South Dakota Science Assessment

2023–2024

Volume 3: Setting Achievement Standards

south dakota
DEPARTMENT OF EDUCATION
Learning. Leadership. Service.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

1. EXECUTIVE SUMMARY

In May 2015, the South Dakota State Board of Education (BOE) adopted the new South Dakota Science Standards. The new standards employ a three-dimensional conceptualization of science understanding, including science and engineering practices, crosscutting concepts, and disciplinary core ideas. With the adoption of the South Dakota Science Standards, and the development of new statewide assessments to measure student achievement relative to those standards, the South Dakota Department of Education (SDDOE) convened a standard-setting workshop to recommend a system of achievement standards for determining whether students have met the learning goals defined by the South Dakota Science Standards.

Under contract to SDDOE, Cambium Assessment, Inc. (CAI) conducted the standard-setting workshop to recommend achievement standards for the South Dakota Science Assessment (SDSA) in grades 5, 8, and 11. The workshop was conducted remotely September 15 – September 16, 2021.

South Dakota's science assessments are designed to measure the attainment of the South Dakota Science Standards adopted by the South Dakota BOE. The assessments are 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 South Dakota's SDSA blueprints, which were constructed to align with the South Dakota Science Standards.

South Dakota 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 assessments, 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 South Dakota Science Standards. SDDOE 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—Level 1, Level 2, Level 3, and Level 4. 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 2021 South Dakota English Language Arts (ELA) and Mathematics Assessments 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 achievement—Level 2, Level 3, and Level 4. Following Round 1, panelists were provided with feedback about the mappings of their fellow panelists and discussed their mappings as a group. Following Round 2, panelists engaged in a moderation session to review and modify recommended achievement standards to facilitate the adoption of an articulated set of achievement standards across grades and assessment systems. A modification to the Level 3 achievement standard was recommended for grade 11 during the moderation session.

Thirteen South Dakota science educators[2] were selected to serve as science standard-setting panelists, with four participants for the grade 5 panel, four participants for the grade 8 panel, and five participants 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 South Dakota Science Standards contributed to the standard-setting- process.

1.1 STANDARD-SETTING WORKSHOP

1.1.1 Overall Structure of the Workshop

The key features of the workshop included the following:

- The standard-setting procedure produced three recommended achievement standards (Level 2, Level 3, and Level 4) that will be used to classify student achievement on the SDSA 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 specified RP value associated with each individual assertion (impact data) and the projected 2021 South Dakota English Language Arts (ELA) and Mathematics Assessments achievement levels of each assertion (benchmark information), were provided to panelists during Round 2 of the Assertion-Mapping Procedure.
- The standard-setting workshop was conducted using CAI's online standard-setting tool. Because the workshop was conducted remotely, each panelist accessed the tool using their own computer.

[2] See Section 5.3.4, Educator Participants for more information on the panelists.

- Following Round 2, panelists engaged in a moderation session for reviewing and modifying recommended achievement standards to achieve an articulated system of standards across grades and assessment systems. A modification to the Level 3 achievement standard was recommended for grade 11 during the moderation session.

1.1.2 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	Level 3	Level 4
5	477	508	527
8	773	810	836
11	1073	1102	1134

Table 2 indicates the percentage of students that will reach or exceed each achievement standard in 2021. Figure 1 represents those values graphically.

Table 2. Percentage of Students Reaching or Exceeding Each Recommended Science Achievement Standard in 2021

Grade	Level 2	Level 3	Level 4
5	79	41	17
8	82	38	9
11	84	48	10

Figure 1. Percentage of Students Reaching or Exceeding Each Recommended Science Achievement Standard in 2021

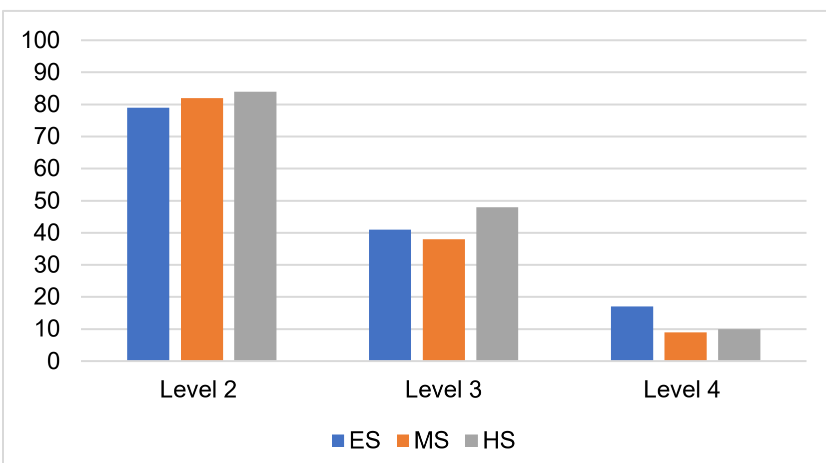
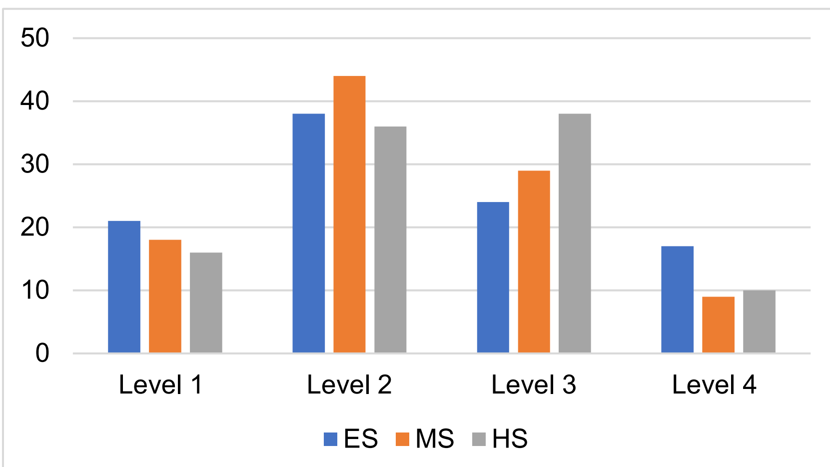


Table 3 indicates the percentage of students classified within each of the achievement levels in 2021. The values are displayed graphically in Figure 2.

Table 3. Percentage of Students Classified Within Each Science Achievement Level in 2021

Grade	Level 1	Level 2	Level 3	Level 4
5	21	38	24	17
8	18	44	29	9
11	16	36	38	10

Figure 2. Percentage of Students Classified Within Each Science Achievement Level in 2021



2. Introduction

South Dakota adopted the 2014 South Dakota Science Standards on May 18, 2015. The South Dakota Department of Education (SDDOE) and its assessment vendor, Cambium Assessment, Inc. (CAI), developed and administered a new assessment to measure the new standards. In spring 2021, they administered new assessments aligned to the South Dakota Science Standards to all grades 5, 8, and 11 students in South Dakota.

South Dakota provides information about the science assessments at: https://doe.sd.gov/assessment/Science.aspx.

New tests require new achievement standards to link achievement on the test to the content standards. SDDOE 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 September 2021.

The purpose of this documentation is to detail the standard-setting process for the South Dakota Science Assessment (SDSA) and resulting achievement standard recommendations.

3. South Dakota Science Standards

The South Dakota Science Assessment assesses the learning objectives described by the South Dakota Science Standards, adopted by South Dakota in 2015.

Information about the South Dakota Science Standards is available at: https://doe.sd.gov/contentstandards/.

The 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 multiple ways. First, rather than describe general knowledge and skills that students should know and be able to do, they describe specific performances that demonstrate what students know and can do. The South Dakota Science Standards refer to these performed knowledge and skills as *performance expectations (PEs)*. Second, the South Dakota Science Standards are intentionally multi-dimensional. Each performance expectation incorporates all three dimensions from *A Framework for K–12 Science Education* (National Research Council, 2012)—a science or engineering practice, a disciplinary core idea, and a crosscutting concept. Another unique feature of the South Dakota Science Standards is the assumption that students should learn all science disciplines, rather than select a few, as is traditionally done in many high schools, where students may elect, for example, to take biology and chemistry but not physics or astronomy.

Figure 3 shows the structure of the South Dakota Science Standards for a single grade 5 PE, 5-PS1-1.

Figure 3. Structure of the South Dakota Science Standards

The Core Ideas of the Fifth Grade standards include:

- Matter and Its Interactions
- Motion and Stability: Forces and Interactions
- Energy
- From Molecules to Organisms: Structures and Processes
- Ecosystems: Interactions, Energy, and Dynamics
- Earth's Place in the Universe
- Earth's Systems
- Earth and Human Activity

Fifth Grade Physical Science Conceptual Understanding:

Because matter exists as particles that are too small to see, matter is always conserved even if it seems to disappear. Measurements of a variety of observable properties can be used to identify particular materials. Chemical reactions that occur when substances are mixed can be identified by the emergence of substances with different properties; the total mass of substances when a reaction occurs remains the same. Energy can be "produced," "used," or "released" by converting stored energy. Plants capture energy from sunlight, which can later be used as fuel or food.

Fifth Grade Physical Science Standards	
5-PS1-1	Develop a model to describe that matter is made of particles too small to be seen. (SEP: 2; DCI: PS1.A; CCC: Scale/Prop.)

Source. https://doe.sd.gov/contentstandards/documents/sdSciStnd.pdf.

4. South Dakota Science Assessment

Due to the unique features of the three-dimensional science standards, items and tests based on the three-dimensional science standards, such as the SDSA, must also incorporate similarly unique features. The most impactful of these changes is that three-dimensional science standards 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.

4.1 Item Clusters and Stand-Alone Items

There are two types of items: item clusters and stand-alone items. An item cluster includes a phenomenon-based stimulus and a series of interactions that allow the student to demonstrate their mastery of the performance expectation (PE) by explaining the phenomenon or designing a solution to a presented engineering problem. The expectation is that item clusters will take students approximately 10 to 12 minutes to complete. Each stimulus ends with a task statement that provides the goal or understanding the student should reach. For example, "In the questions that follow, you will analyze what happens to the train when the brakes are applied." The student may explain, model, investigate, and/or create designs using the knowledge, skills, and abilities described by the PE. For example, in Figure 3, proficiency in this single PE requires activities that demonstrate the ability to analyze and evaluate data, the knowledge of properties and purposes of different forms of matter, and the application of experimental cause and effect. All interactions within an item cluster address the phenomenon presented in the stimulus. Item clusters contain between four and eight interactions.

Most states also utilize stand-alone items. Stand-alone items increase the number of covered PEs per student while being much quicker to complete than item clusters. Incorporating stand-alone items allows the blueprint to cover a greater number of PEs within a limited time. Stand-alone items are also phenomenon-based, contain only one or two interactions, and take students one to three minutes to complete in general.

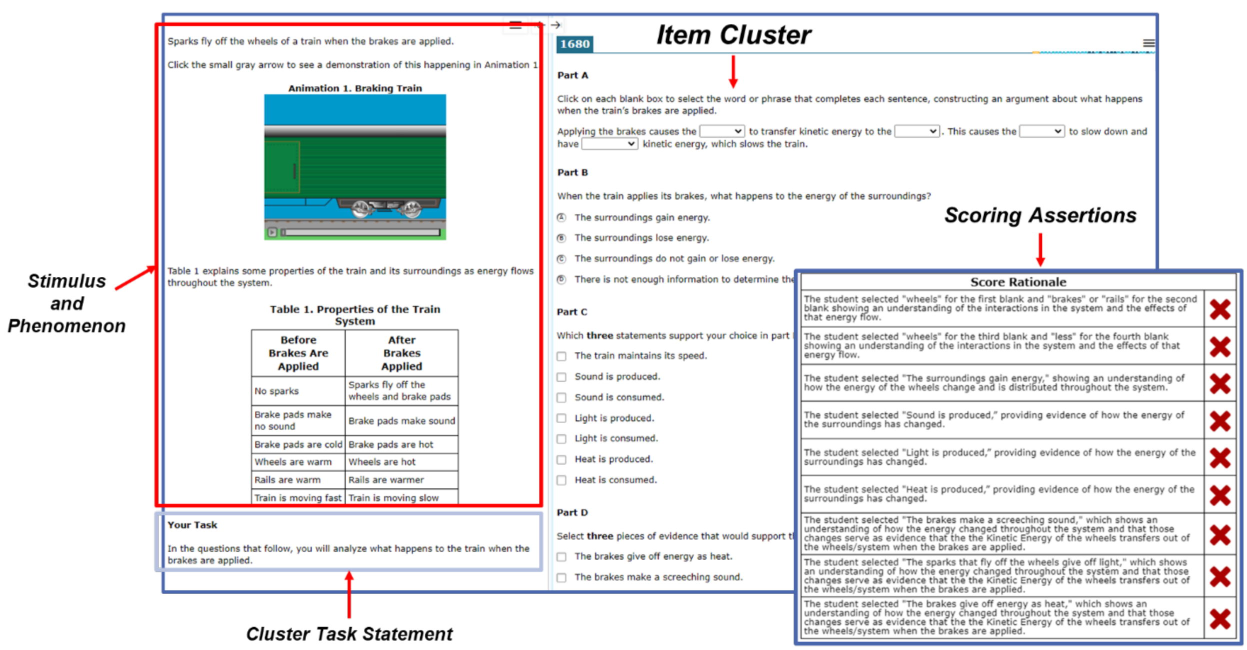
Both item types may use any of the available interaction types, including selected response, multi-select, table match, external copy, edit in-line choice, grids, and/or simulations of scientific investigations. For additional information on interaction types, refer to Volume 2, Appendix 2-C, Style Guide for Science Items, of this technical report.

4.2 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 Item Cluster and Scoring Assertions

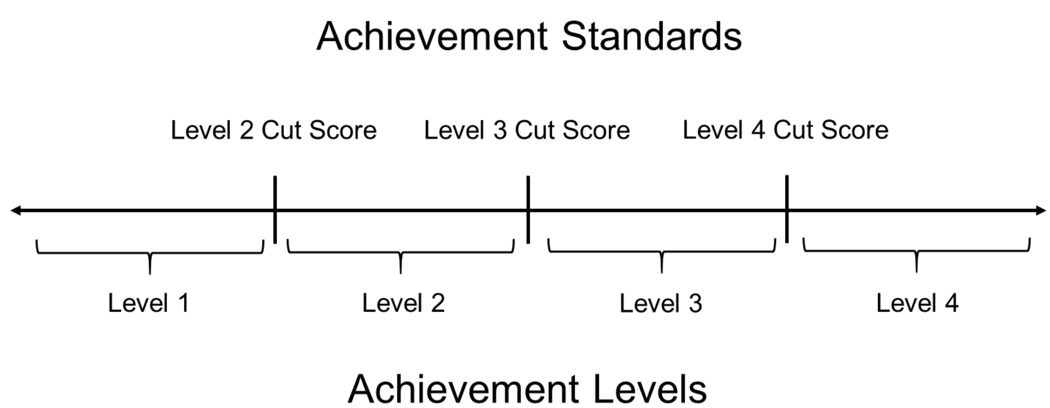


5. Standard Setting

Thirteen educators from South Dakota convened remotely September 15–16, 2021, to complete two rounds of standard setting to recommend three achievement standards for the South Dakota Science Assessment (SDSA).

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

Figure 5. Three Achievement Standards Defining South Dakota'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 to receive credit.

5.1 The Assertion-Mapping Procedure

A modification of traditional approaches to standard setting is necessary for the SDSA due to the structure of the content standards and, subsequently, the structure of the test items assessing the standards. The South Dakota Science Standards adopt a three-dimensional conceptualization of science understanding, including science practices, crosscutting concepts, and disciplinary core ideas. Accordingly, the new SDSA tests are comprised 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 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 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 North Dakota, Hawaii, and Utah statewide assessments in 2021.

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 Bookmark 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]

5.2 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. The three separate panels set achievement standards for each grade.

Table 4. summarizes the composition of the panels and the number of facilitators and panelists assigned to each. The 13 standard-setting participants included table leaders and panelists from South Dakota who taught in the content area and grade for which standards were being set.

Table 4. Panel Assignments

Room	Grade	Panelists	Facilitator	Facilitator Assistant
1	5	4	James McCann Anneka Wiersma	Sydney Brabble Kimberly David
2	8	4	Kevin Dwyer Vanessa Johnson	Melissa Mwai Brody Harkless
3	11	5	Matthew Davis Kam Mangis de Mark	Ethan Yosebashvili Mackenzie Worn

5.3 Participants and Roles

5.3.1 South Dakota Department of Education Staff

Staff from the South Dakota Department of Education (SDDOE) were present throughout the process and provided overall policy context and answered any policy questions that arose.

From SDDOE, attendees included:

- Matt Gill, Director of the Office of Assessment

[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.

- Christina Booth, General Assessment Support, Science Assessment, Science ALT Assessment and MSAA
- Jennifer Fowler, Science Specialist

5.3.2 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. Stephan Ahadi, Managing Director of Psychometrics facilitated and oversaw all AMP processes and tasks and provided training to participants.
- Dr. Frank Rijmen, Senior Director of Psychometrics, supervised all psychometric analyses conducted during and after the workshop.
- Dr. Widad Abdalla, Psychometrician, provided psychometric analyses.
- Alesha Ballman, Psychometric Project Coordinator, oversaw analytics technology and psychometrics.
- Sydney Brabble and Ethan Yosebashvili, Psychometric Support Assistants, provided support as needed.
- Melissa Mwai, Jennifer Chou, Mackenzie Worn, Marie Musumeci, Caroline Lempres, Kimberly David, and Brody Harkless, Program Management Team, managed process and logistics throughout the meeting.
- Floyd Helm, Mark Palomo, Brandon Palomo, and Luis Jorge, System Support Agents, troubleshot technology during the workshop.

5.3.3 Room Facilitators

Two 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:

- James McCann and Anneka Wiersma facilitated the science grade 5 panel
- Kevin Dwyer and Vanessa Johnson facilitated the science grade 8 panel
- Matthew Davis and Kam Mangis de Mark facilitated the grade 11 panel

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

5.3.4 Educator Participants

To establish achievement standards, SDDOE recruited a set of participants from across the state. Panelists included science teachers, administrators, and representatives from other stakeholder groups (e.g., coaches, college faculty) to ensure that a range of perspectives contributed to the standard-setting process and product. In recruiting panelists, SDDOE targeted the recruitment of participants to be representative of the gender and geographic representation of South Dakota's teacher population. All participants also had to be familiar with the South Dakota Science Standards content and test.

SDDOE selected classroom teachers from the resulting potential panelist pool and invited them to participate in the workshop. Due to the Covid-19 pandemic, target numbers fell short from the number of panelists recommended in the standard setting plan. The standard setting plan recommended 12 panelists per grade whereas the number of panelists that participated in the workshop amounted to 4 panelists for grades 5 and 8, and 5 panelists for grade 11.

Overall, the standard-setting workshop panelists were 15% male and 0% non-white. Represented stakeholder groups included Administrators, Coaches, General Education Teachers, Higher Education, and Special Education Teachers, with General Education Teachers comprising 69% of the panels overall. The majority of panelists taught in the grades to which they were assigned to set standards. Overall, 8% of panelists taught elementary school and 31% taught middle school (the remainder taught some combination of grades). Panelists worked in schools (46%), schools and districts (46%), and one worked in university (8%). School district areas included rural (54%), suburban (15%), and urban (31%), and were small (38%), medium (38%), and large (23%). Table 5 summarizes the characteristics of the panels.

