Idaho Standards Achievement Test (ISAT) in Science

2023–2024

Volume 3: Setting Achievement Standards

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

1. EXECUTIVE SUMMARY

In 2018, the Idaho Department of Education (the "Department") adopted the new Idaho State Science Standards. The new standards adopt a three-dimensional conceptualization of science understanding, including science and engineering practices, crosscutting concepts, and disciplinary core ideas. With the adoption of the new science standards, and the development of new statewide assessments to measure student achievement relative to those standards, the Department convened a standard-setting workshop to recommend a system of achievement standards for determining whether students have met the learning goals defined by the Idaho State Science Standards.

Under contract to the Department, Cambium Assessment, Inc. (CAI) conducted the standard-setting workshop to recommend achievement standards for the Idaho Standards Achievement Tests (ISAT) in Science in grades 5, 8, and 11. The workshop was conducted remotely from July 19 to July 20, 2022, following the ALD Workshop that was conducted remotely on July 18, 2022.

The ISAT in Science is designed to measure the attainment of the Idaho State Science Standards adopted by the Department. The assessment is made up of item clusters and stand-alone items. Item clusters represent a series of interrelated student interactions directed toward describing, explaining, and predicting scientific phenomena. Stand-alone items are added to increase the test's coverage of the standards while limiting increases in testing time and burden on students and schools. Test items were developed by CAI, in conjunction with a group of states working to implement three-dimensional science standards. Test items were developed to ensure that each student is administered a test meeting all elements of the ISAT in Science blueprints, which were constructed to align with the Idaho State Science Standards.

Idaho science educators, serving as standard-setting panelists, followed a rigorous standardized procedure to recommend achievement standards demarcating each achievement level. To recommend achievement standards for the new science assessment, panelists participated in the Assertion-Mapping Procedure, an adaptation of the Item-Descriptor (ID) Matching Procedure (Ferrara & Lewis, 2012). Consistent with ordered-item procedures generally (e.g., Mitzel, Lewis, Patz, & Green, 2001), workshop panelists reviewed and recommended achievement standards using an ordered set of scoring assertions[1] derived from student interactions within items. Because the new science items—specifically the item clusters—represent multiple, interdependent interactions through which students engage in scientific phenomena, scoring assertions cannot be meaningfully evaluated independently of the item interactions from which they are derived. Thus, panelists were presented ordered scoring assertions for each item separately rather than for the test overall. Panelists mapped each scoring assertion to the most apt achievement-level descriptor (ALD).

Panelists reviewed ALDs describing the degree to which students have achieved the Idaho State Science Standards. The Department reviewed and revised Range ALDs before the standard-setting workshop. After reviewing the range ALDs, standard-setting panelists worked to identify the

[1] Scoring assertions articulate the evidence the student provides as a means to infer a specific skill or concept, which is aligned to content standards. In other words, scoring assertions capture each measurable action of an item and articulate what evidence the student has provided to infer a specific skill or concept.

knowledge and skills characteristic of students just qualifying for entry into each achievement level.

Working through the ordered scoring assertions for each item, panelists mapped each assertion into one of the four achievement levels—Below Basic, Basic, Proficient, and Advanced. The mapping of scoring assertions was based on the consideration of test content. Panelists were provided additional contextual information, including the percentage of students who performed at or above the achievement level associated with each assertion, as well as the projected National Assessment of Educational Progress (NAEP) science achievement levels of the assertion. The panelists performed the assertion mapping in two rounds of standard setting. Panelists' mapping of the scoring assertions was used to identify the location of the three achievement standards used to classify student performance—Basic, Proficient, and Advanced. Following Round 1, panelists were provided with feedback about the mappings of their fellow panelists and discussed their mappings as a group before they proceeded to Round 2. The achievement standards from Round 2 showed good convergence within each grade-level panel, indicating panelists had a good understanding of the ALDs and the Assertion-Mapping Procedure. Across grades, for a specific achievement level the achievement standards resulted in reasonable percentages. As the standard-setting results from Round 2 concluded the recommended achievement standards, no modifications to the achievement standards were needed.

Twenty-four Idaho science educators were selected to serve as science standard-setting panelists, with ten panelists for the grade 5 panel, seven panelists for the grade 8 panel, and seven panelists for the grade 11 panel. The panelists represented a group of experienced teachers and curriculum specialists, as well as district administrators and other stakeholders. The composition of the panel ensured that a diverse range of perspectives and deep experience with the Idaho State Science Standards contributed to the standard-setting process.

2. STANDARD-SETTING WORKSHOP

The key features of the workshop included the following:

- The standard-setting procedure produced three recommended achievement standards (Basic, Proficient, and Advanced) that will be used to classify student achievement on the ISAT in Science in grades 5, 8 and 11.
- Panelists recommended achievement standards in two rounds.
- Contextual information, including the percentage of students who performed at or above the achievement level associated with each individual assertion (impact data) and the projected NAEP science assessments achievement levels of each assertion (benchmark information), was provided to panelists as part of their review of the ordered assertions.
- The standard-setting workshop was conducted using the CAI's online standard-setting tool. Because the workshop was conducted remotely, the panelists accessed the tool using their own devices. At the end of the workshop, panelists completed online workshop evaluations independently in which they described and evaluated their experience taking part in the standard setting.

3. Results of the Standard-Setting Workshop

Table 1 displays the achievement standards recommended by the standard-setting panelists.

Table 1. Achievement Standards Recommended for Science

Grade	Level 2 Basic Standard	Level 3 Proficient Standard	Level 4 Advanced Standard
5	480	506	534
8	777	807	832
11	1082	1108	1146

Table 2 indicates the percentage of students who will reach or exceed each achievement standard in 2022. Figure 1 represents those values graphically.

Table 2. Percentage of Students Reaching or Exceeding Each Recommended Science Achievement Standard in 2022

Grade	Level 2 Basic Standard	Level 3 Proficient Standard	Level 4 Advanced Standard
5	77	43	10
8	79	41	11
11	73	38	5

Figure 1. Percentage of Students Reaching or Exceeding Each Recommended Science Achievement Standard in 2022

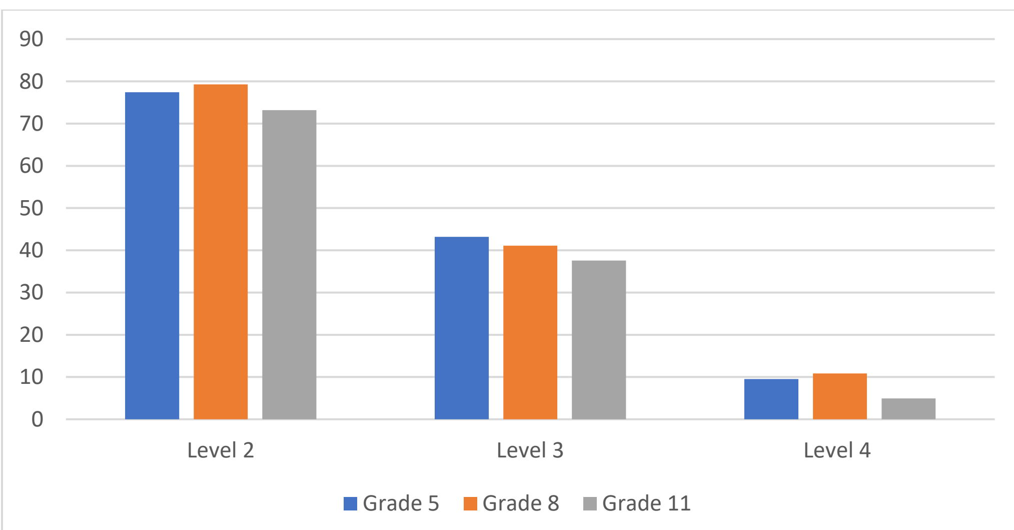


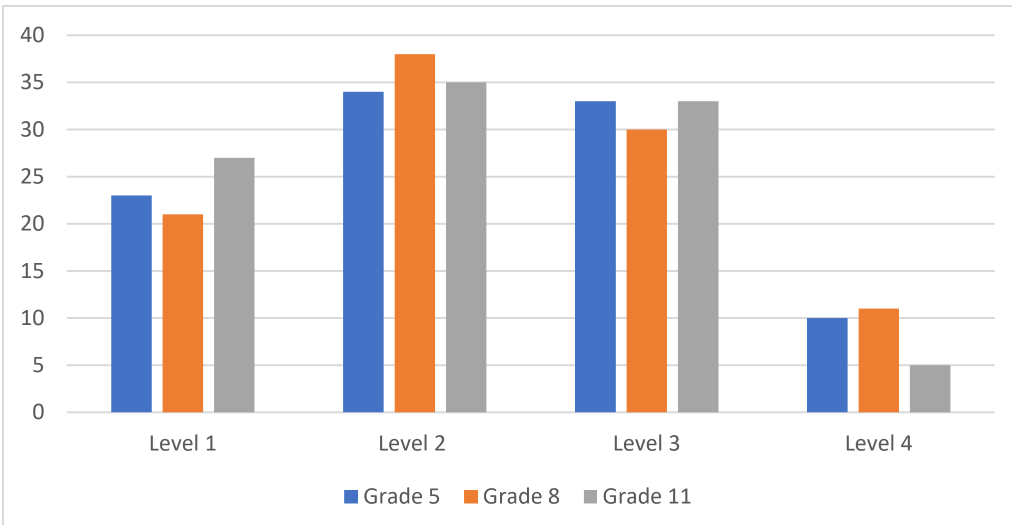
Table 3 indicates the percentage of students classified within each of the achievement levels in 2022. The values are displayed graphically in

Figure 2.

Table 3. Percentage of Students Classified Within Each Science Achievement Level in 2022

Grade	Level 1 Below Basic	Level 2 Basic	Level 3 Proficient	Level 4 Advanced
5	23	34	33	10
8	21	38	30	11
11	27	35	33	5

Figure 2. Percentage of Students Classified Within Each Science Achievement Level in 2022



4. Introduction

In December 2016, the Idaho State Board of Education approved the Idaho State Science Standards, and final legislative approval was obtained in February 2018. The Department and its assessment vendor, Cambium Assessment, Inc. (CAI), developed and administered a new assessment to measure the new standards. In school year 2021–2022, they administered the new assessment aligned to the Idaho State Science Standards to all grades 5, 8, and 11 students in Idaho. Idaho provides information about the ISAT Comprehensive Assessment System at: https://www.sde.idaho.gov/assessment/isat-cas/index.html.

New tests require new achievement standards to link achievement on the test to the content standards. The Department contracted with CAI to establish cut scores for the new tests. To fulfill this responsibility, CAI implemented an innovative, 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. Achievement standards were recommended for grades 5, 8, and 11 in July 2022.

The purpose of this report is to document the standard-setting process for the ISAT in Science and resulting achievement standard recommendations.

5. The Idaho State Science Standards

The ISAT in Science assesses the learning objectives described by the Idaho State Science Standards[2], available at: https://www.sde.idaho.gov/academic/science/. These three-dimensional science standards, based on *A Framework for K–12 Science Education* (National Research Council, 2012), reflect the latest research and advances in modern science education and differ from previous science standards in the following ways. First, rather than describe general knowledge and skills that students should know and be able to do, they describe specific achievements that demonstrate what students know and can do. The Idaho State Science Standards refer to these performed knowledge and skills as *performance standards*. Second, while unidimensionality is a typical goal of standards (and the items that measure them), the Idaho State Science Standards are intentionally multi-dimensional. Each performance standard incorporates three dimensions—a science or engineering practice, a disciplinary core idea, and a crosscutting concept.

Figure 3 shows one of the performance standards for grade 5 in Life Sciences: LS1-5: Molecules to Organisms: Structure and Processes. Further information about the Idaho State Science Standards is available at: https://www.sde.idaho.gov/academic/shared/science/ICS-Science-Legislative.pdf.

Figure 3. Example of Idaho State Science Standards

LS1-5 Molecules to Organisms: Structure and Processes

Performance Standards
Students who demonstrate understanding can:
LS1-5-1. Support an argument that plants get the materials they need for growth chiefly from air and water.
• Further Explanation: Emphasis is on the idea that plant matter comes mostly from air and water, not from the soil.
Supporting Content
LS1.C: Organization for Matter and Energy Flow in Organisms
• Plants acquire their material for growth chiefly from air and water. (LS1-5-1)

Source. Idaho Content Standards : Science (March, 2018) by the Department. Retrieved from: https://www.sde.idaho.gov/academic/shared/science/ICS-Science-Legislative.pdf.

6. ISAT in Science

Due to the unique features of the three-dimensional Idaho State Science Standards, items and tests based on three-dimensional science standards, such as the ISAT in Science, must also incorporate similarly unique features. The most impactful of these changes is that the three-dimensional science tests are multi-dimensional and are thus made up mostly of *item clusters* representing a series of interrelated student interactions directed toward describing, explaining, and predicting scientific phenomena.

[2] The term "Idaho State Content Standards in Science" and the term "Idaho State Science Standards" were used interchangeably in this technical report.

6.1 Item Clusters and Stand-Alone Items

Item clusters include a stimulus and a series of questions that generally take students approximately 6–12 minutes to complete. They consist of a phenomenon—an observable fact or design problem—that an engaged student explains, models, investigates, or designs using the knowledge and skill described by the performance standard to complete a series of activities (made up of multiple interactions). For example, in Figure 3, proficiency in this single performance standard requires activities that demonstrate understanding and knowledge of how plants grow. The stimulus in an item cluster explicitly states a task or goal (for example, "In the questions that follow, you will analyze what happens to the plant when it is put in an open area under the Sun and when water is supplied.") and subsequent interactions build on or relate to the task or response to previous questions. The interactions within an item cluster all address the same phenomenon.

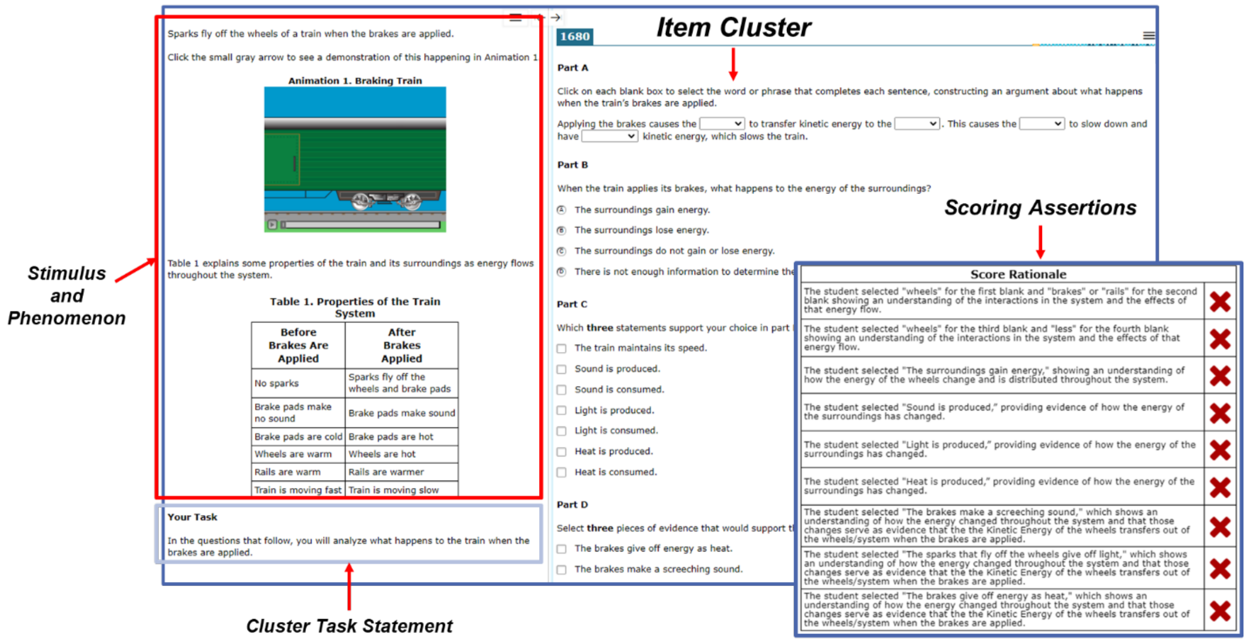
Some added stand-alone items increase the coverage of the test without also increasing testing time or testing burden. Stand-alone items are shorter, are unrelated to other items, and generally take students one to three minutes to complete. Within each item cluster is a variety of interaction types, including selected response, multi-select, table match, edit in-line choice, and simulations of science investigations. Stand-alone items can also be these types.

6.2 Scoring Assertions

Each item cluster and stand-alone item assumes a series of explicit assertions about the knowledge and skills that a student demonstrates based on specific features of the student's responses across multiple interactions. *Scoring assertions* capture each measurable action and articulate what evidence the student has provided to infer a specific skill or concept. Some stand-alone items have more than one scoring assertion, while all item clusters have multiple scoring assertions.

Figure 4 illustrates an item cluster and associated scoring assertions.

Figure 4. Example of the Three-Dimensional Science Standards Item Cluster and Scoring Assertions

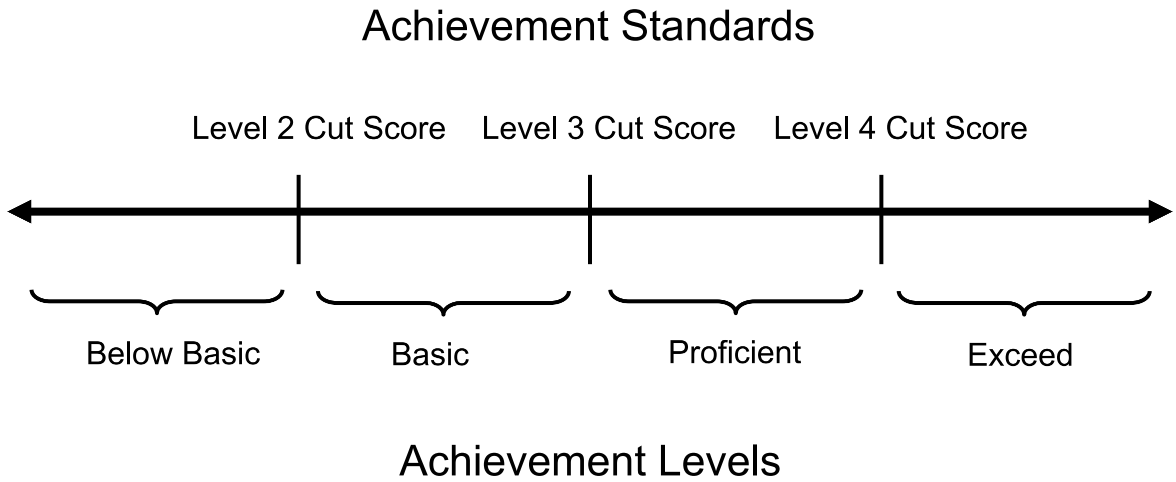


7. STANDARD SETTING

Twenty-four educators from Idaho convened remotely July 19 and 20, 2022, to complete two rounds of standard setting to recommend three achievement standards for the ISAT in Science.

Standard setting is the process to define achievement on the test. Achievement levels are defined by achievement standards, or *cut scores*, that specify how many of the performance standards students must know and be able to fulfill in order to meet the minimum for each achievement level. As shown in Figure 5, three achievement standards are sufficient to define Idaho's four achievement levels.

Figure 5. Three Achievement Standards Defining Idaho's Four Achievement Levels



The cut scores are derived from the knowledge and skills measured by the test item scoring assertions that students at each achievement level are expected to be able demonstrate in order to receive credit.

7.1 THE ASSERTION-MAPPING PROCEDURE

A modification of traditional approaches to setting achievement standards is necessary for tests based on three-dimensional science standards due to the structure of the achievement standards, and subsequently, the structure of test items assessing the achievement standards. While traditional tests and measurement models assume unidimensionality, tests based on three-dimensional science standards adopt a three-dimensional conceptualization of science understanding. Each item cluster or stand-alone item aligns with a science practice, one or more crosscutting concepts, and one disciplinary core idea. Accordingly, the new science assessment is made up mostly of item clusters representing a series of interrelated student interactions directed toward describing, explaining, and predicting scientific phenomena. Some stand-alone items are added to increase the test's coverage of the standards without also increasing testing time or testing burden.

Within each item, a series of explicit assertions are made regarding the knowledge and skills that a student has demonstrated based on specific features of the student's responses across multiple interactions. For example, students may correctly graph data points indicating that they can construct a graph showing the relationship between two variables but may make an incorrect inference regarding the relationship between the two variables, thereby not supporting the assertion that they can interpret relationships expressed graphically.

While some other assessments, especially ELA, comprise items probing a common stimulus, the degree of interdependence among such items is limited and student performance on such items can be evaluated independently of student performance on other items within the stimulus set. This is not the case with the new science items, which may, for example, involve multiple steps in which students interact with products of previous steps. However, unlike traditional stimulus- or passage-based items, the conditional dependencies between the interactions and resulting assertions of an item cluster are too substantial to ignore because those item interactions and assertions are more intrinsically related to each other. The interdependence of student interactions within items has consequences both for scoring and for recommending achievement standards.

To account for the cluster-specific variation of related item clusters, additional dimensions can be added to the item response theory (IRT) model. Typically, these are nuisance dimensions unrelated to student ability. Examples of IRT models that follow this approach are the bi-factor model (Gibbons & Hedeker, 1992) and the testlet model (Bradlow, Wainer, & Wang, 1999). The testlet model is a special case of the bi-factor model (Rijmen, 2010).

Because the item clusters represent performance tasks, the Body of Work (BoW) method (Kingston, Kahl, Sweeny, & Bay, 2001) could also be appropriate for recommending achievement standards. However, the BoW method is manageable only with small numbers of performance tasks and quickly becomes onerous when the number of item clusters approaches 10 or more.

Skaggs, Hein, & Awuor (2007) proposed a standard setting method called the Single-Passage Bookmark method to address challenges presented by passage-based assessments. This method is a variation of the traditional Bookmark method (e.g., Mitzel, Lewis, Patz, & Green, 2001) in which individual ordered item booklets (OIBs) are created for each set of items associated with a passage. Items within each OIB are arranged in order of difficulty. The task of the panelists is to place a bookmark in each OIB as opposed to a single OIB in the traditional Bookmark method. Even though this method showed promise, one limitation and concern expressed by the authors is whether this method can be applied to derive two or more standards.

To address these challenges, Cambium Assessment, Inc. (CAI) psychometricians designed a new method for setting achievement standards on cluster-based assessments. CAI implemented this method for the New Hampshire, Utah, and West Virginia statewide assessments in 2018, for the Connecticut, Oregon, and the joint Multi-State Science Assessment (MSSA) for Rhode Island and Vermont in 2019, and for the Hawaii, North Dakota, South Dakota, and Utah statewide assessments in 2021. The method was also implemented for the Virgin Islands, Montana, and Wyoming statewide assessments in 2022.

The test-centered Assertion-Mapping Procedure (AMP) is an adaptation of the Item-Descriptor (ID) Matching procedure (Ferrara & Lewis, 2012) that preserves the integrity of the item clusters while also taking advantage of ordered-item procedures such as the Bookmarking procedure used frequently for other accountability tests (Rijmen, Cohen, Butcher, & Farley, 2018).

The main distinction between AMP and the Single-Passage Bookmark method is that the panelists evaluate scoring assertions rather than individual items. Scoring assertions are not test items, but inferences that are supported (or not supported) by students' responses in one or more interactions within an item cluster or stand-alone item. Because item clusters represent multiple, interdependent interactions through which students engage in scientific phenomena, scoring assertions cannot be meaningfully evaluated independently of the item from which they are derived. Therefore, the scoring assertions from the same item cluster or stand-alone item are always presented together. Within each item cluster or stand-alone item, scoring assertions are ordered by difficulty (i.e., the IRT difficulty parameter) consistent with the Single-Passage Bookmark method. One can think of the resulting booklet as consisting of different chapters, where each chapter represents an item cluster or stand-alone item. Within each chapter, the (ordered) pages represent scoring assertions. As in ID matching, panelists are asked to map each scoring assertion to the most apt achievement-level descriptor during two rounds of standard setting. As with the Bookmark method, assertion mappings are made independently with the goal of convergence over two rounds of rating, rather than consensus.[3]

7.2 Workshop Structure

One large virtual meeting room served as an all-participant training room. This room broke into three separate virtual working rooms, one for each set of grade-level panels, after the all-group orientation. As shown in Figure 6, three separate panels set achievement standards for each grade.

Figure 6. Virtual Workshop Panels, per Room

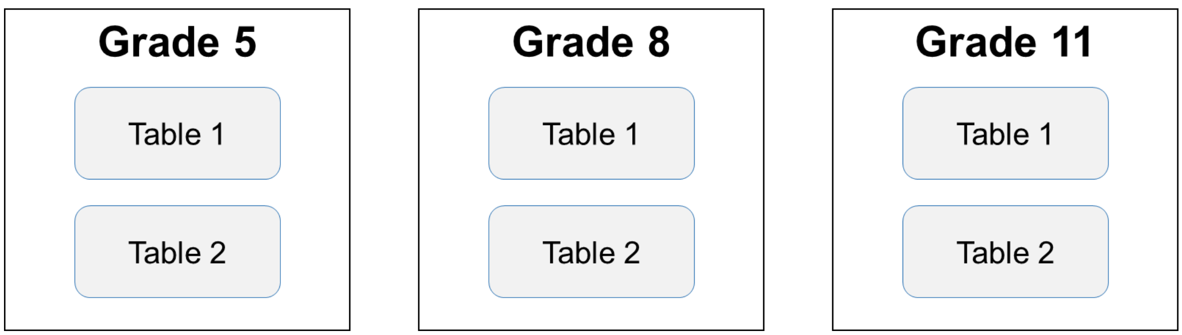


Table 4. Table summarizes the composition of the tables and the number of facilitators and panelists assigned to each. The 24 standard-setting panelists included table leaders and panelists from Idaho who taught in the content area and grade for which standards were being set.

Table 4. Table Assignments

Room	Grade	Tables and Table Leaders (One per Table)	Panelists (per Table)	Facilitator	Facilitator Assistant
1	5	2	5 / 5	Kam Mangis de Mark Hibbah Haddam Kevin Dwyer	Cameron Benham Kylie Dennis
2	8	2	4 / 3	Heather MacRae James McCann	Mary Cochron Sydney Brabble
3	11	2	3 / 4	Matthew Davis Mark Palamo	Alesha Ballman

[3] CAI historically implements two rounds of standard setting as best practice in the Bookmark method and extends this practice to the AMP method. In addition to lessening the panelists' burden of needing to repeat a cognitively demanding task for a third time, using two rounds introduces significant cost efficiency by reducing the number of days needed for standard setting. Panels typically converge in Round 2, and panelists completing two rounds report levels of confidence in the outcomes that are similar to the confidence expressed by panelists participating in three rounds. Psychometric evaluation of the reliability and variability in results from two and three rounds are generally consistent. CAI has used two rounds in standard setting in more than 17 states and 38 assessments, beginning in 2001 with the enactment of the No Child Left Behind (NCLB) Act.

7.3 Participants and Roles

7.3.1 Idaho Department of Education Staff

The Director of Assessment, Kevin Chandler[4], from the Department participated in the process and provided overall policy context and answered any policy questions that arose.

7.3.2 Cambium Assessment, Inc. Staff

CAI facilitated the workshop and each of the content-area rooms, provided psychometric and statistical support, and oversaw technical set-up and logistics. CAI team members were highly qualified to lead the workshop and conduct analyses, and included the following:

- Dr. Frank Rijmen, Senior Director of Psychometrics, supervised all psychometric analyses conducted during and after the workshop and provided training to participants.
- Dr. Yi-Fang Wu, Senior Psychometrician, and Dr. Jiajun Xu, Psychometrician, provided psychometric analyses.
- Alesha Ballman, Senior Psychometric Project Coordinator, oversaw analytics technology and psychometrics.
- Sydney Brabble and Kylie Dennis, Psychometric Support Assistants, provided support as needed.
- Cameron Benham and Mary Cochron[5], Program Management Team, managed process and logistics throughout the meeting.
- Floyd Helm and Mark Palomo, System Support Agents, troubleshot technology during the workshop.

7.3.3 Room Facilitators

Two to three CAI facilitators guided the process in each grade-level room. Facilitators were content experts experienced in leading standard-setting processes, had led standard-setting processes before, and could answer any questions about the workshop or about the items or what the items were intended to measure. They also monitored time and motivated panelists to complete tasks within the scheduled time. Facilitators were:

- Kam Mangis de Mark, Hibbah Haddam, and Kevin Dwyer facilitated the grade 5 panel.
- Heather MacRae and James McCann facilitated the grade 8 panel.
- Matthew Davis and Mark Palamo facilitated the grade 11 panel.

Each facilitator was trained to be extensively knowledgeable of the constructs, processes, and technologies used in standard setting.

[4] The former Director of Assessment during the standard-setting workshop
[5] Former staff at CAI during the standard-setting workshop

7.3.4 Educator Participants

To establish achievement standards, the Department recruited a set of panelists from across Idaho. Panelists included science teachers, administrators, and representatives from other stakeholder groups (e.g., parents, college faculty) to ensure that a range of perspectives contributed to the standard-setting process and product. In recruiting panelists, the Department targeted the recruitment of panelists to be representative of the gender and geographic representation of Idaho's teacher population. All panelists also had to be familiar with the Idaho State Science Standards content and test. During the two-day workshop, all panelists attended each training section and mapping section.

The Department selected classroom teachers from the resulting potential panelist pool and invited them to participate in the workshop. Overall, the standard-setting workshop panelists were 13% male and 4% non-white. Represented stakeholder groups included administrators; coaches; teachers in general education, higher education, and special education, and parents, with general education teachers making up 83% of the panels overall. Most panelists taught in the grade band to which they were assigned to set standards. Overall, 38% of panelists indicated grade 5 as their primary grade taught, 17–21% taught middle school (grades 6 through 8), and 25–29% taught high school (grades 9 through 12). Most panelists worked in schools (92%); some worked in districts and schools (4%) and charter schools (4%). School district areas included rural (38%), suburban (38%), and urban (17%) areas, and were small (33%), medium (38%), and large (25%). Table 5 summarizes the characteristics of the panels.

Table 5. Panelist Characteristics

	Percentage (%) of Panelists, by Panel			
	Grade 5	*Grade 8*	*Grade 11*	*Overall*
Characteristics				
Male	10	0	29	13
Non-White	0	14	0	4
Stakeholder Groups[a]				
General Education Teacher	90	86	71	83
Coach	10	14	0	8
Administrator	0	0	29	8
Special Education Teacher	0	29	14	13
Higher Education	0	0	14	4
Parent	0	29	0	8
Current Position				
School	100	86	86	92
District, School	0	0	14	4
School, Other[b]	0	14	0	4
School District Size				
Large	30	14	29	25

	Percentage (%) of Panelists, by Panel			
	Grade 5	*Grade 8*	*Grade 11*	*Overall*
Medium	40	43	29	38
Small	30	43	29	33
Not Applicable	0	0	14	4
School District Area Urbanicity				
Urban	40	0	0	17
Suburban	30	29	57	38
Rural	30	71	14	38
Not Applicable	0	0	29	8
Primary Grades Taught[c]				
Preschool	0	0	14	4
Kindergarten	20	0	14	13
Grade 1	30	0	14	17
Grade 2	30	14	14	21
Grade 3	20	14	14	17
Grade 4	40	29	14	29
Grade 5	70	14	14	38
Grade 6	10	43	14	21
Grade 7	0	43	14	17
Grade 8	10	29	14	17
Grade 9	0	14	71	25
Grade 10	0	14	86	29
Grade 11	0	14	86	29
Grade 12	0	14	86	29
College	0	0	14	4

Note. [a]The total is more than 100% for "Stakeholder Groups" as panelists had multiple roles in local education systems. [b]Other stakeholder groups include charter schools. [c]The total is more than 100% for "Primary Grades Taught" because panelists had multiple primary grades they taught in local education systems.

For the results of any judgment-based method to be valid, the judgments must be made by individuals who are qualified to make them. Panelists in the ISAT in Science standard-setting workshop were highly qualified. They brought a variety of experience and expertise. Overall, 63% of panelists had earned a master's degree or higher. Nearly half (46%) had taught in their assigned panel's grade and subject for one to 15 years while 30% had taught it for more than 16 years. The average time teaching the Idaho State Science Standards was seven years. Most had experience teaching special populations: 96% taught students eligible to receive free or reduced-price lunch, 88% taught English learners (ELs), and 96% taught students on an Individual Education Plan (IEP). Table 6 summarizes the qualifications of the panels.

Table 6. Panelist Qualifications

	Percentage (%) of Panelists, by Panel			
	Grade 5	*Grade 8*	*Grade 11*	*Overall*
Highest Degree				
Bachelor	20	43	29	29
Master	70	57	57	63
Doctoral	10	0	14	8
Years Teaching Experience				
None	0	0	0	0
Less than 1 year	0	0	0	0
1–5 years	10	43	0	17
6–10 years	10	0	14	8
11–15 years	30	29	14	25
16–20 years	30	14	29	25
More than 20 years	20	14	43	25
Years Teaching Experience in Assigned Grade				
None	30	29	14	25
Less than 1 year	10	0	0	4
1–5 years	10	43	14	21
6–10 years	20	0	14	13
11–15 years	10	14	0	8
16–20 years	20	14	14	17
More than 20 years	0	0	43	13
Subject Areas Currently Teaching[a]				
English Language Arts (ELA)	70	29	0	38
Mathematics	70	43	14	46
Social Studies	70	29	0	38
Science	80	86	71	79
Not Applicable[b]	10	0	14	8
Other[c]	10	29	14	17
Other professional experience in education	30	14	14	21
Years Professional Experience in Education				
None	70	86	86	79
Less than 1 year	0	0	0	0
1–5 years	30	14	0	17
6–10 years	0	0	14	4
11–15 years	0	0	0	0

	Percentage (%) of Panelists, by Panel			
	Grade 5	Grade 8	Grade 11	Overall
16–20 years	0	0	0	0
More than 20 years	0	0	0	0
Experience Teaching Special Student Populations				
Students eligible to receive free/reduced price lunch	100	100	86	96
English Learners (ELs)	90	100	71	88
Students on an Individual Education Plan (IEP)	100	100	86	96
Average years teaching Idaho State Science Standards[d]	6	5	11	7

Note. [a]The total is more than 100% for "Subject Areas Currently Teaching" because many panelists taught multiple subjects. [b]One grade 5 panelist and one grade 11 panelist identified themselves as Coach and Administer, respectively. [c]Other "Subject Areas Currently Teaching" includes Pre-Engineering, Electives, General Studies K–5, and Special Education Teacher and Director. [d]The unit of measures is "year".

Appendix 3-A, Standard-Setting Panelist Characteristics, provides additional information about the individuals participating in the standard-setting workshop.

7.3.5 Table Leaders

Volunteers from the participant pool served as table leaders, who were intimately familiar with students and the subject matter. The day prior to the standard-setting workshop, the group of Idaho educators selected to be the table leaders convened to review, revise, and approve the range achievement-level descriptors (ALDs; see the next section for more details). During the standard-setting workshop, table leaders served as panelists. In addition, table leaders had the responsibility of participating in the moderation session, if moderation was needed.

7.4 MATERIALS

7.4.1 Achievement-Level Descriptors

With the adoption of the new standards in science, and the development of new statewide assessment to assess achievement of those standards, the Department must adopt a similar system of achievement standards[6] to determine whether students have met the learning goals defined by the new standards in science.

Determining the nature of the categories into which students are classified is a prerequisite to standard setting. These categories, or achievement levels, are associated with achievement-level

[6] At the time of the standard setting workshop, performance standards, performance levels, and performance-level descriptors (PLDs) were used in relevant materials and documentation, based on the agreement with the Department prior to the standard setting. In this report, achievement standards, achievement levels, and achievement-level descriptors (ALDs) were used to reflect the most recent decision and request by the Department. In the appendices used during the workshop, however, the term "performance" was used to refer to as "achievement," reflecting the actual settings in July 2022.

descriptors (ALDs) that define the content-area knowledge, skills, and processes that students at each achievement level can demonstrate.

ALDs link the content standards to the achievement standards. There are four types of ALDs:

1. **Policy ALDs.** These are brief descriptions of each achievement level that do not vary across grade or content area.
2. **Range ALDs.** Provided to panelists to review and endorse during the workshop, these detailed grade- and content-area-specific descriptions communicate exactly what students performing at each level know and can do.
3. **Threshold ALDs.** Typically created during and used for standard setting only, these describe what a student just barely scoring into each achievement level knows and can do. They may also be called Target ALDs or Just Barely ALDs.
4. **Reporting ALDs.** These are much-abbreviated ALDs (typically 350 or fewer characters) created following client approval of the achievement standards used to describe student achievement on score reports.

Idaho uses four achievement levels to describe student achievement: “Below Basic”, “Basic”, “Proficient”, and “Advanced”. At the policy level, these achievement levels are defined as follows:

- **Below Basic.** The student has not met the achievement standard and needs substantial improvement to demonstrate the knowledge and skills in science needed for likely success in future coursework.
- **Basic.** The student has nearly met the achievement standard and may require further development to demonstrate the knowledge and skills in science needed for likely success in future coursework.
- **Proficient.** The student has met the achievement standard and demonstrates progress toward mastery of the knowledge and skills in science needed for likely success in future coursework.
- **Advanced.** The student has exceeded the achievement standard and demonstrates advanced progress toward mastery of the knowledge and skills in science needed for likely success in future coursework.

7.4.1.1 Science Range Achievement-Level Descriptor Development

CAI and staff from participating states’ Departments of Education (DOE) reviewed existing range ALDs from several states’ assessments based on three-dimensional science standards, the Next Generation Science Standards (NGSS). States selected the range ALDs based on the standards drafted by the Department as a starting point. Subsequently, CAI, state DOE staff, and educators from multiple states using science assessments based on the Shared Science Assessment Item Bank convened in May 2018 to review and refine the draft range ALDs.[7] The panels created policy

[7] These states included Hawaii, New Hampshire, Oregon, Rhode Island, Utah, Vermont, West Virginia, and Wyoming.

ALDs and reviewed and identified refinements to the range ALDs to describe observable evidence for what student achievement looks like in science at each achievement level and grade. CAI and one of the NGSS authors reviewed and applied recommendations to the ALDs. They ensured consistency, coherence, and articulation across grades and levels. The first part of *Appendix 3-B, Development of Science Range Achievement-Level Descriptors*, provides additional information about the development of the range ALDs prior to states' standard-setting workshops.

7.4.1.2 The Department and Panelist Range Achievement-Level Descriptor Review

The Department then reviewed the ALDs to ensure that the language accurately represented the goals and policies of Idaho. CAI worked with them to make revisions where necessary.

In July 2022, the group of Idaho educators selected to be standard-setting table leaders, who were intimately familiar with students and the subject matter, convened to review, revise, and approve the range ALDs. The second part of *Appendix 3-B, Development of Science Range Achievement-Level Descriptors*, provides training slides of the ISAT in Science ALDs[8] meeting. *Appendix 3-C, ISAT in Science Range Achievement-Level Descriptors*, provides the final range ALDs for the ISAT in Science.

7.4.2 Ordered Scoring Assertion Booklets

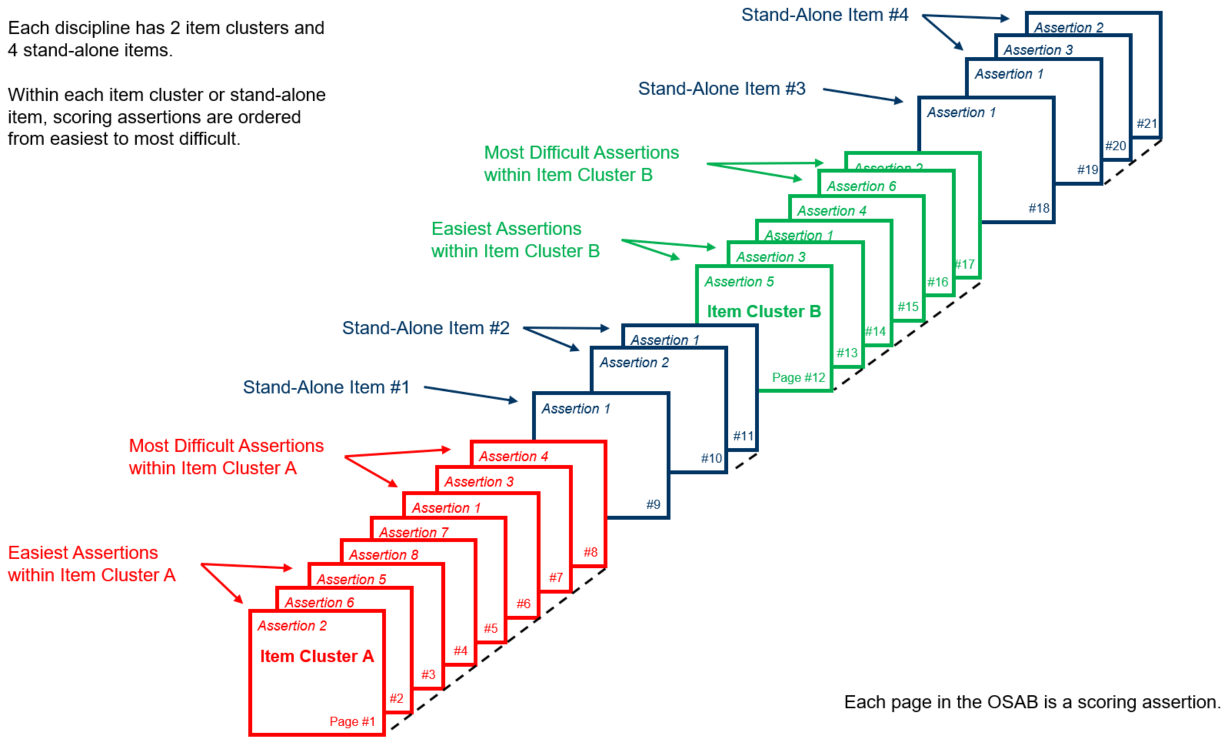
Like the Bookmark method used for establishing achievement standards for traditional science tests, the AMP method uses booklets of ordered test materials for setting standards. Instead of test items, the AMP uses scoring assertions presented in grade-specific booklets called *ordered scoring assertion booklets* (OSABs). Each OSAB represents one possible testing instance resulting from applying the test blueprints to Idaho's item pool.

