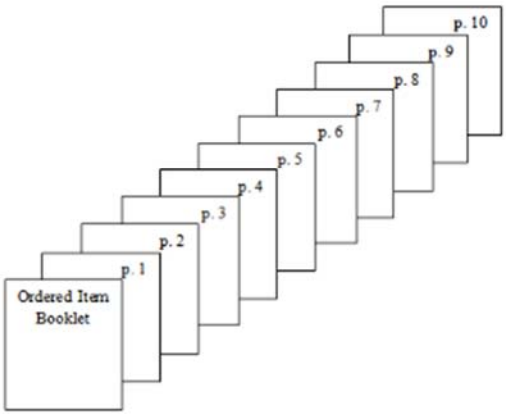
   a) Indicate for yourself where students who are just barely described by the Meets PLD are located.

   b) Indicate for yourself where students who are just barely described by the Approaching PLD are located.

   c) Indicate for yourself where students who are just barely described by the Exceeds PLD are located.

Bookmark Placement Practice Quiz

Here is a hypothetical Ordered Item Book (OIB) that consists of pages 1 through 10:



2. In the Ordered Item Book presented above, is the item on page 5 of the OIB easier, more difficult, or about the same difficulty as the item on page 6?

   ☐ The item on page 5 is easier than the item on page 6

   ☐ The item on page 5 is more difficult than the item on page 6

   ☐ The item on page 5 is about the same as the item on page 6

Bookmark Placement Practice Quiz

1. An Educator Panelist is placing a Round 1 bookmark for Meets using the OIB in the Figure 1. The panelist considers whether students who are just barely described by the Meets PLD have a fifty-fifty likelihood of responding successfully to each page in the OIB. The panelist believes that

   - More than half of just barely Meets students can answer items correctly on pages 1 through 5 of the OIB,
   - Exactly half of just barely Meets students can respond successfully to the item on page 6 of the OIB, and
   - Fewer than half of just barely Meets students can respond successfully to the item on page 7 and beyond.

   Mark the Bookmark Placement form to indicate how the Educator Panelist should indicate their Round 1 Meets recommendation.

Round	Page Number of Bookmark Placement		
	Approaching	Meets	Exceeds
Round 1			

Appendix E: Readiness Form

2019 CTAS Standard Setting Educator Panel – Readiness Form

Preparation for Round 1 Bookmark Method

Panelist ID _____

Committee (e.g., Grades 5 Science) _____

		Yes	No
a.	The training fully explained the concept of a student who just barely meets the criteria described in the Performance Level Descriptors.	❑	❑
b.	The meeting training has prepared me to review the Ordered Item Book (OIB).	❑	❑
c.	The meeting training has prepared me to set bookmarks in the Standard Setting Tool.	❑	❑

I have answered, "Yes" to the above questions and I understand what I need to do to place my bookmarks.

Yes _____ No _____ Initials _____

2019 CTAS Standard Setting Educator Panel – Readiness Form

Preparation for Round 1 Bookmark Method – additional training

Panelist ID _____

Committee (e.g., Grades 5 Science) _____

If I answered "No" to any of the questions on the Round 1 Bookmark Method Readiness Form, I received additional training.

Yes _____ No _____ Initials _____

Following the additional training, I feel sufficiently trained on what I need to do to place my bookmarks.

Yes _____ No _____ Initials _____

2019 CTAS Standard Setting Educator Panel – Readiness Form

Preparation for Round 2

Panelist ID _____

Committee (e.g., Grade 5 Science) _____

I understand the impact and benchmark data provided.

Yes _____ No _____ Initials _____

I understand my task for Round Two.

Yes _____ No _____ Initials _____

I am ready to begin Round Two.

Yes _____ No _____ Initials _____

2019 CTAS Standard Setting Educator Panel – Readiness Form

Preparation for Round 1 Bookmark Method – additional training

Panelist ID _____

Committee (e.g., Grades 5 Science) _____

If I answered "No" to any of the questions on the Round 1 Bookmark Method Readiness Form, I received additional training.

Yes _____ No _____ Initials _____

Following the additional training, I feel sufficiently trained on what I need to do to place my bookmarks.

Yes _____ No _____ Initials _____