Table 5. Panelist Characteristics

	Percentage of Panelists, by Panel			
	Science Grade 5	*Science Grade 8*	*Science Grade 11*	*Overall*
Characteristics				
Male	25%	0%	20%	15%
Non-White	0%	0%	0%	0%
Stakeholder Groups				
Administrator	50%	0%	0%	15%
Coach	0%	50%	0%	15%
General Education Teacher	50%	75%	80%	69%
Higher Education	0%	0%	20%	8%
Special Education Teacher	25%	0%	0%	8%
Current Position				
School	50%	50%	40%	46%
School and District	50%	50%	40%	46%
University	0%	0%	20%	8%
School District Area Size				

	Percentage of Panelists, by Panel			
	Science Grade 5	*Science Grade 8*	*Science Grade 11*	*Overall*
Large	25%	25%	20%	23%
Medium	25%	50%	40%	38%
Small	50%	25%	40%	38%
Not Applicable	0%	0%	0%	0%
School District Area Urbanicity				
Rural	75%	50%	40%	54%
Suburban	0%	0%	40%	15%
Urban	25%	50%	20%	31%
Not Applicable	0%	0%	0%	0%
Primary Grades Taught				
ES (grades 1–5)	25%	0%	0%	8%
MS (grades 6–8)	25%	75%	0%	31%
ES and MS (Preschool, Kindergarten, grades 1–8)	25%	0%	0%	8%
MS and HS (grades 6–12)	25%	25%	60%	38%
HS (grades 9–12) and College	0%	0%	20%	8%
College	0%	0%	20%	8%

For the results of any judgment-based method to be valid, the judgments must be made by individuals who are qualified to make them. Participants in the SDSA standard-setting workshop were highly qualified. They brought a variety of experience and expertise. Overall, 77% of panelists had earned a master's degree or higher. Ten panelists (77%) had taught in their assigned panel's grade and subject. The average time teaching the South Dakota Science Standards was nearly eight years. Many had experience teaching special populations; 54% taught students eligible to receive free or reduced-price lunch, 54% taught English learners (ELs), and 92% 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			
	Science Grade 5	*Science Grade 8*	*Science Grade 11*	*Overall*
Highest Degree				
Bachelor	0%	50%	20%	23%
Master	100%	50%	60%	69%
Doctoral	0%	0%	20%	8%
Years Teaching Experience				
None	0%	0%	0%	0%
Less than 1 year	0%	0%	0%	0%

	Percentage of Panelists, by Panel			
	Science Grade 5	Science Grade 8	Science Grade 11	Overall
1–5 years	25%	0%	0%	8%
6–10 years	25%	25%	20%	23%
11–15 years	50%	25%	20%	31%
16–20 years	0%	25%	0%	8%
More than 20 years	0%	25%	60%	31%
Years Teaching Experience in Assigned Grade				
None	75%	0%	0%	23%
Less than 1 year	0%	0%	0%	0%
1–5 years	0%	25%	0%	8%
6–10 years	25%	50%	40%	38%
11–15 years	0%	25%	0%	8%
16–20 years	0%	0%	0%	0%
More than 20 years	0%	0%	60%	23%
Subject Areas Currently Teaching[a]				
English Language Arts (ELA)	0%	0%	20%	8%
Mathematics	0%	0%	20%	8%
Social Studies	0%	0%	20%	8%
Science	50%	100%	100%	85%
Other[b]	50%	25%	20%	31%
Other professional experience in education	50%	25%	20%	31%
Years Professional Experience in Education				
None	50%	75%	80%	69%
Less than 1 year	0%	0%	0%	0%
1–5 years	25%	25%	0%	15%
6–10 years	0%	0%	20%	8%
11–15 years	0%	0%	0%	0%
16–20 years	25%	0%	0%	8%
More than 20 years	0%	0%	0%	0%
Experience Teaching Special Student Populations				
Students eligible to receive free/reduced price lunch	25%	75%	60%	54%
English Learners (ELs)	50%	75%	40%	54%
Students on an Individual Education Plan (IEP)	75%	100%	100%	92%
Average years teaching the South Dakota Science Standards	9	9	6[c]	8

[a]The total sums to over 100% for "Subject Areas Currently Teaching" as many participants taught multiple subjects. [b]Other Subject Areas Currently Teaching includes Art, ELED, Health, Music, PE, and Special Education. [c]One grade

11 panelist was excluded from the average years teaching the South Dakota Science Standards calculation as a quantifiable year was not given.

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

5.3.5 Table Leaders

Volunteers from the participant pool served as panelist leaders. In addition to serving as panelists and mapping assertions, table leaders had the additional responsibility of participating in the moderation session.

5.4 MATERIALS

5.4.1 Achievement-Level Descriptors

With the adoption of the new standards in science, and the development of new statewide assessments to assess achievement of those standards, SDDOE must adopt a similar system of achievement, or achievement standards, 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 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 state approval of the achievement standards used to describe student achievement on score reports.

South Dakota uses four achievement levels to describe student achievement: “Level 1,” “Level 2,” “Level 3,” and “Level 4.” At the policy level, these achievement levels are defined as follows:

- **Level 1.** Student has not yet met the achievement standard for science expected for this grade. Students performing at this level require substantial improvement toward mastery of science knowledge and skills. Students performing at this level will likely need substantial support to get on track for success in the next grade.

- **Level 2.** Student has nearly met the achievement standard for science expected for this grade. Students performing at this level require further development toward mastery of science knowledge and skills. Students performing at this level will likely need support to get on track for success in the next grade.
- **Level 3.** Student has met the achievement standard for science expected for this grade. Students performing at this level are demonstrating progress toward mastery of science knowledge and skills. Students performing at this level are on track for likely success in the next grade.
- **Level 4.** Student has exceeded the achievement standard for science expected for this grade. Students performing at this level are demonstrating advanced progress toward mastery of science knowledge and skills. Students performing at this level are on track for likely success in the next grade.

Science Range Achievement-Level Descriptor Development

The SDDOE drafted range ALDs that describe observable evidence for what student performance looks like in science at each achievement level and grade. The SDDOE and CAI reviewed the draft range ALDs to ensure that the language accurately represented the goals and policies of the state. CAI worked with them to make revisions where necessary.

The day prior to the standard-setting workshop, the group of South Dakota 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. Appendix 3-B, South Dakota Science Assessment Range Achievement-Level Descriptors, provides the final range ALDs for the SDSA.

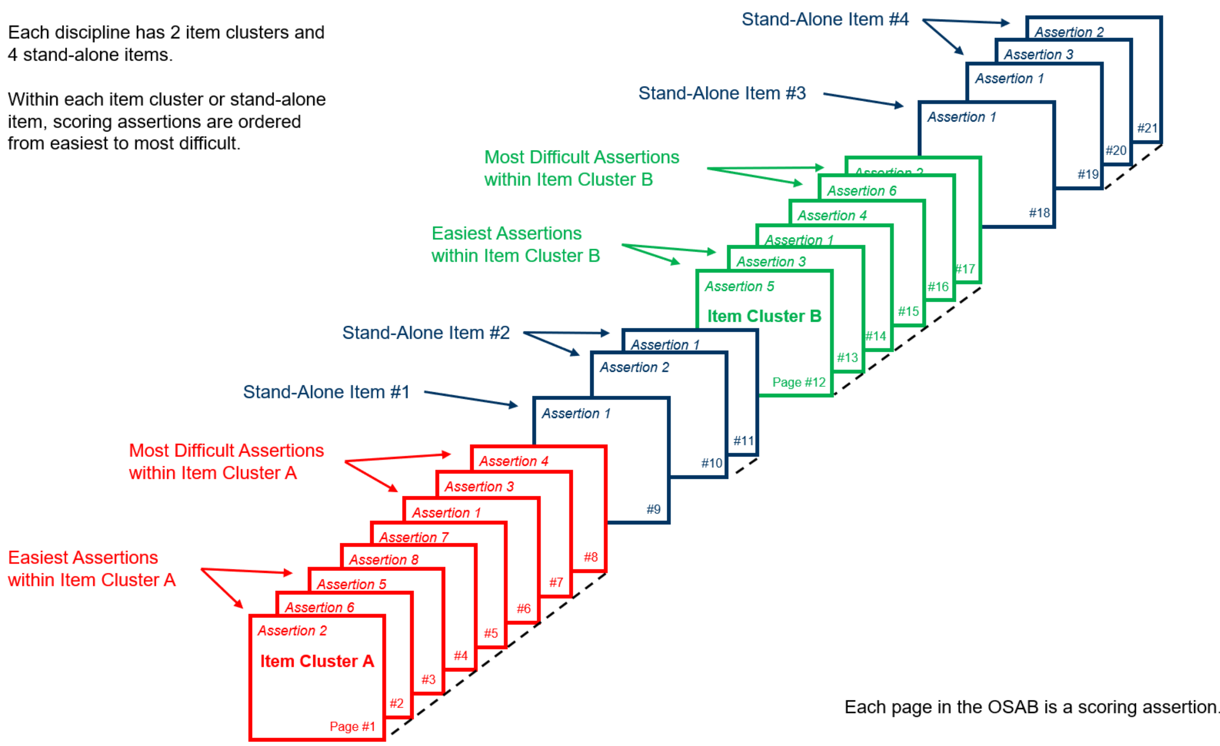
5.4.2 Ordered Scoring Assertion Booklets

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 the state 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 SDDOE and approved without any changes for all three grades.

Figure 6 describes the structure of the OSAB.

Figure 6. Ordered Scoring Assertion Booklet (OSAB)



Since the operational test was adaptive, the order of the items was different over students. The items in the OSABs were grouped by science discipline, so that panelists work through all items associated with one discipline before moving on to the next, allowing panelists to focus on the knowledge and skill requirements for one discipline at a time. For the grade 5 OSAB, the Earth and Space Sciences discipline items were presented first, then Life Sciences items, and then Physical Sciences items. For the grade 8 OSAB, the Physical Sciences discipline items were presented first, then Life Sciences items, and then Earth and Space Sciences items. For the grade 11 OSAB, the Physical Sciences discipline items were presented first, then Life Sciences items, and then Earth and Space Sciences items. Two item clusters and four stand-alone items represent each discipline. 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 within 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 6 represents the order of presentation to the panelists, but assertions were not ordered by overall difficulty across all items. (see Figure 7 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 "Level 1," "Level 2," and "Level 3," but not "Level 4."

Each OSAB contains three disciplines and 18 items (item clusters and stand-alone items). The grade 5 OSAB contained 75 assertions, the grade 8 OSAB contained 77 assertions, and the grade 11 OSAB contained 83 assertions. Each comprised of six item clusters and 12 stand-alone items.

5.4.3 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, 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 in Section 5.7.2.2, Feedback Data). Figure 7 presents the assertion map for grade 5. The assertions maps for all three grades are presented in Appendix 3-C, Standard-Setting Assertion Maps.

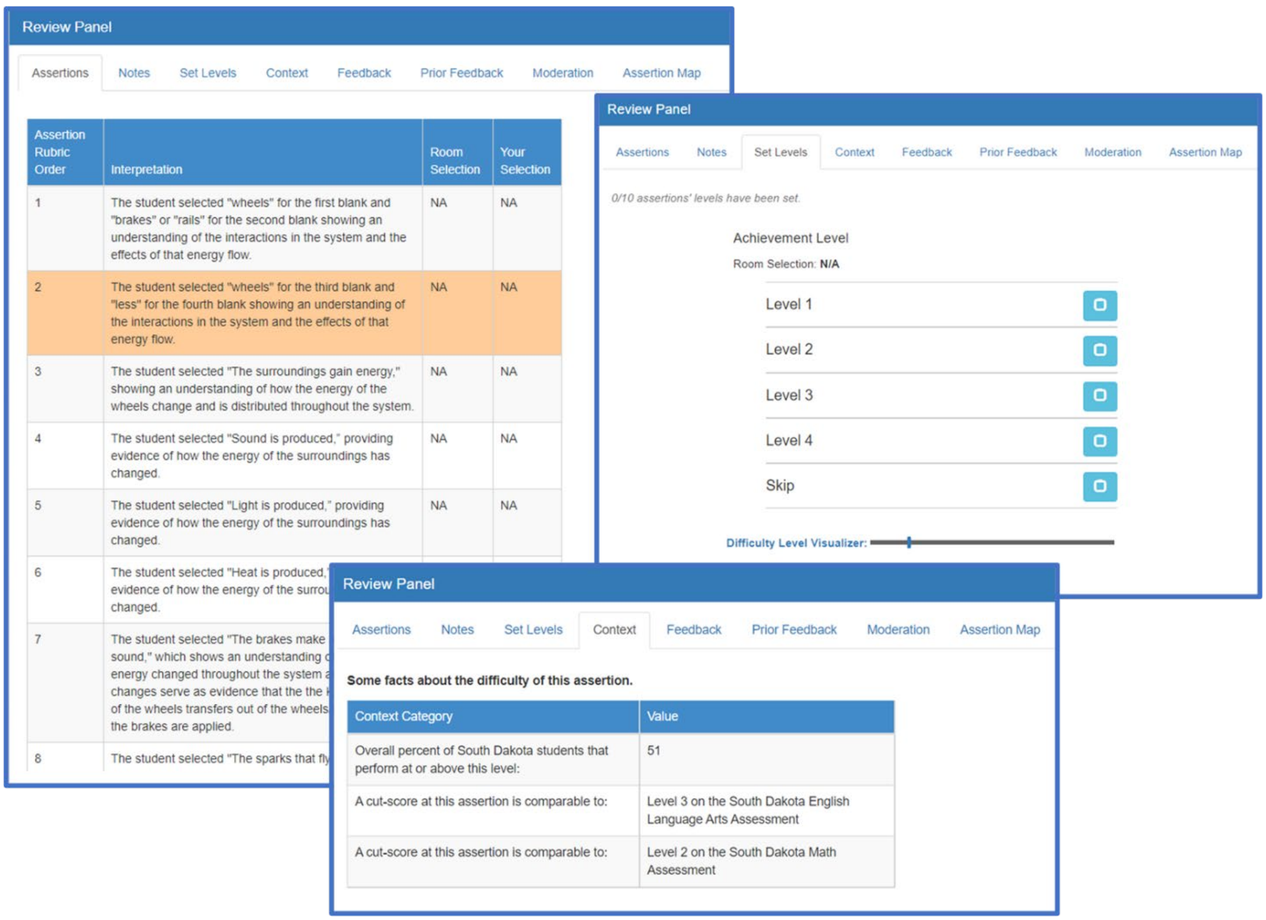
Figure 7. Standard-Setting Assertion Map, Science Grade 5

5.5 WORKSHOP TECHNOLOGY

The standard-setting panelists used CAI's online application for standard setting. Each panelist used their own computer on which they 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 8), 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 content 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).[4] Panelists also reviewed their assertion placement and the overall placement for room.

Figure 8. 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.

[4] The difficulty level visualizer represented the percentage of students who would fall at or above the difficulty level of that assertion.

5.6 EVENTS

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

Table 7. Standard-Setting Agenda Summary

Day 1: Wednesday, September 15, 2021
• Large-Group Orientation • Review and Take the Operational Test • Review Range ALDs • Discuss Threshold ALDs • OSAB Review
Day 2: Thursday, September 16, 2021
• 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

5.6.1 Participant Login

Panelists were required to attend a technical check prior to the standard-setting workshop to ensure they had access to the required sites needed for 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.

5.6.2 Large-Group Introductory Training

Matt Gill and Christina Booth, SDDOE, welcomed panelists to the workshop and provided context and background for the SDSA. Christina Booth outlined the roles and responsibilities of the participants at the workshop: panelists, CAI staff, and SDDOE personnel. Dr. Ahadi 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-E, Standard-Setting Training Slides, provides the slides used during the large-group training.

5.6.3 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 operational SDSA, 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 your computer during the course of 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 your computer

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.

5.6.4 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 2021, 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 as 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. They could score their responses and had 90 minutes to interact with the test.

5.6.5 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 performance, 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 are able to do with respect to the South Dakota Science Standards. Workshop facilitators provided panelists with draft range ALDs, test blueprints, and the South Dakota Science Standards. The facilitators lead 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 participants understood what students in South Dakota should know and be able to do and how much knowledge and skill students are expected to demonstrate at each level of achievement.

5.6.6 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 that 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 within each room.

The AMP employs the range ALDs since panelists are mapping items across the full range of the 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.

5.6.7 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 within 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 also helped to facilitate a common understanding among workshop panelists.

5.6.8 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 5.5, Workshop Technology), the assertion map (described in Section 5.4.3, Assertion Maps), their notes from the OSAB review, and their professional judgments. Figure 9 graphically describes the assertion-mapping process.

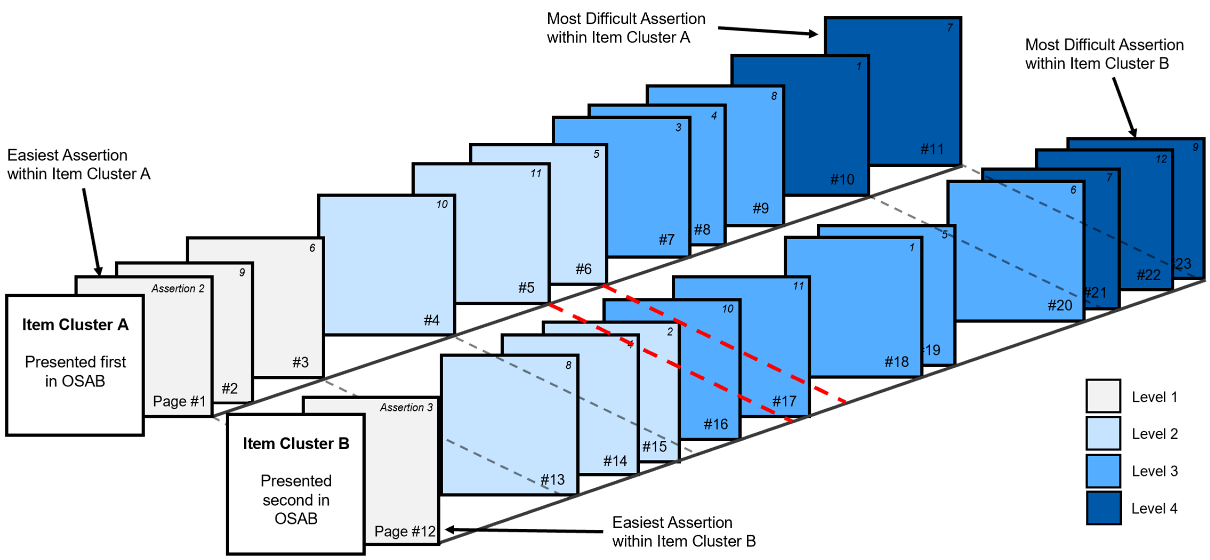
Facilitators provided the following process to guide the mapping of assertions onto ALDs:

1. How does the student interaction give rise to the assertion? Did they plot, select, or write something?
2. Why is this assertion more difficult to achieve than the previous one (within 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 9, 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-E Standard-Setting Training Slides, provides the training slides used during the breakout room training.

Figure 9. Example of Assertion Mapping



Note. Figure 9 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.

5.6.9 Practice Quiz

Panelists completed a practice quiz before beginning a practice round. The quiz assessed panelists' understanding in multiple ways. They must 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-F, Standard-Setting Practice Quiz, provides the quiz that panelists completed before mapping any assertions.

5.6.10 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.

5.6.11 Readiness Assertion

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;
- Understanding how to use the assertion map when reviewing the OSAB and considering assertion mapping decisions;
- Understanding 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-G, Standard-Setting Readiness Forms, provides the forms that panelists completed prior to each round of standard setting.

5.7 Assertion Mapping

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

5.7.1 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 an assertion to the hypothetical cut.[5] 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 assertion mappings for the two adjacent levels were considered (e.g., for the second cut, only the assertion mappings for the "Level 2" and "Level 3" 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

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

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

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

Unlike the Bookmark method, the cut scores for a room were not the median value of the cut scores of the individual panelists. Instead, cut scores at the room (grade) level were computed using the same method but taking into account the assigned levels of all the raters in the room. Applying

[5] 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 middle school and high school in some states, and for South Dakota grades 8 and 11 was because the difficulty of most items exceeded students' abilities. RP50 better aligned with the ALD 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).

these cut scores to the 2021 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 for the room.

5.7.2 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.