The OSABs were assembled using a mixed-integer programming approach. The objective function that was minimized was the number of gaps between the impact values of the assertions across the entire OSAB. A gap was defined as a difference of three percent or more between the impact values of two consecutive assertions ordered by difficulty. The linear constraints of the mixed-integer problem represented the constraints implied by the blueprint. In addition, the total number of assertions was not allowed to exceed 85. A set of feasible solutions was further evaluated based on the distribution of the impact values of assertions across the OSAB. The candidate solution was then reviewed internally by content experts and by the Department and approved without any changes for all three grades.

Figure 7 describes the structure of the OSAB.

[8] As shown in the slides, the term "performance-level descriptor" was used when the Department and CAI conducted the meeting in summer 2022. It is referred to as "achievement-level descriptor" in this technical report.

Figure 7. Ordered Scoring Assertion Booklet (OSAB)



For the operational test, the order of the items was randomized over students. The items in the OSABs were grouped by science content area, so that panelists work through all items associated with one content area before moving on to the next. This allowed panelists to focus on the knowledge and skill requirements for one content area at a time. For each of the grades 5, 8, and 11 OSABs, the Physical Sciences discipline items were presented first, then Life Sciences items, and then Earth and Space Sciences items.

For grades 5 and 8, two item clusters and four stand-alone items represent each discipline; for grade 11, two item clusters and five stand-alone items represent the Physical Sciences, and two item clusters and four stand-alone items represent each of the Life Sciences and Physical Sciences. Within a discipline, the item clusters were presented first, followed by the stand-alone items. The item clusters and stand-alone items were further ordered by mean difficulty of the assertions in the item. This approach may help to reduce some of the cognitive demands on panelists by making clear that some items, and their associated interactions, are easier for students to access, even though the assertions they support are similar in content.

Within each item cluster or stand-alone item, scoring assertions were ordered by difficulty. Easier assertions are those that most students were able to demonstrate, and difficult assertions are those that the fewest students were able to demonstrate. Note that assertions were ordered by difficulty within items only. Across all items, this was generally not the case; for example, the most difficult assertion of an item presented early in the OSAB was typically more difficult than the easiest assertion of the next item in the OSAB. That is, the order of assertions in Figure 7. Ordered Scoring Assertion Booklet (OSAB) represents the order of presentation to the panelists, but assertions were not ordered by overall difficulty across all items (see Figure 8 for a depiction of the overlapping difficulty of assertions in the complete OSAB).

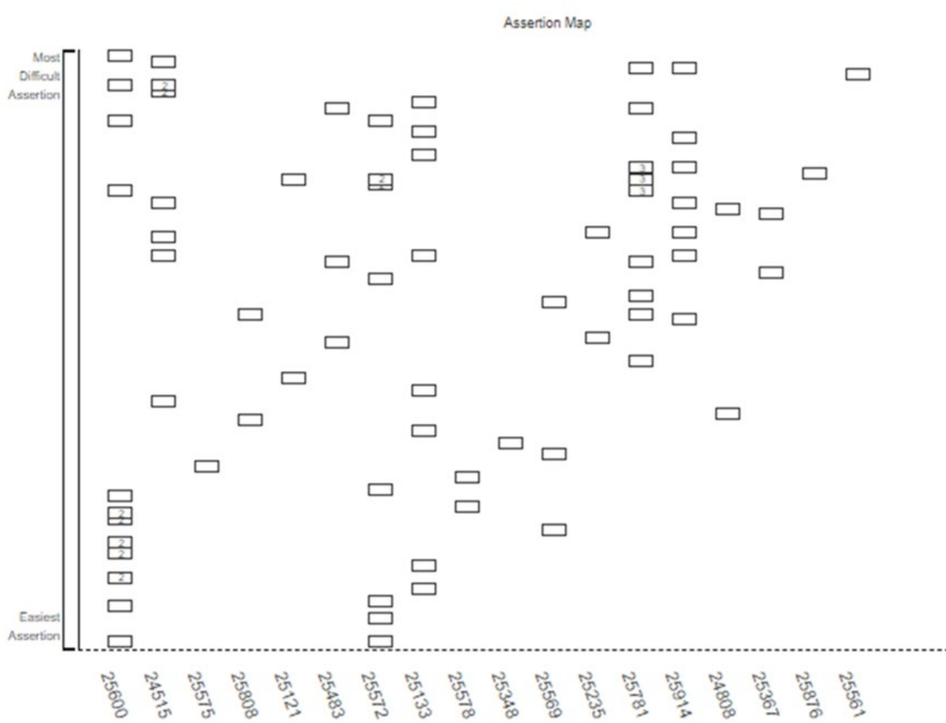
Not all items have assertions that will map onto all achievement levels. For example, an item cluster may have assertions that map onto “Below Basic,” “Basic,” and “Proficient,” but not “Advanced.”

Each of the grade 5 and grade 8 OSABs contained 18 items and the grade 11 OSAB contained 19 items. The grade 5 OSAB contained 75 assertions, the grade 8 OSAB contained 75 assertions, and the grade 11 OSAB contained 75 assertions. Each comprised six item clusters and 12 stand-alone items.

7.4.3 Assertion Maps

Assertion maps were provided to panelists to help reduce the cognitive load of the AMP. The assertion maps were displayed in CAI’s online standard-setting tool and listed all scoring assertions in each OSAB by item ID and assertion and plotted all assertions by difficulty. The assertion maps provided panelists with context about student performance on the assertions in the OSAB, describing the difficulty of each assertion in the underlying OSAB. This was to help panelists easily identify more- or less-difficult assertions and compare the difficulty of assertions across items. The assertion maps were provided during the OSAB review. After Round 1, the assertion maps were updated to also display the tentative standards; more details are provided in Section 7.7.2.2, Feedback Data. Figure 8 presents the assertion map for grade 5. The assertion maps for all three grades are presented in *Appendix 3-D, Standard-Setting Assertion Maps*.

Figure 8. Standard-Setting Assertion Map, Science Grade 5

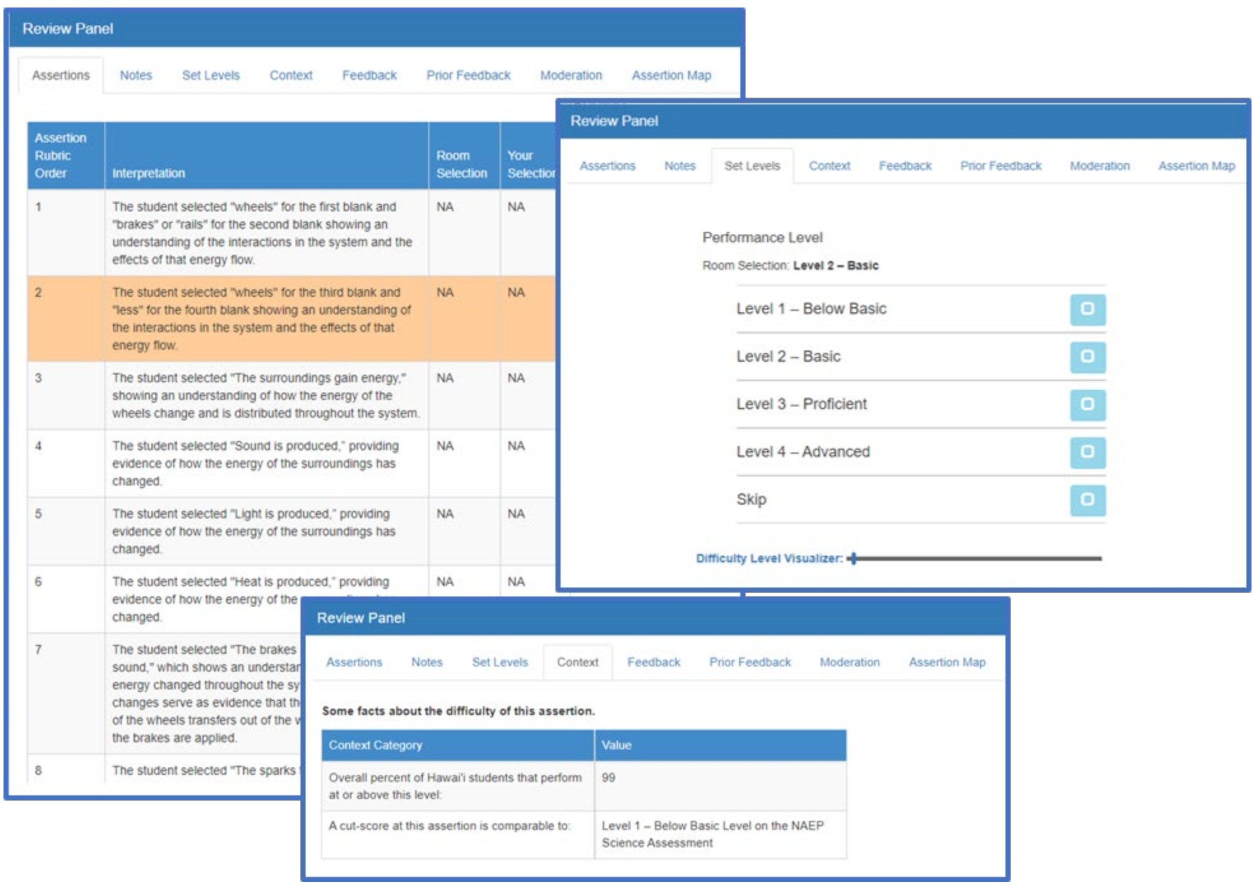


7.5 Workshop Technology

The standard-setting panelists used CAI's online application for standard setting. Each panelist used his or her own device on which he or she took the test, reviewed item clusters and stand-alone items and ancillary materials, and mapped assertions to achievement levels.

Using tabs in the review panel of the tool (see Figure 9), panelists could review the items and scoring assertions, determine the relative difficulty of assertions to other assertions in the same item, examine the content alignment of each item (via the alignment of the assertions within an item, which all align to the same performance standard), assign assertions to achievement levels, add notes and comments on the assertions as they reviewed them, and review contextual information and feedback data. Additionally, they had access to a difficulty level visualizer, a graphic representation of the difficulty of each assertion relative to all other assertions in the OSAB (not just within the item).[9] Panelists also reviewed their assertion placement, their table's placement, the other tables' placement, and the overall placement for all tables.

Figure 9. Example Features in Standard-Setting Tool



Full-time CAI information technology specialists answered questions and ensured that technological processes ran smoothly and without interruption throughout the meeting.

[9] The difficulty level visualizer represented the percentage of students whose ability level would fall at or above the difficulty level of that assertion.

7.6 Events

The standard-setting workshop occurred over two days. Table 7 summarizes each day's events, and this section describes each event listed in greater detail. *Appendix 3-E, Standard-Setting Workshop Agenda*, provides the full workshop agenda.

Table 7. Standard-Setting Agenda Summary

Day 1: Tuesday, July 19, 2022
• Large-Group Orientation • Review and Take the Operational Test • Review Range ALDs* • Discuss Threshold ALDs • OSAB Review
Day 2: Wednesday, July 20, 2022
• Continue OSAB Review • Assertion-Mapping Training • Round 1 Assertion Mapping • Round 1 Feedback and Impact Data Review and Discussion • Round 2 Assertion Mapping • Round 2 Feedback and Impact Data Review • Standard-Setting Workshop Evaluations • Across-Grade Moderation and Articulation

Note. *During the standard-setting workshop, performance standards, performance levels, and performance-level descriptors (PLDs) were used based on the agreement with the Department. In this report, achievement standards, achievement levels, and achievement-level descriptors (ALDs) were used to reflect the most recent decision by the Department.

7.6.1 Participant Login

Panelists were required to attend a technical check prior to the standard-setting workshop to ensure they had access to the sites needed to participate in the workshop. They also received and signed affidavits of non-disclosure at this time, affirming that they would not reveal any secure information they would have access to during the workshop. Panelists arrived at the workshop, virtually, on the first day, and followed the instructions given for joining the workshop via Microsoft Teams.

7.6.2 Large-Group Orientation

Kevin Chandler, Director of Assessment from the Department, welcomed panelists to the workshop and provided context and background for the ISAT in Science. He also outlined the roles and responsibilities of the participants at the workshop: panelists, CAI staff, and the Department personnel. Dr. Rijmen then oriented participants to the workshop by describing the purpose and objectives of the meeting, explaining the process to be implemented to meet those objectives, and outlining the events that would happen each day. He explained that panelists were selected because they were experts, and how the process to be implemented over the two days was designed to elicit and apply 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 3-F, Standard-Setting Training Slides*, provides the slides used during the large-group training.

7.6.3 Confidentiality and Security

Workshop leaders and room facilitators addressed confidentiality and security during orientation and again in each room. Standard setting uses live science test items from the ISAT in Science, requiring confidentiality to maintain their security. Participants were forbidden to do the following either during, or after, the workshop:

- Discuss the test items outside of the meeting
- Discuss judgments or cut scores (their own or others') with anyone outside of the meeting
- Discuss secure materials with non-participants
- Create any form of electronic copy of test content (screenshots, electronic notes, etc.)
- Create any hand-written notes of test content
- Use their devices during the meeting for any purpose other than participating in the standard-setting workshop and item review (e.g., email, web browsing, social media)
- Save notes about item or passage content to their devices

Participants could have general conversations regarding 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.

7.6.4 Take the Test

Following the large-group orientation, panelists broke out into their separate grade-level virtual meeting rooms. As their introduction to the standard-setting process, panelists took a form of the test that students took in 2022, in the grade band to which they would be setting achievement standards. They took the tests online via the same tool used to deliver operational tests to students, and the testing environment closely matched that of students when they took the test.

Taking the same test students take provides the opportunity to interact with and become familiar with the test items and the look and feel of the student experience while testing. The panelists could score their responses and had 90 minutes to interact with the test.

7.6.5 Range Achievement-Level Descriptor Review

After taking the operational test, panelists completed a thorough review of the range ALDs for their assigned grade. Panelists were provided with an overview of the ALDs and their importance to standard setting. The ALDs were used as a reference for evaluating student achievement, so it was important for panelists to understand the critical role of ALDs in the standard-setting process.

Panelists began their review of the range ALDs that define what students in each achievement level know and can do with respect to the Idaho State Science Standards. Workshop facilitators provided panelists with draft range ALDs, test blueprints, and the Idaho State Science Standards.

The facilitators led panelists through a thorough review of the range ALDs for their assigned grade using the materials as references and drawing on the expertise of the panelists.

Panelists identified key words describing the skills necessary for achievement at each level and discussed the skills and knowledge that differentiate achievement in each of the four levels.

Reviewing the range ALDs ensured that panelists understood what students in Idaho should know and be able to do and how much knowledge and skill students are expected to demonstrate at each level of achievement.

7.6.6 Discuss Threshold Achievement-Level Descriptors

After reviewing and discussing the range ALDs, panelists worked in their grade-level groups to develop a shared understanding of the threshold ALDs that describe the skills that students just barely able to score in one achievement level have but students scoring just below the achievement level do not have. Facilitators encouraged panelists to consider the characteristics of students who just barely qualify for entry into the achievement level from those just below. Looking at each ALD, panelists identify the skills needed to just barely perform at that level. The following two questions guide the process:

1. What skills and knowledge must the student demonstrate to qualify for entrance into this achievement level?
2. How does this differ from the upper range of the adjacent (lower) achievement level?

These discussions yielded common descriptions of students just barely characterized by each ALD in each room.

The AMP employs the range ALDs since panelists are mapping items across the full range of each ALD. The purpose of the threshold ALD discussion was to enhance the panelists' understanding of the differences between ALD levels by paying special attention to the transition areas between achievement levels.

7.6.7 Ordered Scoring Assertion Booklet Review

After reviewing and discussing the ALDs, panelists reviewed the item clusters, stand-alone items, and assertions in the OSAB. They took notes on each assertion to document the interactions required by each and described why an assertion might be more or less difficult than the previous assertion in the item. They also noted how each assertion related to the ALDs.

After reviewing the item interactions and scoring assertions individually, panelists engaged in discussion with group members about the skills required and relationships among the reviewed test materials and achievement levels This process ensured that panelists built a solid understanding of how the scoring assertions relate to the item interactions and how the items relate to the ALDs, and it also helped facilitate a common understanding among workshop panelists.

7.6.8 Assertion-Mapping Training

After reviewing the entire OSAB, facilitators described the processes for mapping assertions and determining cut scores. They 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.

Panelists were to match each assertion to the achievement level best supported by the assertion using the ALDs, the difficulty level visualizer (described in Section 7.5, Workshop Technology), the assertion map (described in Section 7.4.3, Assertion Maps) their notes from the OSAB review, and their professional judgments. Figure 10 illustrates the assertion-mapping process.

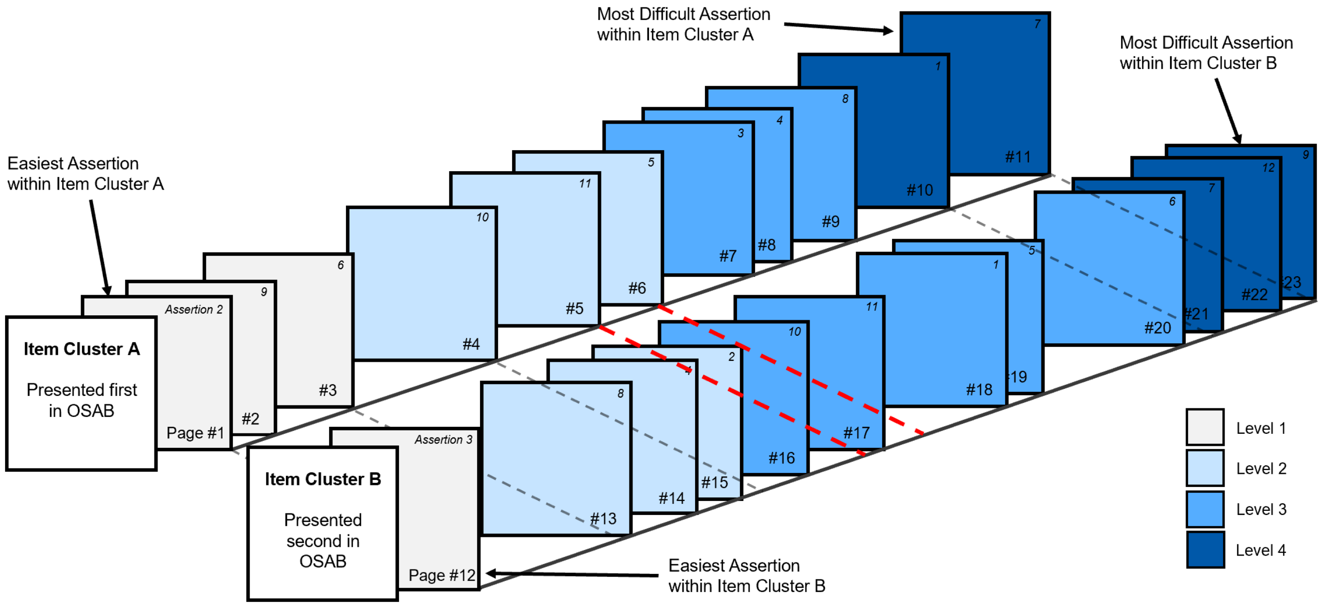
Facilitators provided the following process to guide the mapping of assertions onto ALDs:

1. How does the student interaction give rise to the assertion? Did the student plot, select, or write something?
2. Why is this assertion more difficult to achieve than the previous one (in the item)?
3. Which ALD most ably describes this assertion and the underlying interactions?

It was emphasized that assertions within an item were ordered by difficulty, and therefore, the assigned achievement levels should be ordered, as well. Within each item, panelists were not allowed to place an assertion into a lower achievement level than the level at which the previous assertions had been placed. If panelists felt very strongly that an assertion was out of order in the OSAB, they were asked to skip (not assign any achievement level to) the assertion. However, this was to be used as a last resort.

Because the assertion mapping was done separately for each item, there might have been no perfect ordering of the assigned levels of the assertions across all items as a function of assertion difficulty. It was allowed (and it occurred frequently) that an assertion of one item had a higher difficulty but lower assigned achievement level than another assertion from a different item (i.e., mapping inversions of assertions could occur across items, but mapping inversions of assertions were not allowed within an item). For example, in Figure 10, the difficulty of the assertion on page 6 of item cluster A ("Level 2") has a higher difficulty than the assertion on page 17 of item cluster B ("Level 3"). However, it was expected for the higher achievement levels to be assigned more frequently with increasing assertion difficulty across items. *Appendix 3-F, Standard-Setting Training Slides*, provides the training slides used during the breakout room training.

Figure 10. Example of Assertion Mapping



Note. Figure 10 describes scoring assertion mapping across two item clusters, where the assertions on pages 1, 2, 3, and 12 are mapped onto Level 1; the assertions on pages 4–6 and 13–15 are mapped onto Level 2; the assertions on pages 7–9 and 16–20 are mapped onto Level 3; and the assertions on pages 10, 11, and 21–23 are mapped onto Level 4.

7.6.9 Practice Quiz

Panelists completed a practice quiz before beginning a practice round. The quiz assessed panelists' understanding in multiple ways. They needed to be able to perform the following:

- Describe where "Just Barely" students fall on an achievement scale
- Indicate on a diagram how achievement standards define achievement levels
- Identify more- and less-difficult scoring assertions in the OSAB
- Answer questions about the assertion-mapping process and online application

Room facilitators reviewed the quizzes with the panelists and provided additional training for incorrect responses on the quiz. *Appendix 3-G, Standard-Setting Practice Quiz*, provides the quiz that panelists completed before mapping any assertions.

7.6.10 Practice Round

Following the practice quiz, panelists practiced mapping assertions to ALDs in a short practice OSAB consisting of one item cluster and one stand-alone item. The purpose of the practice round was to ensure that panelists were comfortable with the technology, items, item interactions, and scoring assertions before mapping any assertions in the OSAB. Panelists discussed their practice mappings and asked questions, and the room facilitators provided clarifications and further instructions until everyone had completed the practice round.

7.6.11 Readiness Form

After completing the practice round, and before mapping assertions to achievement levels in Round 1, panelists completed a readiness assertion form. On this form, panelists asserted that their training was sufficient for them to understand the following concepts and tasks:

- The knowledge and skills described by the ALDs and the skills and interactions that differentiate levels
- The structure, use, and importance of the OSAB
- The process to determine and map assertions to ALDs in the standard-setting tool
- How to use the assertion map when reviewing the OSAB and considering assertion mapping decisions
- The contextual information (student impact data and benchmarking data) when considering assertion mapping decisions
- Readiness to begin the Round 1 task

The readiness form for Round 2 focused on affirming an understanding of the feedback data supplied after Round 1. On this form, all panelists affirmed the following:

- Understanding of the feedback data and impact data
- Understanding of the Round 2 task
- Readiness to complete the Round 2 task

Room facilitators reviewed the readiness forms and provided additional training to panelists not asserting understanding or readiness. However, every panelist affirmed readiness before mapping assertions in both rounds of the workshop. *Appendix 3-H, Standard-Setting Readiness Forms*, provides the forms that panelists completed before each round of standard setting.

7.7 Assertion Mapping

Panelists mapped assertions independently, using the ALDs, their notes from reviewing each assertion, the difficulty level visualizer, the assertion map, and contextual information to place each assertion in one of the four achievement levels.

7.7.1 Calculating Cut Scores from the Assertion Mapping

Cut scores were calculated by treating every possible scale value as a hypothetical cut score and evaluating the number of discrepancies between the assertion mappings of the panelists and the achievement levels of the assertions implied by hypothetical cut score. The implied achievement level of an assertion was determined by comparing the response probability of the assertion to the hypothetical cut.[10] Each cut score was defined as the score point that minimized the weighted number of discrepancies. The weights were defined as the inverse of the observed frequencies of each level. For each cut score, only the assertions that were mapped to the two adjacent levels were considered (e.g., for the second cut, only the assertions that were mapped onto “Basic” and “Proficient” were used). Specifically, let n_k be the number of assertions put at achievement level k, t_k be the cut to be estimated, d_i be the assigned achievement level, and θ_i be the RP value of the ith assertion. For each assertion placed at levels k and $k+1$, the misclassification indicator is defined as

$$z_{ik}|t_k = \begin{cases} 1 \text{ if } (d_i = k \text{ and } t_k \le \theta_i) \text{ or } (d_i = k+1 \text{ and } t_k > \theta_i) \\ 0 \text{ otherwise} \end{cases}$$

The cut t_k is then estimated by minimizing a loss function based on the weighted number of misclassifications

$$\arg\min_{t_k} \left(\frac{1}{n_k} \sum_{i \in \{d_i = k\}} z_{ik}|t_k + \frac{1}{n_{k+1}} \sum_{i \in \{d_i = k+1\}} z_{ik}|t_k \right)$$

Unlike the Bookmark method, the cut scores for a table or room were not the median value of the cut scores of the individual panelists. Instead, cut scores at the table and room (grade) level were computed using the same method but taking into account the assigned levels of all the raters at the table and in the room, respectively. Applying these cut scores to the 2022 operational test data created data describing the percentage of students falling into each achievement level. This algorithm calculated cut scores from the assertion mappings by panelist and table and for the room.

7.7.2 Contextual Information and Feedback Data

To be adoptable, achievement standards for a statewide system must be coherent across grades and subjects. They should be orderly across subjects with no dramatic differences in expectation. The following are characteristics of well-articulated standards:

- The cut scores for each achievement level increase smoothly with each increasing grade.
- The cut scores should result in a reasonable percentage of students at each achievement level; reasonableness can be determined by the percentage of students in the achievement levels on historical tests, or contemporaneous tests measuring the same or similar content.
- Barring significant content standard changes (e.g., major changes in rigor), the percentage proficient on new tests should not be radically different from the percentage proficient on historical tests.

[10] Typically, the response probability used in standard setting is .67 (“RP67” [Huynh, 1994]). RP67 is the assertion difficulty point where 67% of the students would earn the score point. The reason to adopt RP50 for grade 11 for Idaho was because the difficulty of most items exceeded students’ abilities. RP50 better aligned with the ALDs and therefore led to more appropriate achievement cut scores. Using RP50 prevented panelists from mapping the first cut score onto the lowest-difficulty assertions on the test. This approach has been adopted for other high-stakes tests, such as the Smarter Balanced Assessments (see Cizek & Koons, 2014).

The standard-setting tool developed by CAI provides feedback data and allows for displaying contextual information to ensure standard-setting recommendations are well articulated.

7.7.2.1 Contextual Information

During OSAB review, panelists were also provided with additional contextual information to help inform their primary content-driven achievement standard recommendations. The standard-setting tool developed by CAI allows for displaying both impact and benchmark data to ensure standard-setting recommendations are well articulated. The contextual information provided included impact data and benchmark data for each of the assertions of the OSAB, as described in the following sections.

Impact Data

The impact data for an assertion was defined as the percentage of students who performed at or above the specified RP value associated with the assertion. Panelists were asked to consider the impact data when making their content-based assertion mappings.

Benchmark Data

The National Assessment of Educational Progress (NAEP) 2015 Idaho science achievement level percentages (NAEP, 2015) provided benchmark data, another source of contextual information that panelists could use to evaluate and adjust their assertion mapping. By comparing the results of each round against the percentage proficient in Idaho on NAEP science, panelists could evaluate the reasonableness of the proposed achievement standards. NAEP science provides the percentages of students in Idaho for grades 4 and 8; benchmark data for grade 5 were interpolated. For each ordered scoring assertion, panelists were provided with the associated achievement level for the NAEP science. An example of the benchmark information provided for each assertion in the review panel of the standard-setting tool is shown in Figure 9. The NAEP 2015 benchmark data were also graphically shown on the left side of the assertion map (see Figure 12).

7.7.2.2 Feedback Data

The online standard-setting tool created feedback data and cut scores corresponding to the assertion mappings for each panelist, for each table, and for the room overall (across all three tables). In addition, panelists were shown impact data based on the cut scores resulting from their assertion mappings. Impact data were defined for panelists as the percentages of students who would reach or exceed each of the achievement standards given the assertion mappings. Percentages were calculated using the student data from the 2022 administration of the ISAT in Science. This information allowed panelists to compare their mappings to other panelists' mappings to evaluate the impact of their current mappings.

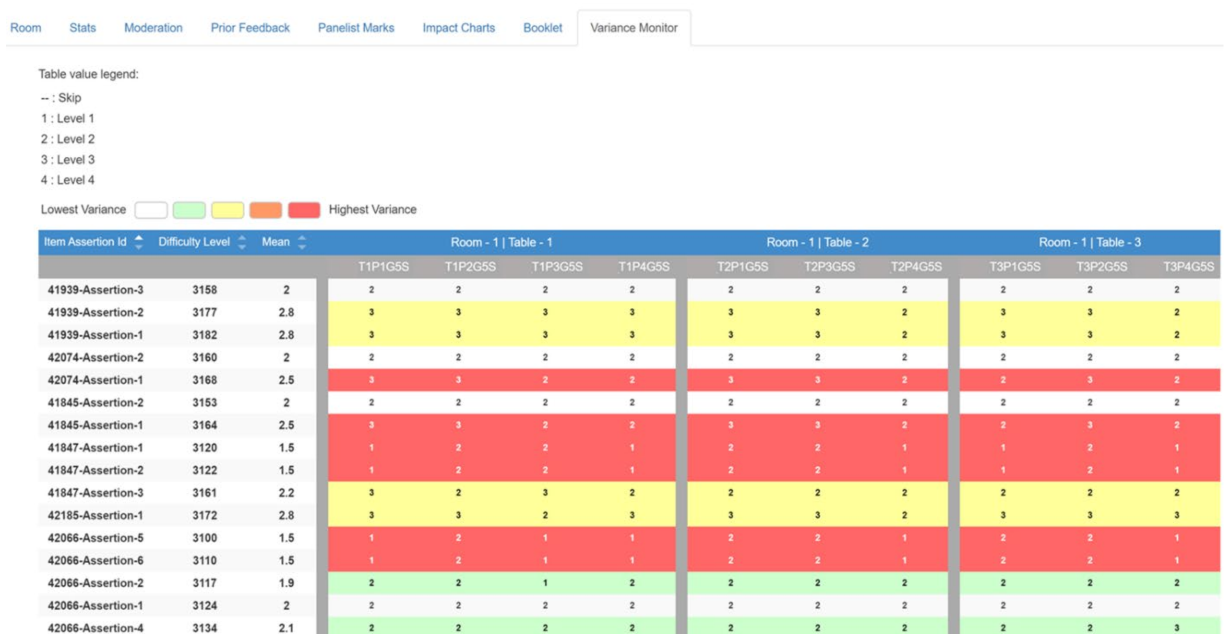
The standard-setting tool also generated variance monitor data and the assertion maps in the tool were updated to display the tentative standards for panelists to evaluate before Round 2 (the variance data and assertion maps are described in more detail in the following sections). All feedback and information served to inform, but not determine, the panelists' Round 2 decisions. Panelists discussed this information and the impact that the Round 1 cut scores may have on students before mapping assertions in Round 2.

After reviewing the feedback data, the workshop facilitators provided panelists with additional instructions for completing Round 2. First, they described the goal of Round 2 as one of convergence, but not consensus, on a common achievement standard. The second goal was to encourage articulation across grade levels. Each room spent time reviewing and discussing assertion mappings and articulation. After completing these discussions, panelists again worked through mapping all OSAB assertions to achievement levels for Round 2.

Variance Monitor Data

Feedback included a review of a variance monitor, part of CAI's online standard-setting tool that color codes the variance of assertion classifications. For all assertions, the variance monitor shows the achievement level to which each panelist assigned the assertion. The tool highlights assertions that panelists have assigned to different achievement levels. Figure 11 illustrates the types of information available in the variance monitor. Room facilitators and panelists reviewed and discussed the assertions with the most variable mappings.

Figure 11. Variance Monitor in CAI's Standard-Setting Tool

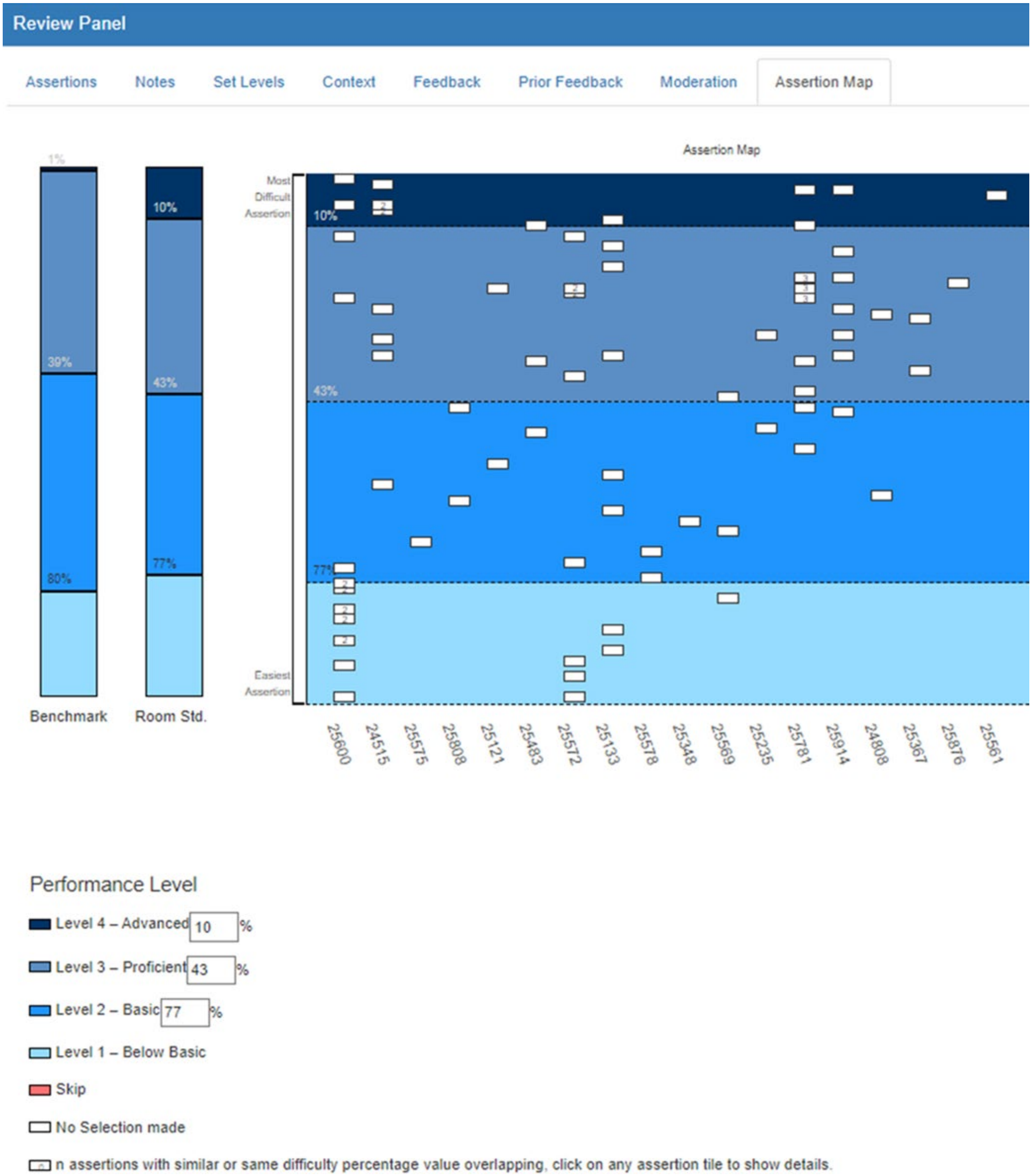


Item Assertion Id	Difficulty Level	Mean	Room - 1 | Table - 1				Room - 1 | Table - 2			Room - 1 | Table - 3		
			T1P1G5S	T1P2G5S	T1P3G5S	T1P4G5S	T2P1G5S	T2P3G5S	T2P4G5S	T3P1G5S	T3P2G5S	T3P4G5S
41939-Assertion-3	3158	2	2	2	2	2	2	2	2	2	2	2
41939-Assertion-2	3177	2.8	3	3	3	3	3	3	2	3	3	2
41939-Assertion-1	3182	2.8	3	3	3	3	3	3	2	3	3	2
42074-Assertion-2	3160	2	2	2	2	2	2	2	2	2	2	2
42074-Assertion-1	3168	2.5	3	3	2	2	3	3	2	2	3	2
41845-Assertion-2	3153	2	2	2	2	2	2	2	2	2	2	2
41845-Assertion-1	3164	2.5	3	3	2	2	3	3	2	2	3	2
41847-Assertion-1	3120	1.5	1	2	2	1	2	2	1	1	2	1
41847-Assertion-2	3122	1.5	1	2	2	1	2	2	1	1	2	1
41847-Assertion-3	3161	2.2	3	2	3	2	2	2	2	2	2	2
42185-Assertion-1	3172	2.8	3	3	2	3	3	3	2	3	3	3
42066-Assertion-5	3100	1.5	1	2	1	1	2	2	1	2	2	1
42066-Assertion-6	3110	1.5	1	2	1	1	2	2	1	2	2	1
42066-Assertion-2	3117	1.9	2	2	1	2	2	2	2	2	2	2
42066-Assertion-1	3124	2	2	2	2	2	2	2	2	2	2	2
42066-Assertion-4	3134	2.1	2	2	2	2	2	2	2	2	2	3

Assertion Maps

In addition to providing the numerical value of the cut scores and impact data, the feedback was also shown on the assertion maps. After Round 1, the assertion maps displayed in CAI's online standard-setting tool are updated with the overall room cut scores and the individual panelist cut scores for Round 1. Figure 12 presents the assertion map for grade 5 with the overall room cut scores for Round 1. The Round 1 and Round 2 assertion maps with overall room cut scores for all three grades are presented in *Appendix 3-I, Round 1 and Round 2 Standard-Setting Assertion Maps*.

Figure 12. Round 1 Standard-Setting Assertion Map, Grade 5



Panelists were instructed to consider their assertion mappings to compare the room cut score and assertions to their cut scores and assertion mappings. They were again reminded to evaluate the relative location of the assertions on the assertion maps.

7.8 Assertion Mapping Results

The CAI online standard-setting tool automatically computes the results and impact data for each round and then CAI room facilitators and psychometricians present the Round 1 results and feedback data for each grade.

7.8.1 Round 1 Results

Table 8 presents the achievement standards and associated impact data (percentage of students falling at or above each of the achievement standards based on the recommended Round 1 cut scores) from Round 1.

Table 8. Round 1 Results

Grade and Table	Cut Score			Impact Data		
	Level 2 Basic	*Level 3 Proficient*	*Level 4 Advanced*	*Level 2 Basic*	*Level 3 Proficient*	*Level 4 Advanced*
Grade 5	**480**	**506**	**534**	**77**	**43**	**10**
Table 1	476	506	534	81	43	10
Table 2	480	506	533	77	43	10
Grade 8	**777**	**807**	**819**	**79**	**41**	**24**
Table 1	777	808	832	79	39	11
Table 2	784	800	819	72	52	23
Grade 11	**1082**	**1108**	**1146**	**73**	**38**	**5**
Table 1	1082	1108	1146	73	38	5
Table 2	1080	1103	1139	76	44	8

Note. The grade row summarizes the room data (across both tables). Impact data describes the percentage of students falling at or above each of the achievement standards based on the recommended Round 1 cut scores.

Reviewing the Round 1 results began with a discussion of the feedback data from Round 1, beginning with table-level feedback and discussion and progressing to the room-level discussion. After reviewing the feedback (i.e., individual cuts, cuts by a table, cuts by a room) and impact data, workshop facilitators provided panelists with additional instructions for completing Round 2. They described the goal of Round 2 as one of convergence, but not consensus on a common achievement standard. The room then spent time reviewing and discussing assertion mappings. After completing these discussions, panelists again worked through the OSAB, mapping assertions for Round 2.

7.8.2 Round 2 Results

Table 9 presents the recommended achievement standards and associated impact data (percentage of students falling at or above each of the achievement standards based on the recommended Round 2 cut scores) for Round 2.

Table 9. Round 2 Results

Grade and Table	Cut Score			Impact Data		
	Level 2 Basic	*Level 3 Proficient*	*Level 4 Advanced*	*Level 2 Basic*	*Level 3 Proficient*	*Level 4 Advanced*
Grade5	**480**	**506**	**534**	**77**	**43**	**10**
Table 1	480	506	536	77	43	8
Table 2	480	506	534	77	43	10
Grade 8	**777**	**807**	**832**	**79**	**41**	**11**
Table 1	777	807	832	79	41	11
Table 2	782	803	815	74	47	29
Grade 11	**1082**	**1108**	**1146**	**73**	**38**	**5**
Table 1	1082	1108	1146	73	38	5
Table 2	1082	1107	1139	73	39	8

Note. The grade row summarizes the room data (across both tables). Impact data describes the percentage of students falling at or above each of the achievement standards based on the recommended Round 2 cut scores.

7.8.3 Convergence Across Rounds

While consensus is not an objective of standard setting, convergence is. Indicators of panelist convergence over rounds are the interquartile range and standard deviations of the standards computed for individual panelists based on their mappings. The interquartile range and standard deviations for each grade and after each round are presented in Table 10. The indicators show that convergence was improved after rounds except for Level 2 Cut at grade 11. After Round 2 results were presented to and discussed with the Department, it was decided not to conduct minor adjustments (i.e., moderation) on the recommended standards.

Table 10. Inter Quartile Range and Standard Deviation of Panelist Recommended Achievement Standards

Grade	Statistic	Level 2 Basic		Level 3 Proficient		Level 4 Advanced	
		Round 1	*Round 2*	*Round 1*	*Round 2*	*Round 1*	*Round 2*
5	**IQR**	9.25	6.00	5.25	2.75	10.25	2.00
	SD	11.40	10.22	7.54	1.94	10.26	3.28
8	**IQR**	9.00	6.50	6.00	3.50	10.50	8.50
	SD	12.66	11.17	7.93	8.19	19.02	6.74
11	**IQR**	3.00	4.75	3.50	1.00	12.50	11.00
	SD	5.07	5.65	5.76	6.63	19.64	11.14

7.8.4 Student Performance

Table 11 displays the achievement standards recommended by the standard-setting panelists.

Table 11. Standard Setting Results for Science

Grade	Cut Score			Impact Data		
	Level 2 Basic	*Level 3 Proficient*	*Level 4 Advanced*	*Level 2 Basic*	*Level 3 Proficient*	*Level 4 Advanced*
5	480	506	534	77	43	10
8	777	807	832	79	41	11
11	1082	1108	1146	73	38	5

Figure 13 displays the percentage of students that will reach or exceed each of the recommended science achievement standards in 2022.

Figure 13. Percentage of Students Reaching or Exceeding Each Recommended Science Achievement Standard in 2022

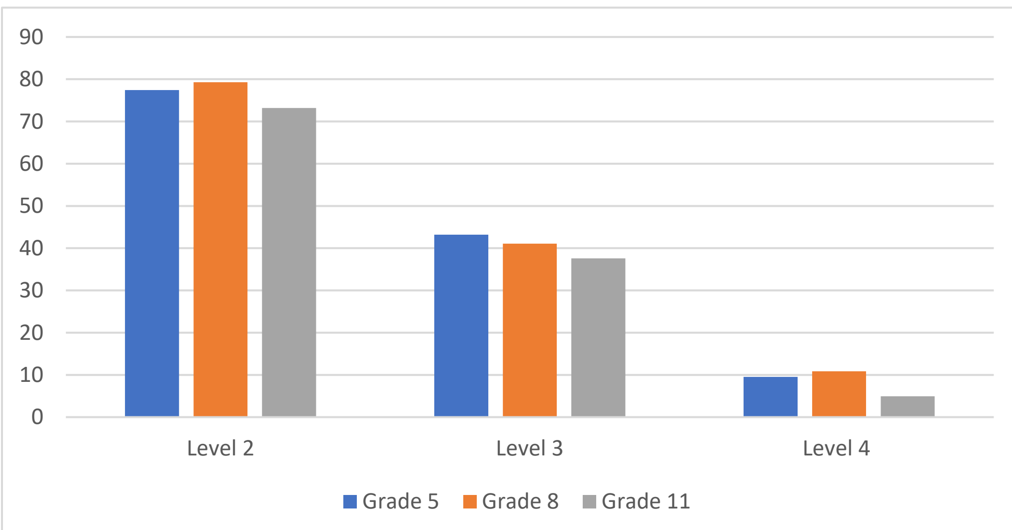
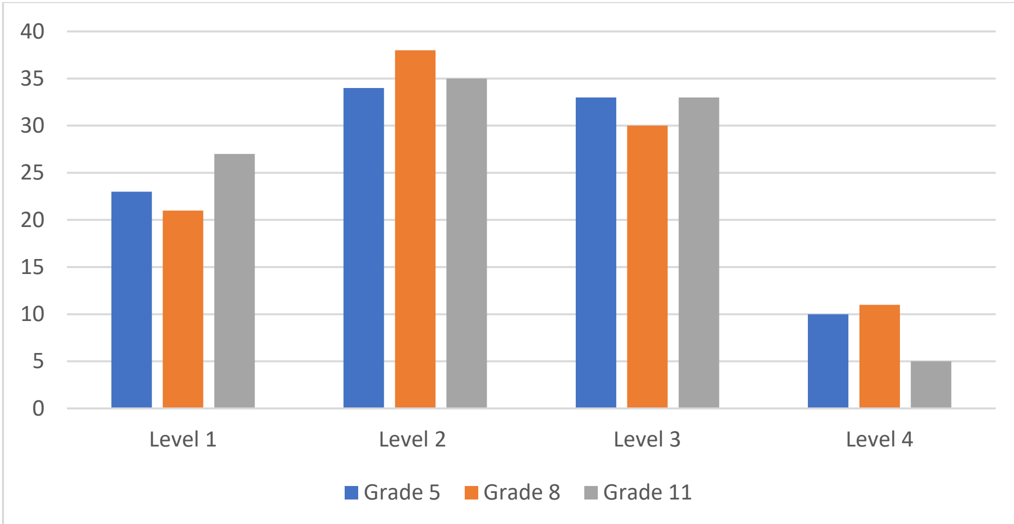


Table 12 indicates the percentage of students classified within each of the achievement levels in 2022. These values are displayed graphically in Figure 14.

Table 12. Percentage of Students Classified Within Each Science Achievement Level in 2022

Grade	Level 1 Below Basic	Level 2 Basic	Level 3 Proficient	Level 4 Advanced
5	23	34	33	10
8	21	38	30	11
11	27	35	33	5

Figure 14. Percentage of Students Classified Within Each Science Achievement Level in 2022



7.9 Workshop Evaluations

After finishing all activities, panelists completed online workshop evaluations independently in which they described and evaluated their experience taking part in the standard setting. Table 13, Table 14, Table 15, Table 16, and Table 17 summarize the results of the evaluations. Evaluation items endorsed by fewer than 90% of panelists are discussed in the text, and the least endorsed items are discussed in terms of the number and type of response.

Workshop panelists overwhelmingly indicated clarity in the instructions, materials, data, and process (see Table 13).

Table 13. Evaluation Results: Clarity of Materials and Process

Please rate the clarity of the following components of the workshop.	Percentage (%) Indicating "Somewhat Clear" or "Very Clear"			
	Grade 5	*Grade 8*	*Grade 11*	*Overall*
Instructions provided by the workshop leader	100	100	100	100
Achievement-Level Descriptors (ALDs)*	100	86	100	96
Ordered Scoring Assertion Booklet (OSAB)	100	100	100	100
Assertion Map	90	86	100	92
Impact Data (percentage of students that would achieve at the level indicated by the assertion difficulty)	90	100	100	96
Panelist Agreement Data	100	100	100	100

Note. Number of responses = 24 (grade 5 responses = 10, grade 8 responses = 7, grade 11 responses = 7). Evaluation response options included "Very Unclear," "Somewhat Unclear," "Somewhat Clear," and "Very Clear."

*During the standard-setting workshop, performance standards, performance levels, and performance-level descriptors (PLDs) were used based on the agreement with the Department. In this report, achievement standards, achievement levels, and achievement-level descriptors (ALDs) were used to reflect the most recent decision by the Department.

As shown in Table 14, most panelists felt that the time allocated to various workshop tasks was about right, though a few panelists had suggestions regarding time allocation:

- Eight panelists reported that the large-group orientation was too long.
- Three panelists indicated there was not enough time to experience the online assessment.
- Seven panelists reported not having enough time to review the ALDs.
- Four panelists indicated having too much time to discuss the skills demonstrated by students who are "just barely" described by each ALD, while another four panelists reported not having enough time for this activity.
- One grade 8 panelist indicated having too much time in reviewing the OSAB, while another grade 8 panelist reported not having enough time for the same activity.
- Two panelists indicated that not enough time was allowed for Round 1 results discussion.

Table 14. 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 (%) Indicating "About Right"**			
	Grade 5	*Grade 8*	*Grade 11*	*Overall*
Large-group orientation	90	14	86	67
Experiencing the online assessment	100	86	71	88
Reviewing the Achievement-Level Descriptors (ALDs)	80	57	71	71
Discussion of the skills demonstrated by students who are "just barely" described by each ALD	70	57	71	67
Reviewing the Ordered Scoring Assertion Booklet (OSAB)	100	71	100	92
Mapping your scoring assertions to achievement levels in each round	80	57	100	79
Round 1 results discussion	90	86	100	92

Note. Number of responses = 24 (grade 5 responses = 10, grade 8 responses = 7, grade 11 responses = 7). Evaluation response options included "Too Little," "Too Much," and "About Right."

Panelists appreciated the importance of the multiple factors contributing to assertion mapping, with nearly all panelists rating each factor as important or very important (see Table 15). Two grade 11 panelists indicated the "Just Barely" ALDs were not important though.

Table 15. Evaluation Results: Importance of Materials

How important were each of the following factors in your mapping of scoring assertions to achievement levels?	**Percentage (%) Indicating "Somewhat Important" or "Very Important"**			
	Grade 5	*Grade 8*	*Grade 11*	*Overall*
Achievement-Level Descriptors (ALDs)	100	100	100	100
"Just Barely" ALDs	100	100	71	92
Your perception of the difficulty of the scoring assertions and items in general	100	100	100	100
Your experience with students	100	100	100	100
Discussions with other panelists	100	100	100	100
Assertion map	90	100	100	96
External benchmark data	100	100	100	100
Impact Data (percentage of students that would achieve at the level indicated by the assertion difficulty)	90	100	100	96
Room agreement data (room, table, and individual standards)	100	100	100	100

Note. Number of responses = 24 (grade 5 responses = 10, grade 8 responses = 7, grade 11 responses = 7). Evaluation response options included "Not Important," "Somewhat Important," and "Very Important."

Participant understanding of the workshop processes and tasks was consistently high (see Table 16) although in the grade 11 panel, two panelists disagreed that the impact data was helpful when mapping assertions to achievement levels.

Table 16. Evaluation Results: Understanding Processes and Tasks

At the end of the workshop, please rate your agreement with the following statements.	**Percentage (%) Indicating "Agree" or "Strongly Agree"**			
	Grade 5	*Grade 8*	*Grade 11*	*Overall*
I understood the purpose of this standard-setting workshop.	100	86	86	92
The procedures used to recommend achievement standards were fair and unbiased.	100	100	100	100
The training provided me with the information I needed to recommend achievement standards.	100	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 assertion correctly.	100	100	100	100
The Achievement-Level Descriptors (descriptions of what students within each achievement level are expected to know and be able to do) provided a clear picture of expectations for student performance at each level.	100	86	100	96
I was able to develop an understanding of the knowledge and skills demonstrated by students who are "just barely" described by the Achievement-Level Descriptors.	90	100	86	92

At the end of the workshop, please rate your agreement with the following statements.	**Percentage (%) Indicating "Agree" or "Strongly Agree"**			
	Grade 5	*Grade 8*	*Grade 11*	*Overall*
I understood how to review each assertion in the Ordered Scoring Assertion Booklet (OSAB) to determine what students must know and be able to do to answer each assertion correctly.	100	100	100	100
I understood how to map assertions to the most apt achievement level.	100	100	100	100
I found the assertion map helpful in my decisions about the assertions I mapped to achievement levels.	100	100	100	100
I found the benchmark data and discussions helpful in my decisions about the assertions I mapped to achievement levels.	100	86	100	96
I found the impact data (percentage of students that would achieve at the level indicated by the assertion difficulty) helpful when mapping assertions to achievement levels.	90	100	71	88
I found the panelist agreement data (room, table, and individual standards) and discussions helpful when mapping assertions to achievement levels.	100	86	100	96
I felt comfortable expressing my opinions throughout the workshop.	100	86	100	96
Everyone was given the opportunity to express his or her opinions throughout the workshop.	100	100	100	100

Note. Number of responses = 24 (grade 5 responses = 10, grade 8 responses = 7, grade 11 responses = 7). Evaluation response options included "Strongly Disagree," "Disagree," "Agree," and "Strongly Agree."

Most panelists agreed that the standards set during the workshop reflected the intended grade-level expectations (see Table 17). However, in the grade 11 panel, four panelists disagreed that students performing at Basic level were below expectations for the grade and two disagreed that students performing at Proficient level met expectations for the grade. Two grade 8 panelists disagreed that students performing at Basic level were below expectations for the grade.

Table 17. Evaluation Results: Student Expectations

Please read the following statement carefully and indicate your response.	**Percentage (%) Indicating "Agree" or "Strongly Agree"**			
	Grade 5	*Grade 8*	*Grade 11*	*Overall*
A student performing at "Basic" is below expectations for the grade.	80	71	43	67
A student performing at "Proficient" meets expectations for the grade.	100	86	71	88
A student performing at "Advanced" is above expectations for the grade.	100	100	86	96

Note. Number of responses = 24 (grade 5 responses = 10, grade 8 responses = 7, grade 11 responses = 7). Evaluation response options included "Strongly Disagree," "Disagree," "Agree," and "Strongly Agree."

7.9.1 Workshop 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.”

Among the 24 panelists, 19 panelists responded to the first question, and 16 responded to the second. Some responses indicated the training was effective and the process was clear. Panelists provided suggestions, such as shortening or lengthening the time allocated for some tasks, engaging the panelists in active discussions, and suggesting a face-to-face standard-setting process to improve effectiveness. Some appreciated the standard-setting materials, process, and technology, and some panelists complimented the professionalism and expertise of the facilitators.

Additional participant comments included the following:

“I found the training included everything I needed to complete the required tasks.”

“Thank you for allowing us to be a part of this important project. I appreciated being able to learn more about how these standardized assessments are created and scored.”

“Matthew did a very good job of moving us forward and keeping everyone on task.”

“I always enjoy and feel like discussions are more cohesive and organic when things are done in person. But this format works when technology cooperates.”

“I feel like my advance experience with the PLD’s was very useful and that more time may need to be given to those individuals who do not participate in the PLD’s days.”

8. Validity Evidence

Validity evidence for standard setting is established in multiple ways. First, standard setting should adhere to the standards established by appropriate professional organizations and be consistent with the recommendations for best practices in the literature and established validity criteria. Second, the process should provide the evidence required of states to meet federal peer review requirements. We describe each of these in the following sections.

8.1 Evidence of Adherence to Professional Standards and Best Practices

The ISAT in Science standard-setting workshop was designed and executed consistent with established practices and best-practice principles (Hambleton & Pitoniak, 2006; Hambleton, Pitoniak, & Copella, 2012; Kane, 2001). The process also adhered to the following professional standards recommended in the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014) related to standard setting:

> *Standard 5.21:* When proposed score interpretation involves one or more cut scores, the rationale and procedures used for establishing cut scores should be documented clearly.

> *Standard 5.22:* When cut scores defining pass-fail or proficiency levels are based on direct judgments about the adequacy of item or test performances, the judgmental process should be designed so that the participants providing the judgments can bring their knowledge and experience to bear in a reasonable way.
>
> *Standard 5.23:* When feasible and appropriate, cut scores defining categories and distinct substantive interpretations should be informed by sound empirical data concerning the relation of test performance to the relevant criteria.

The sections of this report documenting the rationale and procedures used in the standard-setting workshop address Standard 5.21. The AMP standard-setting procedure is appropriate for tests of this type—with interrelated sets of three-dimensional item clusters and scaled using item response theory (IRT). Section 7.1, The Assertion-Mapping Procedure, provides the justification for and the additional benefits of selecting the AMP method to establish the cut scores. Section 7.6, Events, through Section Round 1, Round 1 Results, document the process followed to implement the method.

The design and implementation of the AMP procedure address Standard 5.22. The method directly leverages the subject-matter expertise of the panelists placing assertions into achievement levels and incorporates multiple, iterative rounds of ratings in which panelists modify their judgments based on feedback and discussion. Panelists apply their expertise in multiple ways throughout the process by

- understanding the test, test items, and scoring assertions (from an educator and a student perspective);
- describing the knowledge and skills measured by the test;
- identifying the skills associated with each test item scoring assertion;
- describing the skills associated with student performance at each achievement level;
- identifying which test item scoring assertions students at each achievement level should be able to answer correctly; and
- evaluating and applying feedback and reference data to the Round 2 recommendations and considering the impact of the recommended cut scores on students.

Panelists' understanding of the AMP was assessed with a quiz before the practice round. Additionally, panelists' readiness evaluations provided evidence of a successful orientation to the process and understanding of the process, while their workshop evaluations provide evidence of confidence in the process and resulting recommendations.

The recruitment process resulted in panels that were representative of important regional and demographic groups who were knowledgeable about the subject area and students' developmental level. Section 7.3.4, Educator Participants, summarizes details about the panel demographics and qualifications.

The provision of benchmark, context, and articulation data to panelists after Round 1 addresses Standard 5.23 (see Section 7.7.2, Contextual Information and Feedback Data). These empirical data provide necessary and additional context describing student performance given the recommended standards.

8.2 Evidence in Terms of Peer Review Critical Elements

The United States Department of Education guides the peer review of state assessment systems. This guidance is intended to support states in meeting statutory and regulatory requirements under Title I of the Elementary and Secondary Education Act of 1965 (U.S. Department of Education, 2015). The following critical elements are relevant to standard setting; evidence supporting each element immediately follows.

> **Critical Element 1.5:** Meaningful consultation in the development of challenging state standards and assessments

Idaho educators had heavily involved in all aspects of the assessment development. They played a critical role in establishing achievement levels for the tests. They created the item clusters, reviewed and revised the ALDs, mapped assertions to achievement levels to delineate performance at each achievement level, considered benchmark data and the impact of their recommendations, and formally recommended achievement standards.

Many subject-matter experts contributed to developing Idaho's achievement standards. Contributing educators were subject-matter experts in their content area, in the content standards and curriculum that they teach, and in the developmental and cognitive capabilities of their students. CAI's facilitators were subject-matter experts in the subjects tested and in facilitating effective standard-setting workshops. The psychometricians performing the analyses and calculations throughout the meeting were subject-matter experts in the measurement and statistics principles required of the standard-setting process.

> **Critical Element 6.2**: Performance standards setting. The state used a technically sound method and process that involved panelists with appropriate experience and expertise for setting its academic performance standards.

Evidence to support this critical element includes the following:

1) The rationale for and technical sufficiency of the AMP method selected to establish achievement standards (Section 7.1, The Assertion-Mapping Procedure).

2) Documentation that the method used for setting cut scores allowed panelists to apply their knowledge and experience reasonably and supported the establishment of reasonable and defensible cut scores (Section 7.6, Events; Section 7.6.2, Large-Group Orientation; Section 7.9, Assertion Mapping Results; and Section 8.1, Evidence of Adherence to Professional Standards and Best Practices).

3) Panelists self-reported readiness to undertake the task (Section 7.6.9, Practice Quiz; and Section 7.6.11, Readiness) and confidence in the workshop process and outcomes (Section 7.9, Assertion Mapping Results; and Section Round 1, Round 1 Results) supporting the validity of the process.

4) The standard-setting panels consisted of panelists with appropriate experience and expertise, including content experts with experience teaching Idaho's science content

standards, and individuals with experience and expertise teaching special population and general education students in Idaho (Section 7.3.4, Educator Participants, and *Appendix 3-A, Standard-Setting Panelist Characteristics*).

9. References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for Educational and Psychological Testing*. Washington, DC: Author.

Bradlow, E. T., Wainer, H., & Wang, X. (1999). A Bayesian random effects model for testlets. *Psychometrika*, *64*(2), 153–168.

Cizek, G. J., & Koons, H. (2014). Observation and Report on Smarter Balanced Standard Setting: October 12–20, 2014. Retrieved from https://portal.smarterbalanced.org/library/en/standard-setting-observation-and-report.pdf.

Ferrara, S., & Lewis, D. M. (2012). The item-descriptor (ID) matching method. In G. J. Cizek (Ed.), *Setting Performance Standards. Foundations, Methods, and Innovations* (2nd ed., pp. 255–282). New York: Routledge.

Gibbons, R. D., & Hedeker, D. R. (1992). Full-information bi-factor analysis. *Psychometrika*, *57*, 423–436.

Hambleton, R. K., & Pitoniak, M. J. (2006). Setting performance standards. In R. L. Brennan (Ed.), *Educational Measurement* (4th ed., pp. 433–470). Westport, CT: Praeger.

Hambleton, R. K., Pitoniak, M. J., & Copella, J. M. (2012). Essential steps in setting performance standards on educational tests and strategies for assessing the reliability of results. In G. J. Cizek (Ed.), *Setting Performance Standards: Foundations, Methods, and Innovations* (2nd ed., pp. 47–76). New York: Routledge.

Huynh, H. (1994, Oct.). Some technical aspects in standard setting. In *Proceedings of the Joint Conference on Standard Setting for Large-Scale Assessment Programs* (co-sponsored by National Assessment Governing Board and National Center for Education Statistics), Washington, DC, October 5–7, 1994, pp. 75–91.

Kane, M. T. (2001). So much remains the same: Conception and status of validation in setting standards. In G. J. Cizek (Ed.), *Setting Performance Standards: Concepts, Methods, and Perspectives* (pp. 53–88). Mahwah, NJ: Lawrence Erlbaum.

Kingston, N. M., Kahl, S. R., Sweeney, K. P., & Bay, L. (2001). Setting performance standards using the body of work method. In G. J. Cizek (Ed.), *Setting Performance Standards: Concepts, Methods, and Perspectives* (pp. 219– 248). Mahwah, NJ: Lawrence Erlbaum.

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* (pp. 249–281). Mahwah, NJ: Lawrence Erlbaum Associates.

National Center for Education Statistics. (2015). *The Nation's Report Card*. National Assessment of Educational Progress. Retrieved from: https://nces.ed.gov/nationsreportcard/.

National Research Council. (2012). *A framework for K–12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: The National Academies Press.

Rijmen, F. (2010). Formal relations and an empirical comparison among the bi-factor, the testlet, and a second-order multidimensional IRT model. *Journal of Educational Measurement*, *47*(3), 361−372.

Rijmen, F., Cohen, J., Butcher, T., & Farley, D. (2018, June). Scoring and reporting for assessments developed for the new science standards [Symposium]. National Conference on Student Assessment, San Diego, CA.

U.S. Department of Education. (2015). *Non-Regulatory Guidance for States for Meeting Requirements of the Elementary and Secondary Education Act of 1965, as amended*. Washington, DC. Retrieved from https://www2.ed.gov/policy/elsec/guid/assessguid15.pdf.

Appendix 3-A

Standard-Setting Panelist Characteristics

Standard-Setting Panelist Characteristics

Table A-1. Standard-Setting Panelists, Science Grade 5

Position	Location of Current Position	Gender	Race/ Ethnicity	Level of Education	Years Teaching Experience	Years Professional Experience	Years Teaching/Implementing the Idaho Science Standards	School District Size	School District Area Urbanicity	Table Leader
General Education Teacher	School	Female	White	Master's degree (e.g., M.A., M.S.)	16 to 20	None	15	Large	Urban	
General Education Teacher	School	Female	White	Master's degree	16 to 20	1 to 5	11	Medium	Rural	Yes
General Education Teacher	School	Female	White	Master's degree	16 to 20	None	5	Medium	Urban	
General Education Teacher	School	Female	White	Master's degree	6 to 10	None	4	Medium	Suburban	
General Education Teacher	School	Male	White	Bachelor's degree (e.g., B.A., B.S.)	20+	None	0	Medium	Urban	
Coach	School	Female	White	Doctoral degree (e.g., Ph.D., Ed.)	11 to 15	1 to 5	3 years of middle school standards, 3 years elementary standards	Small	Rural	
General Education Teacher	School	Female	White	Master's degree	11 to 15	None	4	Large	Suburban	

Position	Location of Current Position	Gender	Race/ Ethnicity	Level of Education	Years Teaching Experience	Years Professional Experience	Years Teaching/Implementing the Idaho Science Standards	School District Size	School District Area Urbanicity	Table Leader
General Education Teacher	School	Female	White	Bachelor's degree	1 to 5	None	2	Small	Rural	
General Education Teacher	School	Female	White	Master's degree	20+	None	7	Small	Urban	
General Education Teacher	School	Female	White	Master's degree	11 to 15	1 to 5	9	Large	Suburban	Yes

Table A-2. Standard-Setting Panelists, Science Grade 8

Position	Location of Current Position	Gender	Race/ Ethnicity	Level of Education	Years Teaching Experience	Years Professional Experience	Years Teaching/Implementing the Idaho Science Standards	School District Size	School District Area Urbanicity	Table Leader
Special Education Teacher	School	Female	White	Bachelor's degree (e.g., B.A., B.S.)	1 to 5	None	First timer	Large	Suburban	
General Education Teacher	School	Female	White	Master's degree (e.g., M.A., M.S.)	16 to 20	16 to 20	3	Small	Rural	
General Education Teacher	School	Female	White	Bachelor's degree	1 to 5	None	4	Small	Rural	
General Education Teacher, Special Education Teacher, Parent	School, Charter School	Female	Human	Master's degree	1 to 5	1 to 5	5	Medium	Rural	
General Education Teacher	School	Female	White	Bachelor's degree	11 to 15	None	5+	Medium	Rural	
General Education Teacher, Coach, Parent	School	Female	White	Master's degree	20+	None	20+	Medium	Suburban	Yes
General Education Teacher	School	Female	Hispanic	Master's degree	11 to 15	None	1	Small	Rural	Yes

Table A-3. Standard-Setting Panelists, Science Grade 11

Position	Location of Current Position	Gender	Race/ Ethnicity	Level of Education	Years Teaching Experience	Years Professional Experience	Years Teaching/Implementing the Idaho Science Standards	School District Size	School District Area Urbanicity	Table Leader
General Education Teacher, Higher Education	School	Female	White	Master's degree (e.g., M.A., M.S.)	20+	None	15	Medium	Suburban	
General Education Teacher	School	Female	White	Master's degree	20+	None	4+	Large	Suburban	Yes
General Education Teacher	School	Female	White	Master's degree	16 to 20	None	7	Small	N/A	
General Education Teacher	School	Female	White	Bachelor's degree (e.g., B.A., B.S.)	6 to 10	None	10	N/A	N/A	Yes
General Education Teacher	School	Male	White	Master's degree	20+	None	1	Large	Suburban	
Administrator	School	Male	White	Ed. S.	16 to 20	6 to 10	26	Medium	Suburban	
Administrator, Special Education Teacher	School, District	Female	White	Bachelor's degree	11 to 15	11 to 15	First timer	Small	Rural	

Note. N/A = Not applicable.

Appendix 3-B

Development of Science Range Achievement-Level Descriptors

Exhibit B-1. Development of Three-Dimensional Science Range Performance-Level Descriptors[1]

[1] The term “performance-level descriptors“ was used when they were developed under the three-dimensional science framework.

1. Development of Three-Dimensional Science Range Performance-Level Descriptors

Cambium Assessment, Inc. (CAI) held a meeting on May 18–19, 2018 for the three-dimensional science assessments. Prior to the meeting, CAI and several client states worked together to refine drafts of Policy and Range PLDs created by Washington State's Office of the Superintendent of Public Instruction (OSPI). During the meeting, educators reviewed and provided feedback on these Policy and Range PLDs.

PLDs describe levels or categories of achievement on a large-scale assessment. PLDs are used to inform the evidence required for item development, inform items selected during the form construction process, and support standard-setting panelist recommendations during the standard-setting process. PLDs are then ultimately used to inform stakeholder interpretation of student scores once standards are set. Egan, Schneider, and Ferrara (2012) recommended four stages of PLD development for the following types of PLDs: Policy, Range, Threshold, and Reporting. The focus of the three-dimensional science PLD meeting was on Policy and Range PLDs only.

2. Definitions of Performance-Level Descriptors by Purpose and Intended Audience

2.1 Policy Performance-Level Descriptors

Policy PLDs articulate the overall claims about a student's achievement in each performance level. They are used by policymakers to broadly articulate the goals and rigor for the state's performance standards. Table 1 shows a sample Policy-based PLD.

Table 1. Draft Three-Dimensional Science Policy PLD for Proficient

Level 3
The Level 3 student is proficient in applying three-dimensional science knowledge and skills as specified in the Next Generation Science Standards (NGSS). The student generally performs at the standard for the grade level, is able to access grade-level content, and engages in higher-order thinking skills with some independence and minimal support.

2.2 Range Performance-Level Descriptors

Range PLDs describe the expectations for students across each Disciplinary Core Idea (DCI) and proficiency level, demonstrating how the content represents a progression of knowledge, skills, and processes across performance levels and grade bands. Washington State's Office of the Superintendent of Public Instruction (OSPI) created Range PLDs for Levels 2, 3, and 4, with Level 3 describing Proficiency. Table 2 shows sample Policy PLDs.

Table 2. Draft NGSS Policy PLDs for Grade 8

Level 2	Level 3	Level 4
Use a model and patterns in data to show that the number of tiny particles does not change during chemical reactions and that particle motion changes when thermal energy is added to or removed from a system.	Develop and use models and interpret patterns in data to show that mass is conserved during chemical reactions and to predict changes in particle motion when thermal energy is added to or removed from a system.	Analyze and interpret patterns in data in order to evaluate and revise a model that describes how mass is conserved during chemical reactions and to explain predicted changes in particle motion when thermal energy is added to or removed from a system.

3. Performance-Level Descriptor Workshop

CAI revised OSPI's PLDs to ensure that text sufficiently differentiates between levels. CAI sent for participating states' review and then convened a committee preparation meeting on May 9, 2018, to prepare participating educators and state staff for the May 18–19th, 2018, meeting.

The meeting was divided into three grade-band rooms: elementary, middle, and high school. One CAI facilitator led each grade-band room, and several CAI staff were available to float between rooms to ensure process consistency and answer questions. Each grade-band room included nine educators, enabling room facilitators to divide the rooms into subgroups to complete the work. Table 3 summarizes the composition of facilitators and educators assigned to each grade band. Recruitment included educators representing special populations (English learners [ELs], Special Education).

Table 3. Workshop Panel Assignments

	Elementary School	Middle School	High School
CAI Facilitators	1	1	1
Educators	9	9	9

3.1 Performance-Level Descriptor Workshop

The Performance-Level Descriptor (PLD) workshop occurred over a period of two days. Appendix 1. PLD Workshop Agenda provides the full workshop agenda.

3.1.1 Day One

The workshop began with a welcome from staff from CAI and participating state staff. CAI provided an overview of the policy aspects of the workshop, including how PLD development fits into the overall test development and standard-setting processes. CAI staff provided training on the processes to be used during the workshop. Following the initial overview, CAI provided training on item clusters and scoring assertions. CAI then described the purpose and structure of the Next Generation Science Standards (NGSS) clusters and scoring assertions, and their importance to the standard-setting process.

A facilitator continued training on Policy PLDs. Facilitators walked panelists through several National Reference Point Policy PLDs, outlining the differences in the key descriptors at each performance level. The panelists reviewed the Policy PLDs individually and in small groups. The panelists used the following questions to frame their review of the National Reference PLDs:

- What terms are used to define proficiency?
- Are there certain terms you value over others?
- Are there words or phrases you note that could inform three-dimensional science policy statements going forward?

After small group discussion, facilitators engaged panelists in a room-level discussion and recorded recommendations for Policy PLDs. Facilitators framed discussions by using the following guiding questions:

- What claims should the Policy PLDs make about students at each performance level?
    - Two to five words that provide context for the expectations of students in each performance level
- What general descriptors best articulate the intended rigor for NGSS?
- How should we represent what proficiency means?
    - College and career readiness
    - On grade-level attainment
    - Meeting standards

The goal of the discussion process is to draft Policy PLDs and for the panelists to begin to have a shared sense of the type of student described by each proficiency level. The Policy PLD discussion lasted through the morning of Day 1, ending with lunch.

After lunch, the meeting shifted to Range PLD training within each breakout room. Facilitators described the process for reviewing Range PLDs. Facilitators modeled how to parse out each PLD, focusing on the key words used in each performance level. In modeling how to parse the standards, the facilitator noted the importance of the Level 3 (proficiency) cut score as an anchor for the other descriptors. The facilitator started by parsing a Level 3, then moving to Levels 2 and 4, modeling the sequence panelists would use throughout the workshop. Next, the facilitator led the room through reviewing one Range PLD. They started by reviewing the Level 3 PLD, then moving to Level 2, then Level 4. Depending on how well the panelists understood the task, the facilitator might have reviewed another PLD with the entire group.

Once the facilitators modeled the process for panelists, they split panelists into groups to create Range PLDs. Each room facilitator divided the PLDs among the groups so they could review them more efficiently in the time allotted for the meeting, resulting in three groups of three panelists in each room. Each group tracked any recommended revisions to each PLD. To facilitate discussion, panelists responded to four questions for each PLD:

- Does the PLD reflect the expected achievement exhibited by students at this performance level?
- What revisions were made to the PLD?
- What rationale do you have for any changes?
- What would distinguish an assertion belonging to this PLD from an assertion belonging to the level below?

One member of each group acted as a scribe, using a computer to track changes to the PLDs, and responded to the questions through an online form. CAI created a template for panelists to use when reviewing the Range PLDs.

For the rest of the afternoon, the panelists reviewed the Range PLDs using the following processes:

- The panelists worked through each assigned PLD, ensuring that the PLD showed a clear progression of observable evidence that should be expected from students at each performance level.
- For each PLD, participants began with the Level 3 descriptor, then moved to Level 2, then Level 4.
- Facilitators monitored progress and work to ensure cross-grade coherence and adherence to the expectations set by the Policy PLDs.

This work continued for the duration of Day 1. At the end of Day 1, CAI and state staff reviewed the panelists' work to check for coherence and consistency across grades.

3.1.2 Day 2

Based on results of the review at the end of Day 1, room facilitators and state staff spent time recalibrating groups if necessary. During the morning of Day 2, the panelists completed their assigned standards. Once each group completed its work, the facilitators conducted discussions with their rooms to ensure coherence across PLDs within each grade band. Each group reviewed their grade-band PLDs to ensure consistency and coherence across performance levels and consistency and coherence within each performance level. This discussion extended until lunch.

After lunch, the grade-band groups met for a cross-grade articulation discussion. They compared the expectations across grade bands to ensure a sensible progression of rigor. The committee focused primarily on examining Level 3 to assess if this level is considered the entry point for college-readiness. After the cross-grade articulation discussion, educators were allowed to adjourn.

For the rest of the afternoon, CAI met with participating state staff. The group discussed the results of the meeting and addressed any issues or inconsistencies in the educators' work. The group also discussed next steps for finalizing the PLDs.

4. References

Egan, K. L., Schneider, M. C., & Ferrara, S. (2012). Performance level descriptors: History, practice, and a proposed framework. In G. Cizek (Ed.), *Setting Performance Standards, Second Edition* (pp. 79–106). New York, NY: Routledge.

Science Assessment Team, Office of Superintendent of Public Instruction (2018). *Performance Level Descriptors: Washington Comprehensive Assessment of Science.* Office of Superintendent of Public Instruction.

Schneider, M. C. & Egan, K. L. (2014). A Handbook for Creating Range and Target Performance Level Descriptors. The National Center for the Improvement of Educational Assessment.

Appendix 1. PLD Workshop Agenda

Exhibit 1-A. Day 1 PLD Workshop Agenda

Time	Topic	Lead
7:30–8:30 a.m.	Breakfast	
8:30–9:00 a.m.	Welcome Three-Dimensional Science Item Clusters and Scoring Assertions • The purpose and structure of three-dimensional science item clusters • Scoring Assertions	Jon
9:00–9:30 a.m.	Three-Dimensional Science Performance-Level Descriptors (PLDs) • Describe purposes and uses for Policy and Range PLDs • Describe workshop process	Kevin
9:30–9:45 a.m.	Break	
9:45 a.m.–Noon	Policy PLD Discussion • Review Policy PLDs ○ What are the important elements of the descriptor at each performance level? • Small group discussion • Room discussion • Final recommendations	Kevin
Noon–1:00 p.m.	Lunch	
1:00–2:00 p.m.	Range PLD Training • Purpose of Range PLDs • Tools used in review ○ NGSS Standards ○ Policy PLDs ○ Draft Range PLDs ○ Template for reviewing standards • Parsing standards and draft PLDs to differentiate among performance levels	Room facilitators
2:00–4:30 p.m.	Review draft Range PLDs • Each group reviews assigned PLDs • For each PLD, start with Level 3 (Proficient), then move to Level 2, then Level 4	Room facilitators

Exhibit 1-B. Day 2 PLD Workshop Agenda

Time	Topic	Lead
7:30–8:30 a.m.	Breakfast	
8:30–10:00 a.m.	Continue Range PLD review • Each group reviews assigned PLDs • For each PLD, start with Level 3 (Proficient), then move to Level 2, then Level 4	Room facilitators

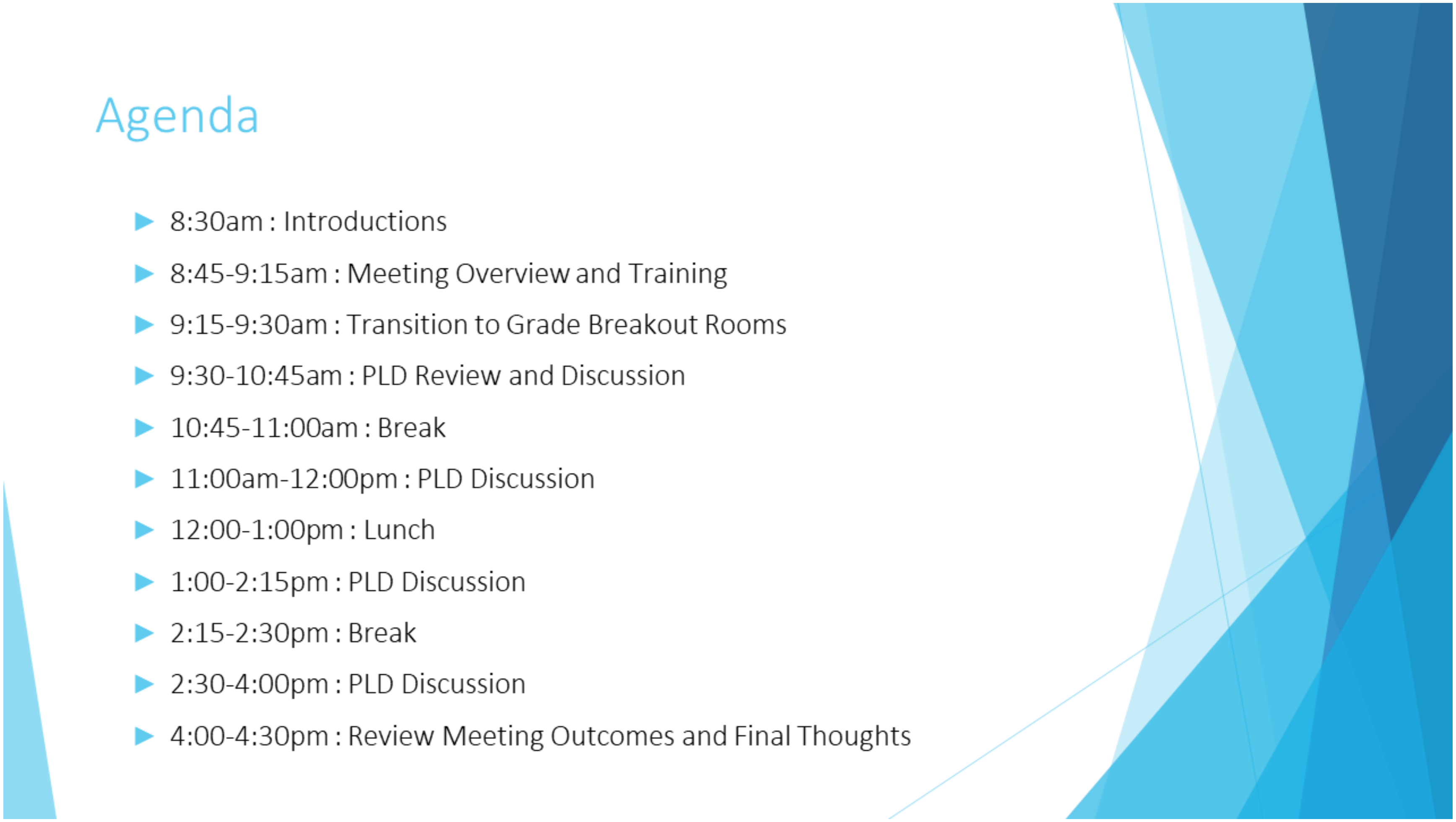
10:00 a.m.–Noon	Room Discussion
	• Room discussion to ensure coherence within the grade-band
	  ○ Ensure consistency and coherence across performance levels throughout the grade band
	  ○ Ensure consistency and coherence within each performance level throughout the grade band
		Room facilitators

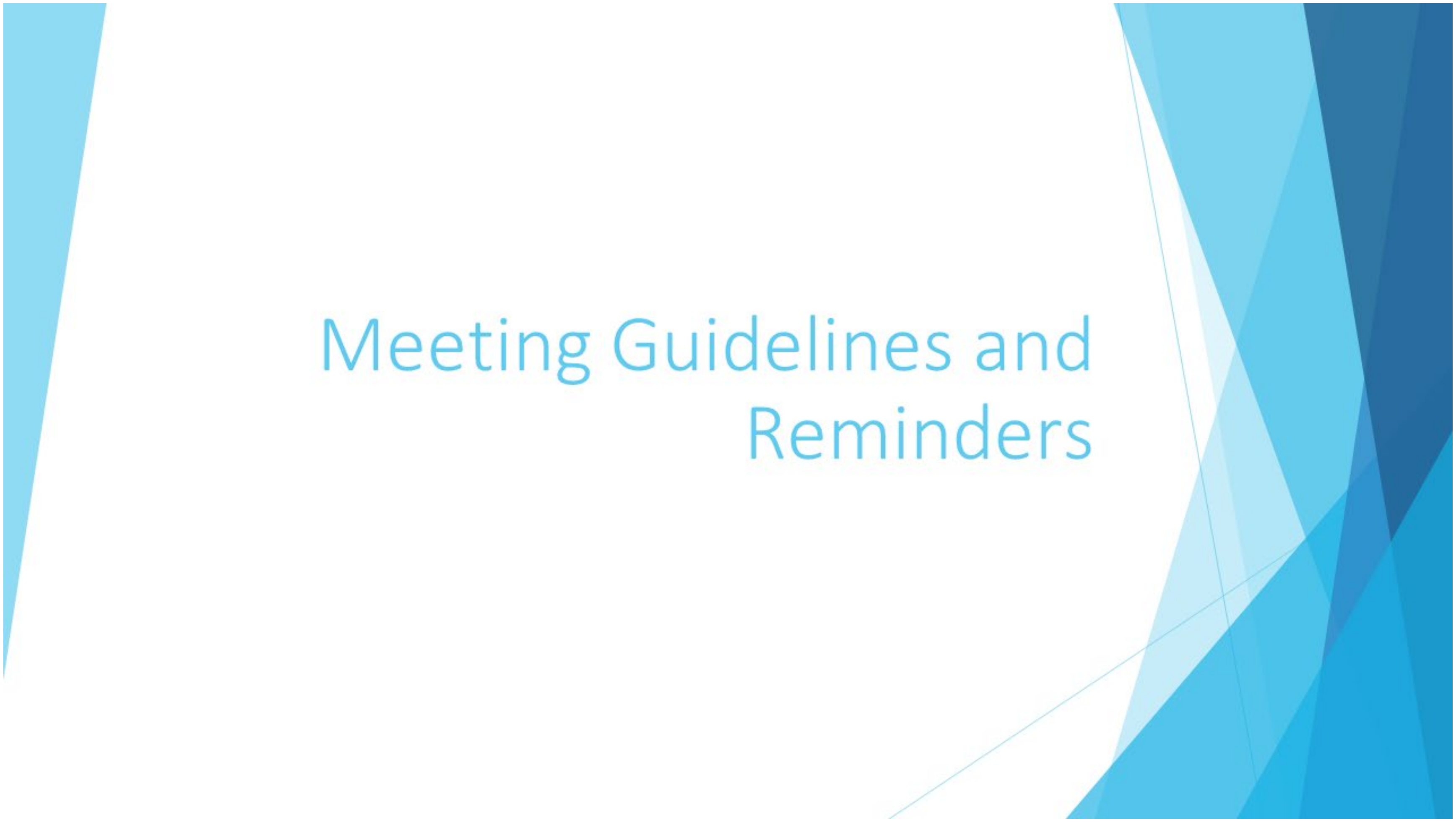
Noon–1:00 p.m.	Lunch	
1:00–2:30 p.m.	Large group: Cross-grade coherence discussion • Ensure cross-grade consistency and coherence across performance levels • Ensure cross-grade consistency and coherence within each performance level	Kevin
2:30 p.m.	Educators adjourn	
2:30–3:00 p.m.	CAI and Department staff • Resolve inconsistencies within or across grades • Discuss next steps	

Exhibit B-2. ISAT in Science Performance-Level Descriptor[2] Meeting Slides

[2] The term "performance-level descriptor" was used when the Idaho SDE and CAI conducted the meeting in summer 2022.