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

5.7.2.1 Contextual Information

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.

Benchmark Data

The 2021 South Dakota English Language Arts (ELA) and Mathematics Assessment scores provided benchmark data, another source of contextual information that panelists could use to evaluate and adjust their assertion mapping. By comparing the results of the round against the South Dakota percentage proficient on ELA and Mathematics, panelists could evaluate the reasonableness of the proposed achievement standards. For each ordered scoring assertion, panelists were provided with the associated achievement level for the South Dakota ELA and Mathematics Assessments. An example of the benchmark information provided for each assertion in the review panel of the standard-setting tool is shown in Figure 8. This provided external evidence of student achievement for panelists to consider when mapping assertions to achievement levels in Round 2.

5.7.2.2 Feedback Data

The online standard-setting tool created feedback data and cut scores corresponding to the assertion mappings for each panelist and for the room overall. 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 2021 administration of the SDSA. This information allowed panelists to compare their mappings to other panelist's mappings to evaluate the impact of their current mappings.

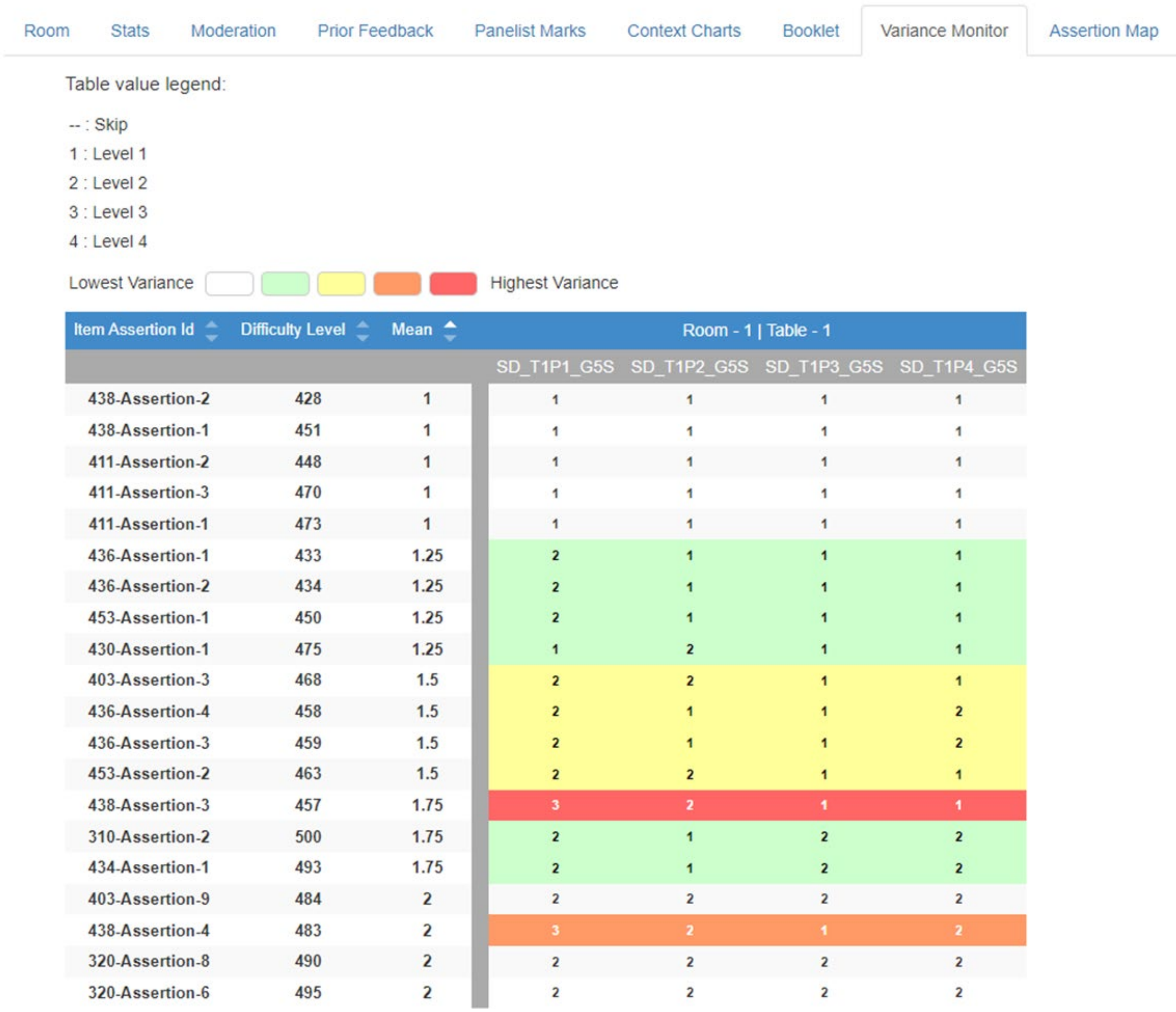
The standard-setting tool also generated variance monitor data and the assertion maps in the tool were updated to display the tentative standards for panelists to evaluate before Round 2 (the variance data and assertion maps are described in more detail below). All feedback and information served to inform, but not determine, their 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 panel spent time reviewing and discussing assertion mappings and articulation, beginning with individual-level feedback and discussion, and progressing to the room-level discussion. 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 10 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 10. Variance Monitor in CAI's Standard-Setting Tool

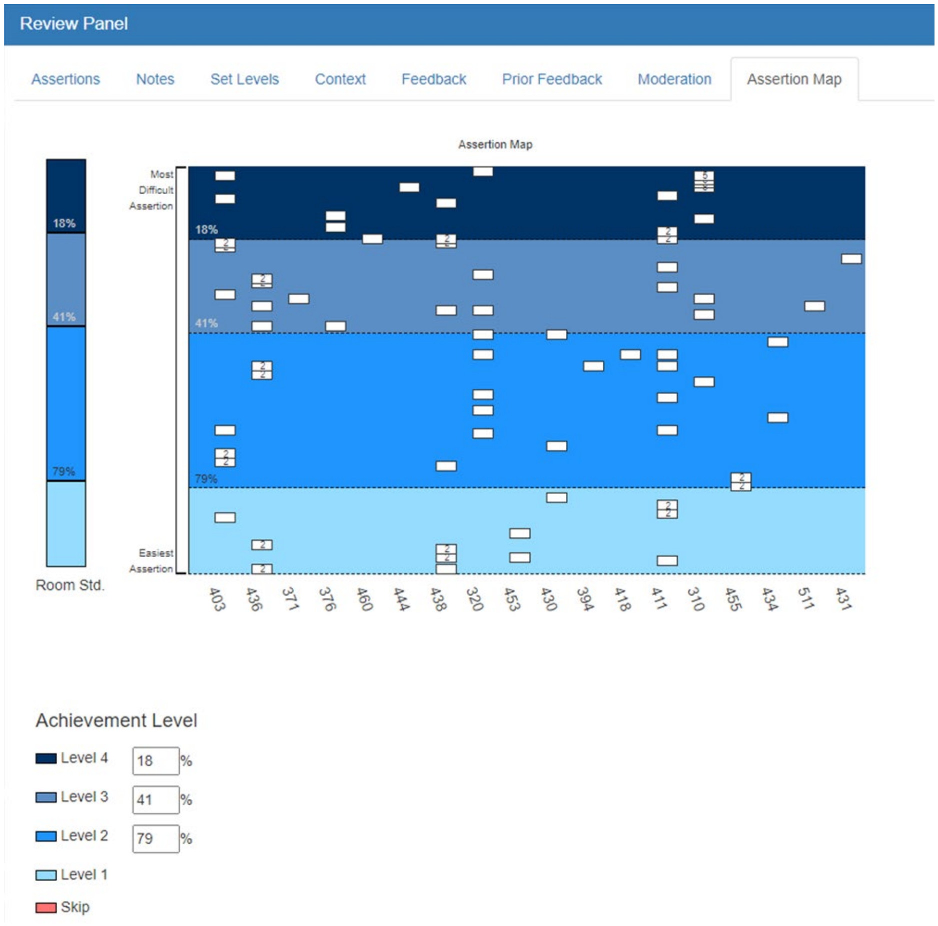


Item Assertion Id	Difficulty Level	Mean	Room - 1 | Table - 1			
			SD_T1P1_G5S	SD_T1P2_G5S	SD_T1P3_G5S	SD_T1P4_G5S
438-Assertion-2	428	1	1	1	1	1
438-Assertion-1	451	1	1	1	1	1
411-Assertion-2	448	1	1	1	1	1
411-Assertion-3	470	1	1	1	1	1
411-Assertion-1	473	1	1	1	1	1
436-Assertion-1	433	1.25	2	1	1	1
436-Assertion-2	434	1.25	2	1	1	1
453-Assertion-1	450	1.25	2	1	1	1
430-Assertion-1	475	1.25	1	2	1	1
403-Assertion-3	468	1.5	2	2	1	1
436-Assertion-4	458	1.5	2	1	1	2
436-Assertion-3	459	1.5	2	1	1	2
453-Assertion-2	463	1.5	2	2	1	1
438-Assertion-3	457	1.75	3	2	1	1
310-Assertion-2	500	1.75	2	1	2	2
434-Assertion-1	493	1.75	2	1	2	2
403-Assertion-9	484	2	2	2	2	2
438-Assertion-4	483	2	3	2	1	2
320-Assertion-8	490	2	2	2	2	2
320-Assertion-6	495	2	2	2	2	2

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 11 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-H, Round 1 and Round 2 Standard-Setting Assertion Maps.

Figure 11. 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.

5.8 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.

5.8.1 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	Cut Score			Impact Data		
	Level 2	*Level 3*	*Level 4*	*Level 2*	*Level 3*	*Level 4*
5	477	508	526	79	41	18
8	771	810	836	84	38	9
11	1067	1099	1134	89	52	10

Note. The grade row summarizes the room data. 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 individual-level feedback and discussion, progressing to the room-level discussion. After reviewing the feedback (i.e., individual cuts and 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.

5.8.2 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	Cut Score			Impact Data		
	Level 2	*Level 3*	*Level 4*	*Level 2*	*Level 3*	*Level 4*
5	477	508	527	79	41	17
8	773	810	836	82	38	9
11	1073	1099	1134	84	52	10

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

5.8.3 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. For all grades and all level standards (with the exception of Level 2 standard in grade 11), the indicators consistently show that there is a convergence in individual standards.

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

Grade	Statistic	Level 2		Level 3		Level 4	
		Round 1	*Round 2*	*Round 1*	*Round 2*	*Round 1*	*Round 2*
5	IQR	6.25	0.75	11.5	6.25	6.75	2.5
	SD	7.63	1.50	12.19	3.77	9.00	3.11
8	IQR	13.5	9.5	3.25	3.25	13.25	2
	SD	13.53	12.03	3.10	3.10	19.33	4.00
11	IQR	5	8	7	2	4	0
	SD	2.83	5.24	4.38	2.19	13.62	9.55

5.8.4 Moderation

Panelists receive the information necessary for articulation prior to Round 2. Often, panelists intuitively create well-articulated sets of achievement standards, but sometimes minor changes might significantly improve articulation.

On the last day of the workshop, table leaders met to discuss and resolve any issues or needs related to cross-grade articulation, resulting in the final recommendations provided in Table 11. Workshop leaders reminded panelists that content is one of multiple considerations in setting achievement standards—perhaps the most important, but not the only consideration; panelists also considered impact and policy in Round 2. After discussion, the moderation panel made a minor adjustment to the grade 11 Level 3 cut for better articulation across the grades.

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

Table 11. Moderated Results for Science

Grade	Cut Score			Impact Data		
	Level 2	*Level 3*	*Level 4*	*Level 2*	*Level 3*	*Level 4*
5	477	508	527	79	41	17
8	773	810	836	82	38	9
11	1073	1102*	1134	84	48*	10

*Minor adjustment made during the moderation session.

Figure 12 displays the percentage of students that will reach or exceed each of the recommended achievement standards in 2021.

Figure 12. Percentage of Students Reaching or Exceeding Each Recommended Science Achievement Standard in 2021

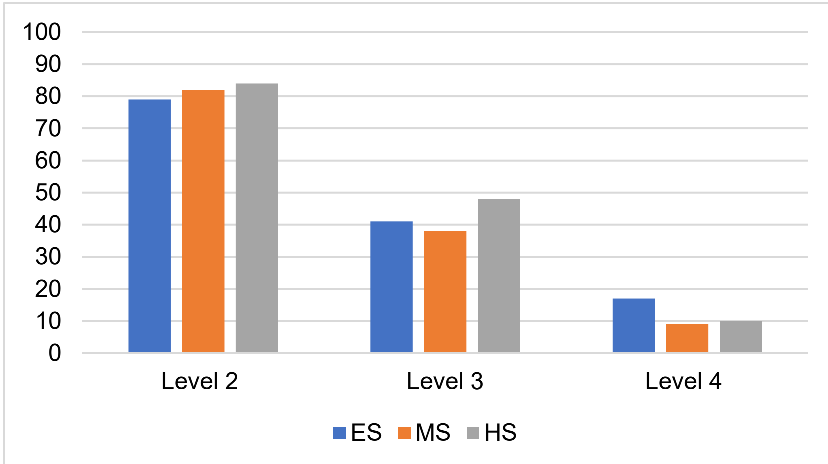
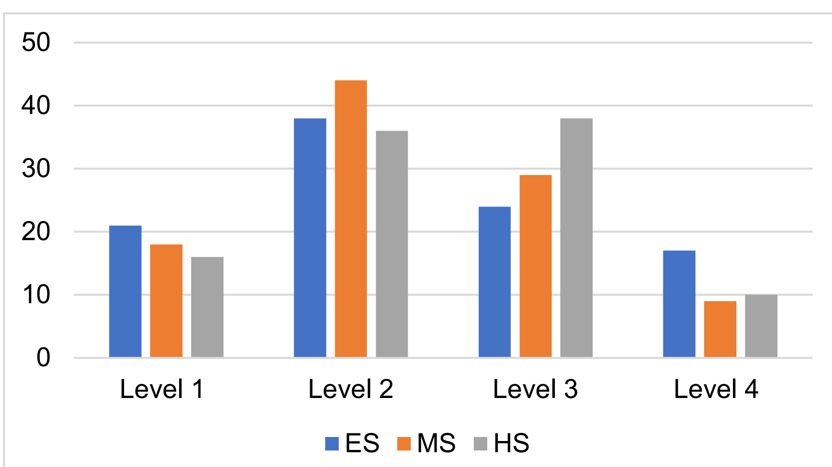


Table 12 indicates the percentage of students classified within each of the achievement levels in 2021. The values are displayed graphically in Figure 13.

Table 12. Percentage of Students Classified Within Each Science Achievement Level in 2021

Grade	Level 1	Level 2	Level 3	Level 4
5	21	38	24	17
8	18	44	29	9
11	16	36	38	10

Figure 13. Percentage of Students Classified Within Each Science Achievement Level in 2021



5.9 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 through 17 summarize the results of the evaluations.

Workshop participants overwhelmingly indicated clarity in the instructions, materials, data, and process (see Table 13). One grade 8 panelist reported some lack of clarity with the ALDs.

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"			
	Science Grade 5	*Science Grade 8*	*Science Grade 11*	*Overall*
Instructions provided by the workshop leader	100%	100%	100%	100%
Achievement-Level Descriptors (ALDs)	100%	75%	100%	92%
Ordered Scoring Assertion Booklet (OSAB)	100%	100%	100%	100%
Assertion Map	100%	100%	100%	100%
Impact Data (percentage of students that would achieve at the level indicated by the assertion difficulty)	100%	100%	100%	100%
Panelist Agreement Data	100%	100%	100%	100%

Note. Number of responses = 13 (grade 5 responses = 4, grade 8 responses = 4, and grade 11 responses = 5). Evaluation response options included "Very Unclear," "Somewhat Unclear," "Somewhat Clear," and "Very Clear."

Panelists felt that the time allocated to various workshop tasks may be adjusted, as shown in Table 14. Of the panelists who did not indicate that the time allocation for a task was "About Right"

- Three indicated that the large-group orientation was too long;
- Two indicated having too much time for taking the assessment;
- Five reported having too much time to review the ALDs;
- One grade 11 participant reported having too little time to discuss the skills demonstrated by students who are "just barely" described by each ALD, and six (one from grade 5, two from grade 8, and three from grade 11) reported having too much time for discussion;
- Three panelists indicated having too much time to review the OSAB;
- Two panelists indicated having too much time to map assertions to achievement levels in each round;
- One panelist reported having too little time for discussing Round 1 results, and three panelists indicated having too much time for 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"			
	Science Grade 5	*Science Grade 8*	*Science Grade 11*	*Overall*
Large-group orientation	75%	50%	100%	77%
Experiencing the online assessment	100%	75%	80%	85%
Reviewing the Achievement-Level Descriptors (ALDs)	100%	50%	40%	62%
Discussion of the skills demonstrated by students who are "just barely" described by each ALD	75%	50%	20%	46%
Reviewing the Ordered Scoring Assertion Booklet (OSAB)	75%	75%	80%	77%
Mapping your scoring assertions to achievement levels in each round	100%	75%	80%	85%
Round 1 results discussion	75%	75%	60%	69%

Note. Number of responses = 13 (grade 5 responses = 4, grade 8 responses = 4, and grade 11 responses = 5). Evaluation response options included "Too Little," "Too Much," and "About Right."

Participants appreciated the importance of the multiple factors contributing to assertion mapping, with all but a single panelist in grade 8 rating each factor as important or very important (see Table 15).

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"			
	Science Grade 5	*Science Grade 8*	*Science Grade 11*	*Overall*
Achievement-Level Descriptors (ALDs)	100%	100%	100%	100%
"Just Barely" ALDs	100%	100%	100%	100%
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	100%	100%	100%	100%
External benchmark data	100%	100%	100%	100%
Impact Data (percentage of students that would achieve at the level indicated by the assertion difficulty)	100%	100%	100%	100%
Room agreement data (room and individual standards)	100%	75%	100%	92%

Note. Number of responses = 13 (grade 5 responses = 4, grade 8 responses = 4, and grade 11 responses = 5). Evaluation response options included "Not Important," "Somewhat Important," and "Very Important."

Participant understanding of the workshop processes and tasks was high (see Table 16). Of the panelists who did not agree

- one participant in grade 11 disagreed that the procedures used were fair and unbiased;
- one panelist in grade 8 and one panelist in grade 11 disagreed that the ALDs provided clear expectations;
- one grade 11 panelist disagreed with the "just barely" ALD statement;
- one grade 8 panelist disagreed with the statement regarding comfortability expressing their opinions throughout the workshop;
- one grade 8 panelist disagreed with the statement that everyone was given the opportunity to express his or her opinions throughout the workshop.

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"			
	Science Grade 5	*Science Grade 8*	*Science Grade 11*	*Overall*
I understood the purpose of this standard-setting workshop.	100%	100%	100%	100%
The procedures used to recommend achievement standards were fair and unbiased.	100%	100%	80%	92%
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%	75%	80%	85%
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.	100%	100%	80%	92%
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%	100%	100%	100%
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.	100%	100%	100%	100%

At the end of the workshop, please rate your agreement with the following statements.	**Percentage Indicating "Agree" or "Strongly Agree"**			
	Science Grade 5	*Science Grade 8*	*Science Grade 11*	*Overall*
I found the panelist agreement data (room and individual standards) and discussions helpful when mapping assertions to achievement levels.	100%	100%	100%	100%
I felt comfortable expressing my opinions throughout the workshop.	100%	75%	100%	92%
Everyone was given the opportunity to express his or her opinions throughout the workshop.	100%	75%	100%	92%

Note. Number of responses = 13 (grade 5 responses = 4, grade 8 responses = 4, and grade 11 responses = 5). Evaluation response options included "Strongly Disagree," "Disagree," "Agree," and "Strongly Agree."