Idaho Standards Achievement Test – Science PLD Meeting
July 18, 2022
Cambium Assessment


Agenda
8:30am : Introductions
8:45-9:15am : Meeting Overview and Training
9:15-9:30am : Transition to Grade Breakout Rooms
9:30-10:45am : PLD Review and Discussion
10:45-11:00am : Break
11:00am-12:00pm : PLD Discussion
12:00-1:00pm : Lunch
1:00-2:15pm : PLD Discussion
2:15-2:30pm : Break
2:30-4:00pm : PLD Discussion
4:00-4:30pm : Review Meeting Outcomes and Final Thoughts


Meeting Guidelines and
Reminders

Guidelines for Group Meetings

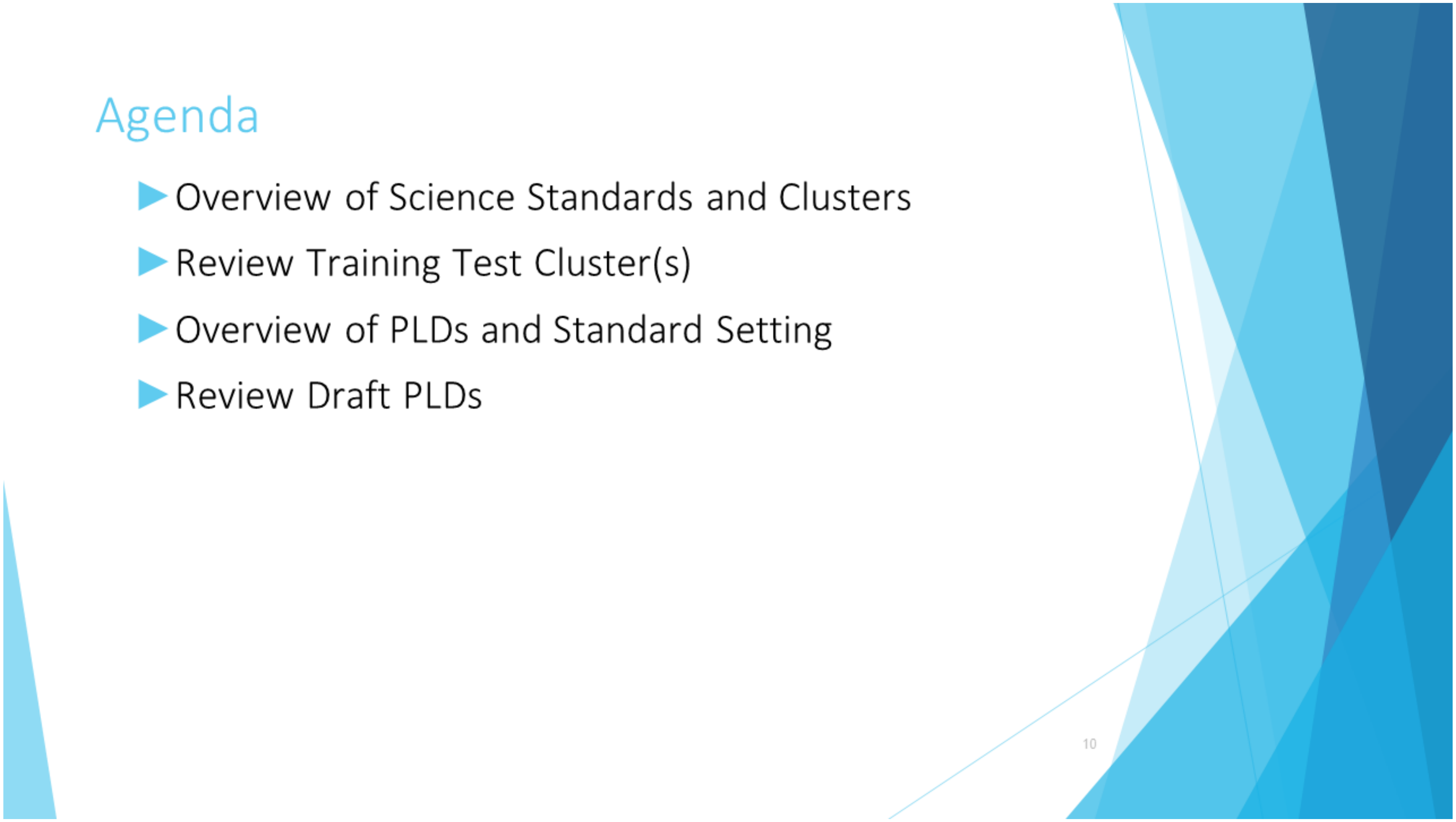
- Please do not use cell phones while in group reviews. If you have an important call you must take, just let us know and we can pause the group discussion as needed.
- If you have any questions about the review or procedures, feel free to reach out to Cambium or SDE staff via the chat box.

Meeting Norms

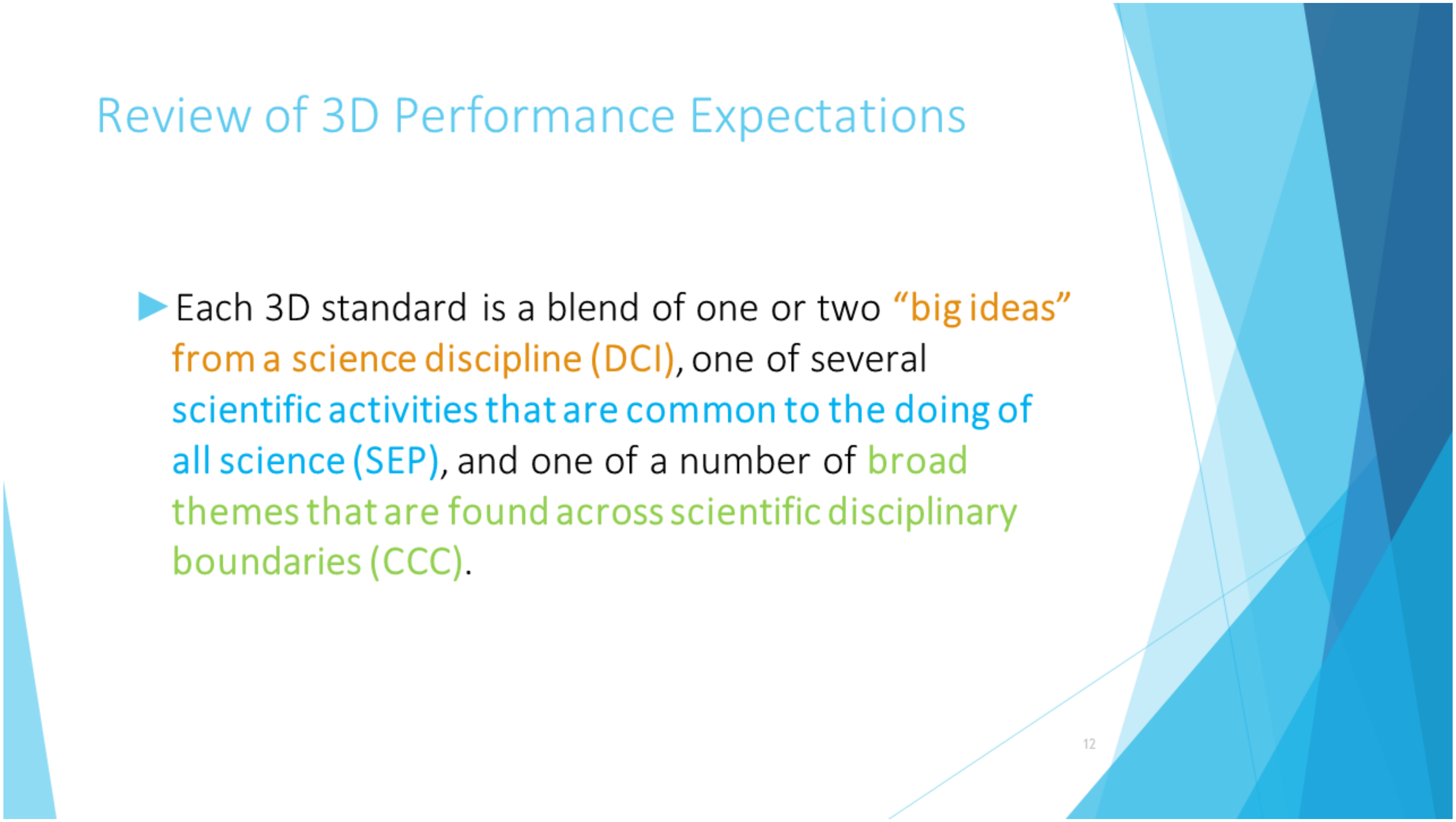
- Listen actively and attentively to each other and the facilitator.
- Take turns providing feedback and encourage everyone to participate.
- Ask for clarification when needed.
- Avoid off-topic and side conversations.
- Build on one another’s statements and offer constructive comments.

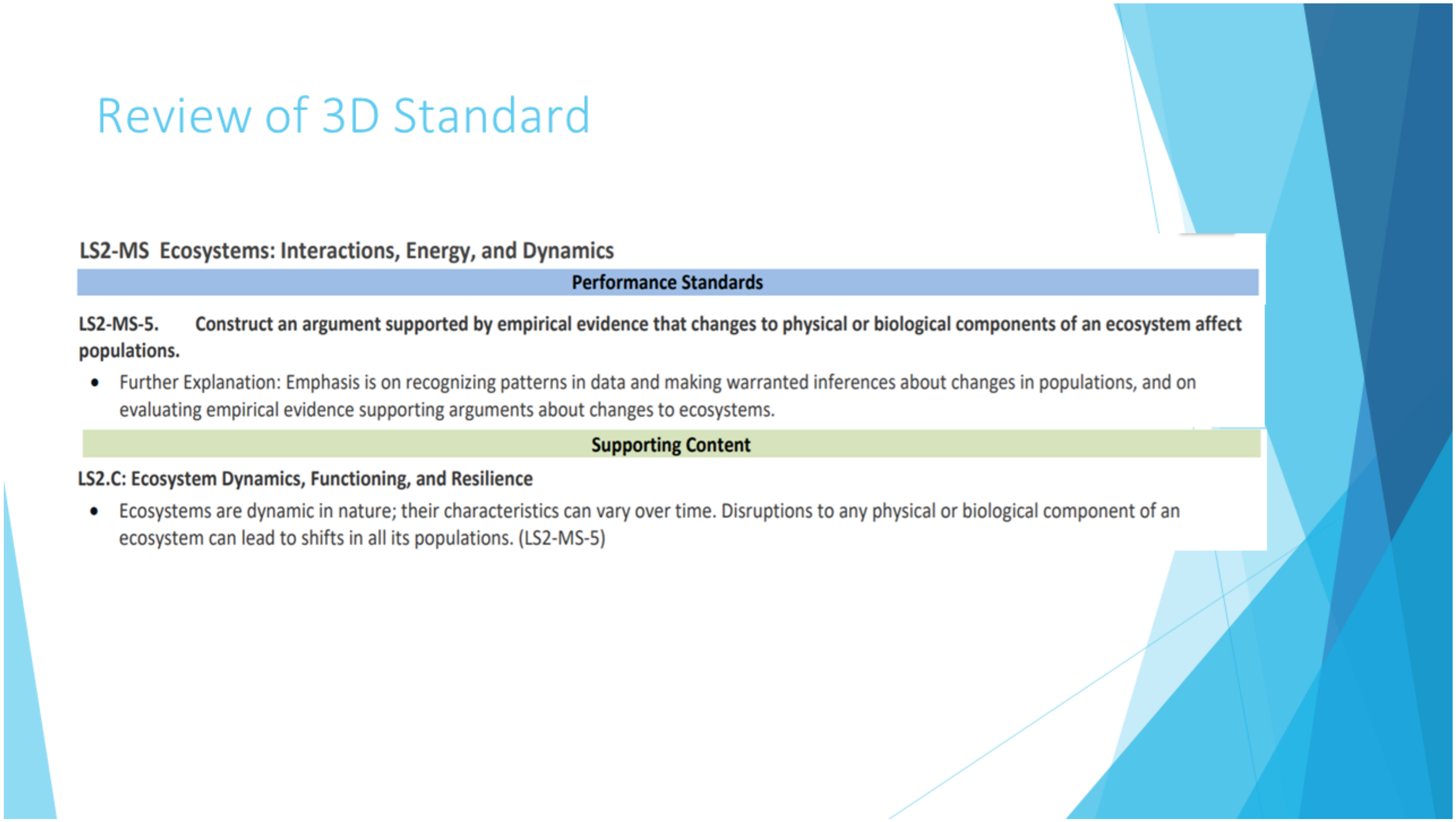
Webinar Protocol

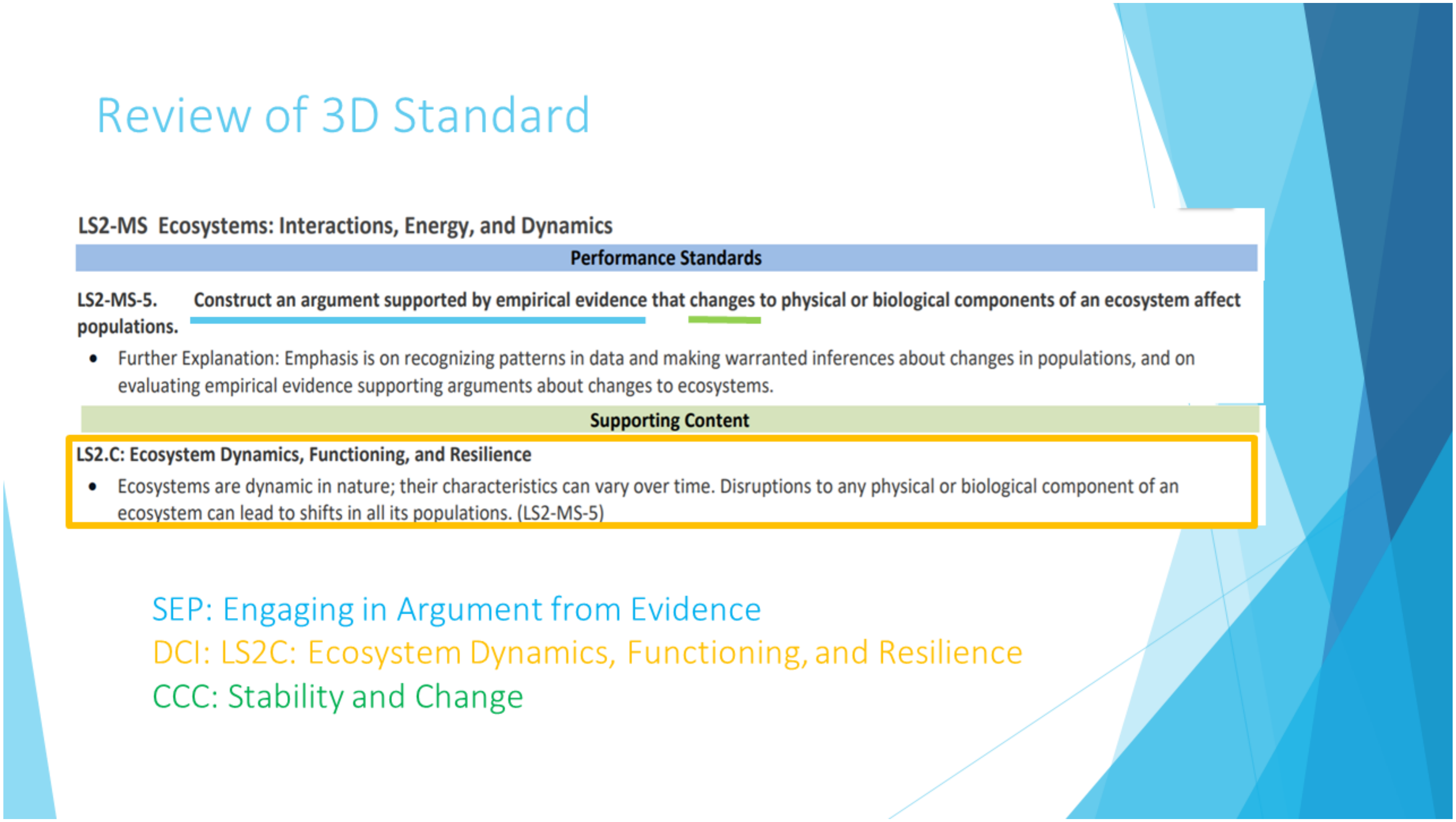
- Use your camera to promote collaboration, when desired.
- Mute your audio when not speaking.
- Use the chat box to communicate during individual working time.
- Use audio to communicate during group collaborative time, unless otherwise directed by the facilitator.
- This is a learning experience for everybody. Please ask questions!


Agenda
Overview of Science Standards and Clusters
Review Training Test Cluster(s)
Overview of PLDs and Standard Setting
Review Draft PLDs
10


Idaho Science Standard
Review


Review of 3D Performance Expectations
Each 3D standard is a blend of one or two "big ideas" from a science discipline (DCI), one of several scientific activities that are common to the doing of all science (SEP), and one of a number of broad themes that are found across scientific disciplinary boundaries (CCC).
12


Review of 3D Standard
LS2-MS Ecosystems: Interactions, Energy, and Dynamics
Performance Standards
LS2-MS-5. Construct an argument supported by empirical evidence that changes to physical or biological components of an ecosystem affect populations.
• Further Explanation: Emphasis is on recognizing patterns in data and making warranted inferences about changes in populations, and on evaluating empirical evidence supporting arguments about changes to ecosystems.
Supporting Content
LS2.C: Ecosystem Dynamics, Functioning, and Resilience
• Ecosystems are dynamic in nature; their characteristics can vary over time. Disruptions to any physical or biological component of an ecosystem can lead to shifts in all its populations. (LS2-MS-5)

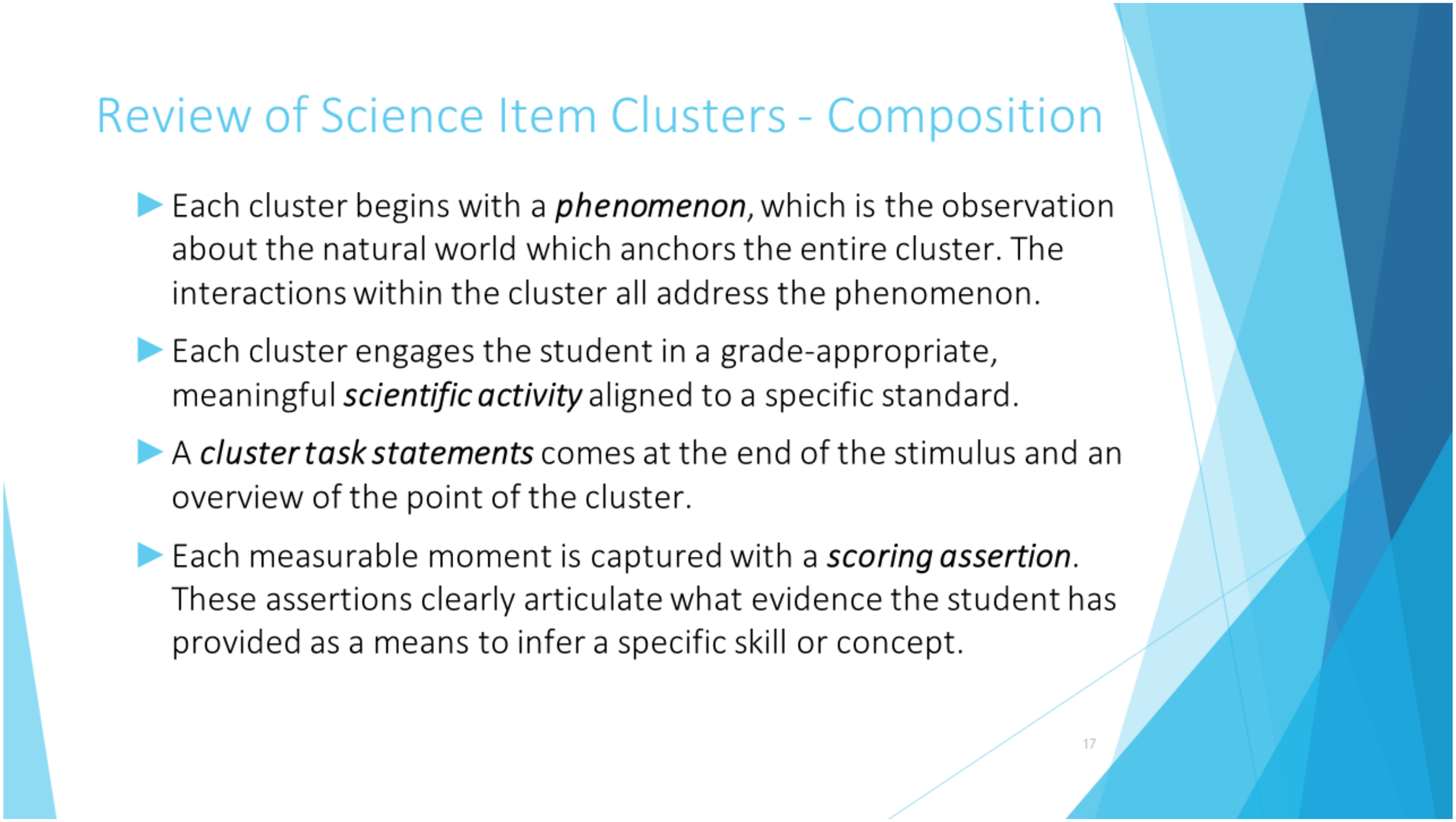

Review of 3D Standard
LS2-MS Ecosystems: Interactions, Energy, and Dynamics
Performance Standards
LS2-MS-5. Construct an argument supported by empirical evidence that changes to physical or biological components of an ecosystem affect populations.
Further Explanation: Emphasis is on recognizing patterns in data and making warranted inferences about changes in populations, and on evaluating empirical evidence supporting arguments about changes to ecosystems.
Supporting Content
LS2.C: Ecosystem Dynamics, Functioning, and Resilience
Ecosystems are dynamic in nature; their characteristics can vary over time. Disruptions to any physical or biological component of an ecosystem can lead to shifts in all its populations. (LS2-MS-5)
SEP: Engaging in Argument from Evidence
DCI: LS2C: Ecosystem Dynamics, Functioning, and Resilience
CCC: Stability and Change

Three Dimensions

Science and Engineering Practices	Disciplinary Core Ideas	Cross-cutting Concepts
Asking Questions and Defining Problems	Physical Science	Patterns
Developing and Using Models	Life Science	Cause and Effect
Planning and Carrying Out Investigations	Earth and Space Science	Scale, Proportion, and Quantity
Analyzing and Interpreting Data	Engineering	Systems and System Models
Using Mathematics and Computational Thinking		Energy and Matter
Constructing Explanations and Designing Solutions		Structure and Function
Engaging in Argument from Evidence		Stability and Change
Obtaining, Evaluating, and Communicating Information		

15

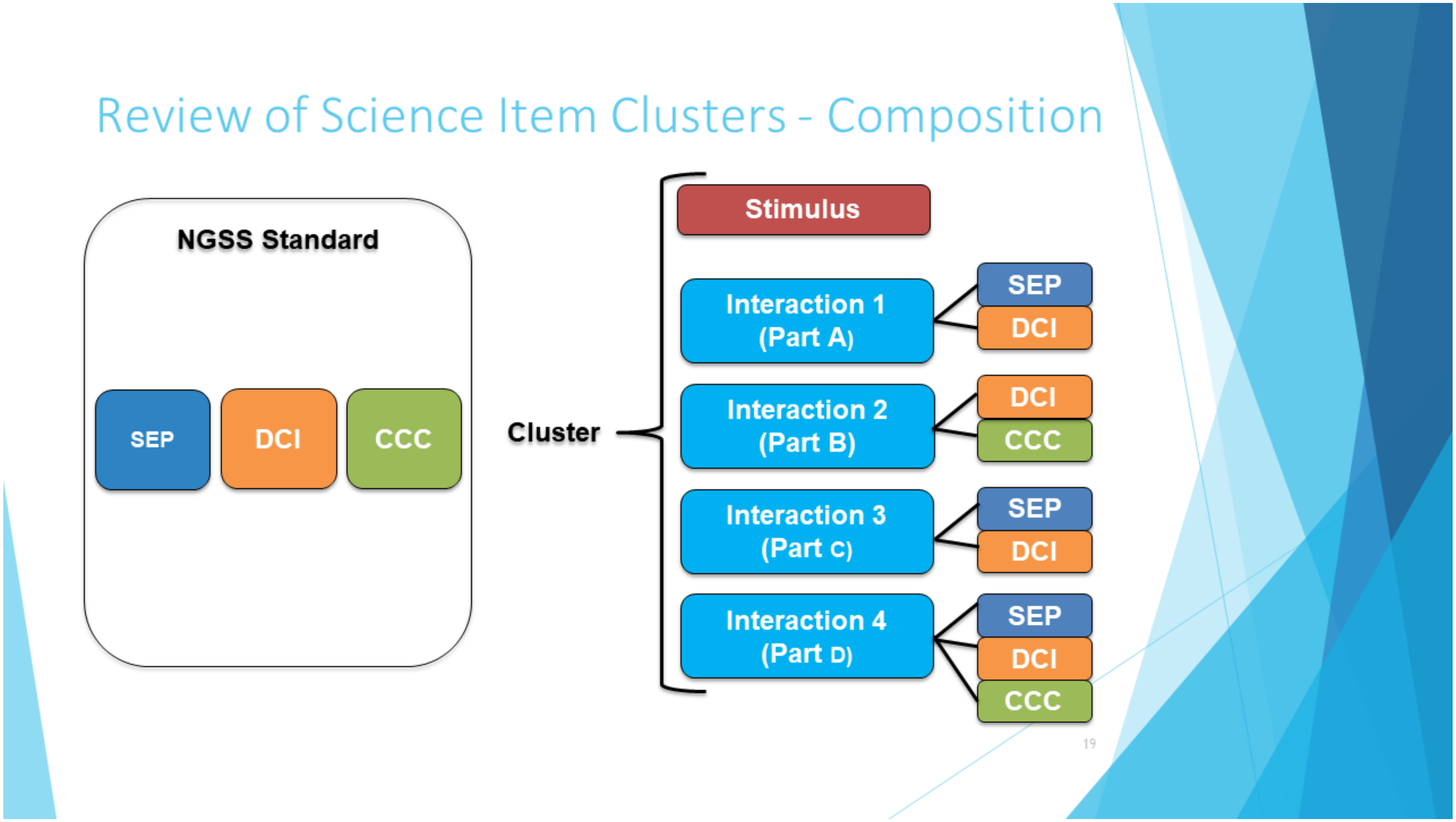

Science Cluster Review


Review of Science Item Clusters - Composition
Each cluster begins with a *phenomenon*, which is the observation about the natural world which anchors the entire cluster. The interactions within the cluster all address the phenomenon.
Each cluster engages the student in a grade-appropriate, meaningful *scientific activity* aligned to a specific standard.
A *cluster task statements* comes at the end of the stimulus and an overview of the point of the cluster.
Each measurable moment is captured with a *scoring assertion*. These assertions clearly articulate what evidence the student has provided as a means to infer a specific skill or concept.
17

Review of Science Item Clusters - Composition

- Evidence centered design
- Multiple interactions in which students engage a phenomenon
  - Identify
  - Describe
  - Model
  - Predict
  - Explain
- Interactions support a set of assertions about what the student has demonstrated they know and are able to do

18


Review of Science Item Clusters - Composition
NGSS Standard
SEP
DCI
CCC
Cluster
Stimulus
Interaction 1
(Part A)
SEP
DCI
Interaction 2
(Part B)
DCI
CCC
Interaction 3
(Part C)
SEP
DCI
Interaction 4
(Part D)
SEP
DCI
CCC
19

Back Next Item Score

Masking Periodic Table Calculator Notes Line Reader Zoom Out Zoom In Custom Settings

Willow populations in Yellowstone National Park have increased since wolves were reintroduced to the park in 1995.

Willows are small trees that grow best in marshlike environments. After studying the Yellowstone food web shown in Diagram 1 and the population data for the park shown in Table 1, students arrive at two different hypotheses.

Diagram 1. Yellowstone Food Web

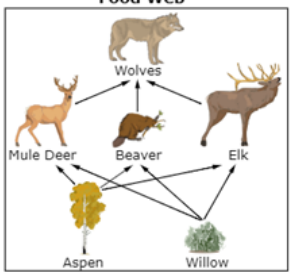


Table 1. Yellowstone Population Data

	Wolves	Elk	Beaver	Mule Deer
1995	31	16,791	10	2,014
2004	171	8,335	120	2,014

Note: These data are approximate.

Hypothesis 1:
When wolves were reintroduced to Yellowstone, the wolves preyed upon the elk, which allowed the beavers to eat more willow. This led to more beavers and beaver dams. Beaver dams create marsh environments that willows do well in, allowing the willow's population to increase.

Part A

Click on each box and select a word/phrase that completes the table with the Yellowstone population data from 1995 and 2004 and the hypothesis those data support.

Table 2. Summary of Yellowstone Population Data and Supported Hypotheses

Data	Hypothesis Supported
Elk population ▾	▾
Beaver population ▾	▾
Mule deer population ▾	▾

Part B

Which hypothesis is best supported by the evidence?

Ⓐ All of the evidence is consistent with Hypothesis 1.

Ⓑ All of the evidence is consistent with Hypothesis 2.

Ⓒ Most of the evidence is consistent with Hypothesis 1.

Ⓓ Most of the evidence is consistent with Hypothesis 2.

Ⓔ The evidence does not favor either hypothesis.

Part C

Aspen trees are shown in Diagram 1. Moose and bison are two plant-eating animal species that are not shown in Diagram 1 but are also part of the Yellowstone food web.

Based on Hypothesis 2, click on each box to select a word/phrase to make a prediction about what would happen to the moose, bison, and aspen tree populations after the reintroduction of wolves.

Table 3. Population Predictions

Species	Population after Wolf Reintroduction	Reason for Impact on Population
Moose	▾	▾
Bison	▾	▾
Aspen tree	▾	▾

Part A

Click on each box and select a word/phrase that completes the table with the Yellowstone population data from 1995 and 2004 and the hypothesis those data support.

Table 2. Summary of Yellowstone Population Data and Supported Hypotheses

Data		Hypothesis Supported
Elk population		
Beaver population		
Mule deer population		

Part B

Which hypothesis is best supported by the evidence?

Ⓐ All of the evidence is consistent with Hypothesis 1.

Ⓑ All of the evidence is consistent with Hypothesis 2.

Ⓒ Most of the evidence is consistent with Hypothesis 1.

Ⓓ Most of the evidence is consistent with Hypothesis 2.

Ⓔ The evidence does not favor either hypothesis.

21

Scoring Assertions

- Within each item are a series of explicit assertions that can be made about the knowledge and skills that a student has demonstrated based on a specific feature of the student's response
- The scoring assertions tell us:
  - The correct response
  - What inference can be made from that correct response

Review of Science Item Clusters – Scoring

Score Rationale	
The student identified that the elk population decreased between 1995 and 2004 giving evidence of understanding of how to interpret the data presented in Table 1.	✗
The student identified that the elk population decreased between 1995 and 2004 and that data supports both hypotheses giving evidence of understanding of how to use evidence to support an argument.	✗
The student identified that the beaver population increased between 1995 and 2004 giving evidence of understanding of how to interpret the data presented in Table 1.	✗
The student identified that the beaver population increased between 1995 and 2004 and that data supports Hypothesis 1 giving evidence of understanding of how to use evidence to support an argument.	✗
The student identified that the mule deer population had no change between 1995 and 2004 giving evidence of understanding of how to interpret the data presented in Table 1.	✗
The student identified that the mule deer population had no change between 1995 and 2004 and that data supports neither hypothesis giving evidence of understanding of how to use evidence to support an argument.	✗
The student identified that most of the evidence is consistent with Hypothesis 1 or the student's conclusion matches their correct inferences, given their characterization of the data, providing some evidence of his or her ability to summarize information and draw a conclusion.	✗

23


Idaho Spring 2022
Science Test
Administration

Spring 2022 Science Test Admin

- Grade band test – Grades 5, 8, and 11
- First operational administration
- Each student received 6 clusters and 12 standalones plus additional field test items

- The operational administration allows us to gather data on student performance in Idaho. This performance data will be used for standard setting.
- Before that happens, Performance Level Descriptors must be reviewed and approved by SDE and this committee.


PLD Workshop

Purpose of the PLD Workshop

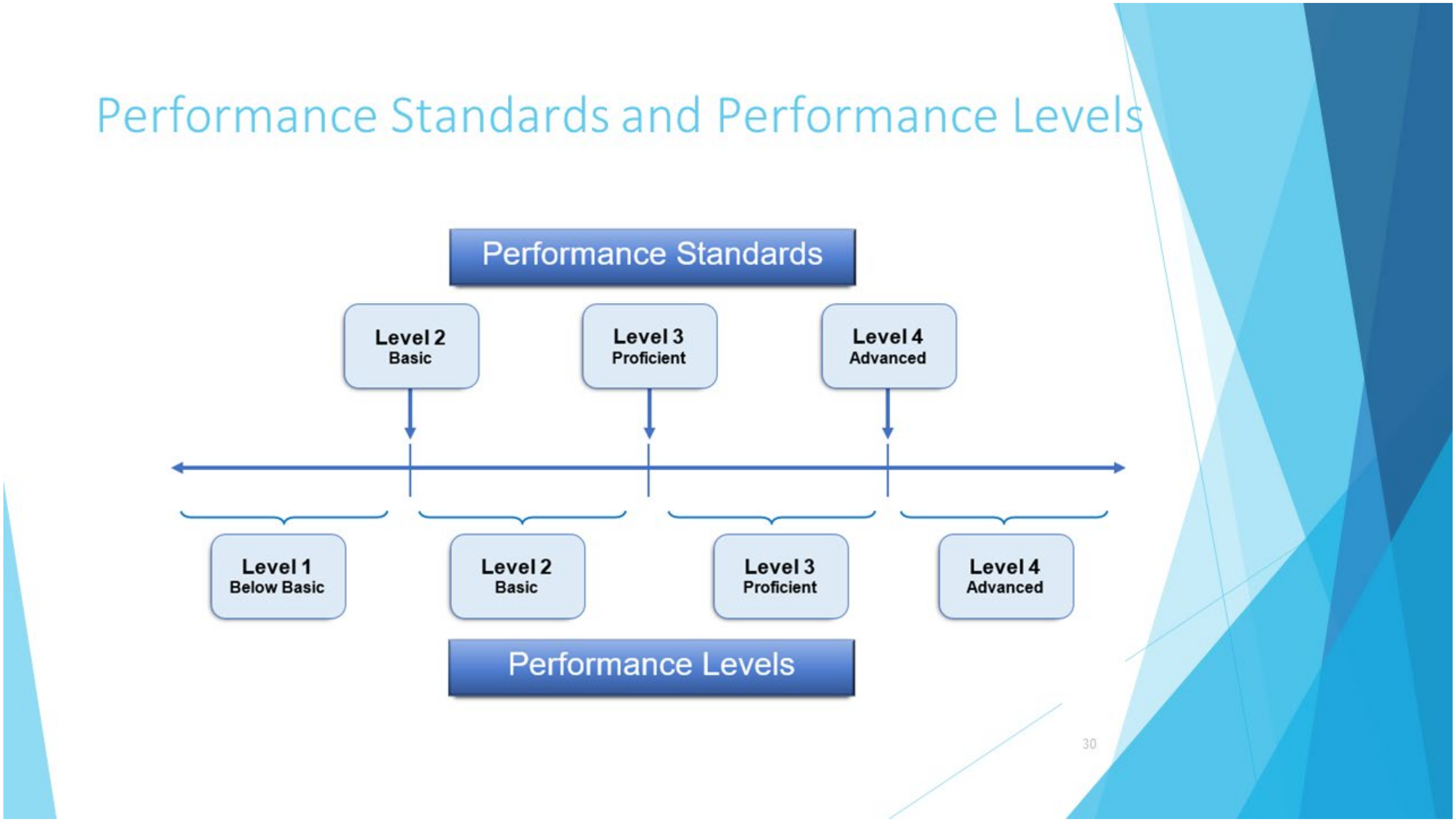
- Recommend to SDE performance level descriptors that differentiate the four achievement levels on the Idaho Standards Achievement Test in science for grades 5, 8, and 11.
- Performance levels will be used during the standard setting workshop to help panelists set performance standards.

28

Performance Level Descriptor (PLD)

- Describe what students within each achievement level are expected to know and be able to do
- PLDs are the link between the content and performance standards

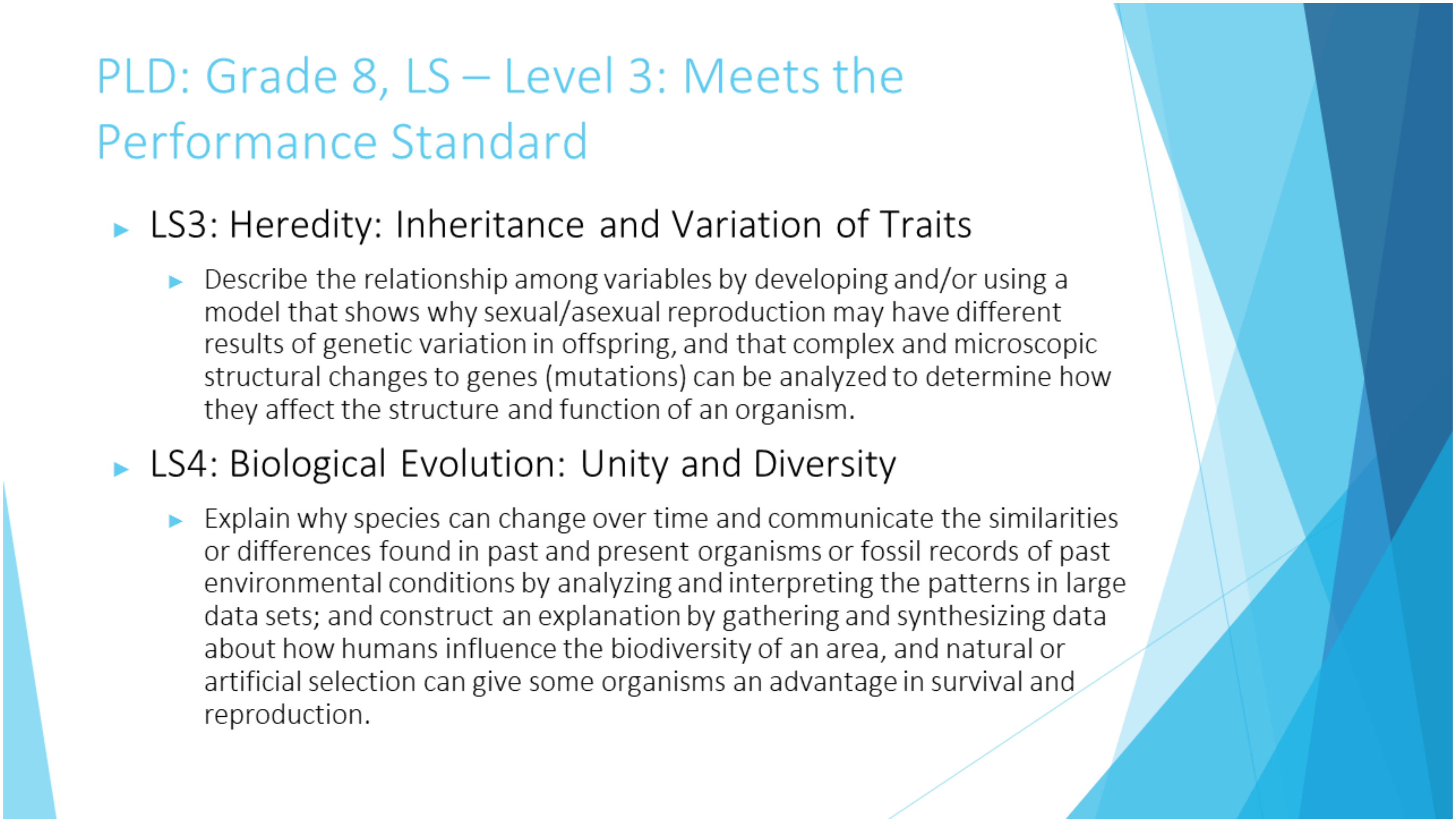
29


Performance Standards and Performance Levels
Performance Standards
Level 2
Basic
Level 3
Proficient
Level 4
Advanced
Level 1
Below Basic
Level 2
Basic
Level 3
Proficient
Level 4
Advanced
Performance Levels
30

PLD: Grade 8, LS – Level 3: Meets the Performance Standard

- LS1: From Molecules to Organisms: Structure and Processes
  - Engage in an argument using evidence by gathering and synthetizing data from an investigation or explain by developing and/or using a model that all living things are made up of cells that work together to form more complex structures and systems; both plants and animals convert energy into food sources but the process to do so is different; characteristic animal behaviors and specialized plant structures affect the probability of reproduction.
- LS2: Ecosystems: Interactions, Energy, and Dynamics
  - Explain and predict the dynamic relationships and interactions between the diverse types of living and nonliving parts of an ecosystem including the flow of energy and the cycling of matter among organisms and abiotic components of an ecosystems by developing and/or using a model; and design and support a solution to mitigate disruptions to any part of an ecosystem by human access to natural resources by analyze and interpret multiple graphical displays of data.