The majority of participants agreed that the standards set during the workshop reflected the intended grade-level expectations (see Table 17). However, two grade 11 panelists disagreed that students performing at Level 2 partially meet expectations for the grade, and one grade 11 panelist disagreed that students performing at Level 3 meet the 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"**			
	Science Grade 5	*Science Grade 8*	*Science Grade 11*	*Overall*
A student performing at "Level 2" has partially met expectations for the grade.	100%	100%	60%	85%
A student performing at "Level 3" has met the expectations for the grade.	100%	100%	80%	92%
A student performing at "Level 4" has exceeded the expectations for the grade.	100%	100%	100%	100%

Note. Number of responses = 13 (grade 5 responses = 4, grade 8 responses = 4, and grade 11 responses = 5). Evaluation response options included "Strongly Disagree," "Disagree," "Agree," and "Strongly Agree."

5.9.1 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."

Twelve panelists responded to the first question, and three responded to the second. Most responses indicated the training was effective and the process was clear. Participants provided minor suggestions, such as shortening or lengthening the time allocated for some tasks and having the workshop in person and with more participants. Many appreciated the organization, well-prepared materials, and technology, and many panelists complimented the professionalism and expertise of the facilitators.

Additional participant comments included:

"This was my first time doing this part of Standards Testing...it was very difficult but I feel like I gained a huge understanding!"

"Well put together. Feel as though the discussion would be richer in person."

"I had a lot of fun doing this. My panel committee and room was very open during discussions and they were great to work with. Even when questions came up everyone was so willing to work together and our leader never made us feel uncomfortable. She made it known that questions were welcomed so that we could learn and understand what was expected. I learned a lot and it was a great opportunity for me and I would love to be able to do this again now that I've seen how it works and had great training over it."

"Vanessa and Kevin did an AMAZING job facilitating the 8th grade group. They were open to questions and discussions and made us feel that all of our opinions and thoughts were important and valid."

6. 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.

6.1 EVIDENCE OF ADHERENCE TO PROFESSIONAL STANDARDS AND BEST PRACTICES

The South Dakota Science Assessment (SDSA) standard-setting workshop was designed and executed consistent with established practices and best-practice principles (Hambleton & Pitoniak, 2006; Hambleton, Pitoniak, & Copella, 2012; Kane, 2001; Mehrens, 1995). 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 interpretations involve 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 with distinct substantive interpretations should be informed by sound empirical data concerning the relation of test performance to the relevant criteria.

The sections of this documentation discussing 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 5.1, The Assertion-Mapping Procedure, provides the justification for and the additional benefits of selecting the AMP method to establish the cut scores; Section 5.6, Events, through Section 5.8.1, 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 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 5.3.4, 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 5.7.2, Contextual Information and Feedback Data). This set of empirical data provides necessary and additional context describing student performance given the recommended standards.

6.2 Evidence in Terms of Peer Review Critical Elements

The United States Department of Education (USDOE) 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.

South Dakota educators 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 South Dakota'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**: Achievement standards setting. The state used a technically sound method and process that involved panelists with appropriate experience and expertise for setting its academic achievement standards.

Evidence to support this critical element includes:

1) The rationale for and technical sufficiency of the AMP method selected to establish achievement standards (Section 5.1, 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 5.6, Events; Section 5.6.2, Large-Group Introductory Training; Section 5.8, Assertion Mapping Results; and Section 6.1, Evidence of Adherence to Professional Standards and Best Practices).
3) Panelists self-reported readiness to undertake the task (Section 5.6.11, Readiness) and confidence in the workshop process and outcomes (Section 5.9, Workshop Evaluations) 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 South Dakota's science content standards, and individuals with experience and expertise teaching special population and general education students in South Dakota (Section 5.3.4, Educator Participants; and Appendix 3-A, Standard-Setting Panelist Characteristics).

7. 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*, 153–168.

Cizek, G. J., & Koons, H. (2014). Observation and Report on Smarter Balanced Standard Setting: October 12–20, 2014. Accessed 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, October). 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 Associates.

Mehrens, W. (1995). Legal and professional bases for licensure testing. In J. C. Impara (Ed.), *Licensure testing: Purposes, procedures, and practices*. Lincoln, NE: Buros Institute of Mental Measurements, University of Nebraska–Lincoln.

Mitzel, H. C., Lewis, D. M., Patz, R. J., & Greene, D. R. (2001). The Bookmark procedure: Psychological perspectives. In G. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives* (pp. 249–281). Mahwah, NJ: Lawrence Erlbaum Associates.

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*, 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.

Skaggs, G., Hein, S. F., & Awuor, R. (2007). Setting passing scores on passage-based tests: A comparison of traditional and single-passage bookmark methods. *Applied Measurement in Education*, *20*, 405–426.

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, D.C. Accessed from https://www2.ed.gov/policy/elsec/guid/assessguid15.pdf.

Appendix 3-A

Standard-Setting Panelist Characteristics

Standard-Setting Panelist Characteristics

Table 3-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 South Dakota Science Standards	School District Size	School District Area Urbanicity	Table Leader
General Education Teacher	School, District	Female	White	Master's degree (e.g., M.A., M.S.), Working presently on my doctoral degree.	11 to 15 years	None	9	Large	Rural	Yes
Administrator, Special Education Teacher	School, District	Female	White	Master's degree (e.g., M.A., M.S.)	6 to 10 years	1 to 5 years	This is my first time	Small	Urban	Yes
General Education Teacher	School	Female	White	Bachelor's degree (e.g., B.A., B.S.), Master's degree (e.g., M.A., M.S.)	1 to 5 years	None	3	Medium	Rural	
Administrator	School	Male	White	Master's degree (e.g., M.A., M.S.)	11 to 15 years	16 to 20 years	15	Small	Rural	

Table 3-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 South Dakota Science Standards	School District Size	School District Area Urbanicity	Table Leader
Coach	School, District	Female	White	Master's degree (e.g., M.A., M.S.)	16 to 20 years	1 to 5 years	18 years	Large	Urban	Yes
General Education Teacher, Coach	School	Female	White	Bachelor's degree (e.g., B.A., B.S.)	11 to 15 years	None	2 time with state assessments	Medium	Rural	Yes
General Education Teacher	School	Female	White	Bachelor's degree (e.g., B.A., B.S.)	More than 20 years	None	More than 10 (Not sure when I first started with standards work)	Small	Rural	
General Education Teacher	School, District	Female	White	Bachelor's degree (e.g., B.A., B.S.), Master's degree (e.g., M.A., M.S.)	6 to 10 years	None	6	Medium	Urban	

Table 3-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 South Dakota 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.)	11 to 15 years	None	5	Medium	Urban	Yes
General Education Teacher	School, District	Male	White	Master's degree (e.g., M.A., M.S.)	More than 20 years	None	15	Large	Suburban	Yes
General Education Teacher	School	Female	White	Master's degree (e.g., M.A., M.S.)	More than 20 years	None	Since way back with Sam Shaw when these standards first came out	Small	Rural	
Higher Education	University	Female	White	Doctoral degree (e.g., Ph.D., Ed.)	More than 20 years	6 to 10 years	3	Medium	Suburban	
General Education Teacher	School, District	Female	White	Bachelor's degree (e.g., B.A., B.S.)	6 to 10 years	None	1 year	Small	Rural	

Appendix 3-B

South Dakota Science Assessment Range Achievement-Level Descriptors

South Dakota Science Assessment Range Achievement-Level Descriptors

Exhibit 3-B-1. South Dakota Science Assessment Range Achievement-Level Descriptors, Grade 5

Students that are a level _______may be able to do things like...	1	2	3	4
Earth and Space Sciences				
ESS1: Earth's Place in the Universe	Identify data, either in graphical displays or in a model, that would help 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.	Represent data in graphical displays and 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.	Analyze and interpret graphical displays of data to use as evidence in order to 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.	Evaluate and revise graphical displays of data to 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.
ESS2: 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 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 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 mathematical thinking, reasoning, and information regarding how the interactions between Earth's four major systems might cause patterned features of Earth, including climate, distribution of water, and physical and biological constructive and deconstructive forces.	Revise a model, analyze data sets from an investigation using mathematical thinking, and research how to better communicate or predict how the interactions between Earth's four major systems might cause patterned features of 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...	1	2	3	4
ESS3: Earth and Human Activity	Use information and observations from sources to identify either weather- related hazards that affect humans or human activity that affects Earth’s resources and environments.	Identify reliable sources and use obtained information to compare multiple solutions to help explain the cause and effect relationship of either weather-related hazards on humans or human activity on Earth’s resources and environments.	Obtain and use evidence from reliable sources to 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 Earth’s resources and environments.	Evaluate, compare, and revise a solution to a problem, using evidence obtained from reliable sources, to predict changes that can occur in the cause and effect relationships of either weather-related hazards on humans or human activity on Earth’s resources and environments.
Life Sciences				
LS1: From Molecules to Organisms: Structures 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 energy and materials to grow and repair their internal and external structures.	Develop and/or use a simple model to represent the life cycles or behavioral systems of organisms; and identify data as evidence to support an argument that organisms need food for energy and materials to grow and repair their internal and external structures.	Develop and/or use a model to describe patterns in the life cycles or behavioral systems of organisms; and use evidence to construct an argument that organisms need food for 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 energy and materials to grow and repair their internal and external structures.
LS2: 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 identify data that can show how an ecosystem changed.	Develop and/or use a simple model to describe the interactions of organisms within an ecosystem and the cycling of matter through those interactions; and collect evidence that shows how an ecosystem can change.	Develop and/or use a model to describe the interactions of organisms within an ecosystem and the cycling of matter through those interactions; and use evidence to explain the effects of a change in one part of the ecosystem.	Evaluate and revise a model that describes the interactions of organisms within an ecosystem and the cycling of matter through those interactions when more information is given; and predict the effects of a change in one part of the ecosystem.

Students that are a level _______may be able to do things like...	**1**	**2**	**3**	**4**
LS3: Heredity: Inheritance and Variation of Traits	Collect and record data from pictures, drawings, and/or text to help explain that organisms inherit the information that dictates how they look and function; and make an observation about an organism when its environment changes.	Use data collected from tables and various graphical displays to support an explanation that organisms inherit the information that dictates how they look and function; and identify information that would help explain what happens to an organism if the environment changes.	Analyze and interpret various forms of data to construct an explanation that organisms inherit the information that dictates how they look and function; and construct an explanation using evidence that supports that an organism has changed in response to environmental changes.	Construct, analyze, and interpret tables and graphical displays of data in order to construct and revise an explanation that organisms inherit the information that dictates how they look and function; and predict what would happen to an organism if its environment continues to change.
LS4: Biological Unity and Diversity	Identify patterns in past or present organism characteristics that can be used as evidence to support 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; and use observations from pictures, drawings, and/or writings to support that current, living organisms can only survive in particular environments or resemble organisms that once lived on Earth.	Demonstrate relationships in past and present organism characteristics that could either provide evidence 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; and identify data that can be used to compare the merits of a solution that can affect a population of organisms.	Analyze and interpret past and present organism characteristics to either provide evidence 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; and analyze and compare the merits of a solution that can affect a population of organisms.	Analyze and interpret past and present organism characteristics to 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; and compare sets of data to help argue the merits of a solution that could affect a population of organisms.
Physical Sciences				

Students that are a level _______may be able to do things like...	1	2	3	4
PS1: Matter and Its Interactions	Make observations about variables that are controlled to determine if a chemical reaction occurs and a new substance is created, measuring and graphing quantities to show that matter is always conserved regardless of the change that occurs; and use a model to show that matter is made of particles too small to be seen.	Organize and test variables that are controlled to determine if a chemical reaction occurs and a new substance is created, measuring and graphing quantities to show that matter is always conserved regardless of the change that occurs; and develop a simple model to show that matter is made of particles too small to be seen.	Plan and conduct an investigation in which variables are controlled to determine if a chemical reaction occurs and a new substance is created, measuring and graphing quantities to show that matter is always conserved regardless of the change that occurs; and develop a model to show that matter is made of particles too small to be seen.	Revise and conduct an investigation in which variables are controlled to 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; and evaluate and revise a model to show that matter is made of particles too small to be seen.
PS2: Motion and Stability: Forces and Interactions	Use questions and components of an investigation to observe the relationship between magnetism and/or gravity and an object's motion.	Use observations from an investigation to provide evidence to support an argument about cause and effect relationships between balanced and unbalanced forces (magnetism and/or gravity) and an object’s motion.	Ask questions, plan and conduct an investigation, and/or use produced data to provide evidence to create and support an argument about cause and effect relationships between balanced and unbalanced forces (magnetism and/or gravity) and an object’s motion.	Ask questions, conduct and compare two different investigations, and/or use produced data to provide evidence to predict cause and effect relationships between balanced and unbalanced forces (magnetism and/or gravity) and an object’s motion.
PS3: Energy	Ask questions based on observations about how energy can be used as a fuel or food or transferred from stored and/or motion energy to different forms like sound, light, and electrical currents.	Make observations using produced data to ask questions about how energy can be used as a fuel or food or transferred from stored and/or motion energy to different forms like sound, light, and electrical currents.	Use models to ask questions and/or use produced data to provide evidence on how energy can be used as a fuel or food or transferred from stored and/or motion energy to different forms like sound, light, and electrical currents.	Evaluate and revise models and/or use produced data to ask questions to make predictions or provide evidence for how energy can be used as a fuel or food or transferred from stored and/or motion energy to different forms like sound, light, and electrical currents.

Students that are a level _______may be able to do things like...	**1**	**2**	**3**	**4**
PS4: Waves and their Applications in Technologies for Information Transfer	Identify parts of a wave model; and identify observations that would help explain how reflected light from objects causes objects to be seen.	Develop and/or use a simple model to make observations about waves and the transfer of information; and record evidence that would help explain how reflected light from objects causes objects to be seen.	Create a solution or develop/and or use a model to describe and compare patterns of waves and the transfer of information; and use evidence to support an explanation for how reflected light from objects causes objects to be seen.	Revise a model to make predictions and compare patterns of waves and transfer of information; and use evidence to construct an explanation for how reflected light from objects causes objects to be seen.

Exhibit 3-B-2. South Dakota Science Assessment Range Achievement-Level Descriptors, Grade 8

Students that are a level__may be able to do things like…	**1**	**2**	**3**	**4**
Earth and Space Sciences				
ESS1: Earth's Place in the Universe	Identify components of a model that measures and collects evidence that explains the similarities and differences in the patterned motions of the Sun-Earth- Moon system, the scale of objects in the solar system, and the role of gravity in the motion of galaxies and the solar system.	Use a model or graphical display to identify data from tables and other graphical displays that can be used as pieces of evidence to describe the patterned motions of the Sun-Earth-Moon system, the scale of objects in the solar system, and the role of gravity in the motion of galaxies and the solar system.	Develop and use a model using graphical displays of data that can be used as pieces of evidence to explain the patterned motions of the Sun-Earth-Moon system, the scale of objects in the solar system, and the role of gravity in the motion of galaxies and the solar system.	Evaluate and revise a model based on constraints and data limitations that explain the patterned motions of the Sun-Earth-Moon system, the scale of objects in the solar system, and the role of gravity in the motion of galaxies and the solar system.

Students that are a level__may be able to do things like...	**1**	**2**	**3**	**4**
ESS2: Earth's Systems	Make measurements and/or observations from graphical data to help identify the components of a model that help explain 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; and identify evidence to explain that Earth’s processes have changed Earth’s surface at varying spatial and time scales.	Use a model or investigation to identify patterns from bar graphs, pictographs, and other graphical data that supports an explanation for how energy and matter flow or cycle throughout Earth’s systems, including the sun and Earth’s interior as primary energy sources; and organize evidence to explain how Earth’s processes have changed Earth’s surface at varying spatial and time scales.	Analyze data from an investigation to develop and use 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; and interpret evidence to construct an explanation that supports how Earth’s processes have changed Earth’s surface at varying spatial and time scales.	Evaluate and revise a model to 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; and evaluate the impact of new data by predicting how Earth’s processes will change Earth’s surface at varying spatial and time scales if a new variable is introduced.
ESS3: Earth and Human Activity	Identify scientific questions using collected and/or graphically represented evidence regarding the dependency of humans on the environment for different natural resources and identify evidence that can help design a simple solution that minimizes the effect of humans on the environment or explain the observed patterns that emerge between natural hazards and their related geological forces.	Ask questions to clarify evidence about data or apply scientific principles about the uneven distribution of natural resources and human dependence on those resources to design a simple solution that minimizes the effect of humans on the environment and explain the patterns in the history of natural hazards and their related geological forces.	Analyze and interpret sets of data regarding the uneven distribution of natural resources and human dependence on those resources to ask questions and design a solution that could minimize the effect of humans on the environment and explain the observable patterns seen in the data from the history of natural hazards and their related geological forces.	Evaluate sets of data regarding the uneven distribution of natural resources and human dependence on the environment for those resources to revise a question or modify a design solution that minimizes the effect of humans on the environment, revise an argument for the effect of humans on the environment, and predict future patterns of natural hazards when considering the impact of humans on the environment.

Students that are a level___may be able to do things like...	1	2	3	4
Life Sciences				
LS1: From Molecules to Organisms: Structures and Processes	Organize information from an investigation to identify components of a model or support an argument using evidence to explain that all living things are made up of cells that work together to form more complex structures and systems, that both plants and animals convert energy into food sources but the process to do so is different, and that characteristic animal behaviors and specialized plant structures affect the probability of reproduction.	Conduct an investigation to support an argument using evidence and use a model to explain that all living things are made up of cells that work together to form more complex structures and systems, that both plants and animals convert energy into food sources but the process to do so is different, and that characteristic animal behaviors and specialized plant structures affect the probability of reproduction.	Plan and conduct an investigation and synthesize data to construct an argument using evidence and develop and use a model to explain that all living things are made up of cells that work together to form more complex structures and systems, that both plants and animals convert energy into food sources but the process to do so is different, and that characteristic animal behaviors and specialized plant structures affect the probability of reproduction.	Evaluate and revise a model or explanation using investigative data as evidence to construct a revised argument that all living things are made up of cells that work together to form more complex structures and systems, that both plants and animals convert energy into food sources but the process to do so is different, and that characteristic animal behaviors and specialized plant structures affect the probability of reproduction.

Students that are a level__may be able to do things like...	**1**	**2**	**3**	**4**
LS2: Ecosystems: Interactions, Energy, and Dynamics	Identify components of a model to 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 biotic and abiotic components of an ecosystem, and organize multiple graphical displays of data to support a solution to mitigate disruptions to any part of an ecosystem by human access to natural resources.	Use a model to explain the dynamic relationships and interactions between the diverse types of living and nonliving parts of an ecosystem, including the flow of energy and cycling of matter among biotic and abiotic components, and organize data in multiple graphical displays to identify patterns which support a solution to mitigate disruptions to any part of an ecosystem by human access to natural resources.	Develop a model to 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 cycling of matter among biotic and abiotic components, and analyze and interpret multiple graphical displays of data to design and evaluate a solution to mitigate disruptions of any part of an ecosystem by human access to natural resources.	Analyze and/or revise a model that explains and supports 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 biotic and abiotic components when a variable in the system is changed, and evaluate limitations of data to propose a revised solution to mitigate disruptions to any part of an ecosystem by human access to natural resources.
LS3: Heredity: Inheritance and Variation of Traits	Identify the components of a model that describes the relationship among variables that show why sexual and asexual reproduction may have different results of genetic variation in offspring and how complex and microscopic structural changes to genes (mutations) can be used to determine how they affect the structure and function of an organism.	Use or manipulate a model to represent cause and effect relationships to describe why sexual and asexual reproduction may have different results of genetic variation in offspring and how complex and microscopic structural changes to genes (mutations) can be analyzed to determine how they affect the structure and function of an organism.	Develop and use a model to describe the relationship among variables that show why sexual and asexual reproduction may have different results of genetic variation in offspring and how complex and microscopic structural changes to genes (mutations) can be analyzed to determine how they affect the structure and function of an organism.	Evaluate and revise a model that explains the relationship among variables as to why sexual/asexual reproduction may have different results of genetic variation in offspring and predicts what changes would occur in the function of an organisms if there is a mutation in the organism’s genes.

Students that are a level__may be able to do things like...	**1**	**2**	**3**	**4**
LS4: Biological Unity and Diversity	Identify evidence in data sets to show that a species has changed over time and identify scientific ideas to support an explanation for how humans influence the biodiversity of an area and how natural or artificial selection can give some organisms an advantage in survival and reproduction.	Organize and identify the patterns in large data sets to explain how species can change over time, communicate the similarities or differences found in past and present organisms or fossil records of past environmental conditions, and gather and use data to construct an explanation for how humans influence the biodiversity of an area, and how natural or artificial selection can give some organisms an advantage in survival and reproduction.	Analyze and interpret the patterns in large data sets to explain how species can change over time, communicate the similarities or differences found in past and present organisms or fossil records of past environmental conditions, and gather and synthesize data using mathematical representations to construct an explanation for how humans influence the biodiversity of an area, and how natural or artificial selection can give some organisms an advantage in survival and reproduction.	Evaluate and revise 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 apply concepts of statistics and probability to form an explanation that as humans influence the biodiversity of an area, natural or artificial selection can give some organisms an advantage in survival and reproduction.
Physical Sciences				

Students that are a level__may be able to do things like...	**1**	**2**	**3**	**4**
PS1: Matter and Its Interactions	Identify the components of a model that explains the conservation of mass when two substances react; and identify data explaining that the properties of matter depend on its atomic and molecular composition and that particle motion changes when thermal energy in a system is changed.	Use a model to explain the conservation of mass when two substances react; and interpret data on the properties of matter to determine if a chemical reaction has occurred, such as the composition of atoms and molecules that make up matter and showing that particle motion changes when thermal energy in a system is changed.	Analyze patterns in graphical displays of data and develop and use a model to explain the conservation of mass when two substances react; and use the properties of matter to determine if a chemical reaction has occurred, including the composition of atoms and molecules that make up matter and showing that particle motion changes when thermal energy in a system is changed.	Evaluate and revise a model to explain the conservation of mass when two substances react; and use evidence to predict how changes to the molecular structure or thermal energy of matter can affect its properties.
PS2: Motion and Stability: Forces and Interactions	Identify components of an investigation, and identify data regarding the relationships between mass, force, and motion, and the attractive and repulsive forces that act at a distance (electric, magnetic, and gravitational forces) that could be used to support a claim.	Identify questions, conduct an investigation, and organize and use data to make a claim regarding the relationships 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 relationships between mass, force, and motion, and the attractive and repulsive forces that act at a distance (electric, magnetic, and gravitational forces).	Evaluate and revise an investigation and analyze and evaluate data to predict and support a claim regarding the relationships 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...	1	2	3	4
PS3: 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 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.	Use a model to describe that kinetic and potential energy interact, transform, or transfer to another object; and interpret data regarding the temperature and total energy of a system and its dependence on a variety of factors, including the types and states of energy, as well as the amount of matter involved to support an argument.	Develop a model or investigation to construct an argument to support a claim about how kinetic and potential energy interact, transform, or transfer to another object; and analyze data from an investigation to provide evidence that the temperature and total energy of a system is dependent on a variety of factors, including the types and states of energy, as well as the amount of matter involved.	Evaluate and/or revise a model to predict changes to the interaction of kinetic and potential energy, including how energy is transformed or transferred to another object; and apply concepts of statistics and probability to construct an argument that the temperature and total energy of a system is dependent on a variety of factors, including the types and states of matter and the amount of matter involved.
PS4: Waves and their Applications in Technologies for Information Transfer	Identify the mathematical components in a model to describe the patterns observed between wave characteristics and wave energy; and identify a claim with evidence to show that waves are reflected, absorbed, or transmitted through various materials.	Use mathematical representations in a model to describe the patterns observed between wave characteristics and wave energy; and support a claim with evidence to show that waves are reflected, absorbed, or transmitted through various materials.	Develop and use mathematical representations in a model to describe the patterns observed between wave characteristics and wave energy; and construct and evaluate a claim supported by evidence to show that waves are reflected, absorbed, or transmitted through various materials.	Evaluate and revise a mathematical model to predict patterns between wave characteristics and wave energy; and integrate qualitative, quantitative, and technical data to provide evidence to support a claim that waves are reflected, absorbed, or transmitted through various materials.

Exhibit 3-B-3. South Dakota Science Assessment Range Achievement-Level Descriptors, Grade 11

Students that are a level___may be able to do things like...	**1**	**2**	**3**	**4**
Earth and Space Sciences				
ESS1: Earth's Place in the Universe	Identify components and limitations of a model that uses mathematical representations to explain the characteristics, processes, and life cycles of objects in the solar system; **and** identify and critique evidence that shows the motion of objects in our solar system and Earth’s early formation and geologic history.	Use existing mathematical concepts and processes to explain algorithms and models that explain the characteristics, processes, and life cycles of objects in the solar system; **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.	Develop and/or use mathematical models to collect data and explain the characteristics, processes, and life cycles of objects in the solar system; **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.	Evaluate and revise a mathematical model to make predictions regarding the characteristics, processes, and life cycles of objects in the solar system; **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.
ESS2: Earth's Systems	Identify components and limitations of a model or investigation 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 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.	Develop and/or use a model to generate and use data from an investigation to analyze and use as 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.	Evaluate and/or revise an investigation or model to predict changes that can occur to Earth’s feedback mechanisms when a variable is either added or changed; **and** analyze the collected data to predict how energy flow into or out of Earth systems will affect other Earth systems, specifically with the planet’s interactions with water, solar radiation, geologic systems, and climate.

Students that are a level___may be able to do things like...	**1**	**2**	**3**	**4**
ESS3: Earth and Human Activity	Identify and construct graphical displays of data that can be used to explain how human activity has been influenced by the availability of natural resources, natural hazards, and climate change; **and** use mathematical representations and/or algorithms to identify the impact of climate change on Earth’s systems and human society and how human society has impacted Earth's systems.	Use data from graphical displays to support a claim that human activity has been influenced by the availability of natural resources, natural hazards, and climate change; **and** use a computational simulation or model to identify the rate of climate change and its impact on Earth’s systems and human society to observe relationships for how human society has impacted Earth's systems.	Evaluate data and construct an explanation for how human activity has been influenced by the availability of natural resources, natural hazards, and climate change; **and** mathematically analyze information from natural resource data with a computational simulation or representation of climate models to predict the rate of climate change and its impact on Earth’s systems and human society to illustrate relationships for how human society has impacted Earth's systems.	Use mathematical thinking to evaluate and/or revise an explanation for how human activity has been influenced by the availability of natural resources, natural hazards, and climate change; **and** create a computational simulation or representation of natural resource data and climate models to use relationships to predict the rate of climate change and its impact on Earth’s systems and human society and how human society has impacted Earth's systems.
Life Sciences				

Students that are a level___may be able to do things like...	**1**	**2**	**3**	**4**
LS1: From Molecules to Organisms: Structures 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 respiration moves energy and matter through the body, forming different products, transferring energy, and replicating DNA and synthesizing proteins.	Conduct an investigation to 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; **and** use collected data to support a claim regarding how cellular respiration moves energy and matter through the body, forming different products, transferring energy, and replicating DNA and synthesizing proteins.	Plan and conduct an investigation and develop and use a model to show that feedback mechanisms maintain homeostasis through the structure, function, and processes of living systems; **and** evaluate data from an investigation to construct an explanation for how cellular respiration moves energy and matter through the body, forming different products, transferring energy, and replicating DNA and synthesizing proteins.	Plan and conduct an investigation and evaluate and revise a model to explain what happens to the feedback mechanisms that maintain homeostasis through the structure, function, and processes of living systems when a variable is changed; **and** apply scientific reasoning, theory and/or models to make and support a claim that cellular respiration moves energy and matter through the body, forming different products, transferring energy, and replicating DNA and synthesizing proteins.

Students that are a level___may be able to do things like...	**1**	**2**	**3**	**4**
LS2: Ecosystems: Interactions, Energy, and Dynamics	Use mathematical representations to identify components or variables in the cycling and flow of matter and energy among organisms in an ecosystem; **and** identify evidence that supports the interactions with biotic and abiotic factors in ecosystems help maintain the population and diversity of organisms.	Use mathematical representations to construct an explanation with data that shows how energy and matter flow and cycle among organisms in an ecosystem; evaluate and identify patterns seen in data that can be used as evidence to explain the interactions of biotic and abiotic factors in maintaining the population and diversity of organisms in an ecosystem; **and** identify biological, physical, or human induced disturbances in conditions that may result in a new ecosystem.	Create and/or use mathematical, computational, and algorithmic representations to support claims about the cycling of matter and flow of energy among organisms in an ecosystem; **and** use evidence and reasoning to construct an explanation for how interactions with biotic and abiotic factors in ecosystems maintain the population and diversity of organisms, but that biological, physical, or human induced disturbances in conditions may result in a new ecosystem.	Evaluate and revise a computational model or simulation that can explain that the cycling of matter and flow of energy among organisms in an ecosystem can be disturbed when a new variable is introduced; use mathematical and computational evidence to argue that interactions with biotic and abiotic factors in ecosystems maintain the population and diversity of organisms; **and** predict how an ecosystem might change with a biological, physical, or human induced disturbance in conditions.
LS3: Heredity: Inheritance and Variation of Traits	Identify an observation or model of DNA, chromosomes, and traits; and use graphical displays of data to identify evidence that supports a claim about genetic and environmental factors that may affect the variation and distribution of traits in a population.	Ask a question that requires sufficient, empirical evidence to answer regarding the relationships between DNA, chromosomes, and traits; **and** analyze data to support a claim defending an argument about genetic and environmental factors and their effect on variation within a population.	Analyze a model or theory and ask questions to determine the relationships between the roles of DNA, chromosomes, and traits; **and** apply concepts of statistics and probability when analyzing evidence to make and defend a claim about genetic and environmental factors that may affect the variation and distribution of traits in a population.	Use a question to analyze and evaluate the relationships between the roles of DNA, chromosomes, and traits; **and** apply concepts of statistics and probability when analyzing evidence to predict the variation and distribution of traits in a population when a genetic and environmental factor is changed.

Students that are a level__may be able to do things like…	**1**	**2**	**3**	**4**
LS4: Biological Unity and Diversity	Identify and use genetic and anatomical evidence obtained from texts and mathematical representations to support that the evolution, extinction, and formation of new species is based on different environmental factors; **and** identify causal and correlational relationships of environmental conditions and population adaptations.	Construct and/or use graphical displays of data to provide genetic and anatomical evidence for how given factors have resulted in diversity through evolution, extinction, and formation of new species; **and** analyze data to distinguish between causal and correlational relationships to support that environmental conditions can lead to adaptations within populations.	Use genetic and anatomical information obtained from texts, mathematical, computational, and/or algorithmic representations to construct an explanation for how given factors have resulted in diversity through evolution, extinction, and formation of new species; **and** generate and analyze mathematical data to support the argument that environmental conditions can lead to adaptations within populations.	Use genetic and anatomical information obtained from texts, mathematical, computational and/or algorithmic representations to evaluate and revise an explanation and predict what would happen to a current species when a given factor is changed; **and** use the generated data to support a prediction of the adaptations a population may experience when environmental conditions are changed.
Physical Sciences				
PS1: Matter and Its Interactions	Recognize the patterns in the periodic table and identify variables that provides an explanation for the properties and characteristics of matter; **and** apply mathematical concepts to an investigation that produces data to identify evidence for an explanation that any chemical process that occurs between matter is due to the collision of molecules, changes in energy, and the atomic configuration of the elements involved.	Use the periodic table to develop a model of atomic structure to support an explanation for the properties and characteristics of matter; **and** collect data from an investigation that can be analyzed for patterned evidence to support the claim that any chemical process that occurs between matter is due to the collision of molecules, changes in energy, and the atomic configuration of the elements involved.	Use the periodic table, atomic structures, and corresponding electrical interactions to construct an investigation and/or mathematical model that explains the properties and characteristics of matter; **and** provide quantitative and qualitative evidence that any chemical processes that occur between matter are due to the collision of molecules, changes in energy, and the atomic configuration of the elements involved.	Use the periodic table, atomic structures, and corresponding electrical interactions to evaluate and/or revise a mathematical model or investigation that predicts the properties and characteristics of matter when a component is changed; **and** construct and/or revise an explanation that any chemical processes that occur between matter are due to the collision of molecules, changes in energy and the atomic configuration of elements.

Students that are a level___may be able to do things like...	**1**	**2**	**3**	**4**
PS2: Motion and Stability: Forces and Interactions	Use mathematical concepts and processes to help 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; **and** interpret graphical displays of data to identify evidence that supports how an object moves.	Collect and/or produce data to distinguish between causal and correlational relationships between force and the distance between interacting objects or force, mass, and acceleration; **and** use mathematical and graphical representations to describe the motion of an object.	Plan and conduct an investigation to collect data 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** use mathematical, graphical, and computational analysis to observe patterns to explain changes in the motion of an object.	Evaluate and revise an investigation, or predict changes to an investigative outcome, when a variable is changed when modeling the relationship between either force and the distance between interacting objects or force, mass, and acceleration; **and** use scientific ideas, principles and/or evidence to revise an explanation and predict changes in the motion of an object when new information is introduced.
PS3: Energy	Identify components and variables of an investigation to describe how energy transfers within and between systems; **and** develop and/or use a model to identify evidence that energy is neither created nor destroyed but converted to less useful forms.	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** develop and/or use a model to support that energy is neither created nor destroyed but converted into less useful forms.	Develop and/or use a mathematical model, using collected or produced data from an investigation, to describe how energy transfers within and between systems; **and** provide empirical data supporting that energy is neither created nor destroyed but converted to less useful forms.	Evaluate and revise a mathematical model, using scientific ideas, principles, theories and/or newly added information or data, to predict how energy transfers within and between systems; **and** apply empirical data to generate quantitative evidence supporting that energy is neither created nor destroyed but converted to less useful forms.

Students that are a level__may be able to do things like...	**1**	**2**	**3**	**4**
PS4: Waves and their Applications in Technologies for Information Transfer	Integrate qualitative and quantitative information to identify data that shows the relationships among wavelength, amplitude, frequency, and other wave features; **and** use mathematical representations to identify components of energy transfer by waves.	Collect and use quantitative data, hypotheses and/or conclusions to collect and use evidence that shows the relationships among wavelength, amplitude, frequency, and other wave features; **and** use mathematics and algorithmic thinking to describe energy transfer by waves.	Analyze technical science information to evaluate a claim regarding the relationships among wavelength, amplitude, frequency, and other wave features; **and** create and/or use computational models to explain how energy transfers and how a wave medium affects the wave.	Evaluate models and technical science information to provide evidence of the relationships among wavelength, amplitude, frequency, and other wave features; **and** use mathematical, computational and/or algorithmic produced data to predict how a change in wave medium would affect a wave.