31

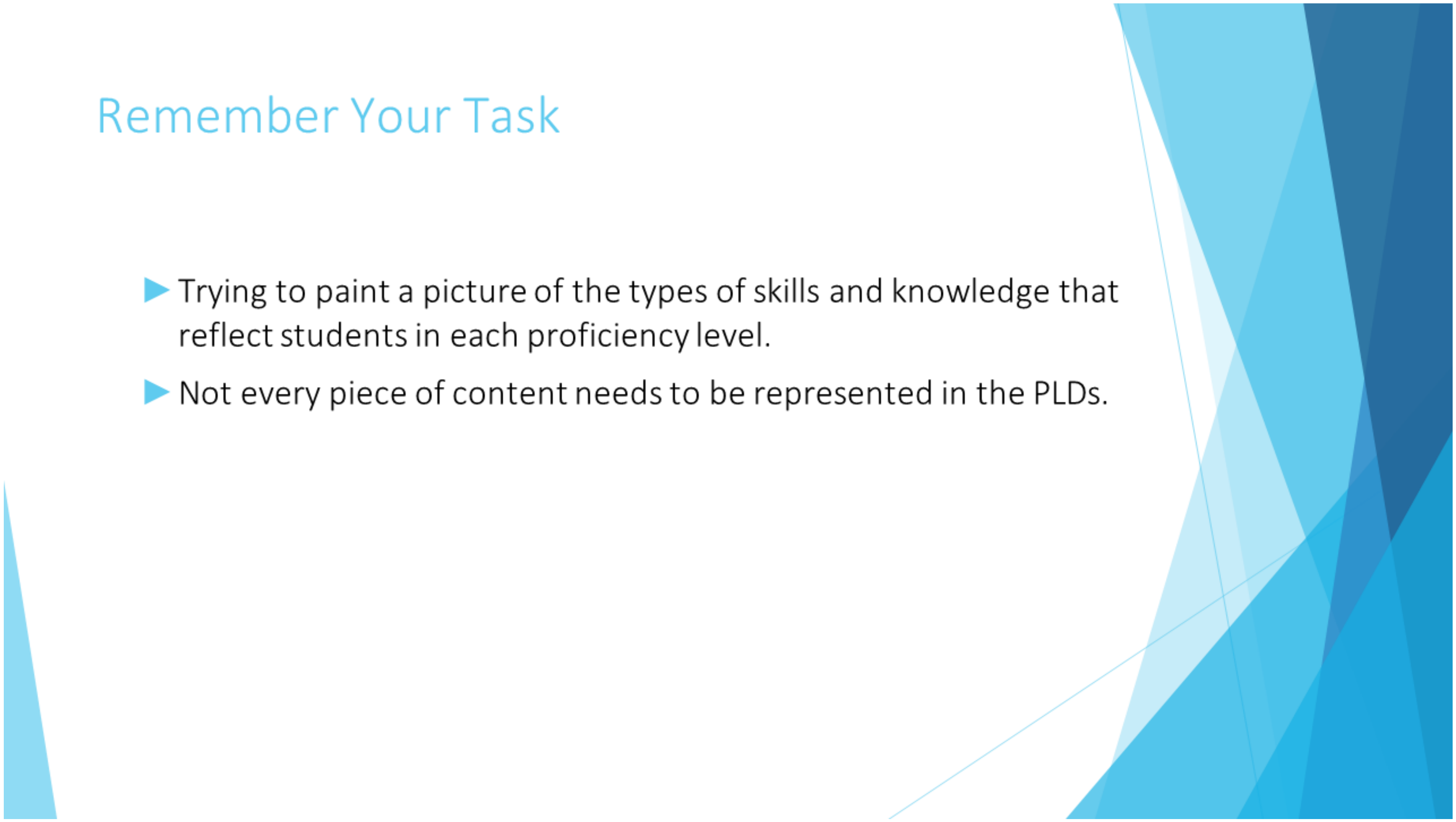

PLD: Grade 8, LS – Level 3: Meets the Performance Standard
LS3: Heredity: Inheritance and Variation of Traits
Describe the relationship among variables by developing and/or using a model that shows why sexual/asexual reproduction may have different results of genetic variation in offspring, and that complex and microscopic structural changes to genes (mutations) can be analyzed to determine how they affect the structure and function of an organism.
LS4: Biological Evolution: Unity and Diversity
Explain why species can change over time and communicate the similarities or differences found in past and present organisms or fossil records of past environmental conditions by analyzing and interpreting the patterns in large data sets; and construct an explanation by gathering and synthesizing data about how humans influence the biodiversity of an area, and natural or artificial selection can give some organisms an advantage in survival and reproduction.

PLD: Grade 8, LS3 – Across Achievement Levels

1	2	3	4
Describe the relationship among variables by identifying the components of a model that shows why sexual/asexual reproduction may have different results of genetic variation in offspring, and that complex and microscopic structural changes to genes (mutations) can be analyzed to determine how they affect the structure and function of an organism.	Describe the relationship among variables by using a model that shows why sexual/asexual reproduction may have different results of genetic variation in offspring, and that complex and microscopic structural changes to genes (mutations) can be analyzed to determine how they affect the structure and function of an organism.	Describe the relationship among variables by developing and/or using a model that shows why sexual/asexual reproduction may have different results of genetic variation in offspring, and that complex and microscopic structural changes to genes (mutations) can be analyzed to determine how they affect the structure and function of an organism.	Describe the relationship among variables by evaluating and revising a model that shows sexual/asexual reproduction may have different results of genetic variation in offspring or predicts what changes would occur in the function of an organisms if there is a mutation in the organism's genes.

Your Task

- Review the Draft PLDs
  - For each DCI, what descriptors for each performance level convey an appropriate sense of rigor?
  - Do the PLDs show a clear progression across performance levels?
  - Are the descriptors for each standard consistent within each performance level?
  - Are the descriptors for each standard appropriate for the grade level? (Even level 4 should be within the grade level.)


Remember Your Task
Trying to paint a picture of the types of skills and knowledge that reflect students in each proficiency level.
Not every piece of content needs to be represented in the PLDs.

Parse & Review the PLDs

- Work individually (30 min) to:
  - Highlight the complexity verbs that show how the expectations change from Level 1 to Level 4
  - Change the font colors to highlight words that reflect the SEPs, DCIs, and CCCs.
  - Begin to think about the skills and knowledge these students can demonstrate at each achievement level
  - Idea is to get a common mental representation of these students
- Once individual reviews are complete, we'll discuss as a group.

36

PLD: Grade 8, LS3 – Across Achievement Levels

1	2	3	4
Describe the relationship among variables by identifying the components of a model that shows why sexual/asexual reproduction may have different results of genetic variation in offspring, and that complex and microscopic structural changes to genes (mutations) can be analyzed to determine how they affect the structure and function of an organism.	Describe the relationship among variables by using a model that shows why sexual/asexual reproduction may have different results of genetic variation in offspring, and that complex and microscopic structural changes to genes (mutations) can be analyzed to determine how they affect the structure and function of an organism.	Describe the relationship among variables by developing and/or using a model that shows why sexual/asexual reproduction may have different results of genetic variation in offspring, and that complex and microscopic structural changes to genes (mutations) can be analyzed to determine how they affect the structure and function of an organism.	Describe the relationship among variables by evaluating and revising a model that shows sexual/asexual reproduction may have different results of genetic variation in offspring or predicts what changes would occur in the function of an organisms if there is a mutation in the organism's genes.


PLD Review
38


PLD Discussion
39

Thank you!

Your input and expertise are vital to this process. Thank you for your time, thoughtful input, and energy!

SDE Contacts:

- Kevin Chandler – kchandler@sde.Idaho.gov
- Brianna Lynch – blynch@sde.Idaho.gov

Appendix 3-C

ISAT in Science Range Achievement-Level Descriptors

Idaho Standards Achievement Test in Science Range Achievement-Level Descriptors

Exhibit C-1. ISAT in Science Range Achievement-Level Descriptors, Grade 5

Students that are a level ____ may be able to do things like...	**Below Basic**	**Basic**	**Proficient**	**Advanced**
Earth Science				
Earth's Place in the Solar System	Explain observable features of Earth's landscape, the appearance of stars in the night sky or the patterns created from the orbit and rotation of the Sun-Earth-Moon system by identifying data in graphical displays or in a model.	Explain the ordered observable features of Earth's landscape, the appearance of stars in the night sky or the patterns created from the orbit and rotation of the Sun-Earth-Moon system by representing data in graphical displays or models.	Explain the ordered, observable features of Earth's landscape, the appearance of stars in the night sky or the patterns created from the orbit and rotation of the Sun-Earth-Moon system by analyzing and interpreting graphical displays of data or models as evidence.	Make a prediction regarding the ordered, observable features of Earth's landscape, the appearance of stars in the night sky or the patterns created from the orbit and rotation of the Sun-Earth-Moon system by evaluating and revising graphical displays of data or models.
Earth's Systems	Make observations from data and/or collect information to identify parts of a model and reveal patterns that would show how the interactions between Earth's four major systems might cause patterned features of the Earth, including climate, distribution of water, and physical and biological constructive and deconstructive forces.	Represent data sets or graphs, and/or carry out investigations using models or information that shows how the interactions between Earth's four major systems might cause patterned features of the Earth, including climate, distribution of water, and physical and biological constructive and deconstructive forces.	Develop and/or use simple models, carry out investigations or evaluate evidence using reasoning and information regarding how the interactions between Earth's four major systems might cause patterned features of the Earth, including climate, distribution of water, and physical and biological constructive and deconstructive forces.	Develop and/or revise a model, analyze the data sets from an investigation using research methods to better communicate or predict how the interactions between Earth's four major systems might cause patterned features of the Earth, including climate, distribution of water, and physical and biological constructive and deconstructive forces.

Students that are a level ____ may be able to do things like...	Below Basic	Basic	Proficient	Advanced
Earth and Human Activity	Identify either weather-related hazards on humans or human activity on the Earth's resources and environments by using information and observations from sources.	Explain the cause and effect relationship of either weather-related hazards on humans or human activity on the Earth's resources and environments by using obtained information from a reliable source to compare multiple solutions.	Generate and evaluate the merits or accuracy of a solution that could explain and reduce the cause and effect relationship of either weather-related hazards on humans or human activity on the Earth's resources and environments by obtaining and using evidence from reliable sources.	Predict changes that can occur in the cause and effect relationships of either weather-related hazards on humans or human activity on the Earth's resources and environments by evaluating, comparing and revising a solution to a problem using evidence obtained from reliable sources.
Life Science				
From Molecules to Organisms: Structure and Processes	Identify components of a model that represent parts of a life cycle or behavioral system of organisms; and make observations about organisms that need food for the energy and materials to grow and repair their internal and external structures.	Support an argument by using a model that describes or represents the life cycles or behavioral systems of organisms; and support that organisms need food for the energy and materials to grow and repair their internal and external structures by identifying data as evidence.	Describe patterns in the life cycles or behavioral systems of organisms by developing and/or using a model; and construct an argument by using evidence that organisms need food for the energy and materials to grow and repair their internal and external structures.	Evaluate and revise a model that describes patterns in the life cycles or behavioral systems of organisms when a variable changes; and compare and refine arguments that organisms need food for the energy and materials to grow and repair their internal and external structures.
Ecosystems: Interactions, Energy, and Dynamics	Identify the parts of a model that represents interactions of organisms within an ecosystem and the cycling of matter through those interactions; and identifying data that can show how an ecosystem changed.	Describe the interactions of organisms within an ecosystem and the cycling of matter through those interactions by using a model; and collecting evidence that shows how an ecosystem can change.	Describe the interactions of organisms within an ecosystem and the cycling of matter through those interactions by developing and/or using a model; and using evidence to show the effect that occurs when one	Describes the interactions of organisms within an ecosystem and the cycling of matter through those interactions when more information is given by evaluating and revising a model; and predicting the effects of an ecosystem

Students that are a level ____ may be able to do things like...	Below Basic	Basic	Proficient	Advanced
			part of the ecosystem is changed.	when one part of the ecosystem is changed.
Heredity: Inheritance and Variation of Traits	Explain that organisms inherit the information that dictates how they look and function by collecting and recording data from pictures, drawings, and/or text; and making an observation about an organism when its environment changes.	Support an explanation that organisms inherit the information that dictates how they look and function by using data collected from tables and various graphical displays; and identifying information that would help explain what happens to an organism if the environment changes.	Construct an explanation that organisms inherit the information that dictates how they look and function by analyzing and interpreting various forms of data to; and construct an explanation using evidence that supports that an organism has changed in response to environmental changes.	Construct and revise an explanation that organisms inherit the information that dictates how they look and function by constructing, analyzing and interpreting tables and graphical displays of data; and predicting what would happen to an organism if its environment continues to change.
Biological Evolution: Unity and Diversity	Explain that when there is a change in the environment, certain individual organisms could have variations in traits that lead to advantages in survival and reproduction by identifying patterns in past or present organism characteristics that can be used as evidence; and explain that current, living organisms can only survive in particular environments or resemble organisms that once lived on Earth by using observations from pictures, drawings, and/or writings.	Support an explanation that when there is a change in the environment, certain individual organisms could have variations in traits that lead to advantages in survival and reproduction, or that living organisms resemble organisms that once lived on earth by identifying and/or recording past and present observations; and identify data that can be used to compare the merits of a solution that can affect a population of organisms.	Provide evidence that the when there is a change in the environment, certain individual organisms could have variations in traits that lead to advantages in survival and reproduction, or that living organisms resemble organisms that once lived on earth by analyzing and interpreting past and present organism characteristics; and analyze and compare the merits of a solution that can affect a population of organisms.	Evaluate and revise a constructed explanation that states that with a change in the environment, certain individual organisms could have variations in traits that lead to advantages in survival and reproduction, or that living organisms resemble organisms that once lived on earth by analyzing and interpreting past and present organism characteristics; and argue the merits of a solution that could affect a population of organisms by comparing sets of data.

Students that are a level ____ may be able to do things like...	Below Basic	Basic	Proficient	Advanced
Physical Science				
Matter and Its Interactions	Determine if a chemical reaction occurs and a new substance is created, measuring and graphing quantities to show matter is always conserved regardless of the change that occurs by making observations about variables that are controlled; and show matter exists made of particles too small to be seen by developing a model.	Determine if a chemical reaction occurs and a new substance is created, measuring and graphing quantities to show matter is always conserved regardless of the change that occurs by using models to test variables that are controlled; and to show matter exists made of particles too small to be seen by developing a model.	Determine if a chemical reaction occurs and a new substance is created, measuring and graphing quantities to show matter is always conserved regardless of the change that occurs by conducting an investigation in which variables are controlled; and to show matter exists made of particles too small to be seen by developing a model.	Determine if a chemical reaction occurs and a new substance is created, measuring and graphing quantities to show matter is always conserved regardless of the change that occurs by evaluating and revising a model using quantitative data in which variables are controlled; and to show matter exists made of particles too small to be seen by developing a model.
Motion and Stability: Forces and Interactions	Identify questions from an investigation about cause and effect relationships between balanced and unbalanced forces (magnetism and/or gravity) and an object's motion.	Provide evidence supporting an argument about cause and effect relationships between balanced and unbalanced forces (magnetism and/or gravity) and an object's motion by using observations from an investigation.	Provide evidence to construct an argument about cause and effect relationships between balanced and unbalanced forces (magnetism and/or gravity) and an object's motion by asking questions, planning and conducting an investigation and/or using produced data.	Provide evidence to predict cause and effect relationships between balanced and unbalanced forces (magnetism and/or gravity) and an object's motion by asking questions, conducting and comparing two different investigations and/or using produced data.

Students that are a level ____ may be able to do things like...	Below Basic	Basic	Proficient	Advanced
Energy	Make observations using produced data to ask questions on how energy can be used as a fuel or food; or be transferred from stored and/or motion energy to different forms like sound, light, and electrical currents.	Make observations using produced data to provide evidence on how energy can be used as a fuel or food; or be transferred from stored and/or motion energy to different forms like sound, light, and electrical currents.	Make predictions by using models or asking questions and provide evidence by using produced data on how energy can be used as a fuel or food, be transferred from stored and/or motion energy to different forms like sound, light, and electrical currents.	Make predictions by evaluating and developing and/or revising models or asking questions and/or provide evidence by using produced data on how energy can be used as a fuel or food; or be transferred from stored and/or motion energy to different forms like sound, light, and electrical currents.
Waves and their Applications in Technologies for Information Transfer	Make observations about patterns of light or mechanical waves using models; and explain using evidence how reflected light from objects causes objects to be seen.	Describe the patterns of light or mechanical waves by using a given model; and explain using evidence how reflected light from objects causes objects to be seen. Compare multiple solutions to transfer information.	Describe the patterns of light or mechanical waves by creating a solution or developing and/or using a model; and explain using evidence how reflected light from objects causes objects to be seen. Construct and compare multiple solutions to transfer information.	Make predictions and describe the patterns of light or mechanical wave by developing and/or revising a model; and explain using evidence how reflected light from objects causes objects to be seen. Construct and compare multiple solutions to transfer information.

Exhibit C-2. ISAT in Science Range Achievement-Level Descriptors, Grade 8

Students that are a level ____ may be able to do things like...	Below Basic	Basic	Proficient	Advanced
Earth Science				
Earth's Place in the Solar System	Identify components of a model that measures and collects evidence to explain the similarities and differences in the patterned motions of the Sun-Earth-Moon system, the role of gravity in the motion of galaxies and the solar system, or the relative occurrence of events in the Earth's and solar system's history.	Identify data from tables and other graphical displays by developing and/or using a simple model that can be used as pieces of evidence to explain the patterned motions of the Sun-Earth-Moon system, the role of gravity in the motion of galaxies and the solar system, or the relative occurrences of events in the Earth's and solar system's history.	Explain the patterned motions of the Sun-Earth-Moon system, the role of gravity in the motion of galaxies and the solar system, or the relative occurrence of events in the Earth's and solar system's history by developing and/or using a model or by using graphical displays of data.	Explain the patterned motions of the Sun-Earth-Moon system, the role of gravity in the motion of galaxies and the solar system, or the relative occurrence of events in the Earth's and solar system's history by evaluating and/or revising a model based on constraints and data limitations.
Earth's Systems	Identify the patterns in the flow or cycles of energy and matter throughout Earth's systems, including the sun and Earth's interior as primary energy sources by making measurements and/or observations from graphical data to help identify the components of a model; and explain that Earth's processes have changed the Earth's surface at varying spatial and time scales by identifying evidence.	Explain patterns using a model or using an investigation or using bar graphs, pictographs, and other various graphical data that supports how energy and matter flow or cycle throughout Earth's systems, including the sun and Earth's interior as primary energy sources; and explain that Earth's processes have changed the earth's surface at varying spatial and time	Develop, use and/or revise a model that shows patterns in the flow or cycles of energy and matter throughout Earth's systems, including the sun and Earth's interior as primary energy sources by analyzing data from an investigation; and construct an explanation for how Earth's processes have changed the Earth's surface at varying spatial and time scales by interpreting evidence.	Generate data that supports an explanation that shows patterns in how energy and matter flow or cycle throughout Earth's systems, including the sun and Earth's interior as primary energy sources by evaluating and revising a model; and evaluate the impact of new data by predicting how the Earth's processes will change the earth's surface at varying spatial and time scales if a new variable is introduced.

Students that are a level ____ may be able to do things like...	Below Basic	Basic	Proficient	Advanced
		scales by organizing evidence.		
Earth and Human Activity	Identify scientific questions using collected and/or graphically represented evidence regarding the dependency of humans on the environment for different resources; and identify evidence that can help design a simple solution that minimizes the effect of humans on the environment or identify the observed patterns that emerge between natural hazards and their related geological forces.	Ask questions about data or apply scientific ideas about the uneven distribution of natural resources and human dependence on the environment for those resources to design a simple solution that minimizes the effect of humans on the environment; and to explain the history of natural hazards and their related geological forces.	Ask questions and/or design a solution that could minimize the effect of humans on the environment by analyzing and interpreting sets of data regarding the uneven distribution of natural resources and human dependence on the environment for those resources; and explain the observable patterns seen in the data from the history of natural hazards and their related geological forces.	Evaluate and revise a question that can modify a design solution that minimizes the effect of humans on the environment by analyzing and interpreting sets of data regarding the uneven distribution of natural resources and human dependence on the environment for those resources; and explain the effect of humans on the environment; and predicts future patterns of natural hazards when considering the impact of humans on the environment.
Life Science				
From Molecules to Organisms: Structure and Processes	Organize information from an investigation to support an argument using evidence or identify the components of a model to explain that all living things are made up of cells that work together to form more complex structures and systems; both plants and animals convert energy into food sources but the	Support an argument using evidence by gathering and organizing information from an investigation or explain by using a model that all living things are made up of cells that work together to form more complex structures and systems; both plants and animals convert energy into food sources but the process to	Engage in an argument using evidence by gathering and synthesizing data from an investigation or explain by developing and/or using a model that all living things are made up of cells that work together to form more complex structures and systems; both plants and animals convert energy into	Evaluate and revise a model or explanation using investigative data as evidence to support an argument that all living things are made up of cells that work together to form more complex structures and systems; both plants and animals convert energy into food sources but the process to do so is

Students that are a level ____ may be able to do things like...	**Below Basic**	**Basic**	**Proficient**	**Advanced**
	process to do so is different; characteristic animal behaviors and specialized plant structures affect the probability of reproduction.	do so is different; characteristic animal behaviors and specialized plant structures affect the probability of reproduction.	food sources but the process to do so is different; characteristic animal behaviors and specialized plant structures affect the probability of reproduction.	different; characteristic animal behaviors and specialized plant structures affect the probability of reproduction.
Ecosystems: Interactions, Energy, and Dynamics	Explain the dynamic relationships and interactions between the diverse types of living and nonliving parts of an ecosystem including the flow of energy and the cycling of matter among organisms and abiotic components of an ecosystems by identifying components of a model; and suppport a solution to mitigate disruptions to any part of an ecosystem by human access to natural resources by organizing multiple graphical displays of data.	Explain the dynamic relationships and interactions between the diverse types of living and nonliving parts of an ecosystem including the flow of energy and the cycling of matter among organisms and abiotic components of an ecosystems by using a model; and suppport a solution to mitigate disruptions to any part of an ecosystem by human access to natural resources by organizing and identifying patterns from multiple graphical displays of data.	Explain and predict the dynamic relationships and interactions between the diverse types of living and nonliving parts of an ecosystem including the flow of energy and the cycling of matter among organisms and abiotic components of an ecosystems by developing and/or using a model; and design and suppport a solution to mitigate disruptions to any part of an ecosystem by human access to natural resources by analyzing and interpreting multiple graphical displays of data.	Explain and support that the dynamic relationships and interactions between the diverse types of living and nonliving parts of an ecosystem, including the flow of energy and the cycling of matter among producers, consumers, and decomposers when a variable in the system is changed by analyzing and/or revising a model; and design a solution to mitigate disruptions to any part of an ecosystem by human access to natural resources by evaluating limitations of data when analyzing and interpreting multiple graphical displays of data.

Students that are a level ____ may be able to do things like...	**Below Basic**	**Basic**	**Proficient**	**Advanced**
Heredity: Inheritance and Variation of Traits	Describe the relationship among variables by identifying the components of a model that shows why sexual/asexual reproduction may have different results of genetic variation in offspring, and that complex and microscopic structural changes to genes (mutations) can be analyzed to determine how they affect the structure and function of an organism.	Describe the relationship among variables by using a model that shows why sexual/asexual reproduction may have different results of genetic variation in offspring, and that complex and microscopic structural changes to genes (mutations) can be analyzed to determine how they affect the structure and function of an organism.	Describe the relationship among variables by developing and/or using a model that shows why sexual/asexual reproduction may have different results of genetic variation in offspring, and that complex and microscopic structural changes to genes (mutations) can be analyzed to determine how they affect the structure and function of an organism.	Describe the relationship among variables by evaluating and revising a model that shows sexual/asexual reproduction may have different results of genetic variation in offspring or predicts what changes would occur in the function of an organisms if there is a mutation in the organism's genes.
Biological Evolution: Unity and Diversity	Explain why species can change over time and communicate the similarities or differences found in past and present organisms or fossil records of past environmental conditions by identifying the patterns in large data sets; and construct an explanation using data about how humans influence the biodiversity of an area, and natural or artificial selection can give some organisms an advantage in survival and reproduction.	Explain why species can change over time and communicate the similarities or differences found in past and present organisms or fossil records of past environmental conditions by organizing and identifying the patterns in large data sets; and construct an explanation by gathering and using data about how humans influence the biodiversity of an area, and natural or artificial selection can give some organisms an advantage in survival and reproduction.	Explain why species can change over time and communicate the similarities or differences found in past and present organisms or fossil records of past environmental conditions by analyzing and interpreting the patterns in large data sets; and construct an explanation by gathering and synthesizing data about how humans influence the biodiversity of an area, and natural or artificial selection can give some organisms an advantage in survival and reproduction.	Analyze and evaluate an explanation using large data sets that show the similarities or differences found in past and present organisms or fossil records of past environmental conditions; and form an explanation by applying concepts of statistics and probability (variability) and synthesizing the data that as humans influence the biodiversity of an area, natural or artificial selection can give some organisms an advantage in survival and reproduction.

Students that are a level ____ may be able to do things like...	**Below Basic**	**Basic**	**Proficient**	**Advanced**
Physical Science				
Matter and Its Interactions	Explain the conservation of mass when two substances react by identifying the components a model; and construct an explanation by interpreting data and using evidence that supports that the properties of matter are a function of the composition of atoms and molecules that make up matter, as well as the thermal energy.	Explain the conservation of mass when two substances react by using a model; and construct an explanation by gathering and interpreting data and using evidence that supports that the properties of matter are a function of the composition of atoms and molecules that make up matter, as well as the thermal energy.	Explain the conservation of mass when two substances react by developing and/or using a model; and construct an explanation by analyzing data and using evidence that supports that the properties of matter are a function of the composition of atoms and molecules that make up matter, as well as the thermal energy.	Explain the conservation of mass when two substances react by evaluating and revising a model; and predict how changes to the molecular structure or thermal energy of matter can affect its properties by using data and evidence.
Motion and Stability: Forces and Interactions	Identify questions, conduct an investigation, and identify data, regarding the relationship between mass, force, and motion, and the attractive and repulsive forces that act at a distance (electric, magnetic, and gravitational forces.)	Identify questions, conduct an investigation, and organize and use data to make a claim regarding the relationship between mass, force, and motion, and the attractive and repulsive forces that act at a distance (electric, magnetic, and gravitational forces.)	Ask questions, plan and conduct an investigation, and analyze and interpret data to make and support a claim regarding the relationship between mass, force, and motion, and the attractive and repulsive forces that act at a distance (electric, magnetic, and gravitational forces.)	Ask questions, conduct, evaluate, and revise an investigation, and analyze and evaluate data to make and support a claim regarding the relationship between mass, force, and motion, and the attractive and repulsive forces that act at a distance (electric, magnetic, and gravitational forces.)

Students that are a level ____ may be able to do things like...	Below Basic	Basic	Proficient	Advanced
Energy	Identify components of a model that investigates how kinetic and potential energy interact, transform, or transfer to another object; and collect and record data for an investigation that provides data regarding the temperature and total energy of a system and its dependency on a variety of factors, including the types and states of matter, as well as the amount of matter involved.	Describe kinetic and potential energy interact, transform, or transfer to another object by using a given model; and collect and record data regarding the temperature and total energy of a system and its dependence on a variety of factors, including the types and states of matter, as well as the amount of matter involved.	Describe how kinetic and potential energy interact, transform, or transfer to another object by developing and using a model or investigation; and provide evidence by analyzing data from an investigation that the temperature and total energy of a system is dependent on a variety of factors, including the types and states of matter, as well as the amount of matter involved.	Predict changes to the interaction of kinetic and potential energy, including how energy is transformed, or transferred to another object by evaluating and/or revising a model; and provide evidence that the temperature and total energy of a system is dependent on a variety of factors, including the types and states of matter, as well as the amount of matter involved by applying concepts of statistics and probability.
Waves and their Applications in Technologies for Information Transfer	Describe the patterns observed between wave characteristics and wave energy by identifying the mathematical components in a model; and show that waves are reflected, absorbed, or transmitted through various materials by selecting a claim with evidence.	Describe the patterns observed between wave characteristics and wave energy by using given mathematical representations in a model; and show that waves are reflected, absorbed, or transmitted through various materials by supporting a claim with evidence.	Describe the patterns observed between wave characteristics and wave energy by developing and using mathematical representations in a model; and show that waves are reflected, absorbed, or transmitted through various materials by constructing a claim supported by evidence.	Predict patterns between wave characteristics and wave energy by evaluating and revising a mathematical model; and provide evidence to support a claim that waves are reflected, absorbed, or transmitted through various materials by integrating qualitative, quantitative, and technical data.

Exhibit C-3. ISAT in Science Range Achievement-Level Descriptors, Grade 11

Students that are a level ____ may be able to do things like...	Below Basic	Basic	Proficient	Advanced
Earth Science				
Earth's Place in the Solar System	Identify the characteristics, processes and life cycles of objects in the solar system by identifying components and limitations of a model that uses mathematical representations; and identify and critique evidence that shows the motion of objects in our solar system and Earth's early formation and geologic history.	Explain algorithms and models that describe the characteristics, processes, and life cycles of objects in the solar system by using existing mathematical concepts and processes; and construct an explanation, which uses the relationship between different variables, for the motion of objects in our solar system and Earth's early formation and geologic history.	Collect data and explain the characteristics, processes, and life cycles of objects in the solar system by developing and/or using mathematical models; and construct an explanation based on qualitative and quantitative evidence for the motion of objects in our solar system and Earth's early formation and geologic history.	Make predictions regarding the characteristics, processes, and life cycles of objects in the solar system by evaluating and revising a mathematical model; and construct and revise an explanation based on evidence, scientific theories and laws for the motion of objects in our solar system and Earth's early formation and geologic history.
Earth's Systems	Identify components and limitations of a model or investigation, including mathematical algorithms and computations, to show that energy flows into and out of one Earth system and how energy flow can cause feedback effects to occur with other Earth systems, specifically with the planet's interactions with water, solar radiation, geologic systems, and climate.	Conduct an investigation or use an existing model, including mathematical algorithms and computations, to show that energy flows into and out of one Earth system, and how energy flow can cause feedback effects with other Earth systems, specifically with the planet's interactions with water, solar radiation, geologic systems, and climate.	Analyze and use evidence as support that variations in energy flow into or out of Earth systems will cause feedback effects with other Earth systems, specifically with the planet's interactions with water, solar radiation, geologic systems, and climate by developing and/or using a model to generate and use quantitative data from an investigation.	Predict changes that can occur to the Earth's feedback mechanisms when a variable is either added or changed by evaluating and/or revising an investigation or computational model; and analyze the collected data by applying concepts of statistics and probability to show how energy flow into or out of an Earth system, specifically with the planet's interactions with water, solar radiation, geologic

Students that are a level ____ may be able to do things like...	Below Basic	Basic	Proficient	Advanced
				systems and climate, affect those feedback effects.
Earth and Human Activity	Explain how human activity has been influenced by the availability of natural resources, natural hazards, and climate change by identifying and constructing graphical displays of data; and identify the impact of climate change on Earth's systems and human society and how human society has impacted the Earth's systems by using simple mathematical representations and/or algorithms.	Support a claim that human activity has been influenced by the availability of natural resources, natural hazards, and climate change by using data from graphical displays; and identify the rate of climate change and its impact on Earth's systems and human society and how human society has impacted the Earth's systems by using a computational model.	Construct an explanation by evaluating data for how human activity has been influenced by the availability of natural resources, natural hazards, and climate change; and predict the rate of climate change and its impact on Earth's systems and human society and how human society has impacted the Earth's systems by mathematically analyzing information from natural resource data and climate models.	Evaluate and/or revise an explanation for how human activity has been influenced by the availability of natural resources, natural hazards, and climate change by using mathematical thinking; and predict the rate of climate change and its impact on Earth's systems and human society and how human society has impacted the Earth's systems by applying techniques of algebra and functions to natural resource data and climate models.
Life Science				
From Molecules to Organisms: Structure and Processes	Identify the relationships between variables that contribute to the feedback mechanisms that maintain homeostasis through the structure, function, and processes of living systems; and identify the components and limitations of a model that can be used to support an explanation for how cellular	Collect data which will serve as evidence for a model that shows that feedback mechanisms maintain homeostasis through the structure, function, and processes of living systems by conducting an investigation; and support a claim by evaluating collected data regarding	Demonstrate that feedback mechanisms maintain homeostasis through the structure, function, and processes of living systems by developing and/or using a model; and construct an explanation for cellular respiration moves energy and matter through the body, forming different products, transferring	Explain what happens to the feedback mechanisms that maintain homeostasis through the structure, function, and processes of living systems when a variable is changed by evaluating and revising a model; and make and support a claim that cellular respiration moves energy and matter through the

Students that are a level ____ may be able to do things like...	Below Basic	Basic	Proficient	Advanced
	respiration moves energy and matter through the body, forming different products, transferring energy, and replicating DNA and protein synthesis.	how cellular respiration moves energy and matter through the body, forming different products, transferring energy, and replicating DNA and protein synthesis.	energy, and replicating DNA and protein synthesis by evaluating data from an investigation.	body, forming different products, transferring energy, and replicating DNA and protein synthesis by applying scientific reasoning, theory and/or models.
Ecosystems: Interactions, Energy, and Dynamics	Identify components or variables in the cycling and flow of matter and energy among organisms in an ecosystem by using mathematical thinking; and support that the interactions with biotic and abiotic factors in ecosystems help maintain the population and diversity of organisms by identifying patterns within the evidence.	Provide an explanation with data that shows how energy and matter flow and cycle among organisms in an ecosystem using mathematical representations; explain the interactions of biotic and abiotic factors in maintaining the population and diversity of organisms in an ecosystem by evaluating and identifying patterns seen in data that can be used as evidence; and identify disturbances in conditions, biological, physical, or human induced, that may result in a new ecosystem.	Support claims about the cycling of matter and flow of energy among organisms in an ecosystem by creating and/or using mathematical, computational and algorithmic representations; and construct an explanation for how interactions with biotic and abiotic factors in ecosystems maintain the population and diversity of organisms, but that disturbances in conditions, biological, physical or human induced, may result in a new ecosystem by using patterns, evidence and reasoning.	Explain that the cycling of matter and flow of energy among organisms in an ecosystem can be disturbed when a new variable is introduced by evaluating and revising a computational model or simulation; argue that interactions with biotic and abiotic factors in ecosystems maintain the population and diversity of organisms by using patterns as well as mathematical and computational evidence; and predict how an ecosystem might change with a disturbance in conditions, biological, physical or human induced.

Students that are a level ____ may be able to do things like…	Below Basic	Basic	Proficient	Advanced
Heredity: Inheritance and Variation of Traits	Identify an observation or model of DNA, chromosomes, and traits; and identify evidence which supports a claim about genetic and environmental factors that may affect the variation and distribution of traits in a population by using graphical displays of data.	Ask a question that requires sufficient, empirical evidence to answer regarding the relationship of DNA, Chromosomes, and traits; and make a claim about genetic and environmental factors and their effect on variation within a population by analyzing data.	Ask and investigate a question which determines the relationship between the role of DNA and chromosomes, and traits by analyzing a model or theory; and construct an argument about genetic and environmental factors that may affect the variation and distribution of traits in a population by applying mathematical concepts to analyze evidence.	Analyze and evaluate the relationship between the role of DNA and chromosomes, and traits by using a question; and predict the variation and distribution of traits in population when a genetic and environmental factor is changed by applying concepts of statistics and probability to analyze evidence.
Biological Evolution: Unity and Diversity	Identify and use genetic and anatomical evidence to support that evolution, extinction, and formation of new species is based on different environmental factors by obtaining evidence from texts and mathematical representations.; and identify causal and correlational relationships of environmental conditions and population adaptations.	Provide genetic and anatomical evidence for how given factors have resulted in diversity through evolution, extinction, and formation of new species by constructing and/or using graphical displays of data; and support that environmental conditions can lead to adaptations within populations by analyzing data to distinguish between causal and correlational relationships.	Construct an explanation for how given factors have resulted in diversity through evolution, extinction, and formation of new species by using genetic and anatomical information obtained from texts, mathematical, computational, and/or algorithmic representations; and support the argument that environmental conditions can lead to adaptations within populations by generating and analyzing mathematical data.	Evaluate and revise an explanation to predict what would happen to a current species when a given factor is changed by using genetic and anatomical information obtained from texts and/or mathematical, computational and/or algorithmic representations; and predict and support the adaptations a population may experience when environmental conditions are changed.

Students that are a level ____ may be able to do things like...	Below Basic	Basic	Proficient	Advanced
Physical Science				
Matter and Its Interactions (Chemistry)	Identify the patterns in the periodic table as well as variables and limitations of a model that provides an explanation for the properties and characteristics of matter; and identify evidence for an explanation that any chemical process that occurs between matter is due to a collision of molecules, change in energy, and atom configuration of the elements involved by applying mathematical concepts to an investigation that produces data.	Provide an explanation for the properties and characteristics of matter by developing a model of atomic structure, including simple computations and algorithms, using the periodic table; and support the claim that any chemical process that occurs between matter is due to a collision of molecules, change in energy, and atom configuration of the elements involved by collecting data from an investigation that can be analyzed for patterned evidence.	Construct an investigation and/or mathematical model that explains the properties and characteristics of matter by using the periodic table, subatomic structures and corresponding electrical interactions; and provide quantitative and qualitative evidence that any chemical processes that occur between matter is due to a collision of molecules, change in energy and atom configuration of the elements involved.	Evaluate and/or revise a mathematical model or investigation that predicts the properties and characteristics of matter when a component is changed by using the periodic table, subatomic structures and corresponding electrical interactions; and construct and/or revise an explanation that any chemical processes that occur between matter is due to the collision of molecules, change in energy and atom configuration of elements.
Motion and Stability: Forces and Interactions (Physics)	Identify limitations or components of an investigation that shows the relationship between either force and the distance between interacting objects, or force, mass, and acceleration by collecting and/or producing data; and identify evidence that supports how an object moves by interpreting graphical displays of data.	Distinguish between causal and correlational relationships between force and the distance between interacting objects, or force, mass, and acceleration by using mathematical concepts and processes; and describe the motion of an object using mathematical and graphical representations.	Plan and conduct an investigation to serve as the basis for a model that explains the relationship between either force and the distance between interacting objects, or force, mass, and acceleration; and explain changes in the motion of an object by using mathematical, graphical, and	Evaluate and revise an explanation or predict changes to an investigative outcome, when a variable is changed when modeling the mathematical relationship between either force and the distance between interacting objects, or force, mass, and acceleration; and revise an explanation and predict changes in the motion of an

Students that are a level ____ may be able to do things like...	Below Basic	Basic	Proficient	Advanced
			computational analysis to observe patterns.	object when new information is introduced using scientific ideas, principles, and/or evidence.
Energy (Chemistry and Physics)	Describe how energy transfers within and between systems by calculating quantities of energy or identifying components and variables of an investigation; and identify evidence that energy is not created nor destroyed, but converted to less useful forms by using a model.	Collect and/or use mathematical data from an investigation to serve as the basis for a model that provides evidence of energy transfer within and between systems; and support the claim that energy is not created nor destroyed, but converted into less useful forms by developing and/or using a model.	Create and use a mathematical model to describe how energy transfers within and between systems by using collected or produced data; and support the claim that energy is neither created nor destroyed, but converted to less useful forms by gathering empirical data.	Predict how energy transfers within and between systems by evaluating and revising a mathematical model using scientific ideas, principles, theories and/or newly added information or data; and support the claim that energy is neither created nor destroyed, but converted to less useful forms by analyzing, empirical data.
Waves and their Applications in Technologies for Information Transfer (Physics)	Identify data that shows the relationship between wavelength, amplitude, and frequency, and other wave phenomena by integrating qualitative and quantitative information; and identify components of energy transfer by waves by using mathematical representations.	Apply quantitative data, hypotheses, and/or conclusions that shows the relationship between wavelength, amplitude, and frequency, and other wave phenomena; and describe energy transfer by waves by using mathematic and algorithmic thinking.	Analyze a claim regarding the relationship between wavelength, amplitude, and frequency, and other wave phenomena by using technical, scientific information; and explain how energy transfers and the effects on the wave due to the nature of a wave medium by creating or using computational models.	Evaluate evidence regarding the relationship between wavelength, amplitude, and frequency, and other wave phenomena by using models and technical scientific information; and predict the effects on the wave due to the nature of a wave medium use mathematical, computational and/or algorithmic produced data.