Appendix 3-C

Standard-Setting Assertion Maps

Standard-Setting Assertion Maps

Exhibit 3-C-1. Standard-Setting Assertion Map, Science Grade 5

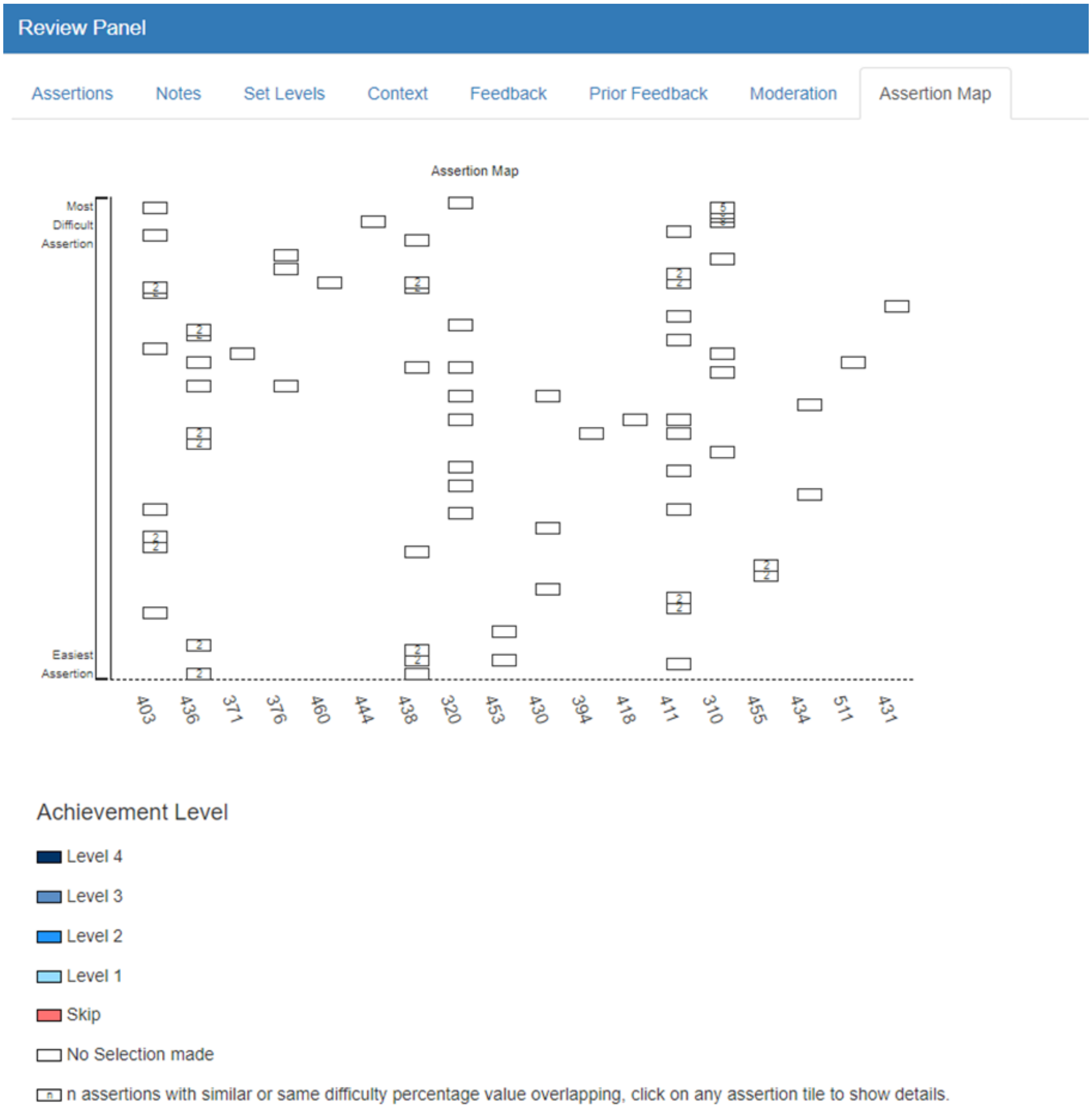


Exhibit 3-C-2. Standard-Setting Assertion Map, Science Grade 8

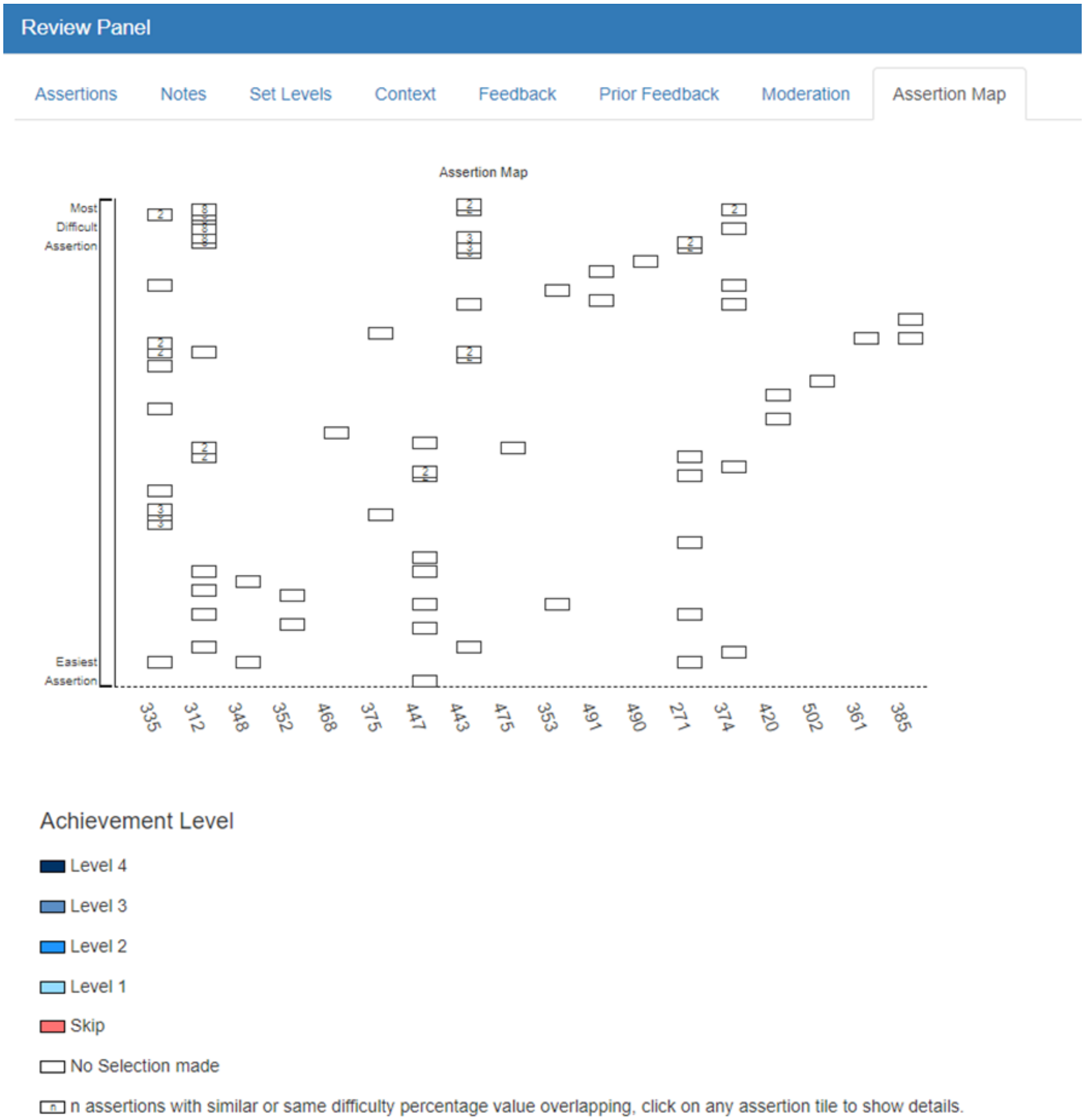
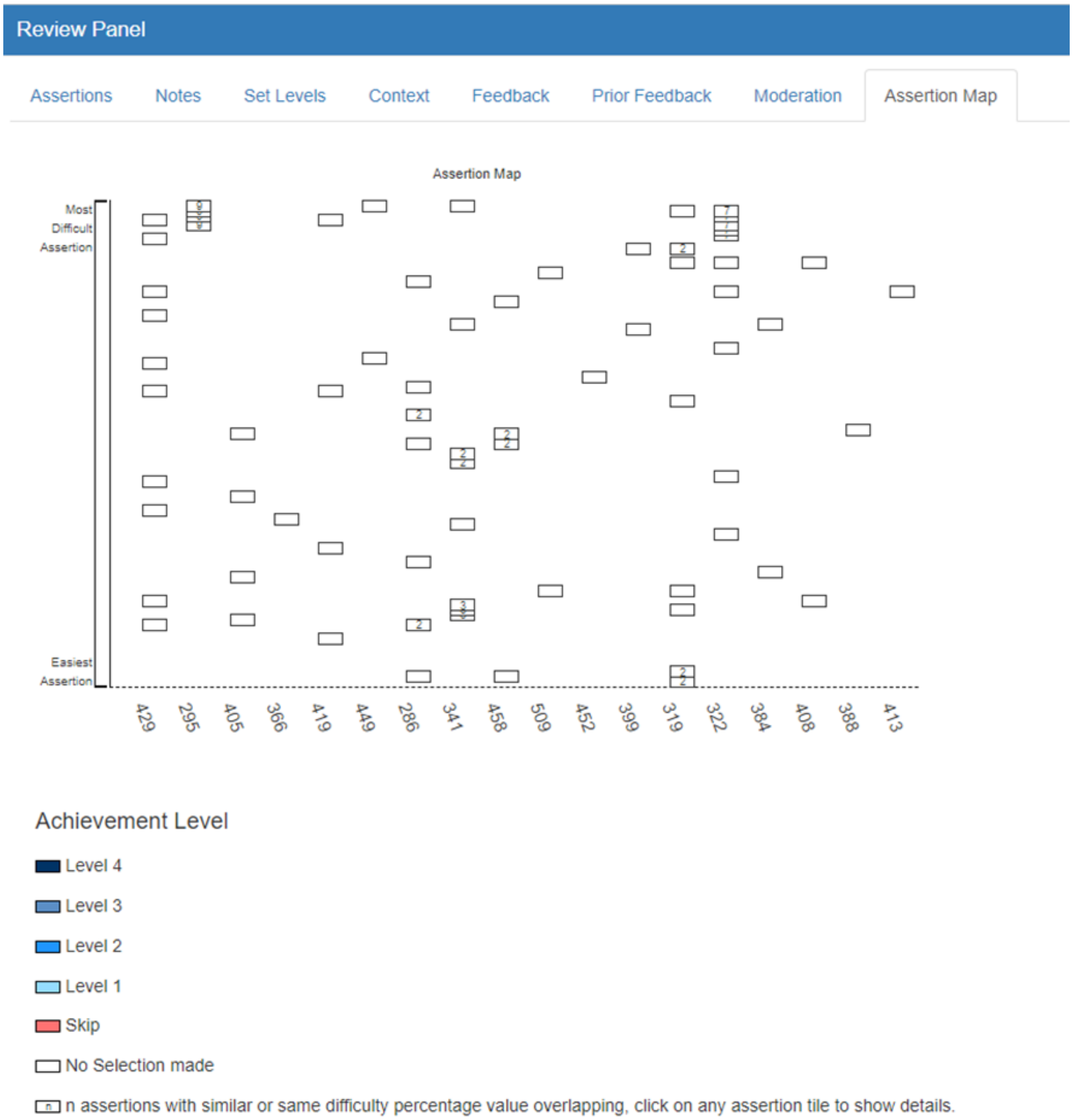


Exhibit 3-C-3. Standard-Setting Assertion Map, Science Grade 11



Appendix 3-D

Standard-Setting Workshop Agenda

Standard-Setting Workshop Agenda

Exhibit 3-D-1. Day 1 Standard-Setting Workshop Agenda

south dakota
DEPARTMENT OF EDUCATION
Learning. Leadership. Service.

2021 Standard Setting for the South Dakota Science Assessment (SDSA)

SCIENCE PANEL AGENDA

September 15 – 16, 2021

**Cameras must remain on throughout the meeting. Notes must be destroyed.*

Standard-Setting Workshop Day 1 – Wednesday, September 15, 2021 Day 1 Meeting Times: 8 a.m. – 5 p.m. CDT | 7 a.m. – 4 p.m. MDT | 9 a.m. – 6 p.m. EDT	
8:00 – 8:30 a.m. CDT	**Participant Login**
8:30 – 8:45 a.m. CDT	**Welcome and Introductions from the South Dakota Department of Education (SDDOE)**
8:45 – 9:30 a.m. CDT	**Large-Group Orientation** Welcome and introductions Purpose of standard-setting workshop General overview of standard-setting procedures and key concepts • Achievement-level descriptors (ALDs) • 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. CDT	**Break, and Separate into Small Group Rooms**
9:45 – 11:15 a.m. CDT	**Panelists Experience Online Operational Assessment and Test Environment**
11:15 – 12:15 p.m. CDT	**Review Range ALDs and Discuss Threshold ALDs** Parse range ALDs to identify specific claims within achievement levels Identify knowledge and skills differentiating student performance between levels
12:15 – 1:00 p.m. CDT	**Lunch (on your own)**
1:00 – 2:00 p.m. CDT	**Continue Discussions of ALDs**
2:00 – 5:00 p.m. CDT	**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

Cambium Assessment, Inc. 1

Standard-Setting Workshop Day 1 – Wednesday, September 15, 2021	
Day 1 Meeting Times: 8 a.m. – 5 p.m. CDT | 7 a.m. – 4 p.m. MDT | 9 a.m. – 6 p.m. EDT	
	• 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 ALDs? 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. CDT	**Adjourn**

Exhibit 3-D-2. Day 2 Standard-Setting Workshop Agenda

*Cameras must remain on throughout the meeting. Notes must be destroyed.

Standard-Setting Workshop Day 2 – Thursday, September 16, 2021

Day 1 Meeting Times: 8:30 a.m. – 5 p.m. CDT | 7:30 a.m. – 4 p.m. MDT | 9:30 a.m. – 6 p.m. EDT

8:30 – 10:00 a.m. CDT	**Continue Review of OSAB Items**
10:00 – 11:00 a.m. CDT	**Training on Assertion-Mapping Task** Review of assertion-mapping key concepts • Achievement-level descriptors (ALDs) • Ordered scoring assertions • Assertion map Training on assertion-mapping tool Practice assertion-mapping task and standard-setting quiz
11:00 – 11:15 a.m. CDT	**Break**
11:15 – 12:30 p.m. CDT	**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. CDT	**Lunch (on your own)**
1:30 – 2:30 p.m. CDT	**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 Training on usage of contextual information – benchmark and impact data
2:30 – 3:30 p.m. CDT	**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. CDT	**Workshop Evaluations and Educator Panel Adjourn**
4:00 – 5:00 p.m. CDT	**Across Grade Moderation with All Science Table Leaders**
5:00 p.m. CDT	**Table Leader Adjourn**

Appendix 3-E

Standard-Setting Training Slides

Exhibit 3-E-1. Large-Group Orientation Slides

Cambium Assessment

south dakota
DEPARTMENT OF EDUCATION
Learning. Leadership. Service.

Standard Setting: Science

September 15 – 16, 2021
South Dakota Science Assessment (SDSA)

2

Welcome and Introductions

South Dakota Department of Education

State Education Representatives

3

- **South Dakota Department of Education (SDDOE)**
  - Matthew Gill, Office of Assessment Administrator
  - Chris Booth, Program Specialist
  - Beth Schiltz, Special Education Program Specialist
  - Jennifer Fowler, Science Specialist

4

Large-Group Orientation

Cambium Assessment, Inc.

Cambium Assessment

south dakota
DEPARTMENT OF EDUCATION
Learning. Leadership. Service.

Workshop Leaders

5

- **Cambium Assessment, Inc.**
  - Psychometrics
    - Stephan Ahadi
    - Frank Rijmen
    - Widad Abdalla
  - Room Facilitators
    - Grade 5: Jim McCann and Anneka Wiersma
    - Grade 8: Kevin Dwyer and Vanessa Johnson
    - Grade 11: Matt Davis and Kam Mangis de Mark

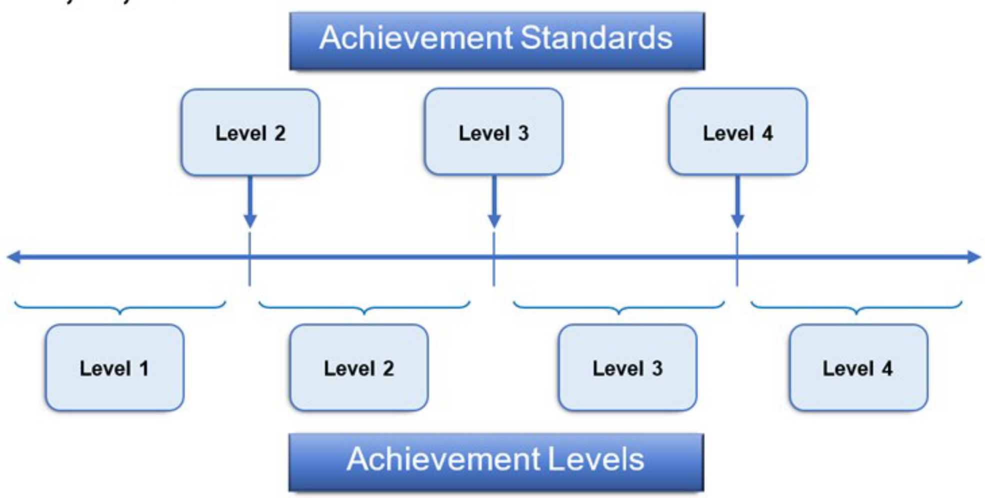
Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Purpose of the Standard-Setting Workshop

6

- Recommend to the South Dakota State Board of Education three achievement standards to differentiate the four achievement levels on the South Dakota Science Assessment in grades 5, 8, and 11



Cambium Assessment

south dakota DEPARTMENT OF EDUCATION
Learning. Leadership. Service.

Main Workshop Activities

7

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

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Importance of Security

8

- Cameras are required for participants
- 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

south dakota DEPARTMENT OF EDUCATION
Learning. Leadership. Service.

Reason for New Science Standards

9

- The South Dakota State Board of Education adopted the new South Dakota Science Standards in May 2015
- New science assessments, aligned with the South Dakota Science Standards, were developed and administered to grade 5, 8, and 11 students in South Dakota in spring 2021

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Description of the Science Test Design

10

- Grades 5, 8, and 11 tests assess students' understanding of the South Dakota Science Standards
- The SDSA at grades 5, 8, and 11 includes 6 item clusters and 12 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

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

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

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

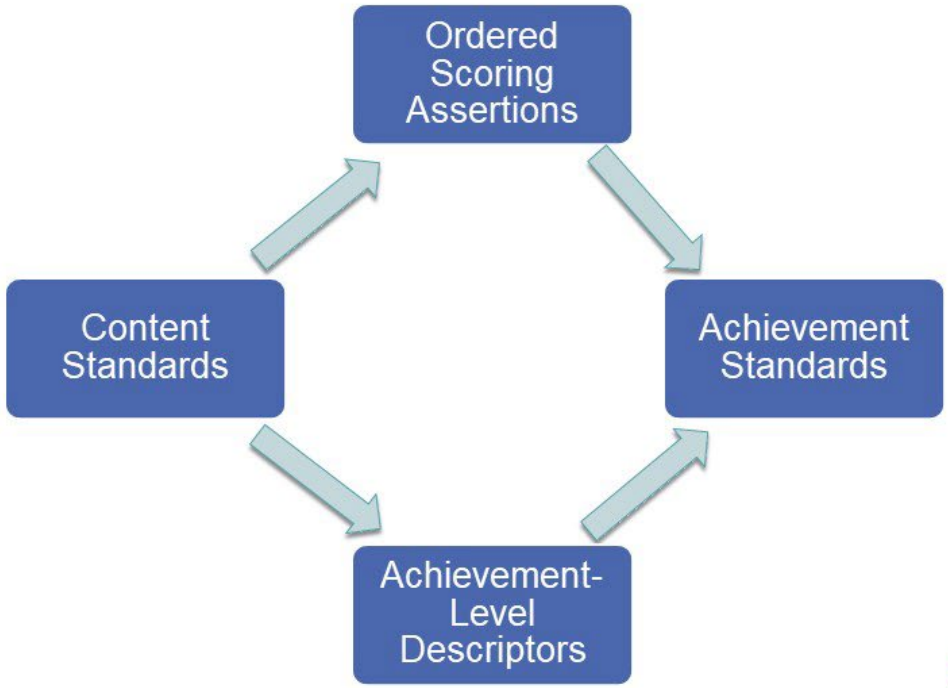
Standard Setting

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 achievement level on the test

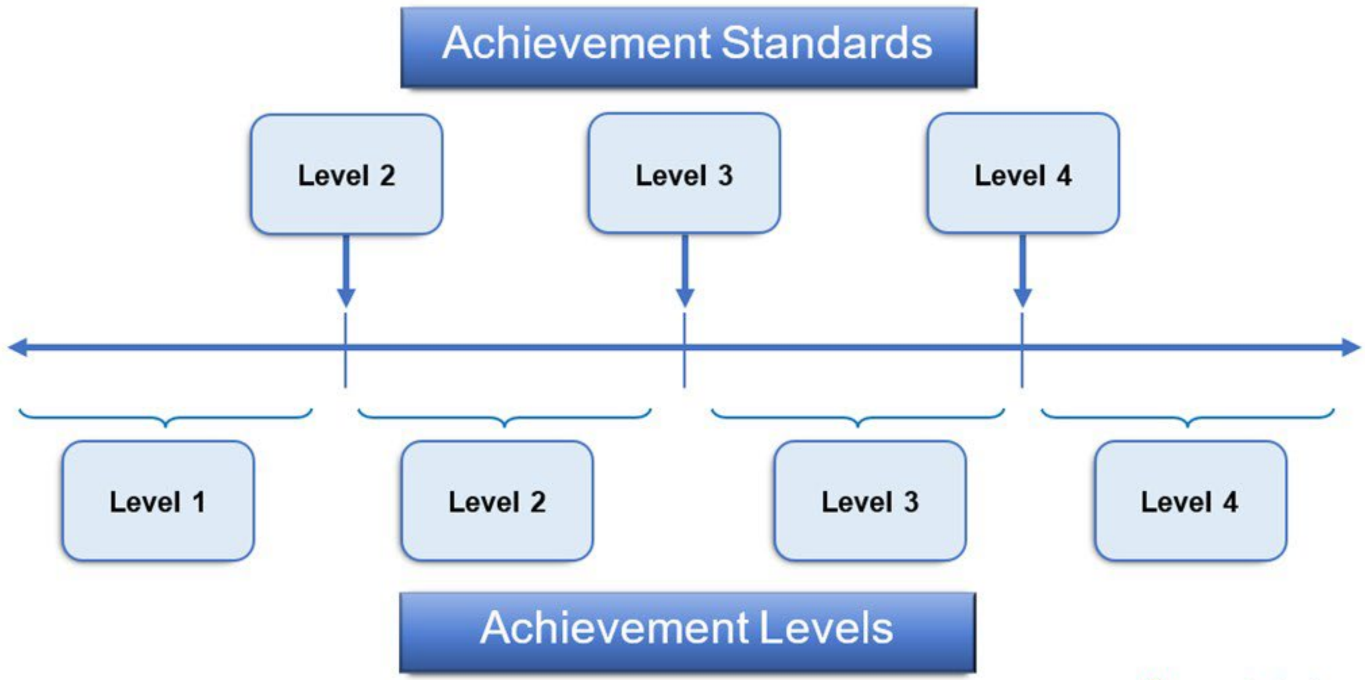
From Content Standards to Achievement Standards

13



Achievement Standards and Achievement 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 achievement levels
- Is being employed to recommend achievement standards in multiple states assessing three-dimensional science standards

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION
Learning. Leadership. Service.

Key Elements of the AMP

16

- Achievement-level descriptors (ALDs)
  - Range ALDs
  - Threshold ALDs (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

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION
Learning. Leadership. Service.