Appendix 3-D

Standard-Setting Assertion Maps

Standard-Setting Assertion Maps

Exhibit D-1. Standard-Setting Assertion Map, Science Grade 5

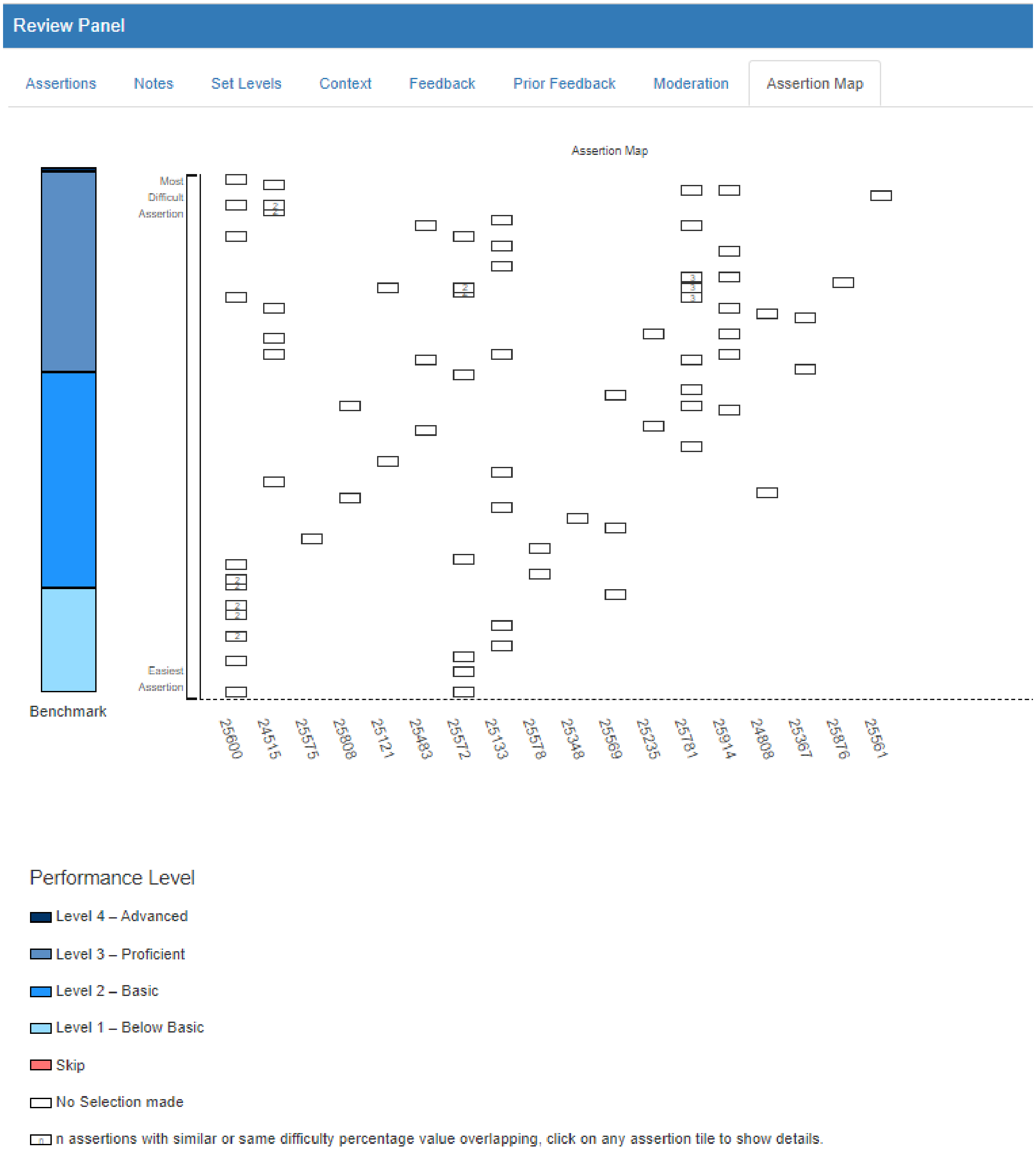


Exhibit D-2. Standard-Setting Assertion Map, Science Grade 8

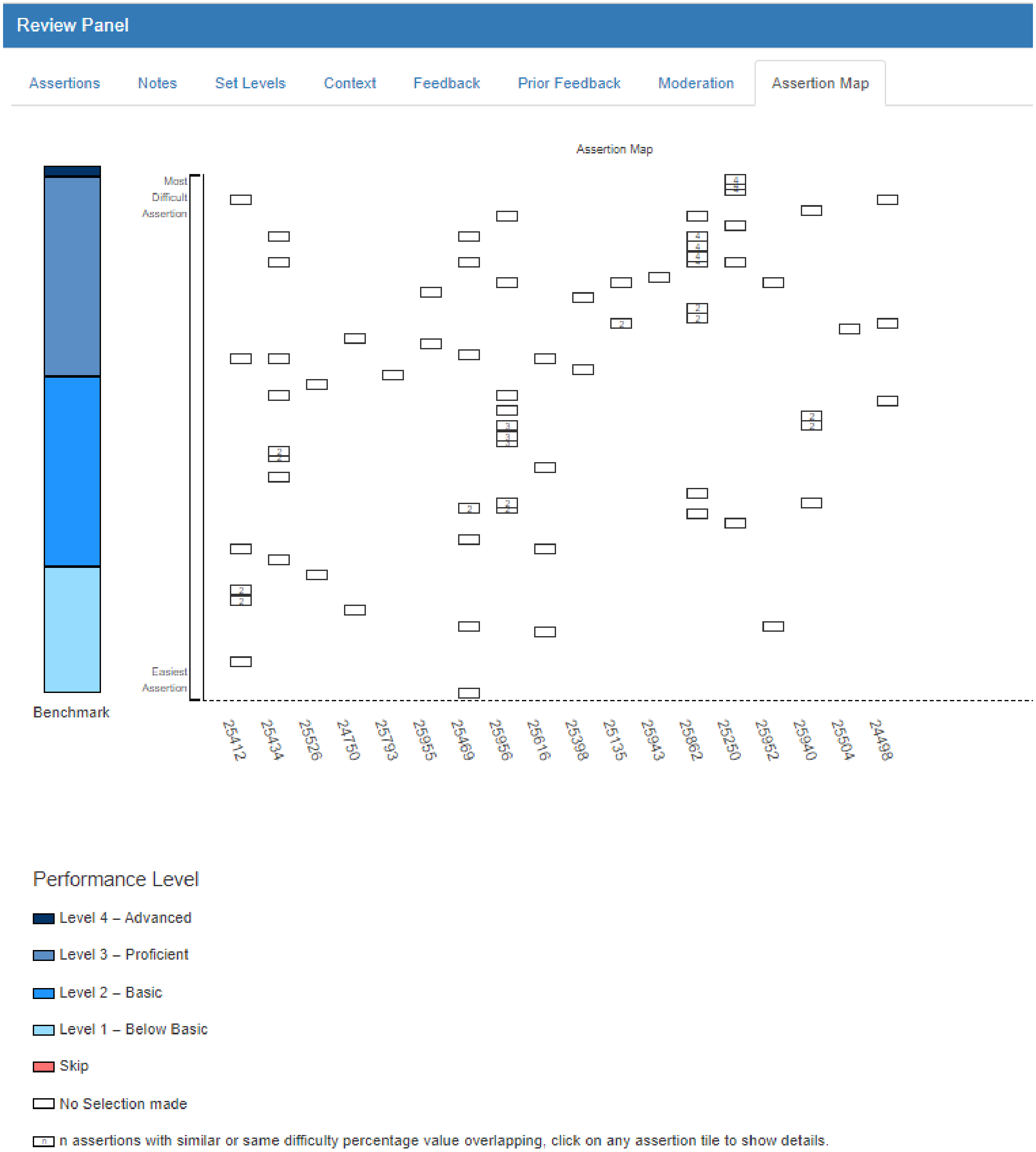
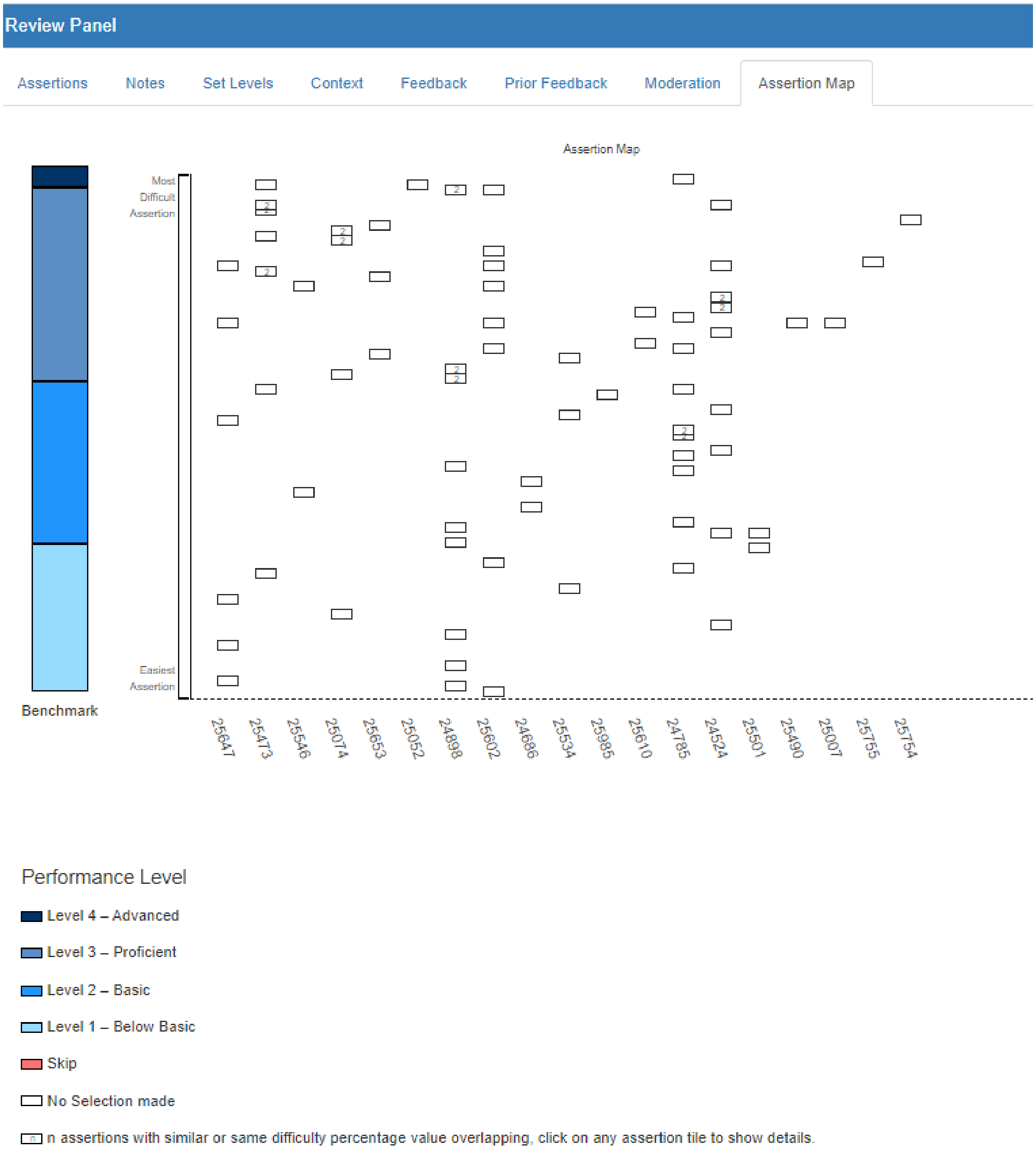


Exhibit D-3. Standard-Setting Assertion Map, Science Grade 11



Appendix 3-E

Standard-Setting Workshop Agenda

Standard-Setting Workshop Agenda

Exhibit E-1. Day 1 Standard-Setting Workshop Agenda

IDAHO
STATE DEPARTMENT OF EDUCATION

2022 Standard Setting for the ISAT in Science

SCIENCE PANEL AGENDA

July 19 – 20, 2022

Standard-Setting Workshop Day 1 – Tuesday, July 19, 2022	
8:00 – 8:30 a.m.	**Participant Login**
8:30 – 8:45 a.m.	**Welcome and Introductions from the Idaho State Department of Education (SDE)**
8:45 – 9:30 a.m.	**Large-Group Orientation and Introductory Training** Welcome and introductions Purpose of standard-setting workshop General overview of standard-setting procedures and key concepts • Performance-level descriptors (PLDs) • Item clusters and stand-alone items Item interactions Scoring assertions • Item cluster review • Assertion mapping – two rounds • Contextual information – benchmark and impact data • Panelist feedback and impact data
9:30 – 9:45 a.m.	**Break, and Separate into Small Group Rooms**
9:45 – 11:15 a.m.	**Panelists Experience Online Operational Assessment and Test Environment**
11:15 – 12:15 p.m.	**Review Range PLDs and Discuss Threshold PLDs** Parse range PLDs to identify specific claims within performance levels Identify knowledge and skills differentiating student performance between levels
12:15 – 1:00 p.m.	**Lunch (on your own)**
1:00 – 2:00 p.m.	**Continue Discussions of PLDs**
2:00 – 5:00 p.m.	**Review of Ordered Scoring Assertion Booklet (OSAB) Items** Composition of the item clusters and stand-alone items Training on how to review item clusters and stand-alone items • How do the item interactions support the scoring assertion? • Why is this assertion more difficult than the previous assertion? • How does the scoring assertion and the underlying interactions relate to the PLDs? Training on usage of contextual information – benchmark and impact data Instruction in accessing the item clusters and stand-alone items Review of item clusters and stand-alone items in the OSAB
5:00 p.m.	**Adjourn**

Cambium Assessment, Inc. 1

Exhibit E-2. Day 2 Standard-Setting Workshop Agenda

2022 Standard Setting for the ISAT in Science

SCIENCE PANEL AGENDA

July 19 – 20, 2022

Standard-Setting Workshop Day 2 – Wednesday, July 20, 2022	
8:30 – 10:00 a.m.	**Continue Review of OSAB Items**
10:00 – 11:00 a.m.	**Training on Assertion-Mapping Task** Review of assertion-mapping key concepts • Performance-level descriptors (PLDs) • Ordered scoring assertions • Assertion map • Contextual information – benchmark and impact data Training on assertion-mapping tool Practice assertion-mapping task and standard-setting quiz
11:00 – 11:15 a.m.	**Break**
11:15 – 12:30 p.m.	**Round 1 Assertion Mapping** Review of assertion-mapping procedures and key concepts Completion of assertion-mapping readiness form Round 1 assertion mapping
12:30 – 1:30 p.m.	**Lunch (on your own)**
1:30 – 2:30 p.m.	**Review Panelist Feedback Data and Discuss Round 1 Results** How to use panelist agreement feedback data and impact data Presentation and discussion of Round 1 panelist agreement feedback data and impact data
2:30 – 3:30 p.m.	**Round 2 Assertion Mapping** Review of assertion-mapping procedures and key concepts Completion of assertion-mapping readiness form Round 2 assertion mapping
3:30 – 4:00 p.m.	**Workshop Evaluations and Educator Panel Adjourn**
4:00 – 5:00 p.m.	**Across Grade Moderation with All Science Table Leaders**
5:00 p.m.	**Table Leader Adjourn**

Cambium Assessment, Inc. 2

Appendix 3-F

Standard-Setting Training Slides

Exhibit F-1. Large-Group Orientation Slides

Cambium Assessment

IDAHO
STATE DEPARTMENT OF EDUCATION

Standard Setting: Science

July 19 – 20, 2022
Idaho Standards Performance Test (ISAT) in Science

2

Welcome and Introductions

Idaho State Department of Education (SDE)

State Representatives

3

Idaho State Department of Education (SDE)

- Kevin Chandler, Director of Assessment

4

Large-Group Orientation

Cambium Assessment, Inc.

Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

Workshop Leaders

5

- **Cambium Assessment, Inc.**
  - Psychometrics
    - Frank Rijmen
    - Yi-Fang Wu
    - Jiajun Xu
  - Room Facilitators
    - Grade 5: Kam Mangis de Mark, Hibbah Haddam, Kevin Dwyer
    - Grade 8: Heather MacRae, Jim McCann
    - Grade 11: Matt Davis, Mark Palamo

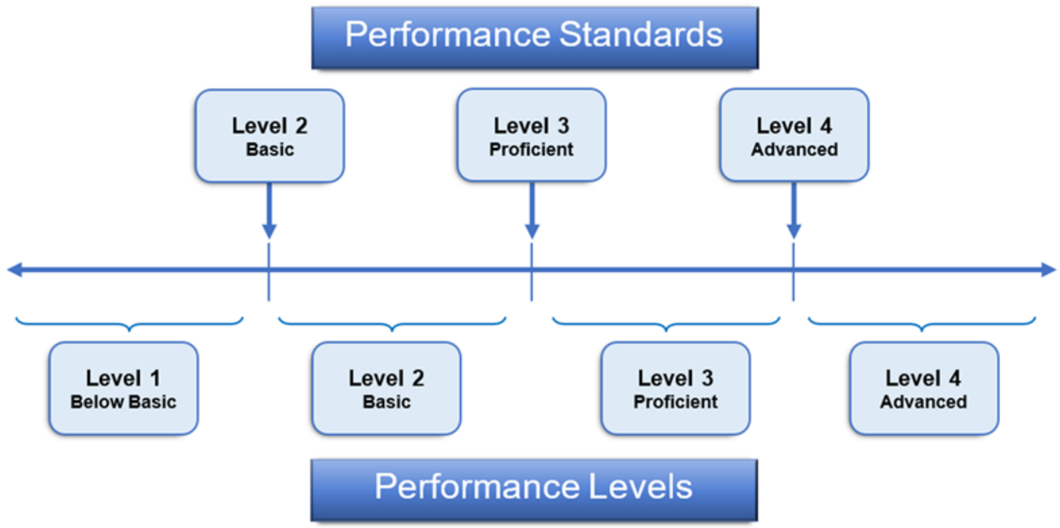
Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

Purpose of the Standard-Setting Workshop

6

- Recommend to the Idaho State Board of Education three performance standards to differentiate the four performance levels on the ISAT in Science



Main Workshop Activities

7

- Large-Group Orientation
- Panel Training
  - Take the Online Operational Assessment
  - Review Range PLDs
  - Discuss Just Barely PLDs
  - Review the Ordered Scoring Assertion Booklet
  - Training on Assertion-Mapping Procedure
- Recommend Performance Standards
  - Two rounds
  - Panelist feedback following Round 1
  - Vertical Articulation
- Workshop Evaluation

Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

Importance of Security

8

- Please do not:
  - Create any form of electronic copy of test content (screenshots, electronic notes, etc.)
  - Create any hand-written notes of test content
  - Discuss test content with anyone outside the meeting
  - Use your computer during the course of the meeting for any purpose other than participating in the item review (e.g., email, web browsing, social media)
  - Save notes about item or passage content to your computer

Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

Reason for New Science Standards

9

- The Idaho State Science Standards (ISSS) were approved in 2018
- In Spring 2022, assessments aligned to the ISSS were administered to all students Idaho in grades 5, 8, and 11

Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

Description of the Science Test Design

10

- Grades 5, 8, and 11 tests assess students' understanding of the ISSS
- ISAT in Science at grades 5, 8, and 11 includes 6 item clusters and 12 or 13 stand-alone items
  - **Item clusters** include a stimulus and a series of questions that generally take students about 6–12 minutes to complete
  - **Stand-alone items** are shorter and generally take 1–3 minutes to complete
- All items ask students to use science and engineering practices and apply their understanding of disciplinary core ideas and crosscutting concepts to make sense out of real-world phenomena

Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

Scoring Assertions

11

- Within each item cluster, a series of explicit assertions can be made about the knowledge and skills that a student has demonstrated based on specific features of the student's responses
- Scoring assertions can be supported based on students' responses in one or more interactions within an item cluster.
- For example:
  - A student correctly graphs data points indicating that (s)he can construct a graph showing the relationship between two variables,
  - Makes an incorrect inference about the relationship between the two variables, thereby not supporting the assertion that the student can interpret relationships expressed graphically

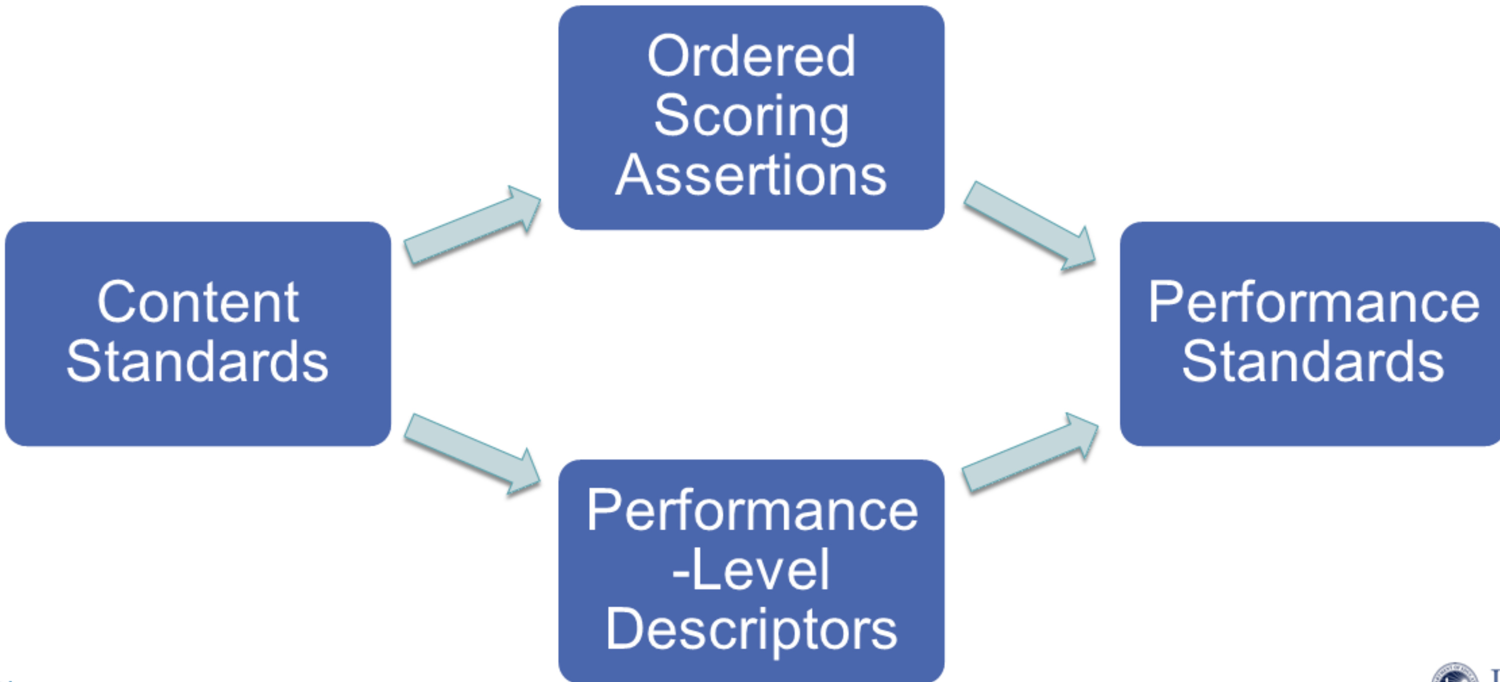
Standard Setting

12

- Systematic process by which trained participants use their knowledge of academic content standards, test items, and student performance to recommend cut-scores associated with each performance level on the test

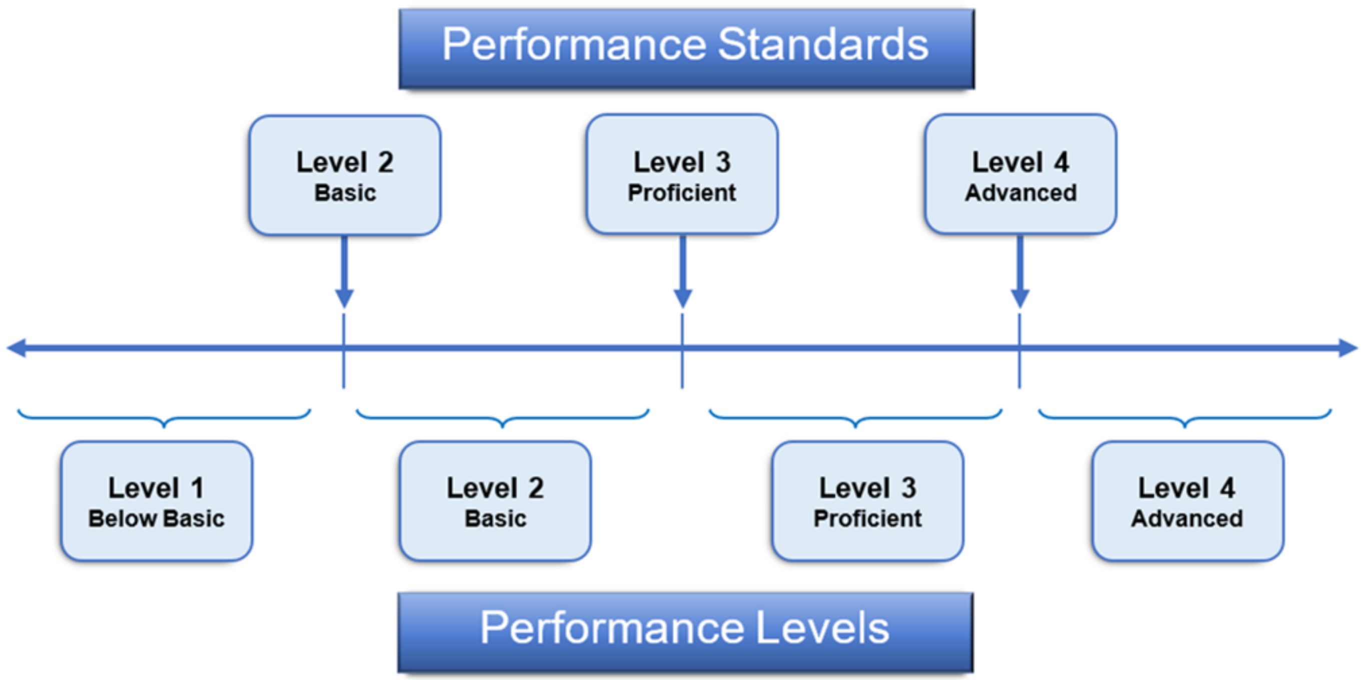
From Content Standards to Performance Standards

13



Performance Standards and Performance Levels

14



Assertion-Mapping Procedure (AMP)

15

- Test-centered procedure
- Employs an ordered item procedure adapted to accommodate new multiple interaction item types
- Map ordered scoring assertions to performance levels
- Is being employed to recommend performance standards in multiple states assessing three-dimensional science standards

Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

Assertion-Mapping Procedure (AMP)

15

- Test-centered procedure
- Employs an ordered item procedure adapted to accommodate new multiple interaction item types
- Map ordered scoring assertions to performance levels
- Is being employed to recommend performance standards in multiple states assessing three-dimensional science standards

Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

Performance-Level Descriptors (PLDs)

17

- Describe what students within each performance level are expected to know and be able to do
- PLDs are the link between the content and performance standards

Grade 8 Range PLDs – Level 3 Proficient Standard

18

Earth and Space Sciences

- **Earth's Place in the Solar System:** Explain the patterned motions of the Sun-Earth-Moon system, the role of gravity in the motion of galaxies and the solar system, or the relative occurrence of events in the Earth's and solar system's history by developing and/or using a model or by using graphical displays of data.
- **Earth's Systems:** Develop, use and/or revise a model that shows patterns in the flow or cycles of energy and matter throughout Earth's systems, including the sun and Earth's interior as primary energy sources by analyzing data from an investigation; and construct an explanation for how Earth's processes have changed the Earth's surface at varying spatial and time scales by interpreting evidence.
- **Earth and Human Activity:** Ask questions and/or design a solution that could minimize the effect of humans on the environment by analyzing and interpreting sets of data regarding the uneven distribution of natural resources and human dependence on the environment for those resources; and explain the observable patterns seen in the data from the history of natural hazards and their related geological forces.

Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

Grade 8 Range PLDs Across Performance Levels

19

Earth's Place in the Solar System

- **Level 1: Identify components of a model** that measures and collects evidence to explain the similarities and differences in the patterned motions of the Sun-Earth-Moon system, the role of gravity in the motion of galaxies and the solar system, or the relative occurrence of events in the Earth's and solar system's history.
- **Level 2: Identify data from tables and other graphical displays by developing and/or using a simple model** that can be used as pieces of evidence to explain the patterned motions of the Sun-Earth-Moon system, the role of gravity in the motion of galaxies and the solar system, or the relative occurrences of events in the Earth's and solar system's history.
- **Level 3: Explain** the patterned motions of the Sun-Earth-Moon system, the role of gravity in the motion of galaxies and the solar system, or the relative occurrence of events in the Earth's and solar system's history **by developing and/or using a model or by using graphical displays of data.**
- **Level 4: Explain** the patterned motions of the Sun-Earth-Moon system, the role of gravity in the motion of galaxies and the solar system, or the relative occurrence of events in the Earth's and solar system's history **by evaluating and revising a model based on constraints and data limitations.**

"Just Barely" Meets the Performance Standard

20

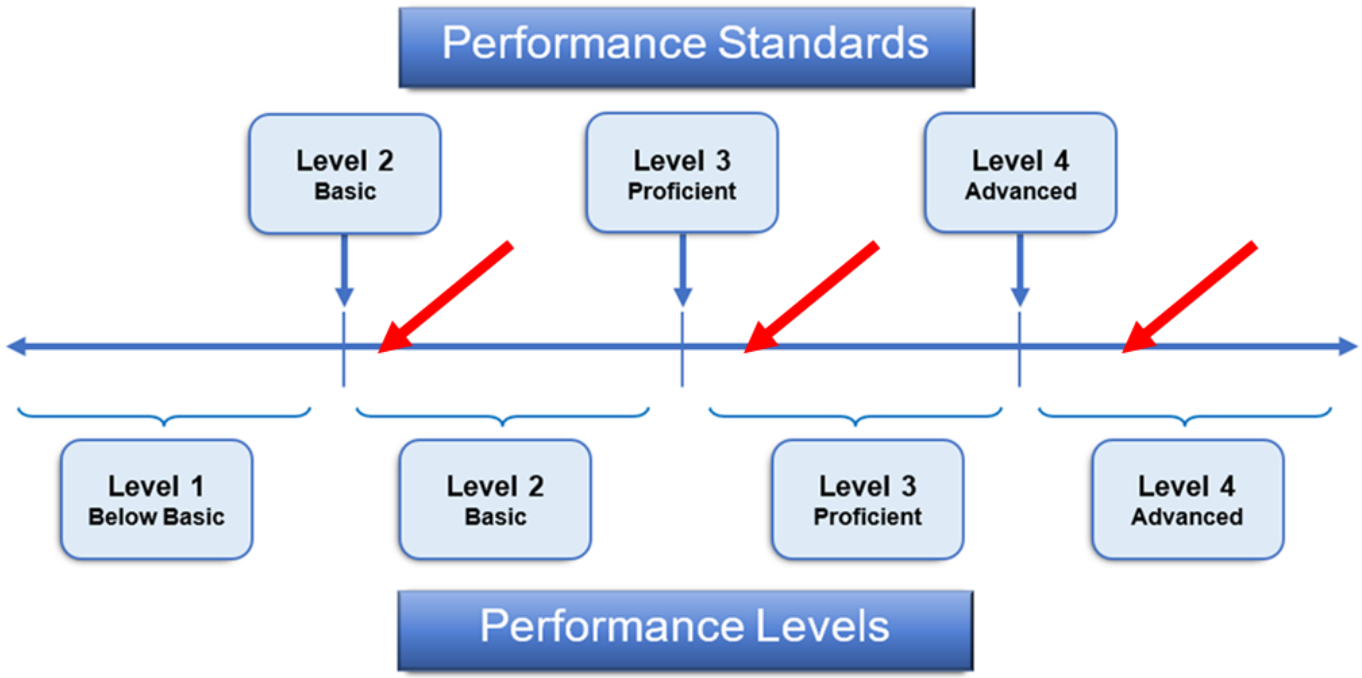
- When considering each performance level, we are especially interested in the transition areas between performance levels
- Pay attention to characteristics of students who ***just barely*** qualify for entry into the performance level from those just below
  - Not a typical example of students in the performance level
  - Although they are not good examples of the performance level, they do still meet the standard, or description in the PLD

Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

Threshold "Just Barely" PLDs

21



Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

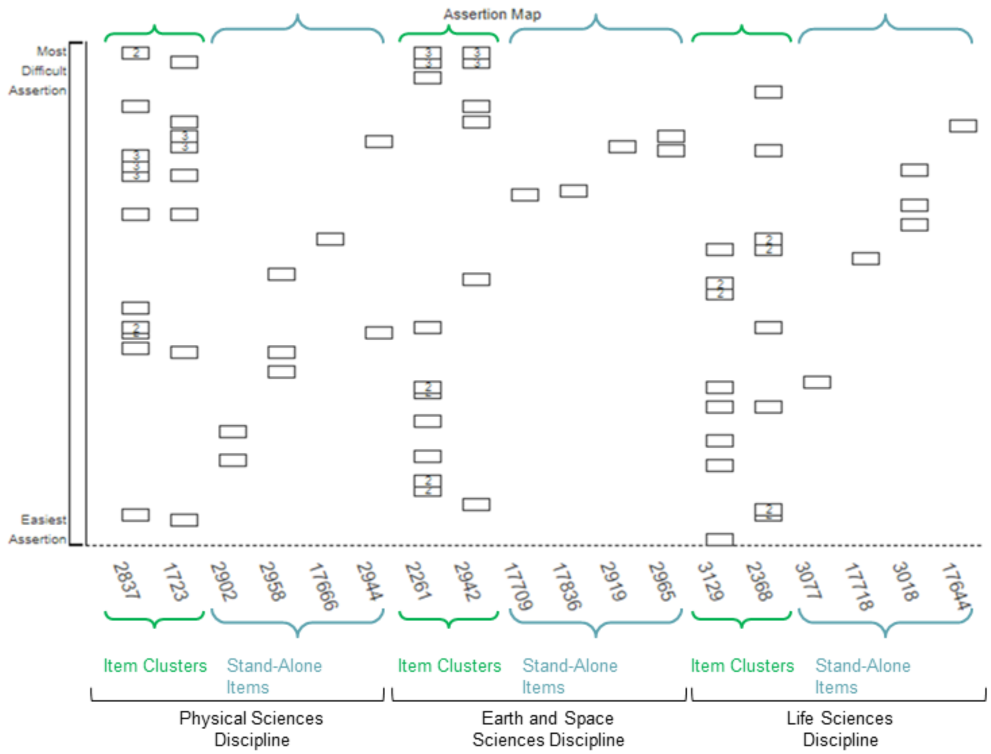
Ordered Scoring Assertions

22

- The ordered scoring assertion booklet (OSAB) constitutes a test administration:
  - A test form that meets test blueprint specifications
- It is important to evaluate scoring assertions as they relate to the item interactions
- Assertions within items are ordered by difficulty
  - Assertions within an item may not represent all PLDs

Assertion Map

23



Studying the Items and Scoring Assertions

24

- Working individually, for each scoring assertion ask yourself:
  1. *How do the item interactions support the scoring assertion?*
  2. *Why is this assertion more difficult than the previous assertions (within the item)?*
  3. *How does the scoring assertion and the underlying interactions relate to the PLDs?*
- Working as a group
  - Discuss how item interactions support scoring assertions
  - Discuss ordering of scoring assertions
  - Discuss how scoring assertions are related to the PLDs

Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

What If an Assertion Seems Out of Order?

25

- Assertion ordering is based on student performance
- Assertions may seem out of order because they are ordered by difficulty, and not by content or cognitive process
- Identify why a scoring assertion is more difficult than the assertions before it, and easier than the assertions following it
  - Pay special attention to the interactions supporting the assertions
  - Assertions may be more or less difficult because of the underlying interactions

What If an Item Seems Wrong or Unfair?

26

- Do not let yourself get distracted – this is not an item review meeting
- If you believe something is wrong with an item interaction or scoring assertion, tell the Workshop Leader, then skip over the assertion as you review the rest of the assertions within the item

Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

Assertion-Mapping Task

27

- Map assertions to performance levels
  - Map each assertion to the performance level that the assertion best supports
  - Consider what differentiates students who just barely qualify for entry into the performance level from those not quite ready for entry into the performance level
  - Evidence that the student has demonstrated knowledge and skills necessary for entry into the performance level

Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

Assertion-Mapping Task

28

- Contextual information for each assertion
  - Benchmarking data: 2015 Idaho NAEP Science results
    - Shows the associated performance level for the NAEP Science Assessment if the Science standards were to exhibit the same rigor
  - Student impact data: percentage of students performing at or above the level associated with the assertion

Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

Assertion-Mapping Task

29

- Map assertions in the online standard-setting tool
  - Mapping of assertions to performance levels should respect the ordering of assertions by difficulty within an item
    - No inversions within an item

Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

Group Feedback and Discussion

30

- Goals
  - Add important information to your thinking
  - Develop common understandings
  - Inform possible re-evaluation of assertion mappings
- Expectation is converging judgments
  - Consensus is not a requirement or goal

Feedback and Impact Data

31

- Percentage of students reaching or exceeding the standard based on assertion mappings
- Group discussion
  - Does the percentage of students reaching or exceeding the current recommended performance standard seem reasonable?
  - What are the implications for the performance standards?
  - All performance standard recommendations should be based on content rationales drawn from the Idaho State Science Standards

Creating a System of Performance Standards

32

- Performance standards for a statewide system must be coherent across grades and subjects
  - Articulation
  - Benchmarking
  - Moderation

Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

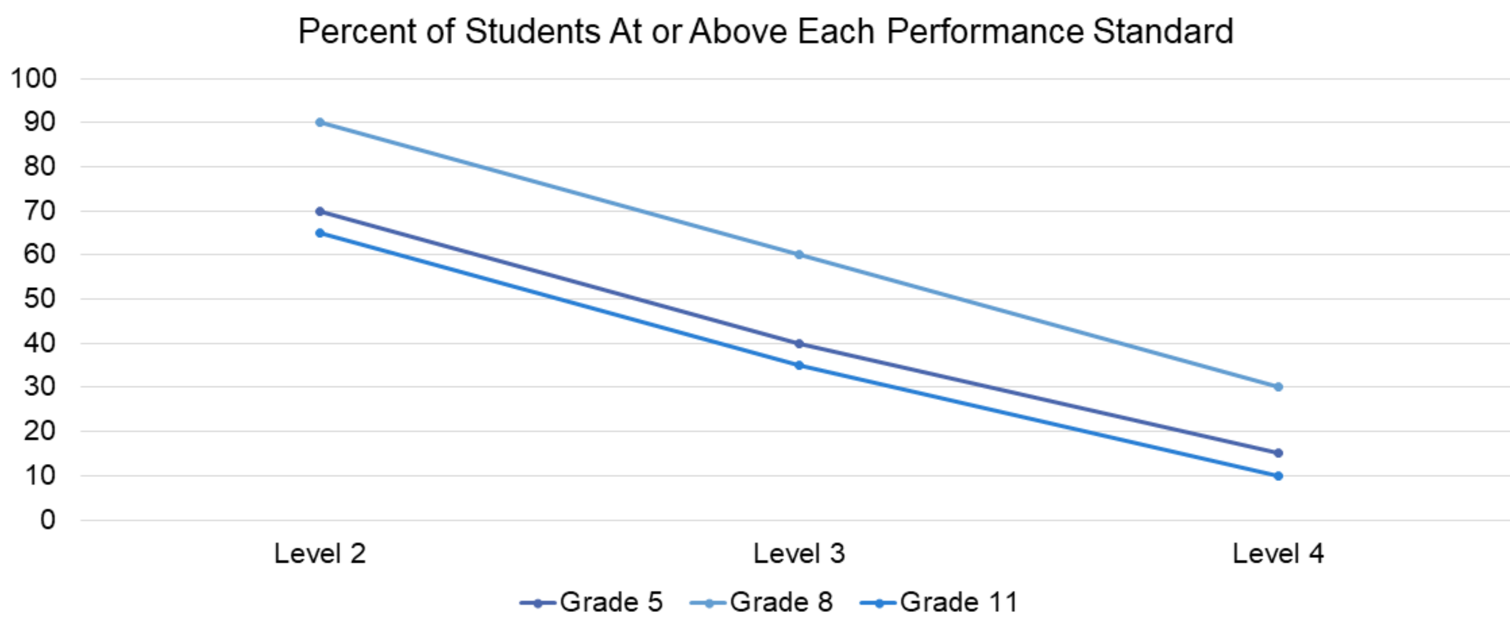
Benchmarking

33

- Are performance standards nationally competitive and represent on track for college readiness?
- Performance levels for benchmark assessments will provide context about the general neighborhood in which performance standards likely reside

Articulation

34



Moderation

35

- After the standards have been recommended by the panelists, the Table Leaders meet to review the outcomes
  - All members are invited to observe this meeting but only the Table Leaders participate
- If there are anomalies across grades or subjects the Table Leaders are permitted to adjust the performance standards (assuming there is a good content reason for doing so)

Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

Break Into Groups

36

Panel	Facilitators
Grade 5 Science	Kam Mangis de Mark Hibbah Haddam Kevin Dwyer
Grade 8 Science	Heather MacRae Jim McCann
Grade 11 Science	Matt Davis Mark Palamo

Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

Exhibit F-2. Breakout Room Slides

Cambium Assessment

IDAHO
STATE DEPARTMENT OF EDUCATION

Standard Setting: Science Breakout Rooms

July 19 – 20, 2022
Idaho Standards Performance Test (ISAT) in Science

2

Welcome and Introductions

Idaho State Department of Education (SDE)

Cambium Assessment

IDAHO
STATE DEPARTMENT OF EDUCATION

3

Breakout Room Training

Cambium Assessment, Inc.

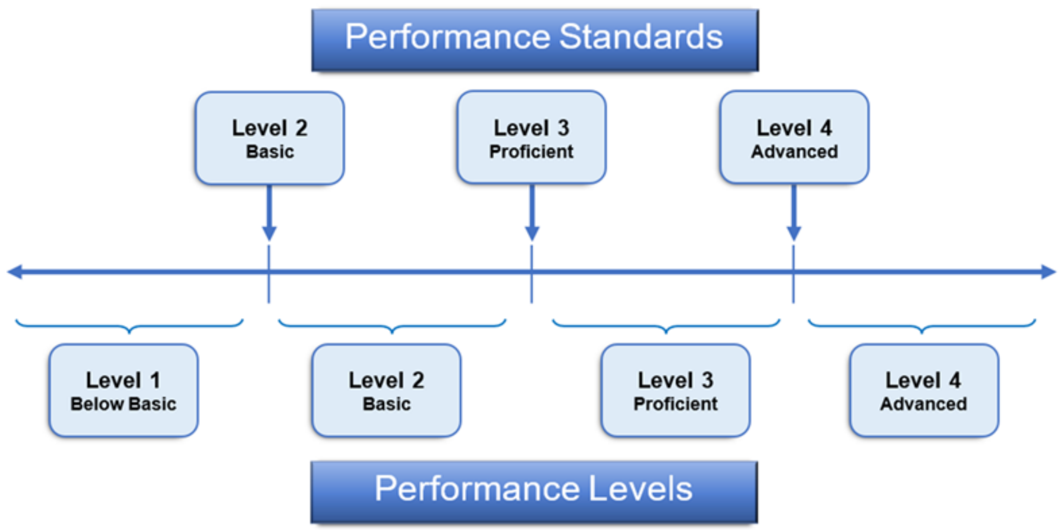
Introductions

4

Purpose of the Standard-Setting Workshop

5

- Recommend to the Idaho State Board of Education three performance standards to differentiate the four performance levels on the ISAT in Science



Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

Main Workshop Activities

6

- Large-Group Orientation
- Panel Training
  - Take the Online Operational Assessment
  - Review Range PLDs
  - Discuss Just Barely PLDs
  - Review the Ordered Scoring Assertion Booklet
  - Training on Assertion-Mapping Procedure
- Recommend Performance Standards
  - Two rounds
  - Panelist feedback following Round 1
  - Vertical Articulation
- Workshop Evaluation

Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

Importance of Security

7

- Please do not:
  - Create any form of electronic copy of test content (screenshots, electronic notes, etc.)
  - Create any hand-written notes of test content
  - Discuss test content with anyone outside the meeting
  - Use your computer during the course of the meeting for any purpose other than participating in the item review (e.g., email, web browsing, social media)
  - Save notes about item or passage content to your computer

Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

Reason for New Science Standards

8

- The Idaho State Science Standards (ISSS) were approved in 2018
- In Spring 2022, assessments aligned to the ISSS were administered to all students Idaho in grades 5, 8, and 11

Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

Description of the Science Test Design

9

- Grades 5, 8, and 11 tests assess students' understanding of the ISSS
- ISAT in Science at grades 5, 8, and 11 includes 6 item clusters and 12 or 13 stand-alone items
  - **Item clusters** include a stimulus and a series of questions that generally take students about 6–12 minutes to complete
  - **Stand-alone items** are shorter and generally take 1–3 minutes to complete
- All items ask students to use science and engineering practices and apply their understanding of disciplinary core ideas and crosscutting concepts to make sense out of real-world phenomena

Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

Review of 3D Science Standards

10

- Each 3D "standard" is a blend of one or two "big ideas" from a science discipline (DCI), one of several scientific activities that are common to the doing of all science (SEP), and one of a number of broad themes that are found across scientific disciplinary boundaries (CCC).

Review of Items – 3D Composition

11

For States, By States

MS-LS1-1 From Molecules to Organisms: Structures and Processes

Students who demonstrate understanding can:

MS-LS1-1. Conduct an investigation to provide evidence that living things are made of cells; either one cell or many different numbers and types of cells. [Clarification Statement: Emphasis is on developing evidence that living things are made of cells, distinguishing between living and non-living things, and understanding that living things may be made of one cell or many and varied cells.]

The performance expectation above was developed using the following elements from the NRC document *A Framework for K-12 Science Education*:

Science and Engineering Practices	Disciplinary Core Ideas	Crosscutting Concepts
Planning and Carrying Out Investigations Planning and carrying out investigations in 6-8 builds on K-5 experiences and progresses to include investigations that use multiple variables and provide evidence to support explanations or solutions. • Conduct an investigation to produce data to serve as the basis for evidence that meet the goals of an investigation.	**LS1.A: Structure and Function** • All living things are made up of cells, which is the smallest unit that can be said to be alive. An organism may consist of one single cell (unicellular) or many different numbers and types of cells (multicellular).	**Scale, Proportion, and Quantity** • Phenomena that can be observed at one scale may not be observable at another scale. - - - - - - - - - - - - - - ***Connections to Engineering, Technology and Applications of Science*** **Interdependence of Science, Engineering, and Technology** • Engineering advances have led to important discoveries in virtually every field of science, and scientific discoveries have led to the development of entire industries and engineered systems.

Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

Review of Items – 3D Composition

12

- Three-dimensional science standards

Scientific and Engineering Practices	Crosscutting Concepts	Disciplinary Core Ideas
▸ Asking questions or defining problems ▸ Developing and using models ▸ Planning and carrying out investigations ▸ Analyzing and interpreting data ▸ Using mathematics and computational thinking ▸ Constructing explanations and designing solutions ▸ Engaging in argument from evidence ▸ Obtaining, evaluating, and communicating information	▸ Patterns ▸ Cause and effect: mechanism and explanation ▸ Scale, proportion, and quantity ▸ Systems and system models ▸ Energy and matter: flows, cycles, and conservation ▸ Structure and function ▸ Stability and change	▸ Earth and Space Science ▸ Life Science ▸ Physical Science ▸ Engineering

Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

Item Clusters and Stand-Alone Items

13

- Item clusters
  - Designed to engage the student in grade-appropriate, meaningful scientific activity aligned to a specific standard
  - Item clusters include a stimulus and a series of questions that generally take students about 6–12 minutes to complete
- Stand-alone items are shorter and generally take students 1–3 minutes to complete

Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

Structure of Item Clusters

14

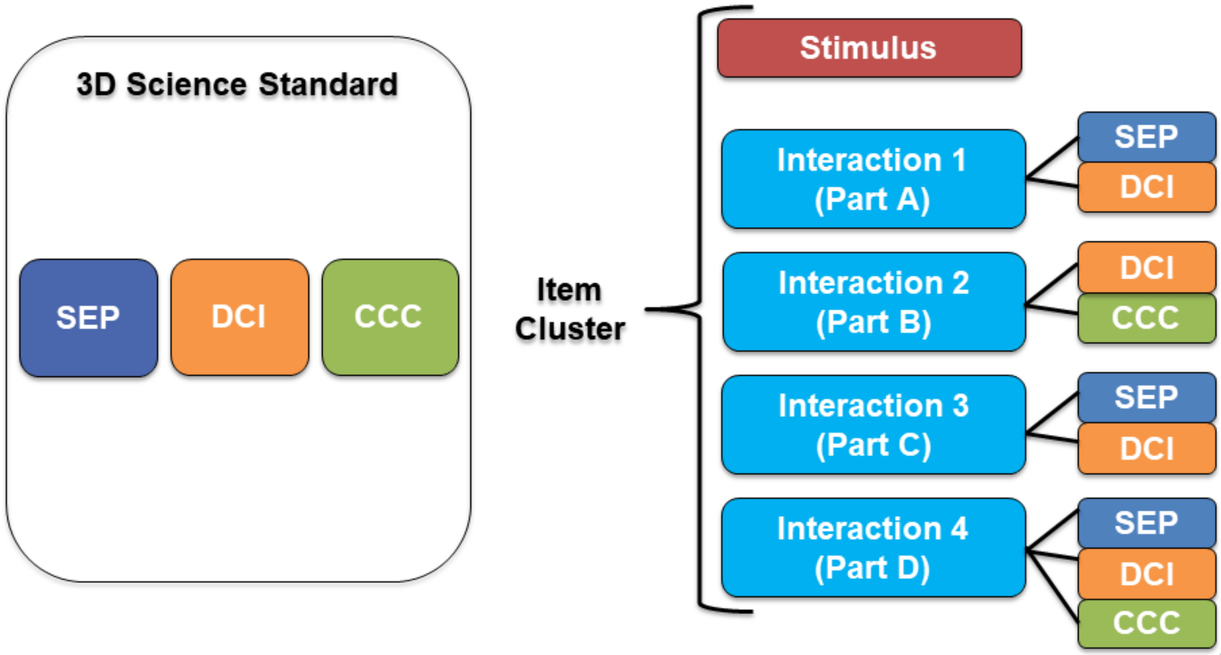
- Each item cluster begins with a ***phenomenon***, which is the observation about the natural world which anchors the entire item cluster. The interactions within the item cluster all address the phenomenon.
- Each item cluster engages the student in a grade-appropriate, meaningful ***scientific activity*** aligned to a specific standard.
- A ***cluster task statement*** comes at the end of the stimulus and an overview of the point of the item cluster.
- Each measurable moment is captured with a ***scoring assertion***. These assertions clearly articulate what evidence the student has provided as a means to infer a specific skill or concept.

Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

Review of Item Clusters – Composition

15



Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

Review of Items – Composition Example

16

Sparks fly off the wheels of a train when the brakes are applied.

Click the small gray arrow to see a demonstration of this happening in Animation 1.

Animation 1. Braking Train

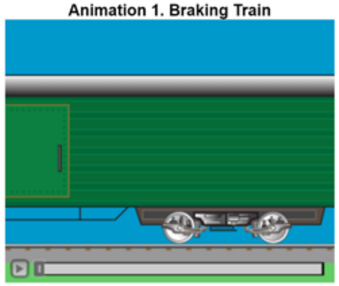


Table 1 explains some properties of the train and its surroundings as energy flows throughout the system.

Table 1. Properties of the Train System

Before Brakes Are Applied	After Brakes Applied
No sparks	Sparks fly off the wheels and brake pads
Brake pads make no sound	Brake pads make sound

1680

Part A

Click on each blank box to select the word or phrase that completes each sentence, constructing an argument about what happens when the train's brakes are applied.

Applying the brakes causes the [▾] to transfer kinetic energy to the [▾]. This causes the [▾] to slow down and have [▾] kinetic energy, which slows the train.

Part B

When the train applies its brakes, what happens to the energy of the surroundings?

- (A) The surroundings gain energy.
- (B) The surroundings lose energy.
- (C) The surroundings do not gain or lose energy.
- (D) There is not enough information to determine the energy of the surroundings.

Part C

Which **three** statements support your choice in part B?

- ☐ The train maintains its speed.
- ☐ Sound is produced.
- ☐ Sound is consumed.
- ☐ Light is produced.
- ☐ Light is consumed.
- ☐ Heat is produced.
- ☐ Heat is consumed.

Scoring Assertions

17

- Within each item cluster, a series of explicit assertions can be made about the knowledge and skills that a student has demonstrated based on specific features of the student's responses
- Scoring assertions can be supported based on students' responses in one or more interactions within an item cluster.
- For example:
  - A student correctly graphs data points indicating that (s)he can construct a graph showing the relationship between two variables,

Review of Items – Scoring Assertions

18

Score Rationale	
The student selected "wheels" for the first blank and "brakes" or "rails" for the second blank showing an understanding of the interactions in the system and the effects of that energy flow.	✖
The student selected "wheels" for the third blank and "less" for the fourth blank showing an understanding of the interactions in the system and the effects of that energy flow.	✖
The student selected "The surroundings gain energy," showing an understanding of how the energy of the wheels change and is distributed throughout the system.	✖
The student selected "Sound is produced," providing evidence of how the energy of the surroundings has changed.	✖
The student selected "Light is produced," providing evidence of how the energy of the surroundings has changed.	✖
The student selected "Heat is produced," providing evidence of how the energy of the surroundings has changed.	✖
The student selected "The brakes make a screeching sound," which shows an understanding of how the energy changed throughout the system and that those changes serve as evidence that the the Kinetic Energy of the wheels transfers out of the wheels/system when the brakes are applied.	✖
The student selected "The sparks that fly off the wheels give off light," which shows an understanding of how the energy changed throughout the system and that those changes serve as evidence that the the Kinetic Energy of the wheels transfers out of the wheels/system when the brakes are applied.	✖
The student selected "The brakes give off energy as heat," which shows an understanding of how the energy changed throughout the system and that those changes serve as evidence that the the Kinetic Energy of the wheels transfers out of the wheels/system when the brakes are applied.	✖

Cambium Assessment

IDAHO
STATE DEPARTMENT OF EDUCATION

Experience the Online Assessment

19

- Time to "Take the Test"
- Items administered in spring 2021
- Interface is similar to the online test environment that the students experienced
- This is an opportunity to interact with the items
- No need to "complete" the test, you will have more time later to become very familiar with the items
- You can score your responses
- You have ~90 minutes (stop at 11:45 am)

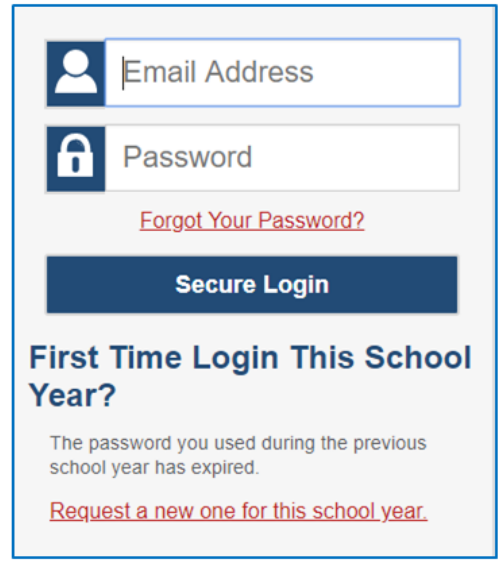
Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

Accessing the Online Assessment

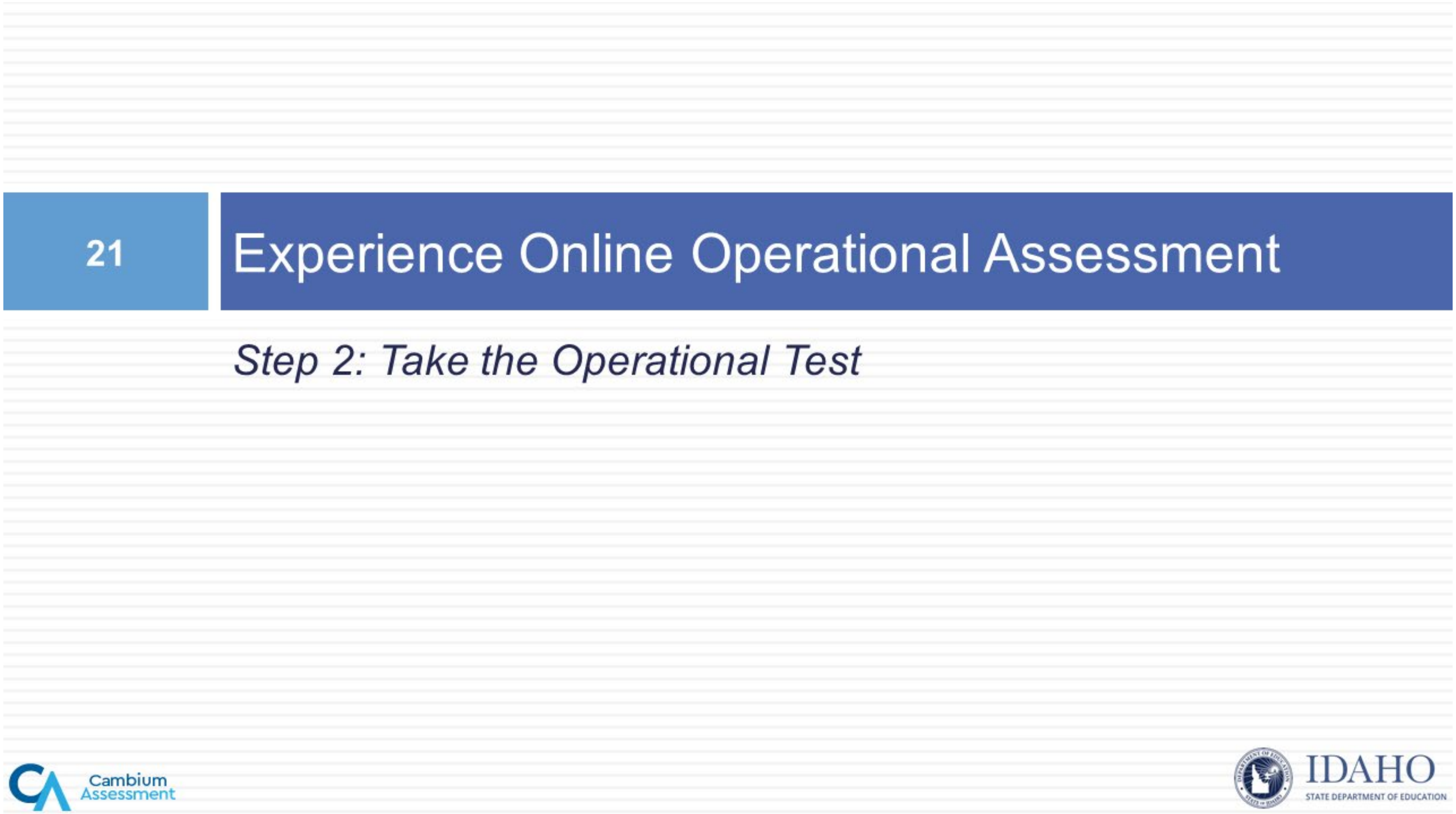
20

- Open the Chrome browser
- Sign in with your Username and Password



Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION


21
Experience Online Operational Assessment
Step 2: Take the Operational Test
Cambium Assessment
IDAHO
STATE DEPARTMENT OF EDUCATION

Standard Setting

22

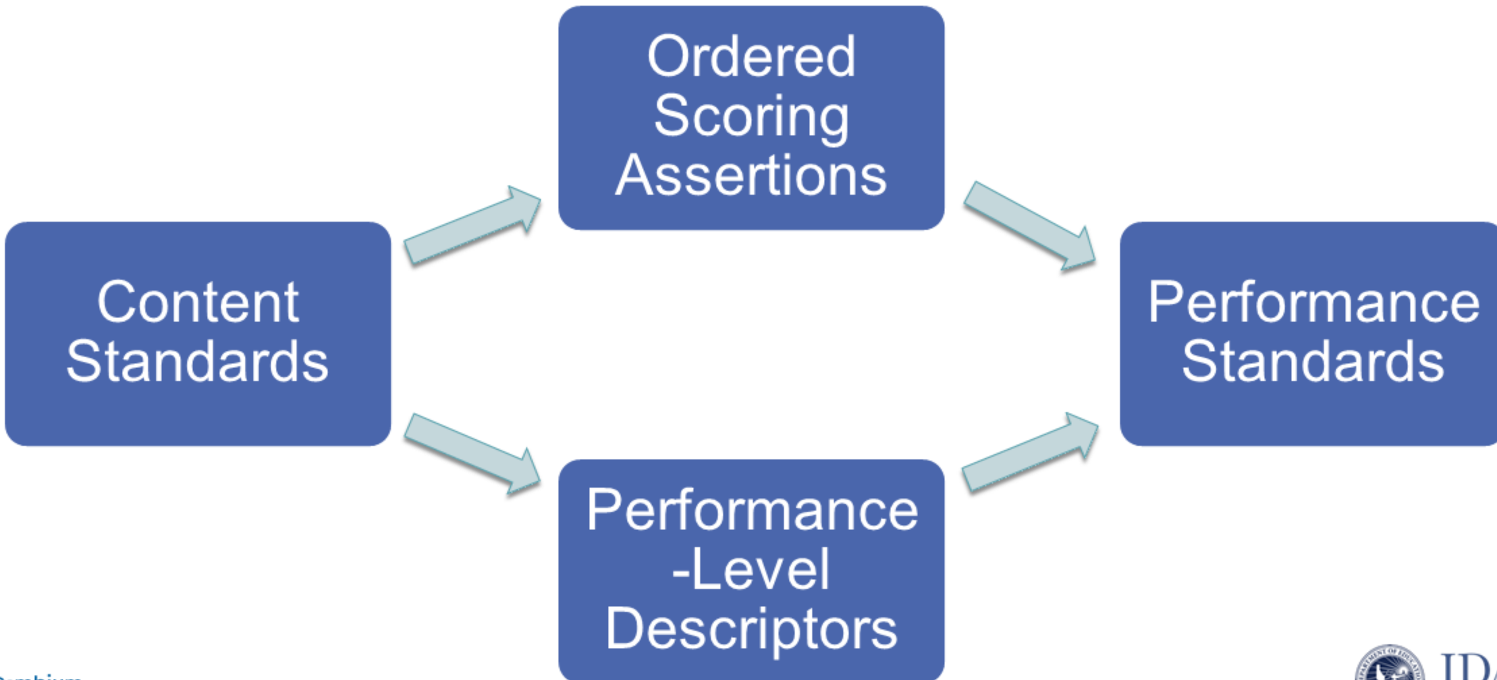
- Systematic process by which trained participants use their knowledge of academic content standards, test items, and student performance to recommend cut-scores associated with each performance level on the test

Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

From Content Standards to Performance Standards

23

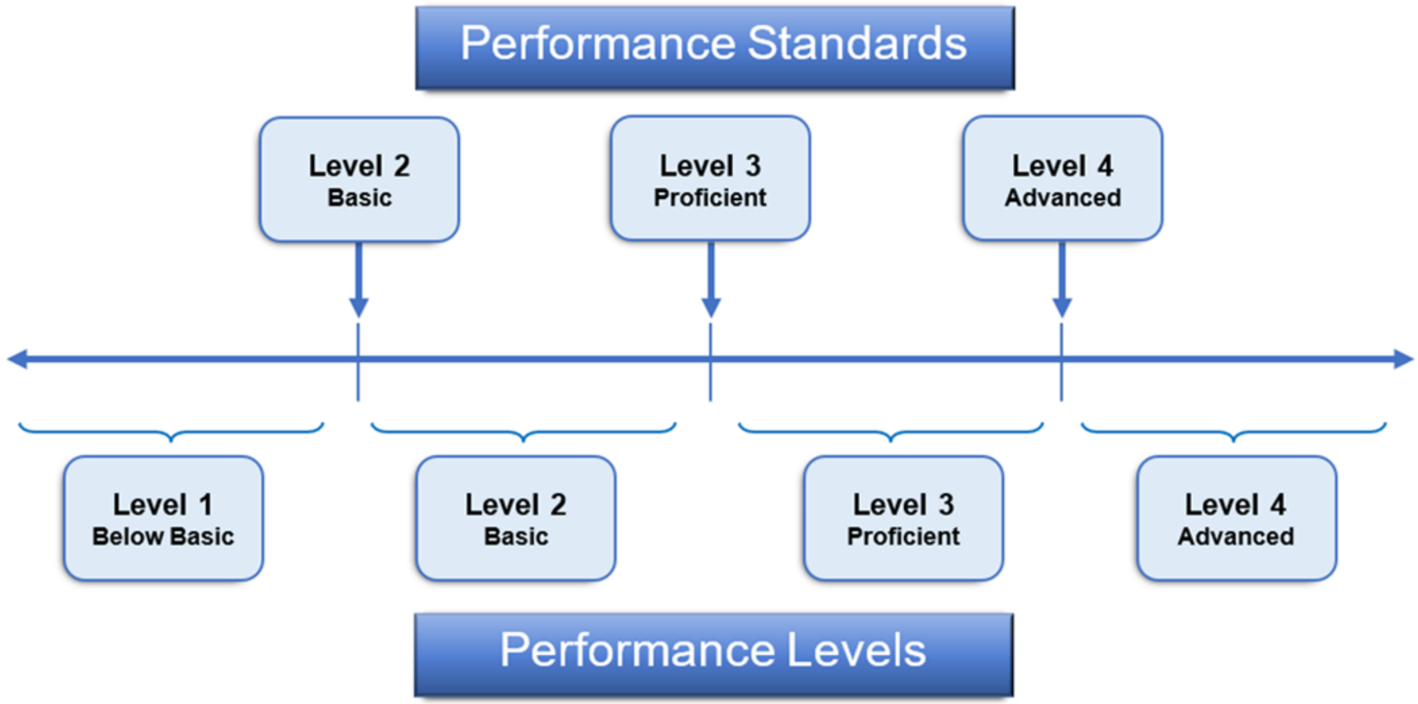


Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

Performance Standards and Performance Levels

24



CamI Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

Assertion-Mapping Procedure (AMP)

25

- Test-centered procedure
- Employs an ordered item procedure adapted to accommodate new multiple interaction item types
- Map ordered scoring assertions to performance levels
- Is being employed to recommend performance standards in multiple states assessing three-dimensional science standards

Key Elements of the AMP

26

- Performance-level descriptors (PLDs)
  - Range PLDs
  - Threshold PLDs (just barely meets)
- Ordered scoring assertions
- Assertion map
- Assertion mapping in multiple rounds
  - Contextual information – benchmarking data and student impact data
  - Panelist feedback and group discussion
- Vertical articulation and moderation

Performance-Level Descriptors (PLDs)

27

- Describe what students within each performance level are expected to know and be able to do
- PLDs are the link between the content and performance standards

Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

Grade 8 Range PLDs – Level 3 Proficient Standard

28

Physical Sciences

- **MS-PS1:**
- **MS-PS2:**
- **MS-PS3:**
- **MS-PS4:**

Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

Grade 8 Range PLDs Across Performance Levels

29

MS-PS3: Energy

- **Level 1:**
- **Level 2:**
- **Level 3:**
- **Level 4:**

Parse and Review the PLDs

30

- Take a few minutes to review the PLDs taking notice of the verbs and skills that differentiate the performance levels
  - Think about how the skills change from Below Basic to Advanced
  - Think about the skills and knowledge these students can demonstrate
  - Idea is to get a common mental representation of these students

  REMEMBER: Not every piece of content will be represented in the PLDs
- PLD Discussion

“Just Barely” Meets the Performance Standard

31

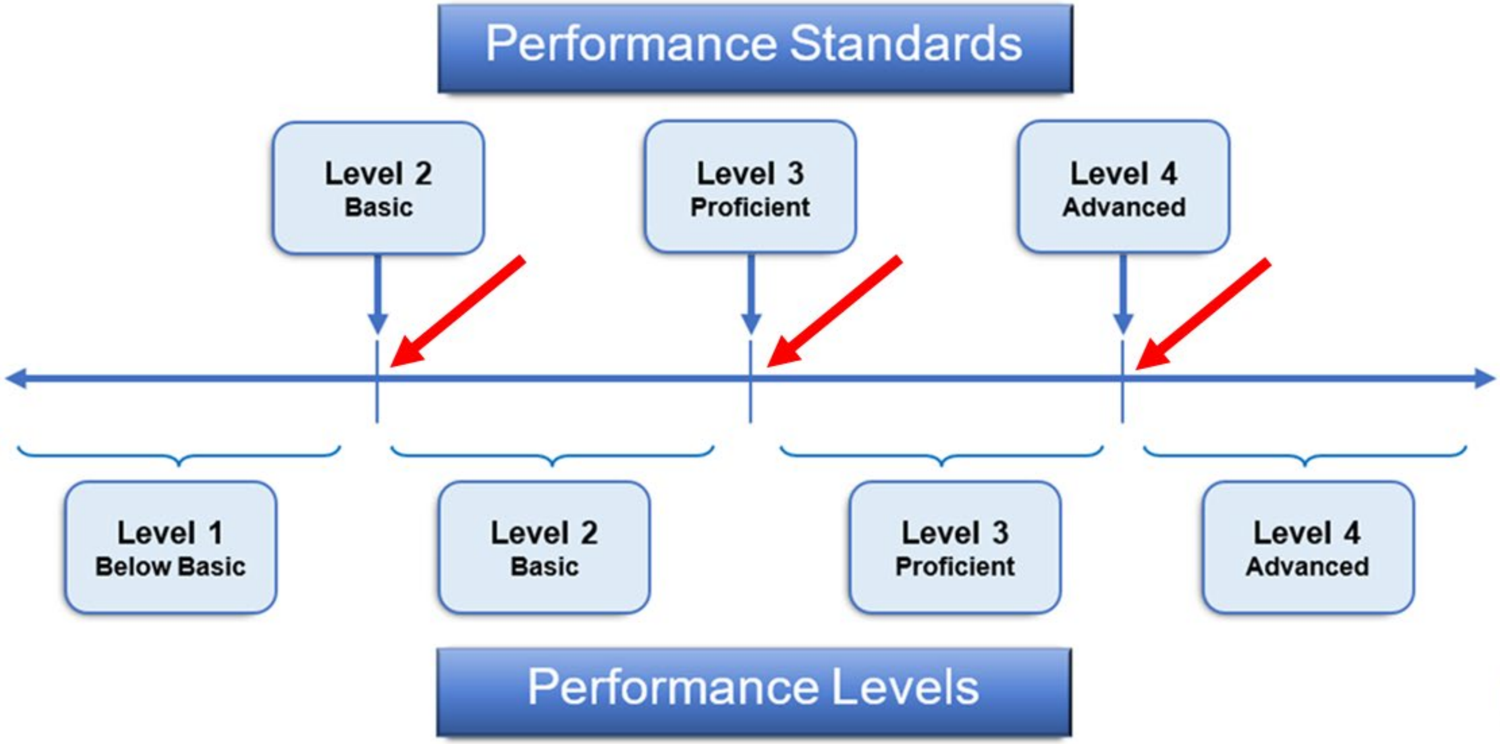
- When considering each performance level, we are especially interested in the transition areas between performance levels
- Pay attention to characteristics of students who ***just barely*** qualify for entry into the performance level from those just below
  - Not a typical example of students in the performance level
  - Although they are not good examples of the performance level, they do still meet the standard, or description in the PLD

Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

Threshold "Just Barely" PLDs

32



Purpose of Just Barely Discussion

33

- Identify the types of skills these students can demonstrate
- Come to a common understanding of these skills and big ideas

Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

Just Barely Discussion

34

- 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 discuss the skills that a just barely student needs to have to gain entry into each of the four levels
- For each performance level think about:
  - What skills and knowledge must the student demonstrate to qualify for entrance into this performance level?
  - How does this differ from the upper range of the adjacent performance level?

Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

35

Review of Ordered Scoring Assertion Booklet

Step 4: Review of Ordered Scoring Assertion Booklet

Important Concepts

36

- "Just barely" meets the performance level
  - Differentiate students who just barely qualify for entry into a performance level from those just below
- Assertion mapping
  - Map each scoring assertion to the performance level that the assertion best supports
- Ordering of assertions
  - Assertions are ordered by difficulty within an item
  - Mapping of assertions to performance levels should reflect the ordering – no inversions within an item

Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

Ordered Scoring Assertions

37

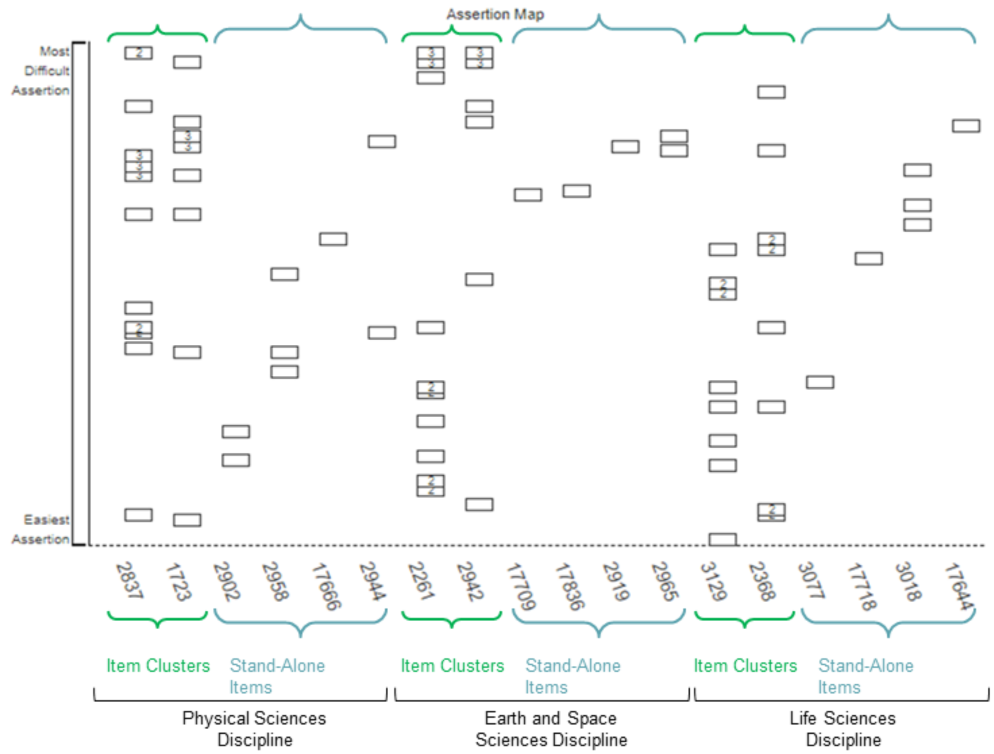
- The ordered scoring assertion booklet (OSAB) constitutes a test administration:
  - A test form that meets test blueprint specifications
- It is important to evaluate scoring assertions as they relate to the item interactions
- Assertions within items are ordered by difficulty
  - Assertions within an item may not represent all PLDs

Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

Assertion Map

38



Ordered Scoring Assertion Booklet

39

- See the **Difficulty Level Visualizer** – graphic representation of the difficulty of each assertion relative to the student population

- Example of how to use this:
  - After reviewing the item and scoring assertion you believe this is a relatively difficult concept. However, you see it is on the far left of the scale, ask yourself:
    - What made this so easy for the student?
    - Is the student really "analyzing" or perhaps it is a concept that is very familiar to students, and it is more of a rote concept?

Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

Studying the Items and Scoring Assertions

40

- For each scoring assertion ask yourself:
  1. *How do the item interactions support the scoring assertion?*
  2. *Why is this assertion more difficult than the previous assertions (within the item)?*
  3. *How does the scoring assertion and the underlying interactions relate to the PLDs?*
- Working as a group
  - Discuss how item interactions support scoring assertions
  - Discuss ordering of scoring assertions
  - Discuss how scoring assertions are related to the PLDs

Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

What If an Assertion Seems Out of Order?

41

- Assertion ordering is based on student performance
- Assertions may seem out of order because they are ordered by difficulty, and not by content or cognitive process
- Identify why a scoring assertion is more difficult than the assertions before it, and easier than the assertions following it
  - Pay special attention to the interactions supporting the assertions
  - Assertions may be more or less difficult because of the underlying interactions

Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

What If an Item Seems Wrong or Unfair?

42

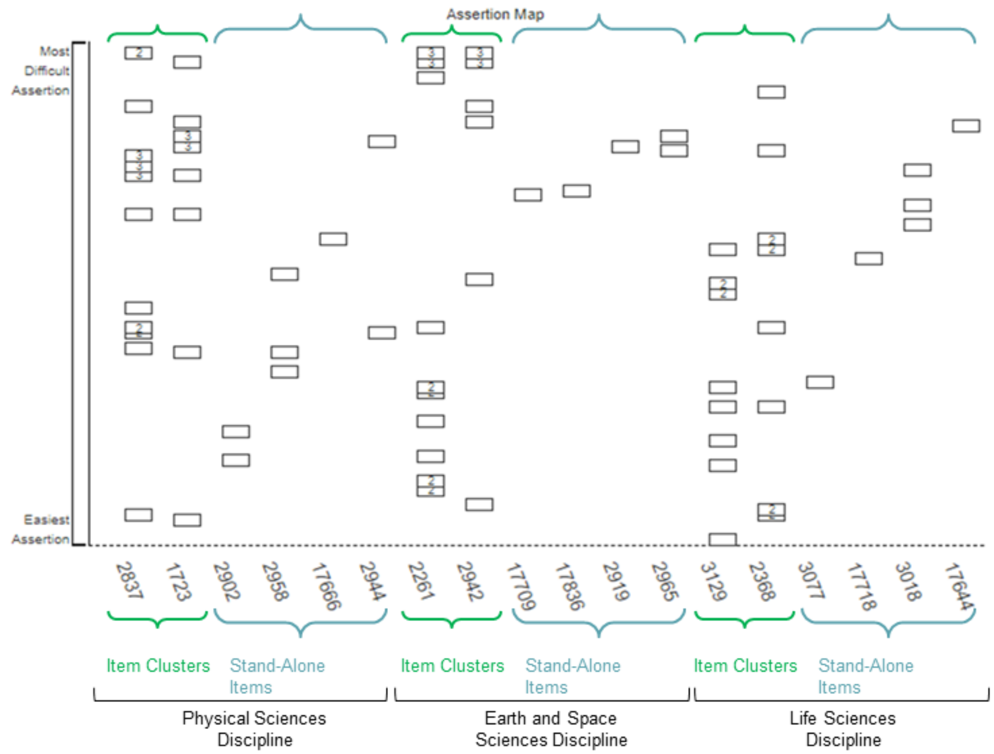
- Do not let yourself get distracted – this is not an item review meeting
- If you believe something is wrong with an item interaction or scoring assertion, tell the Workshop Leader, then skip over the assertion as you review the rest of the assertions within the item

Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

Assertion Map

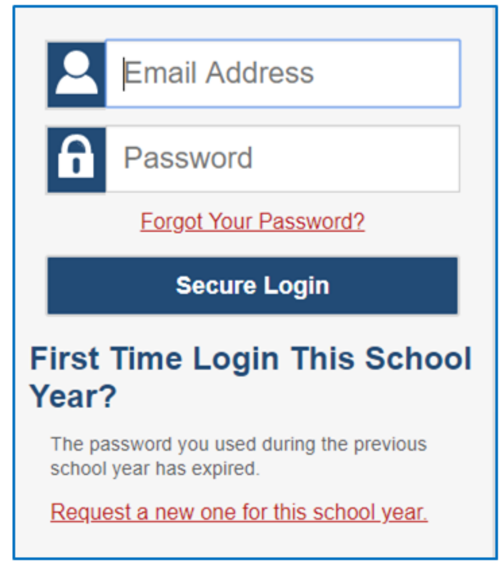
43



Accessing the OSAB

44

- Open the Chrome browser
- Sign in with your Username and Password



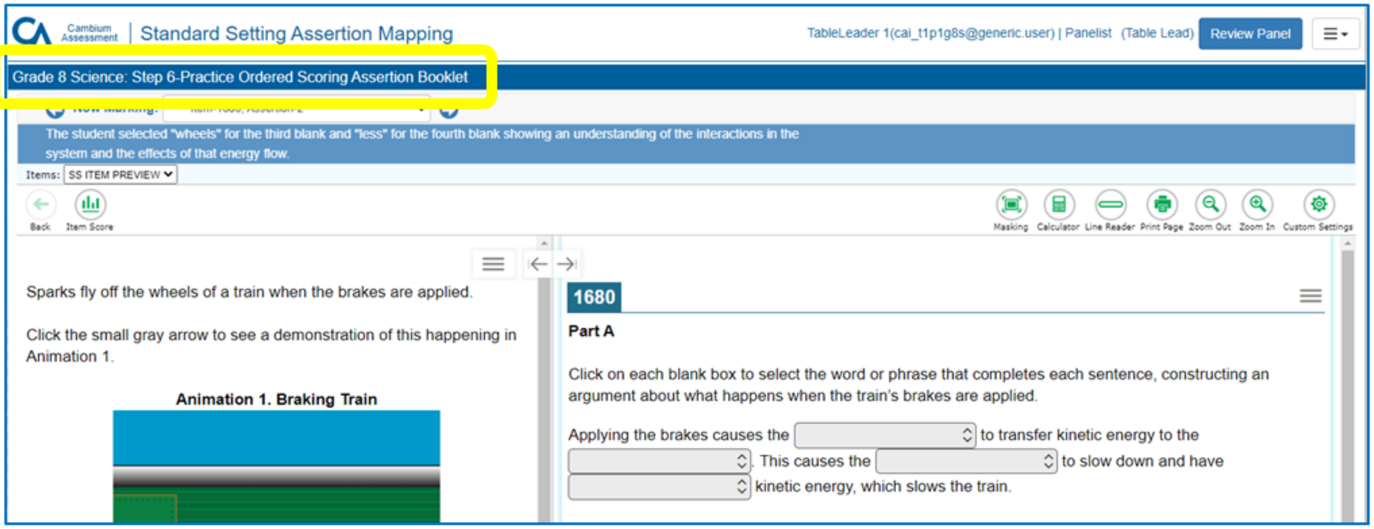
Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