Achievement-Level Descriptors (ALDs)

17

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

Grade 8 Range ALDs – Level 3

18

Physical Sciences

- **MS-PS1:** Analyze patterns in graphical displays of data and develop and/or use a model to explain the conservation of mass when two substances react…
- **MS-PS2:** Ask questions, plan and conduct an investigation, and analyze and interpret data to make and support a claim regarding the relationships between mass, force, and motion, and the attractive and repulsive forces that act at a distance (electric, magnetic, and gravitational forces).
- **MS-PS3:** Develop and/or use a model or investigation to construct an argument to support a claim about how kinetic and potential energy interact, transform, or transfer to another object…
- **MS-PS4:** Develop and/or use mathematical representations in a model to describe the patterns observed between wave characteristics and wave energy…

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Grade 8 Range ALDs Across Achievement Levels

19

MS-PS2 Motion and Stability: Forces and Interactions

- **Level 1: Identify components of an investigation, and identify data** regarding the relationships between mass, force, and motion, and the attractive and repulsive forces that act at a distance (electric, magnetic, and gravitational forces) that could be used to **support a claim.**
- **Level 2: Identify questions, conduct an investigation, and organize and use data to make a claim** regarding the relationships between mass, force, and motion, and the attractive and repulsive forces that act at a distance (electric, magnetic, and gravitational forces).
- **Level 3: Ask questions, plan and conduct an investigation, and analyze and interpret data to make and support a claim** regarding the relationships between mass, force, and motion, and the attractive and repulsive forces that act at a distance (electric, magnetic, and gravitational forces).
- **Level 4: Ask questions to conduct, evaluate, and revise an investigation; and analyze and evaluate data to predict and support a claim** regarding the relationships between mass, force, and motion, and the attractive and repulsive forces that act at a distance (electric, magnetic, and gravitational forces).

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Important Concepts

20

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

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Ordered Scoring Assertions

21

- 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 ALDs

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

What If an Assertion Seems Out of Order?

22

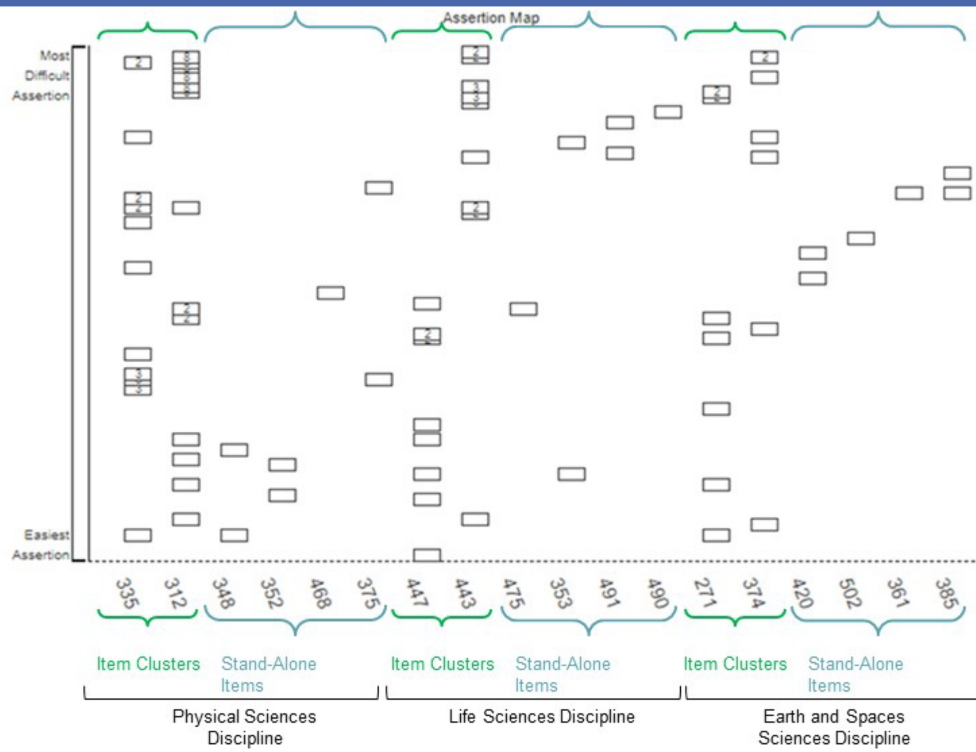
- Assertion ordering is based on student performance
- Assertions may seem out of order because they are ordered by difficulty, and not by content or cognitive process
- Identify why a scoring assertion is more difficult than the assertions before it, and easier than the assertions following it
  - Pay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

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Assertion-Mapping Task

23



Cambium Assessment

south dakota DEPARTMENT OF EDUCATION
Learning. Leadership. Service.

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 ALDs?*
- Working as a group
  - Discuss how item interactions support scoring assertions
  - Discuss ordering of scoring assertions
  - Discuss how scoring assertions are related to the ALDs

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

What If an Item Seems Wrong or Unfair?

25

- 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

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

"Just Barely" Meets the Achievement Standard

26

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

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Threshold “Just Barely” ALDs

27

Assertion-Mapping Task

28

- Map assertions to achievement levels
  - Consider what differentiates students who just barely qualify for entry into the achievement level from those not quite ready for entry into the achievement level
  - Evidence that the student has demonstrated knowledge and skills necessary for entry into the achievement level
- Map assertions in the online standard-setting tool

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Group Feedback and Discussion

29

- 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

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION
Learning. Leadership. Service.

Feedback and Impact Data

30

- 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 achievement standard seem reasonable?
  - What are the implications for the achievement standards?
  - All achievement standard recommendations should be based on content rationales

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION
Learning. Leadership. Service.

Creating a System of Achievement Standards

31

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

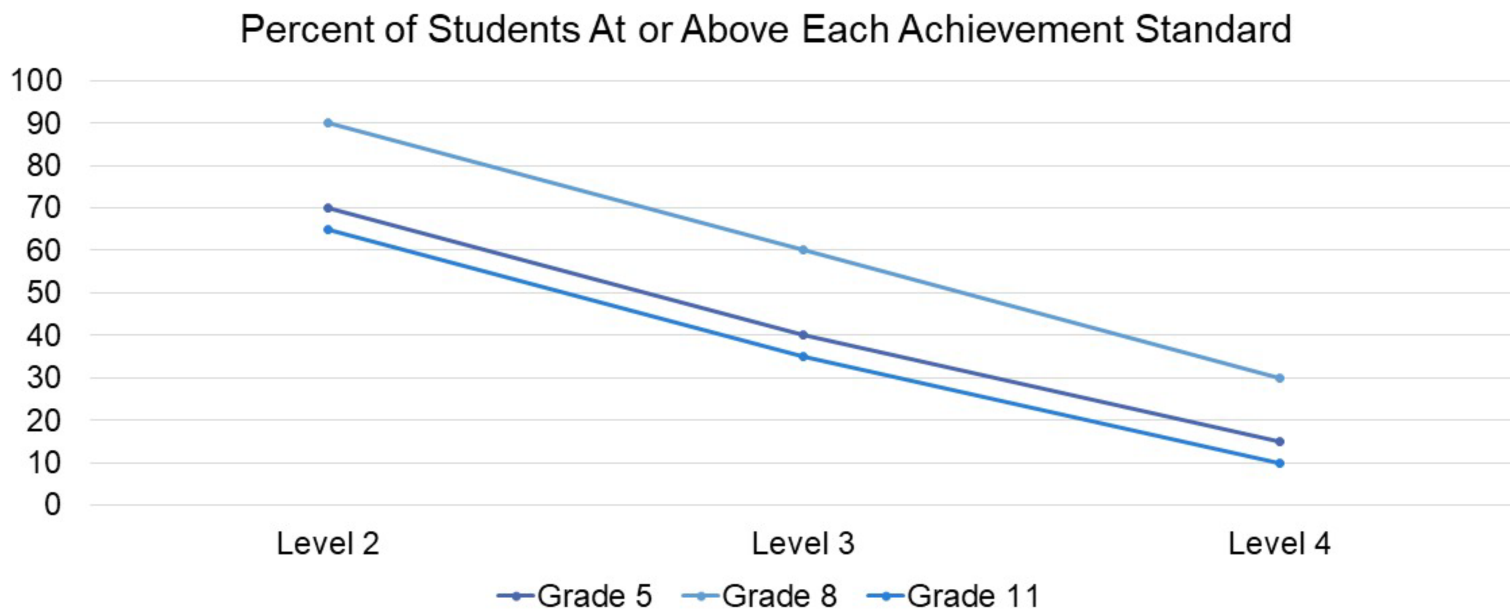
Benchmarking

32

- Are achievement standards nationally competitive and represent on track for college readiness?
  - Smarter Balanced ELA and Mathematics
- Achievement levels for benchmark assessments will provide context about the general neighborhood in which achievement standards likely reside

Articulation

33



Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Moderation

34

- 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 achievement standards (assuming there is a good content reason for doing so)

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Break Into Groups

35

Panel	Facilitators
Grade 5 Science	Jim McCann Anneka Wiersma
Grade 8 Science	Kevin Dwyer Vanessa Johnson
Grade 11 Science	Matt Davis Kam Mangis de Mark

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION
Learning. Leadership. Service.

Exhibit 3-E-2. Breakout Room Slides

Standard Setting: Science

September 15 – 16, 2021
South Dakota Science Assessment

2

Standard-Setting Workshop Day 1

Recommending Achievement Standards for Grade 5 Science

Welcome!

3

- Introductions
- Housekeeping
  - Please stay on camera unless we are at lunch or on a break.
  - Let us know if you need to step away from the meeting.

Standard-Setting Workshop Day 1 Agenda

4

- Experience Online Operational Assessment and Test Environment
- Review Range ALDs and Discuss Threshold ALDs
- Review Ordered Scoring Assertion Booklet (OSAB)

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION
Learning. Leadership. Service.

5

Operational Test Review

Description of the Science Test Design

6

- Grades 5, 8, and 11 tests assess students' understanding of the South Dakota Science Standards
- The SDSA at grades 5, 8, and 11 includes 6 item clusters and 12 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

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

SDSA Grade 5 Blueprint

7

Discipline	DCI	Standards Assessed		Percentage
Physical Science	PS1	5-PS1-1 5-PS1-2	5-PS1-3 5-PS1-4	33%
	PS2	3-PS2-1 3-PS2-2 3-PS2-3	3-PS2-4 5-PS2-1	
	PS3	4-PS3-1 4-PS3-2 4-PS3-3	4-PS3-4 5-PS3-1	
	PS4	4-PS4-1 4-PS4-2	4-PS4-3	
Life Science	LS1	3-LS1-1 4-LS1-1	4-LS1-2 5-LS1-1	33%
	LS2	3-LS2-1	5-LS2-1	
	LS3	3-LS3-1	3-LS3-2	
	LS4	3-LS4-1 3-LS4-2	3-LS4-3 3-LS4-4	
Earth and Space Science	ESS1	4-ESS1-1 5-ESS1-1	5-ESS1-2	33%
	ESS2	3-ESS2-1 3-ESS2-2 4-ESS2-1	4-ESS2-2 5-ESS2-1 5-ESS2-2	
	ESS3	3-ESS3-1 4-ESS3-2	4-ESS3-1 5-ESS3-1	

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Review of 3D Science Standards

8

- 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).

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION
Learning. Leadership. Service.

Review of Items – 3D Composition

9

The Core Ideas of the Third Grade standards include:

- Motion and Stability: Forces and Interactions
- From Molecules to Organisms: Structures and Processes
- Ecosystems: Interactions, Energy, and Dynamics
- Heredity: Inheritance and Variation of Traits
- Biological Unity and Diversity
- Earth's Systems
- Earth and Human Activity

The Core Ideas of the Fourth Grade standards include:

- Energy
- Waves and Their Applications in Technologies for Information Transfer
- From Molecules to Organisms: Structures and Processes
- Earth's Place in the Universe
- Earth's Systems
- Earth and Human Activity

The Core Ideas of the Fifth Grade standards include:

- Matter and Its Interactions
- Motion and Stability: Forces and Interactions
- Energy
- From Molecules to Organisms: Structures and Processes
- Ecosystems: Interactions, Energy, and Dynamics
- Earth's Place in the Universe
- Earth's Systems
- Earth and Human Activity

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Review of Items – 3D Composition

10

Fifth Grade Physical Science Conceptual Understanding:

Because matter exists as particles that are too small to see, matter is always conserved even if it seems to disappear. Measurements of a variety of observable properties can be used to identify particular materials. Chemical reactions that occur when substances are mixed can be identified by the emergence of substances with different properties; the total mass of substances when a reaction occurs remains the same. Energy can be “produced,” “used,” or “released” by converting stored energy. Plants capture energy from sunlight, which can later be used as fuel or food.

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Review of Items – 3D Composition

11

Fifth Grade Physical Science Standards	
5-PS1-1	Develop a model to describe that matter is made of particles too small to be seen. (SEP: 2; DCI: PS1.A; CCC: Scale/Prop.)
5-PS1-2	Measure and graph quantities to provide evidence that regardless of the type of change that occurs when heating, cooling, or mixing substances, the total weight of matter is conserved. (SEP: 5; DCI: PS1.A, PS1.B; CCC: Scale/Prop.)
5-PS1-3	Make observations and measurements to identify materials based on their properties. (SEP: 3; DCI: PS1.A; CCC: Scale/Prop.)
5-PS1-4	Conduct an investigation to determine whether the mixing of two or more substances results in new substances. (SEP: 3; DCI: PS1.B; CCC: Cause/Effect)
5-PS2-1	Support an argument that the gravitational force exerted by Earth on objects is directed down. (SEP: 7; DCI: PS2.B; CCC: Cause/Effect)
5-PS3-1	Use models to describe that energy in animals' food (used for body repair, growth, motion, and to maintain body warmth) was once energy from the sun. (SEP: 2; DCI: PS3.D, LSI.C ; CCC: Energy/Matter)

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

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

south dakota DEPARTMENT OF EDUCATION
Learning. Leadership. Service.

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

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

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

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Review of Science Item Clusters – Composition

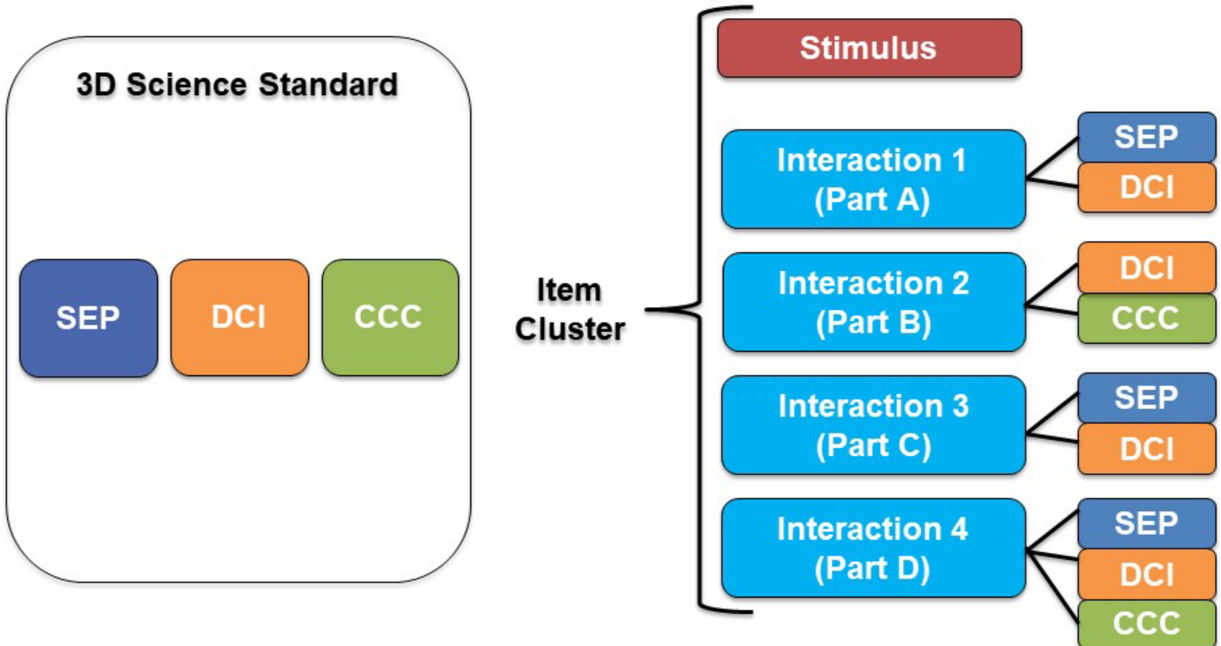
- Evidence-centered design
- Multiple interactions in which students engage a phenomenon
  - Identify
  - Describe
  - Model
  - Predict
  - Explain
- Interactions support a set of assertions about what the student has demonstrated they know and are able to do

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Review of Item Clusters – Composition

16



Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Review of Item Clusters – Composition Example

17

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

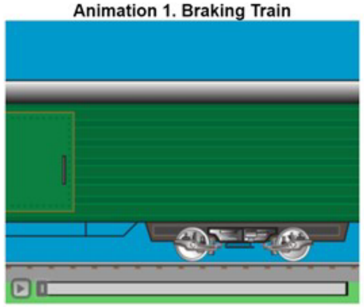


Table 1 explains some properties of the train and its surroundings as energy flows throughout the system.

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?

- Ⓐ The surroundings gain energy.
- Ⓑ The surroundings lose energy.
- Ⓒ The surroundings do not gain or lose energy.
- Ⓓ 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.

Review of Item Clusters – 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

Experience the Online Assessment

19

- Time to "Take the Test"
- Item clusters administered in spring 2021
- Interface is similar to the online test environment that 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 11:15 a.m.)
- Please complete the ***Panelist Demographic Survey*** if you finish early!

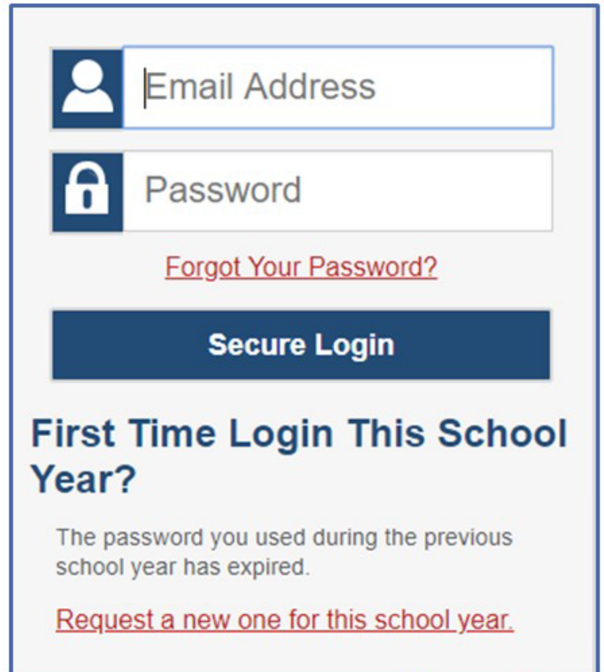
Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Accessing the Online Assessment

20

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



Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

21

Experience Online Operational Assessment

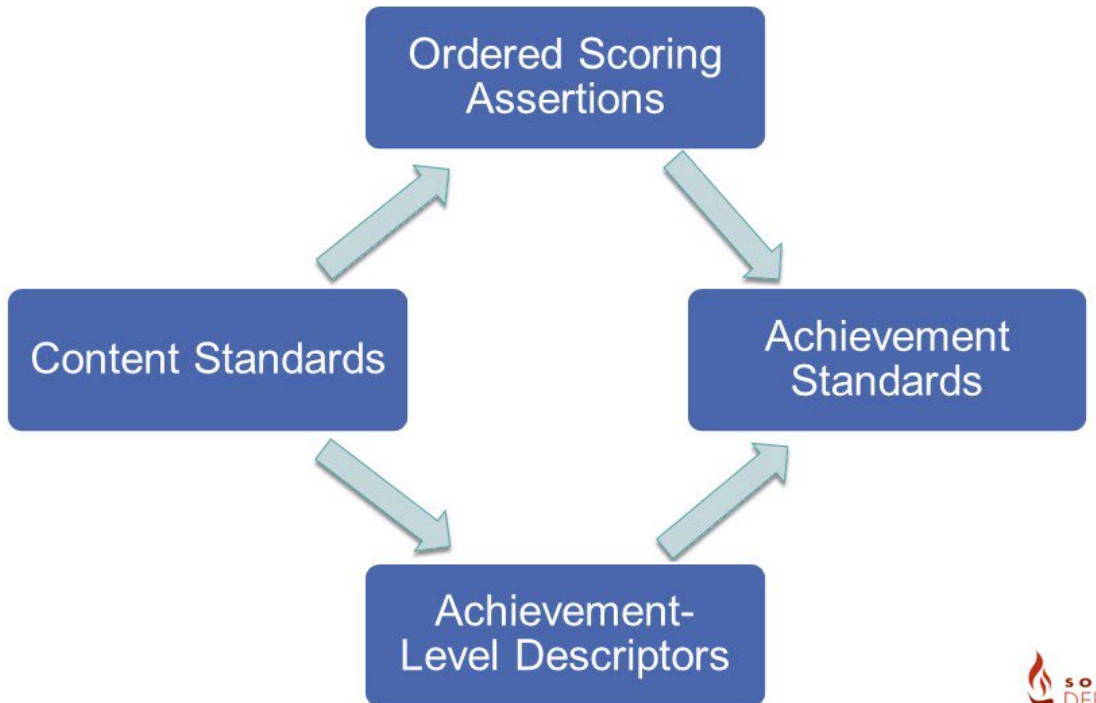
Step 2: Take the Operational Test

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

From Content Standards to Achievement Standards

22



Achievement Standards and Achievement Levels

23

24

Review of Achievement-Level Descriptors

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION
Learning. Leadership. Service.