Navigating the OSAB

45

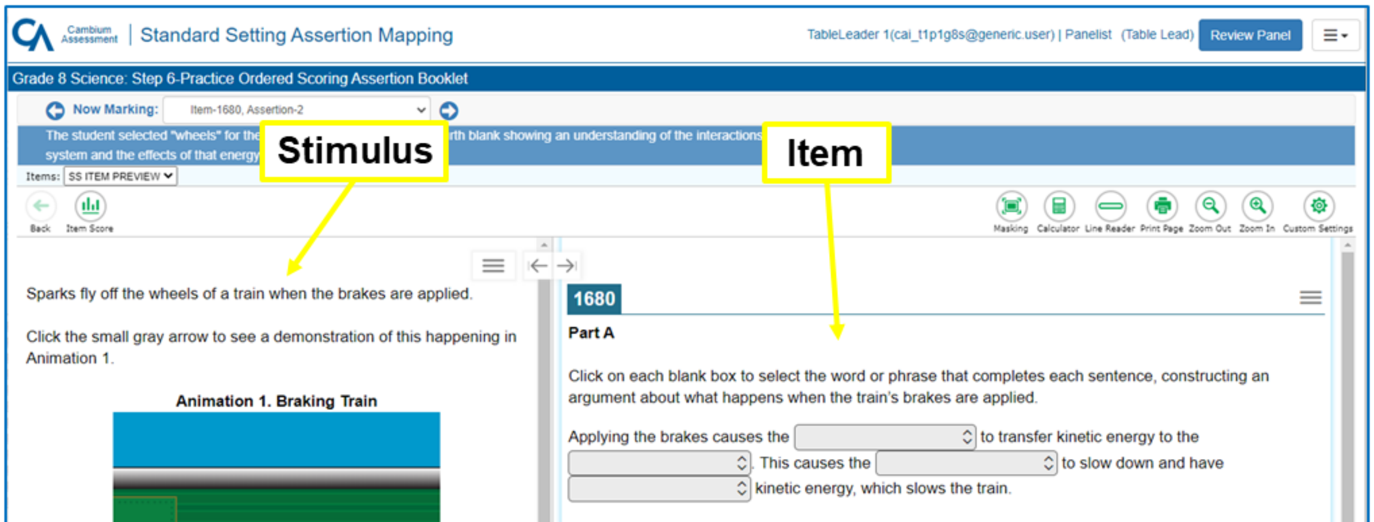
- Test and step we are working on shown at the top of the screen



Navigating the OSAB

46

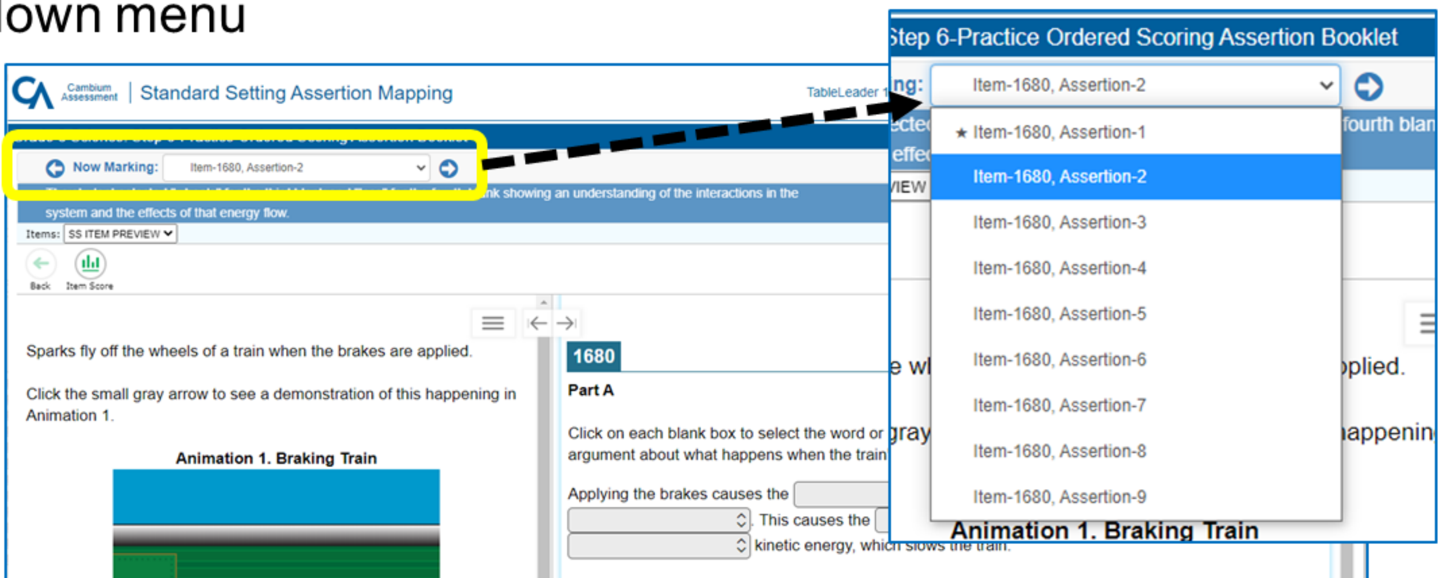
- View the stimulus on the left side of the screen and the item on the right



Navigating the OSAB

47

- Move forward in the OSAB or select an assertion from the drop-down menu



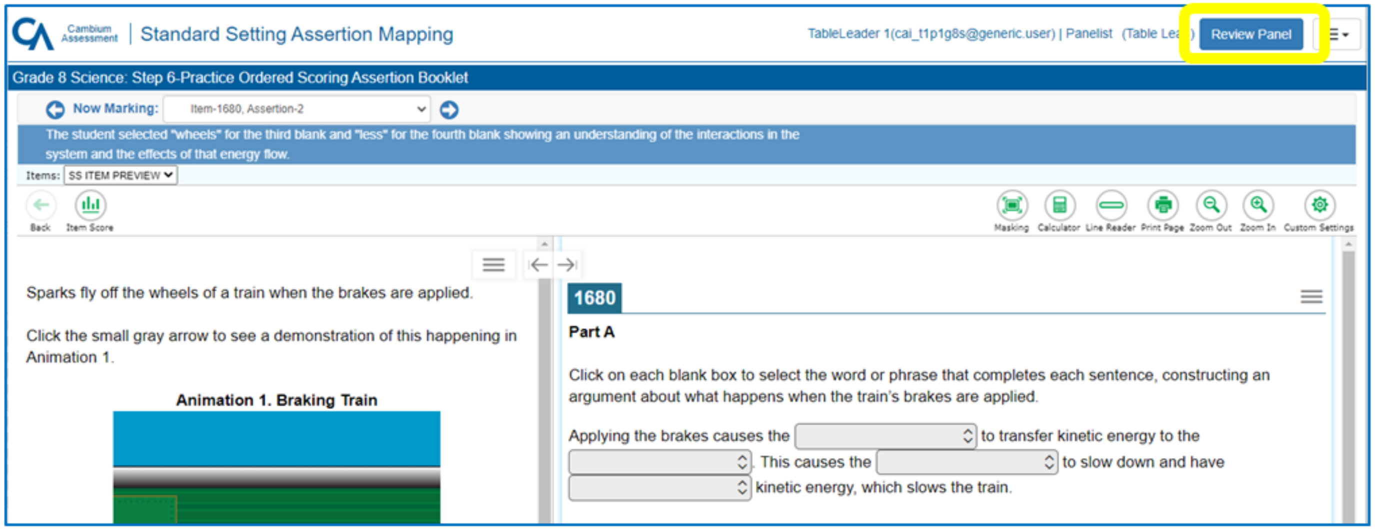
Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

Navigating the OSAB

48

- Access the Review Panel

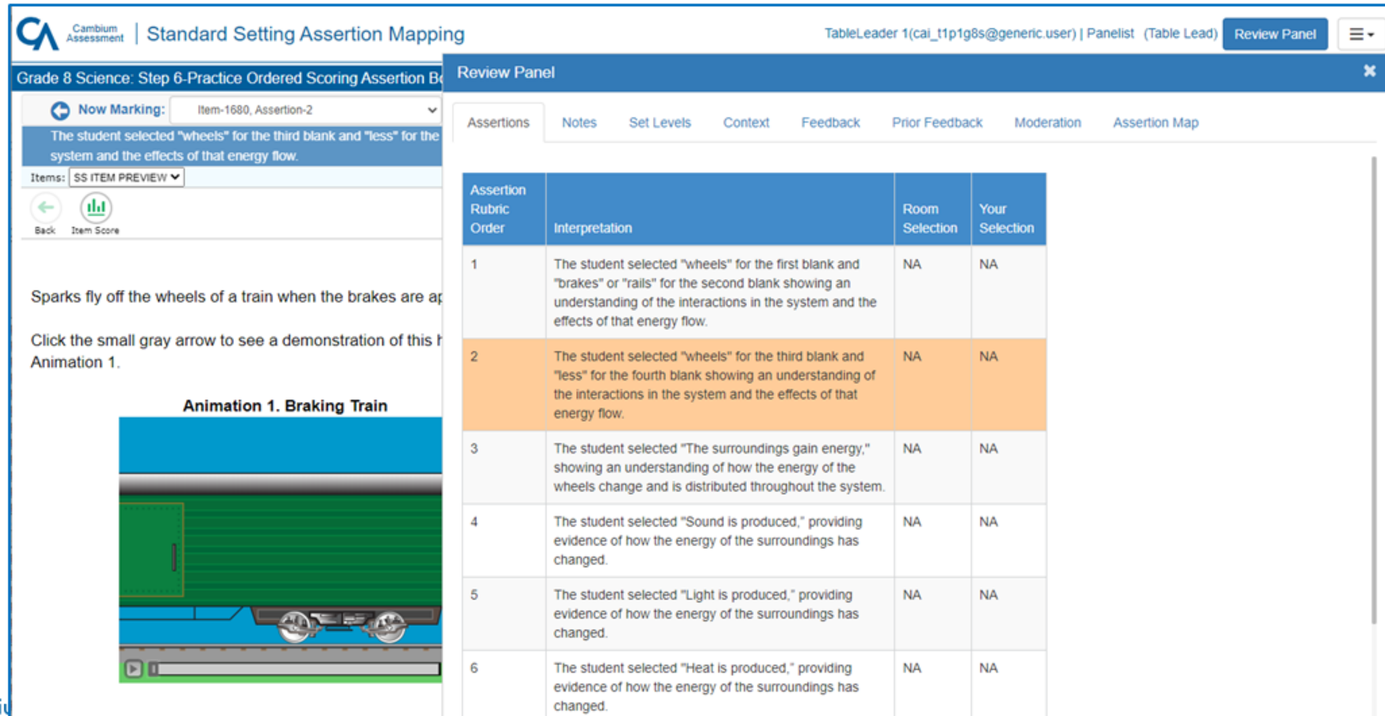


Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

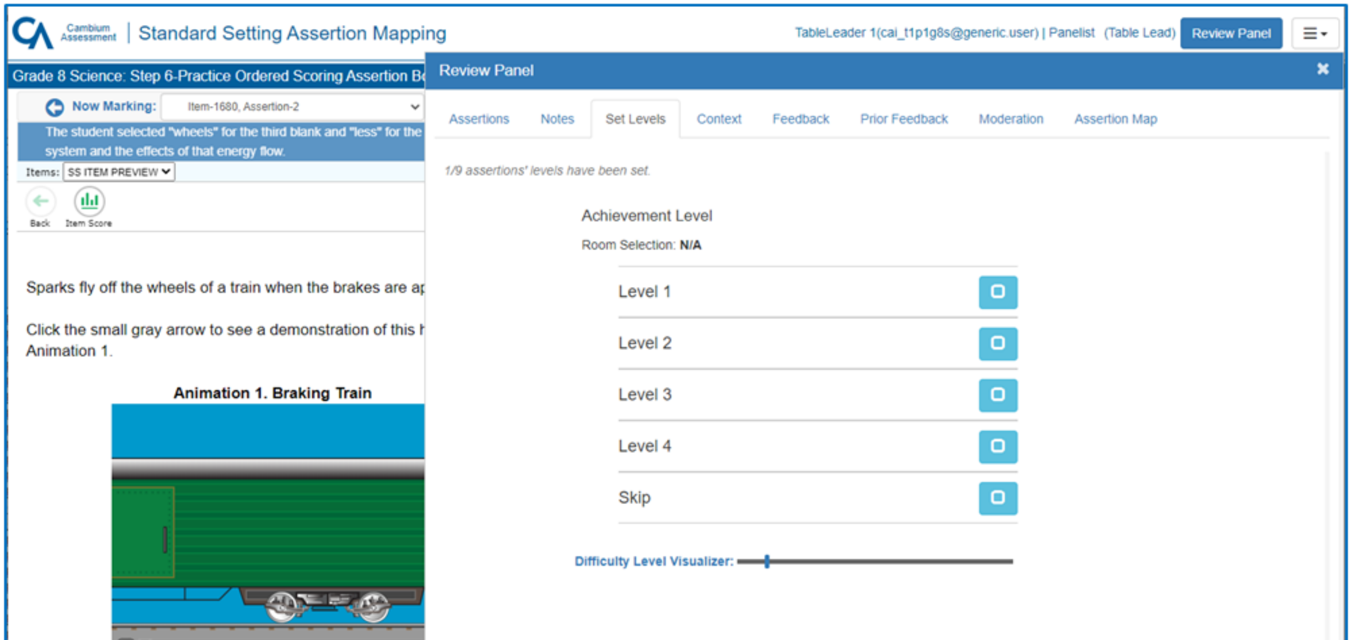
Navigating the OSAB – Review Panel

49



Navigating the OSAB – Review Panel

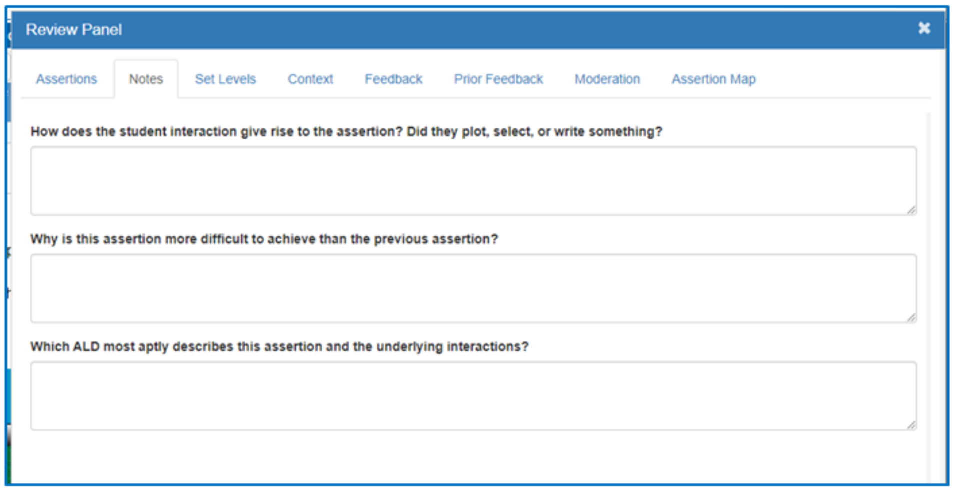
50



Navigating the OSAB – Review Panel

51

- “Context” tab – presents benchmarking and student impact data
- “Notes” tab – this is for your reference



Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

Studying the Items and Scoring Assertions

52

- We will work together on a set of items, asking and answering the following for each scoring assertion:
  1. *How do the item interactions support the scoring assertion?*
  2. *Why is this assertion more difficult than the previous assertions?*
  3. *How does the scoring assertion and the underlying interactions relate to the PLDs?*
- Then review the stand-alone items.

Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

Review of the OSAB

53

- Let's review the items together

54

Standard Setting Day 2

Recommending Performance Standards for Grade 8 Science

Standard Setting Day 2 Agenda

55

- Continued review of OSAB
- Training on Assertion-Mapping Task
- Round 1 Assertion Mapping
- Review Feedback Data and Discuss Round 1 Results
- Round 2 Assertion Mapping
- Across Grade Moderation

Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

Continue review of OSAB

56

Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

57

Training on Assertion-Mapping Task

Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

Important Concepts

58

- "Just barely" meets the performance level
  - Differentiate students who just barely qualify for entry into a performance level from those just below
- Assertion mapping
  - Map each scoring assertion to the performance level that the assertion best supports
- Ordering of assertions
  - For assertion mapping, assertions are ordered by difficulty within an item
  - **Assertions within an item may not represent all PLDs**
  - Mapping of assertions to performance levels should reflect the ordering – **no inversions within an item***
  - Pay attention to the Difficulty Level Visualizer and Assertion Map across items

Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

Assertion-Mapping Task

59

- Map assertions to performance levels
  - Map each assertion to the performance level that the assertion best supports
  - Consider “neighborhoods”—where does Level 1 clearly become Level 2, Level 2 become Level 3, etc.
  - Consider what differentiates students who just barely qualify for entry into the performance level from those not quite ready for entry into the performance level
  - Evidence that the student has demonstrated knowledge and skills necessary for entry into the performance level

Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

Benchmark and Impact Data

60

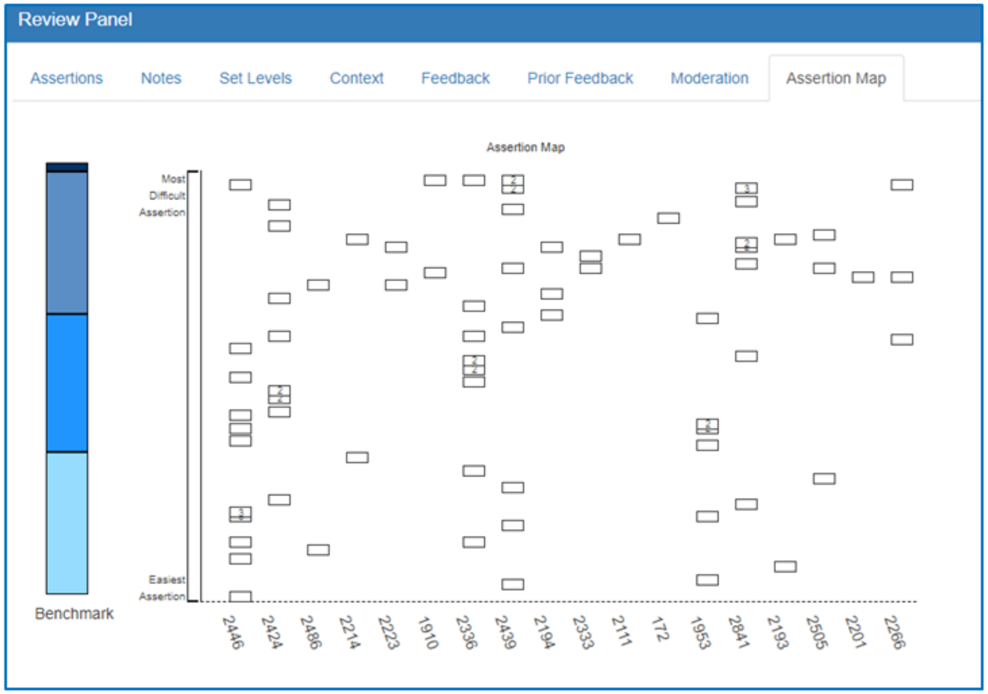
- Contextual information for each assertion
  - Benchmarking data: 2015 Idaho NAEP Science results
    - Shows the associated performance level for the NAEP Science Assessment if the Science standards were to exhibit the same rigor
  - Student impact data: percentage of students performing at or above the level associated with the assertion

Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

Assertion Map with Benchmark Information

61



Mapping Ordered Assertions to Performance Levels

62

- You will map each scoring assertion to a performance level using the following tools:
  - PLDs
  - Difficulty Level Visualizer
  - Assertion Map
  - Contextual Information – benchmarking data and student impact data
  - Your professional judgement (and notes)
- Remember, scoring assertions are ordered from easiest to most difficult within each item
- **If you think that a subsequent assertion is at a lower level than a previous assertion, you might have been premature at mapping the level for the earlier assertion**
- You may “Skip” if an assertion seems to be out of place
  - Only use as a last resort

Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

Assertion-Mapping Task

63

- Map assertions in the online standard-setting tool
  - Mapping of assertions to performance levels should reflect the ordering of assertions by difficulty within an item
    - No inversions within an item

64

Practice Assertion-Mapping Task and Standard-Setting Quiz

Step 6: Practice Ordered Scoring Assertion Booklet

Practice Online Assertion-Mapping Task

65

- Purpose of this activity is to practice mapping assertions in the online environment. This is meant to help you become familiar with the tool and process.
  - Shortened version of the OSAB
  - One item cluster
- Log into the system and review the item cluster and ordered scoring assertions answering the three questions as you go
- Then, map each scoring assertion to a performance level and click "confirm"
- This is meant to help you become familiar with the tool and process

Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

Practice quiz!

66

- Reminders for some key concepts

Round 1 Readiness Form

67

- Any questions?
- Is everyone ready for Round 1?
- If so, please fill out the readiness form

Round 1 Assertion Mapping

68

- You will map each assertion to a performance level
- Use the tools and documents along with your professional judgment
- Scoring assertions are ordered from easiest to most difficult within each item
- If you feel that a subsequent assertion is at a lower level than a previous assertion, then you might have been premature at mapping the level for the earlier assertion
- Should be a logical progress of performance levels (within an item)
  - No inversions
- You may “Skip” if, after consideration, the assertion seems to be out of place
  - Use as last resort
- When you have assigned all assertions click on the “Confirm” button
- This is an individual task
- Lunch is at 12:30 pm

69

Round 1 Assertion Mapping

Step 8: Round 1 Assertion Mapping

Group Feedback and Discussion

70

- Goals
  - Add important information to your thinking
  - Develop common understandings
  - Inform possible re-evaluation of assertion mappings
- Expectation is converging judgments
  - Consensus is not a requirement or goal

Feedback and Impact Data

71

- Percentage of students reaching or exceeding the standard based on assertion mappings
- Group discussion
  - Does the percentage of students reaching or exceeding the current recommended performance standard seem reasonable?
  - What are the implications for the performance standards?
  - All performance standard recommendations should be based on content rationales drawn from the Idaho State Science Standards

Variance Monitor

72

- Consensus is NOT required, convergence is a goal
- Let's see where we have the most variance
- Discuss within each table for 15 minutes
- Then, we will come together for group conversation for 15 minutes

Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

Variance Monitor

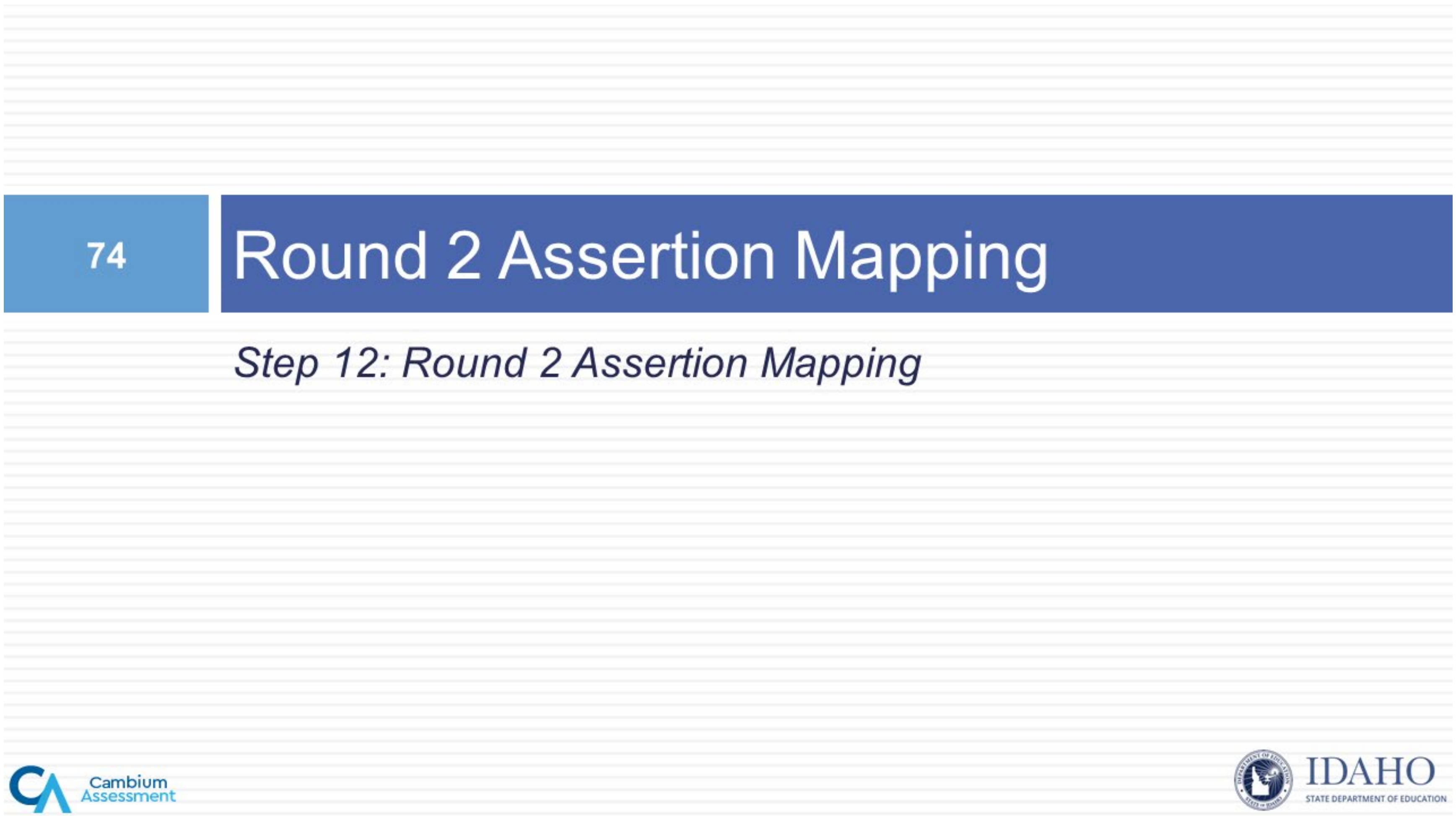
73

Lowest Variance ☐ ☐ ☐ ☐ ☐ Highest Variance

Item Assertion Id	Difficulty Level	Mean		Room - 1 | Table - 1					Room - 1 | Table - 2				
				VI_T1P3 _G5S	VI_T1P4 _G5S	VI_T1P5 _G5S	VI_T1P7 _G5S		VI_T2P2 _G5S	VI_T2P3 _G5S	VI_T2P4 _G5S	VI_T2P6 _G5S	VI_T2P7 _G5S
18027-Assertion-7	505	1.89		2	2	3	1		3	3	1	1	1
18027-Assertion-8	517	2.11		2	2	3	1		3	3	2	2	1
18027-Assertion-1	525	2.33		2	2	3	2		3	3	2	2	2
18027-Assertion-5	526	2.38		--	2	3	2		3	3	2	2	2
18027-Assertion-3	535	2.67		3	2	3	2		3	3	2	3	3
18027-Assertion-9	545	3.11		3	3	3	3		3	3	3	4	3
18027-Assertion-2	550	3.11		3	3	3	3		3	3	3	4	3
18027-Assertion-6	552	3.22		3	4	3	3		3	3	3	4	3
2931-Assertion-1	518	1.89		3	1	1	2		2	3	1	2	2

Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION


74
Round 2 Assertion Mapping
Step 12: Round 2 Assertion Mapping
Cambium Assessment
IDAHO
STATE DEPARTMENT OF EDUCATION

Round 2 Readiness Form

75

- Any questions?
- Is everyone ready for Round 2?
- If so, please fill out the readiness form

Round 2 Assertion Mapping

76

- You will use the next hour to map each scoring assertion to a performance level
- Use the tools and documents along with your professional judgment, **contextual information – benchmark ad impact data**, and **feedback data**
- Scoring assertions are ordered from easiest to most difficult within each item
- If you feel that a subsequent assertion is at a lower level than a previous assertion, then you might have been premature at setting the level for the earlier assertion
  - No inversions
- You may “Skip” if, after consideration, the assertion seems to be out of place
  - Use as a last resort
- When you have assigned all assertions click on the “Confirm” button
- This is an individual task
- You have until 3:00 pm
- **Complete evaluations**

Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

77

Round 2 Results

Step 14: Results of Round 2

78

Moderation

Step 16: Moderation

Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

Creating a System of Performance Standards

79

- Performance standards for a statewide system must be coherent across grades and subjects
  - Articulation
  - Benchmarking
  - Moderation

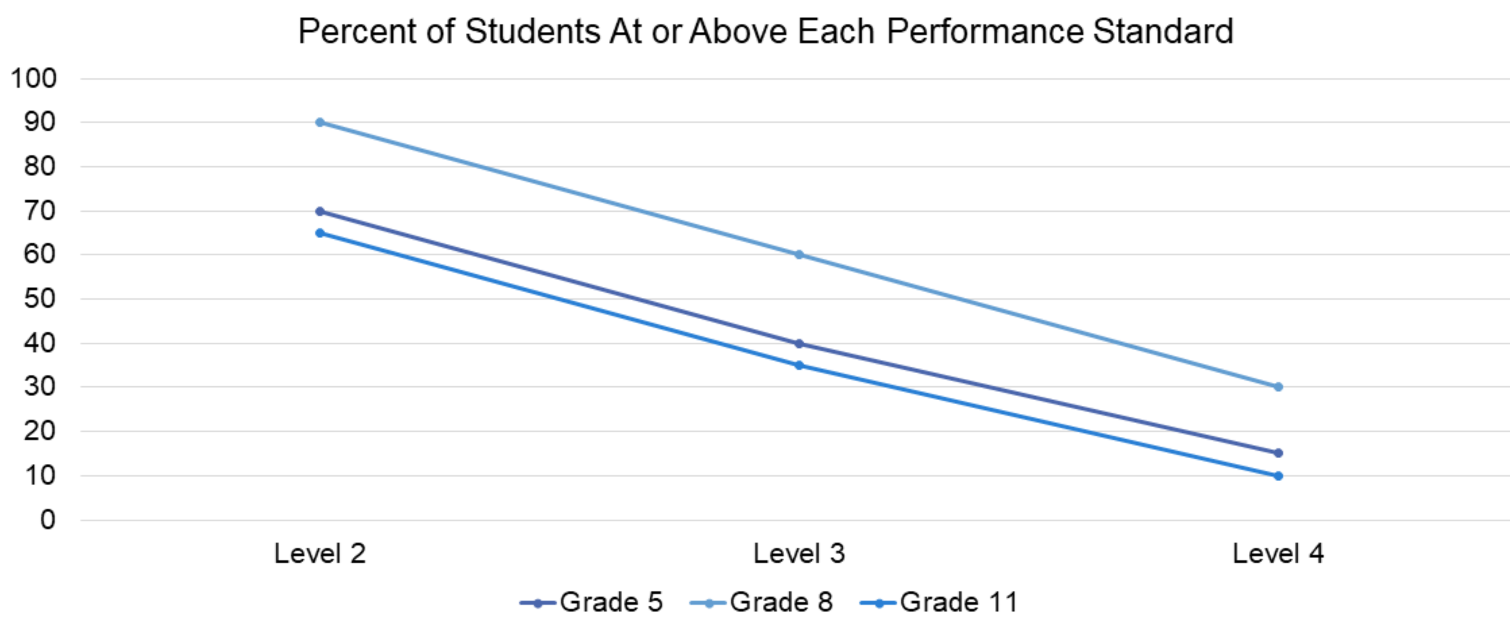
Benchmarking

80

- Are performance standards nationally competitive and represent on track for college readiness?
- Performance levels for benchmark assessments will provide context about the general neighborhood in which performance standards likely reside

Articulation

81



Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

Moderation

82

- After the standards have been recommended by the panelists, the Table Leaders meet to review the outcomes
  - All members are invited to observe this meeting but only the Table Leaders participate
- If there are anomalies across grades or subjects the Table Leaders are permitted to adjust the performance standards (assuming there is a good content reason for doing so)

Cambium Assessment

IDAHO
STATE DEPARTMENT OF EDUCATION

Break Into Groups

83

Panel	Facilitators
Grade 5 Science	Kam Mangis de Mark Hibbah Haddam Kevin Dwyer
Grade 8 Science	Heather MacRae Jim McCann
Grade 11 Science	Matt Davis Mark Palamo

Cambium Assessment

IDAHO STATE DEPARTMENT OF EDUCATION

Appendix 3-G

Standard-Setting Practice Quiz

Standard-Setting Practice Quiz

Exhibit G-1. Standard-Setting Practice Quiz

2022 ISAT Science Assessment Standard Setting - Assertion Mapping Practice Quiz

* Required

1. Name: *

2. Panelist ID (e.g., ID_T1P1_G5S): *

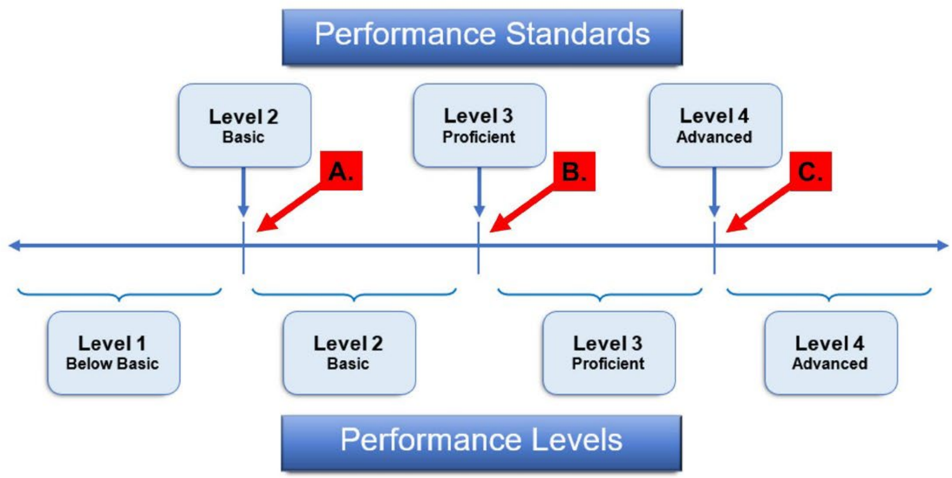
3. Assigned Committee: *

   Mark only one oval.

   ◯ Science Grade 5
   ◯ Science Grade 8
   ◯ Science Grade 11

Performance Standards and Performance Levels

The graphic below illustrates the relationship between the performance standards that you will recommend and the performance levels that they demarcate:



4. Which red box on the performance continuum graphic above illustrates students who are just barely described by the Level 3 Proficient PLD? *

Mark only one oval.

- Box A
- Box B
- Box C

5. Which red box on the performance continuum graphic above illustrates students who are just barely described by the Level 2 Basic PLD? *

Mark only one oval.

- Box A
- Box B
- Box C

6. Which red box on the performance continuum graphic above illustrates students who are just barely described by the Level 4 Advanced PLD? *

   Mark only one oval.

   - Box A
   - Box B
   - Box C

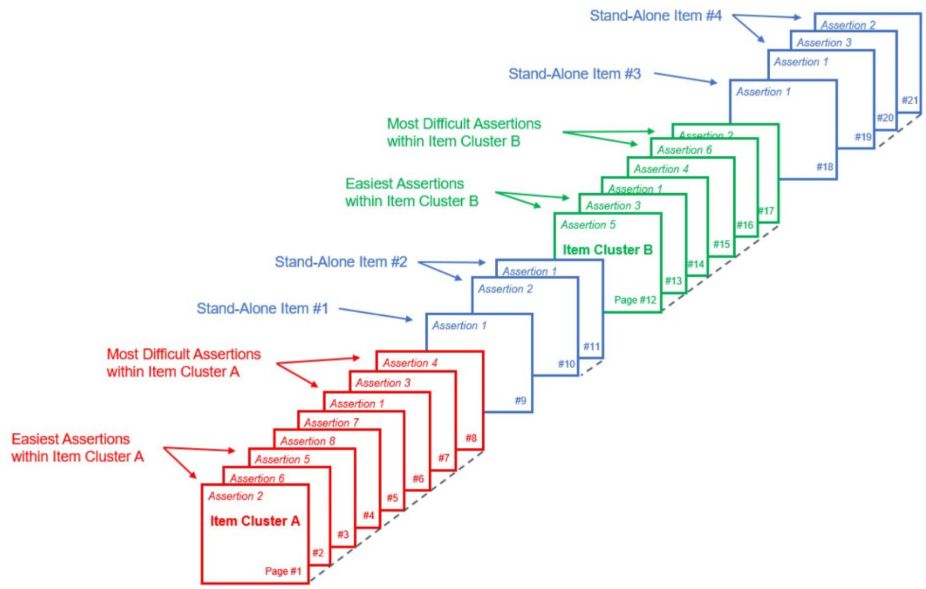
7. Which performance standard differentiates between the Level 2 Basic performance level and the Level 3 Proficient performance level? *

   Mark only one oval.

   - Level 2 Basic
   - Level 3 Proficient
   - Level 4 Advanced

Ordered Scoring Assertion Booklet (OSAB)

Here is a hypothetical Ordered Scoring Assertion Booklet (OSAB) that consists of pages 1 through 21:



8. Within each item cluster or stand-alone item within the OSAB, scoring assertions are ordered by difficulty. In the hypothetical OSAB presented above, is the assertion on page 7 of the OSAB easier, more difficult, or about the same as the assertion on page 3? *

Mark only one oval.

- ◯ The assertion on page 7 is easier than the assertion on page 3
- ◯ The assertion on page 7 is more difficult than the assertion on page 3
- ◯ The assertion on page 7 is about the same as the assertion on page 3
- ◯ The difficulty of the assertions on pages 7 and 3 cannot be compared in this graphic because they are not within the same item

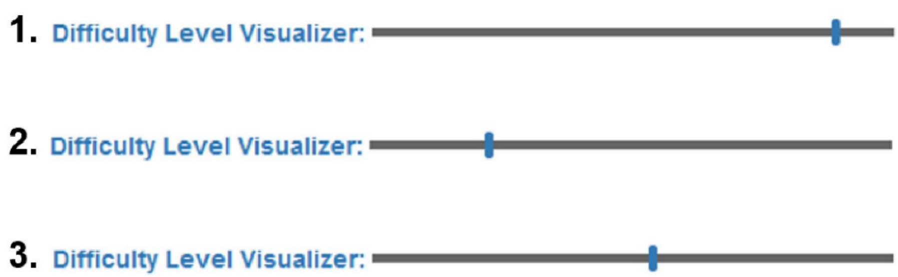
Standard-Setting Assertion Mapping Tool

9. Do you have to assign each scoring assertion to a performance level (or use the skip button)? *

Mark only one oval.

- Yes
- No

Below are three different scoring assertions' Difficulty Level Visualizers.



10. Which Difficulty Level Visualizer in the image above represents the most difficult scoring assertion? *

Mark only one oval.

- Difficulty Level Visualizer 1
- Difficulty Level Visualizer 2
- Difficulty Level Visualizer 3

11. Which Difficulty Level Visualizer in the image above represents the least difficult scoring assertion? *

Mark only one oval.

- ⬭ Difficulty Visualizer 1
- ⬭ Difficulty Visualizer 2
- ⬭ Difficulty Visualizer 3

Appendix 3-H

Standard-Setting Readiness Forms

Standard-Setting Readiness Forms

Exhibit H-1. Standard-Setting Round 1 Readiness Form

2022 ISAT Science Assessment Standard Setting Educator Panel - Readiness Form

Preparation for Round 1 Assertion Mapping

* Required

1. Name: *

____

2. Panelist ID (e.g., ID_T1P1_G5S): *

____

3. Assigned Committee: *

Mark only one oval.

- ◯ Science Grade 5
- ◯ Science Grade 8
- ◯ Science Grade 11

Preparation for Round 1 Assertion Mapping

4. The workshop training has prepared me to review the Performance-Level Descriptors (PLDs) and fully explained the concept of threshold PLDs. *

Mark only one oval.

- ◯ Yes
- ◯ No

5. The workshop training has prepared me to review the Ordered Scoring Assertion Booklet (OSAB). *

Mark only one oval.

- Yes
- No

6. The workshop training has clearly explained how to use the assertion map when reviewing the OSAB. *

Mark only one oval.

- Yes
- No

7. The workshop training has clearly explained the task of mapping assertions in the OSAB to the performance levels in the standard-setting tool. *

Mark only one oval.

- Yes
- No

8. The workshop training has fully explained how to use the contextual information (student impact data and benchmarking data) when mapping assertions to performance levels. *

Mark only one oval.

- Yes
- No

9. I have answered "Yes" to the above questions and I understand what I need to do to map assertions to performance levels. (Please initial below.) *

Mark only one oval.

- Yes
- No

10. Initial: *

11. If I answered "No" to any of the above questions, I received additional training. (Please initial below.) *

Mark only one oval.

- Yes
- No
- Not applicable

12. Initial: *

13. Following the additional training, I feel sufficiently trained on what I need to do to map assertions to performance levels. (Please initial below.) *

Mark only one oval.

- Yes
- No
- Not applicable

14. Initial: *

____

Exhibit H-2. Standard-Setting Round 2 Readiness Form

2022 ISAT Science Assessment Standard Setting Educator Panel - Readiness Form

Preparation for Round 2 Assertion Mapping

* Required

1. Name: *

____

2. Panelist ID (e.g., ID_T1P1_G5S): *

____

3. Assigned Committee: *

Mark only one oval.

◯ Science Grade 5
◯ Science Grade 8
◯ Science Grade 11

Preparation for Round 2 Assertion Mapping

4. The workshop training has clearly explained how to use the assertion map when reviewing the Ordered Scoring Assertion Booklet (OSAB). *

Mark only one oval.

◯ Yes
◯ No

5. The workshop training has clearly explained the task of mapping assertions in the OSAB to the performance levels in the standard-setting tool. *

Mark only one oval.

- Yes
- No

6. The workshop training has fully explained how to use the contextual information (student impact data and benchmarking data) when mapping assertions to performance levels. *

Mark only one oval.

- Yes
- No

7. The training fully explained the panel feedback data and impact data that was presented. *

Mark only one oval.

- Yes
- No

8. I understand my task for Round 2. *

Mark only one oval.

- Yes
- No

9. I have answered “Yes” to the above questions and I understand what I need to do to map assertions to performance levels. (Please initial below.) *

   Mark only one oval.

   - Yes
   - No

10. Initial: *

    ____

11. If I answered "No" to any of the above questions, I received additional training. (Please initial below.) *

    Mark only one oval.

    - Yes
    - No
    - Not applicable

12. Initial: *

    ____

13. Following the additional training, I feel sufficiently trained on what I need to do to map assertions to performance levels. (Please initial below.) *

    Mark only one oval.

    - Yes
    - No
    - Not applicable

14. Initial: *