Achievement-Level Descriptors (ALDs)

25

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

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION
Learning. Leadership. Service.

Grade 5 ALDs – Level 3

26

- **PS1:** Plan and conduct an investigation in which variables are controlled to determine if a chemical reaction occurs and a new substance is created, measuring and graphing quantities to show that matter is always conserved regardless of the change that occurs; and develop a model to show that matter is made of particles too small to be seen.
- **PS2:** Ask questions, plan and conduct an investigation, and/or use produced data to provide evidence to create and support an argument about cause and effect relationships between balanced and unbalanced forces (magnetism and/or gravity) and an object's motion.
- **PS3:** Use models to ask questions and/or use produced data to provide evidence on how energy can be used as a fuel or food or transferred from stored and/or motion energy to different forms like sound, light, and electrical currents.
- **PS4:** Create a solution or develop/and or use a model to describe and compare patterns of waves and the transfer of information; **and** use evidence to support an explanation for how reflected light from objects causes objects to be seen.

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION
Learning. Leadership. Service.

Grade 5 Range ALDs Across Achievement Levels

27

1	2	3	4
Make observations about variables that are controlled to determine if a chemical reaction occurs and a new substance is created, measuring and graphing quantities to show that matter is always conserved regardless of the change that occurs; **and** use a model to show that matter is made of particles too small to be seen.	Organize and test variables that are controlled to determine if a chemical reaction occurs and a new substance is created, measuring and graphing quantities to show that matter is always conserved regardless of the change that occurs; **and** develop a simple model to show that matter is made of particles too small to be seen.	Plan and conduct an investigation in which variables are controlled to determine if a chemical reaction occurs and a new substance is created, measuring and graphing quantities to show that matter is always conserved regardless of the change that occurs; **and** develop a model to show that matter is made of particles too small to be seen.	Revise and conduct an investigation in which variables are controlled to 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; **and** evaluate and revise a model to show that matter is made of particles too small to be seen.

Cambium Assessment

DEPARTMENT OF EDUCATION
Learning. Leadership. Service.

Parse and Review the ALDs

28

- Take a few minutes to review the ALDs taking notice of the verbs and skills that differentiate the achievement levels
  - Think about how the skills change from Level 1 to Level 4
  - 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 ALDs
- ALD Discussion

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Threshold "Just Barely" ALDs

29

- When considering each achievement level, we are especially interested in the transition areas between achievement levels
- Pay attention to characteristics of students who ***just barely*** qualify for entry into the achievement level from those just below
  - Not a typical example of students in the achievement level
  - Although they are poor examples of the achievement level, they do meet the standard, or description in the ALD
    - Just barely Level 2
    - Just barely Level 3
    - Just barely Level 4

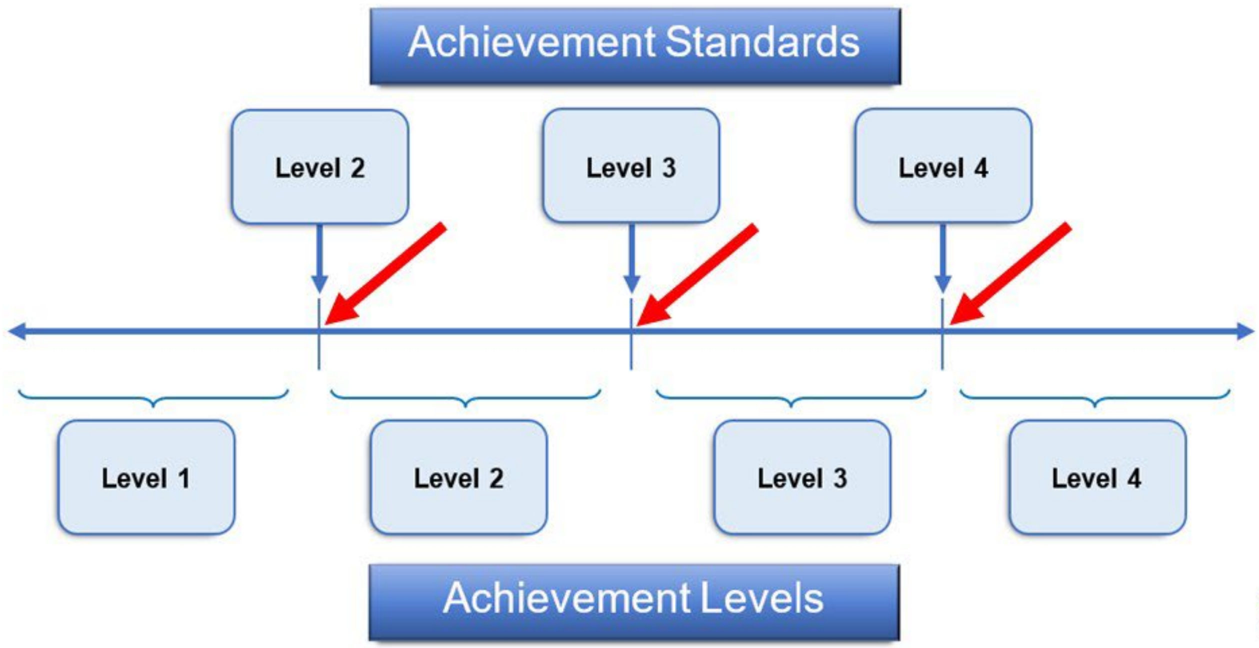
Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Threshold "Just Barely" ALDs

30

- Although "just barely," they do meet the standard



Purpose of Just Barely Discussion

31

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

Just Barely Discussion

32

- Think about what skills, concepts, or knowledge a just barely student would need to have to enter 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 achievement level think about:
  - What skills and knowledge must the student demonstrate to qualify for entrance into this achievement level?
  - How does this differ from the upper range of the adjacent achievement level?

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

33

Review of Ordered Scoring Assertion Booklet

Step 4: Review of Ordered Scoring Assertion Booklet

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION
Learning. Leadership. Service.

Ordered Scoring Assertions

34

- The Ordered Scoring Assertion Booklet (OSAB) represents the full range of standards assessed by the blueprint
- It is important to evaluate scoring assertions as they relate to the item interactions
- Within the OSAB, the scoring assertions are ordered from easiest to most difficult, within an item
  - Assertions within an item may not represent all ALDs

Review of Item Clusters – Composition Example

35

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

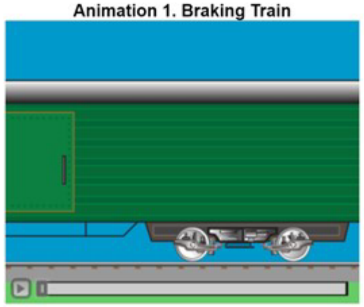


Table 1 explains some properties of the train and its surroundings as energy flows throughout the system.

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?

Ⓐ The surroundings gain energy.

Ⓑ The surroundings lose energy.

Ⓒ The surroundings do not gain or lose energy.

Ⓓ 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.

Cambium Assessment

Review of Item Clusters – Scoring Assertions

36

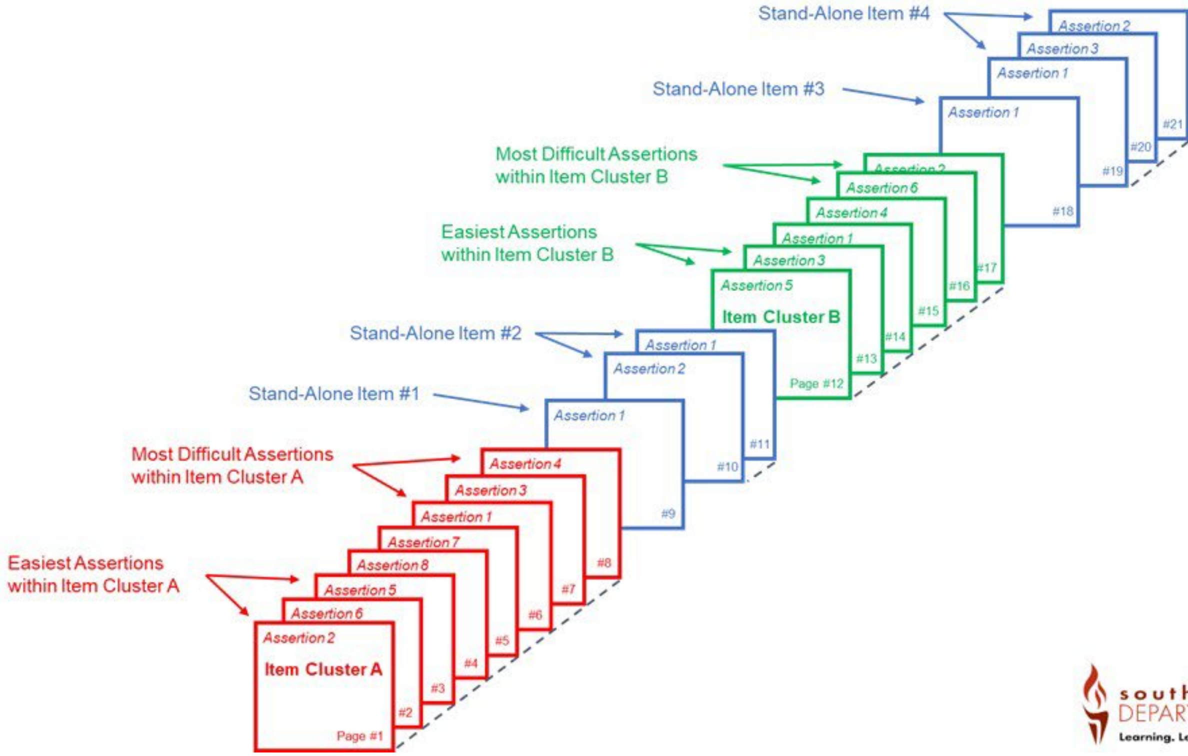
Score Rationale	
The student selected "wheels" for the first blank and "brakes" or "rails" for the second blank showing an understanding of the interactions in the system and the effects of that energy flow.	✗
The student selected "wheels" for the third blank and "less" for the fourth blank showing an 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

uth dakota
EPARTMENT OF EDUCATION
rning. Leadership. Service.

Ordered Scoring Assertion Booklet

37



What If an Assertion Seems Out of Order?

38

- Assertion ordering is based on student performance
- Assertions may seem out of order because they are ordered by difficulty, 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 (within an item)
  - Pay special attention to the interactions supporting the assertions
  - Assertions may be more or less difficult because of the underlying interactions
  - Think about how the phenomenon may affect the difficulty of the task (difficulty of similar tasks between items may vary)

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Ordered Scoring Assertion Booklet: Difficulty Level Visualizer

39

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

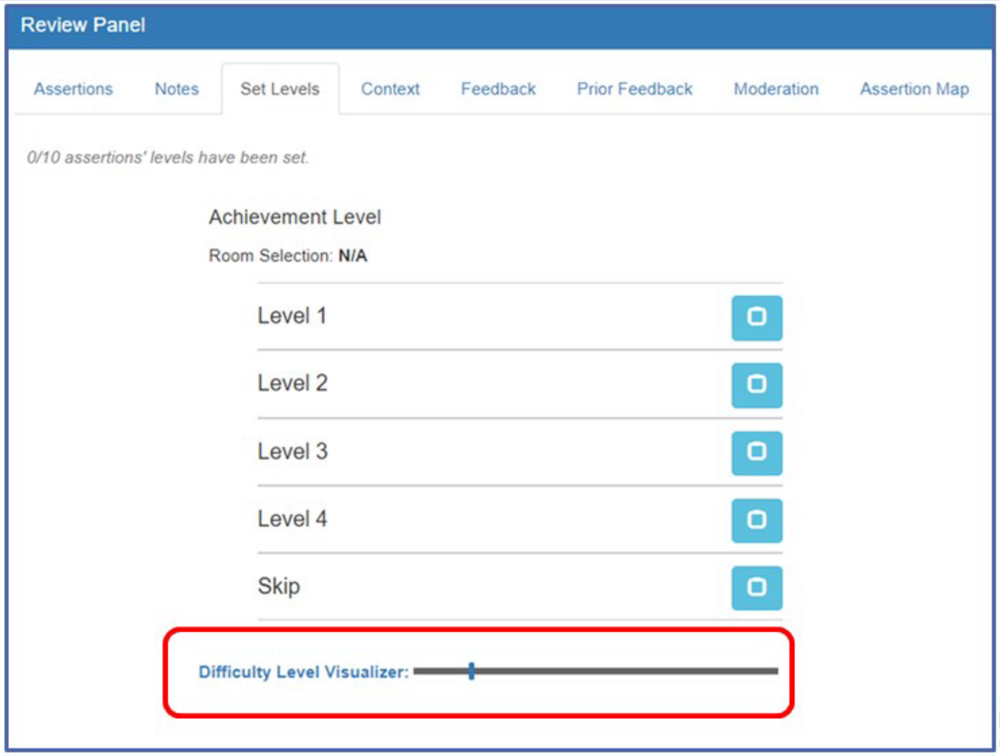
- Example of how to use this:
  - After reviewing the item and scoring assertion you believe this is a relatively difficult concept. However, you see it is on the far left of the scale, ask yourself:
    - What made this so easy for the student?
    - Is the student really “analyzing” or perhaps it is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

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Ordered Scoring Assertion Booklet: Difficulty Level Visualizer

40

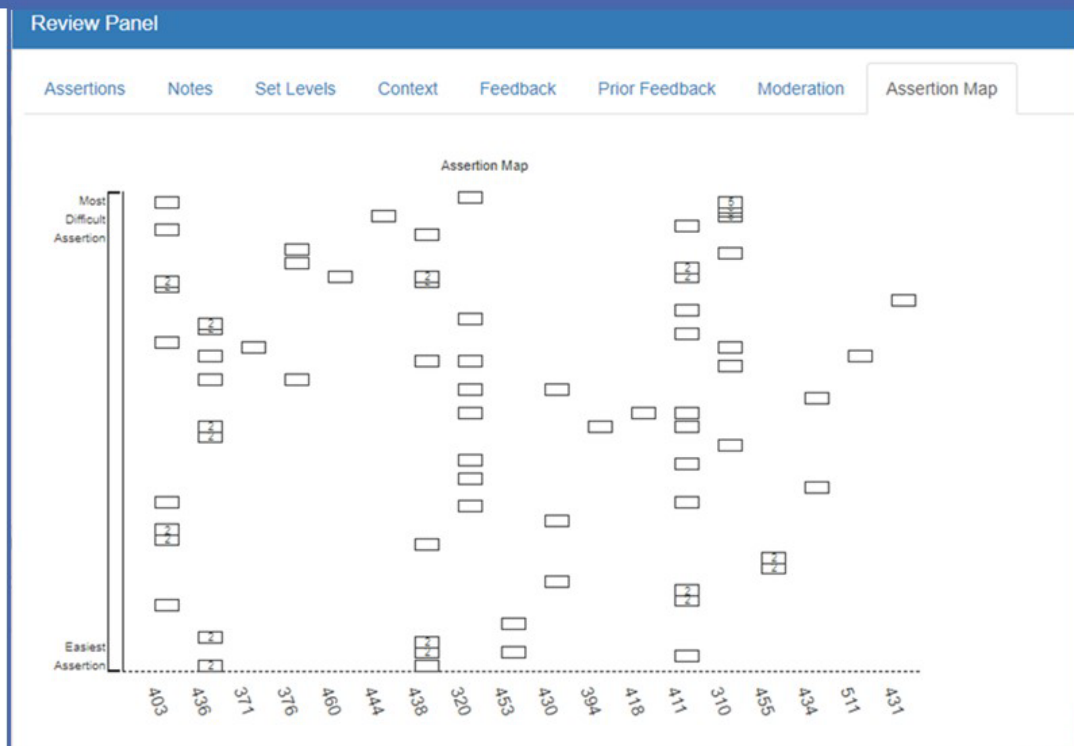


Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Ordered Scoring Assertion Booklet: Grade 5 OSAB Assertion Map

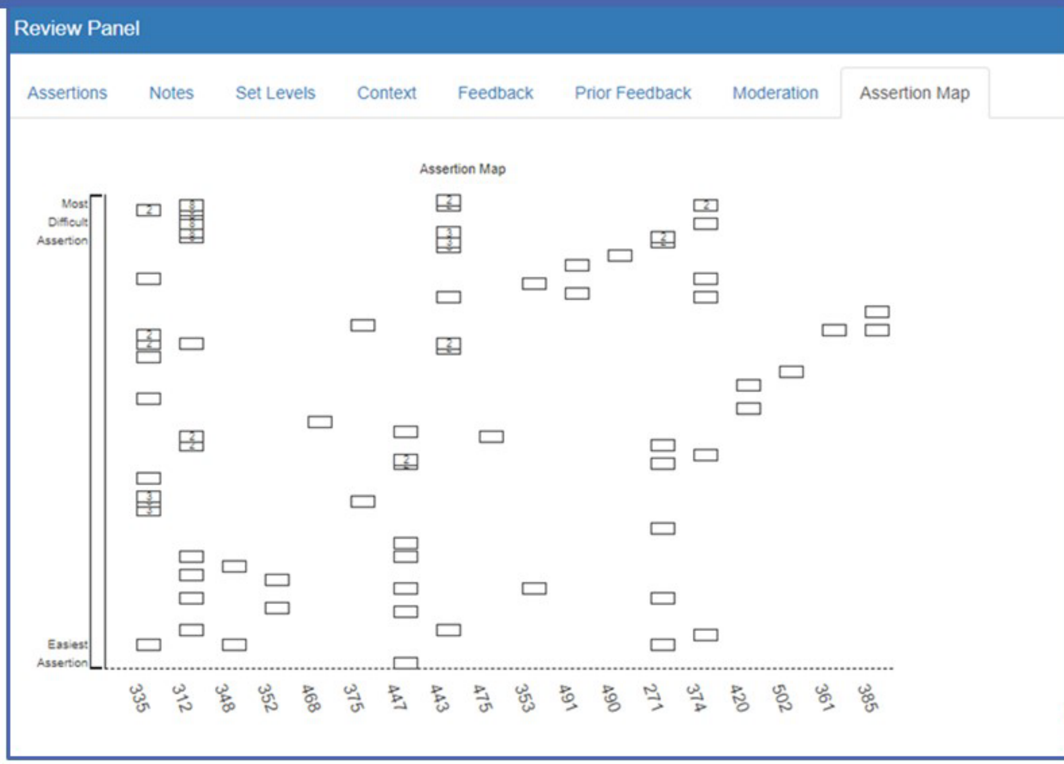
41



Cambium Assessment

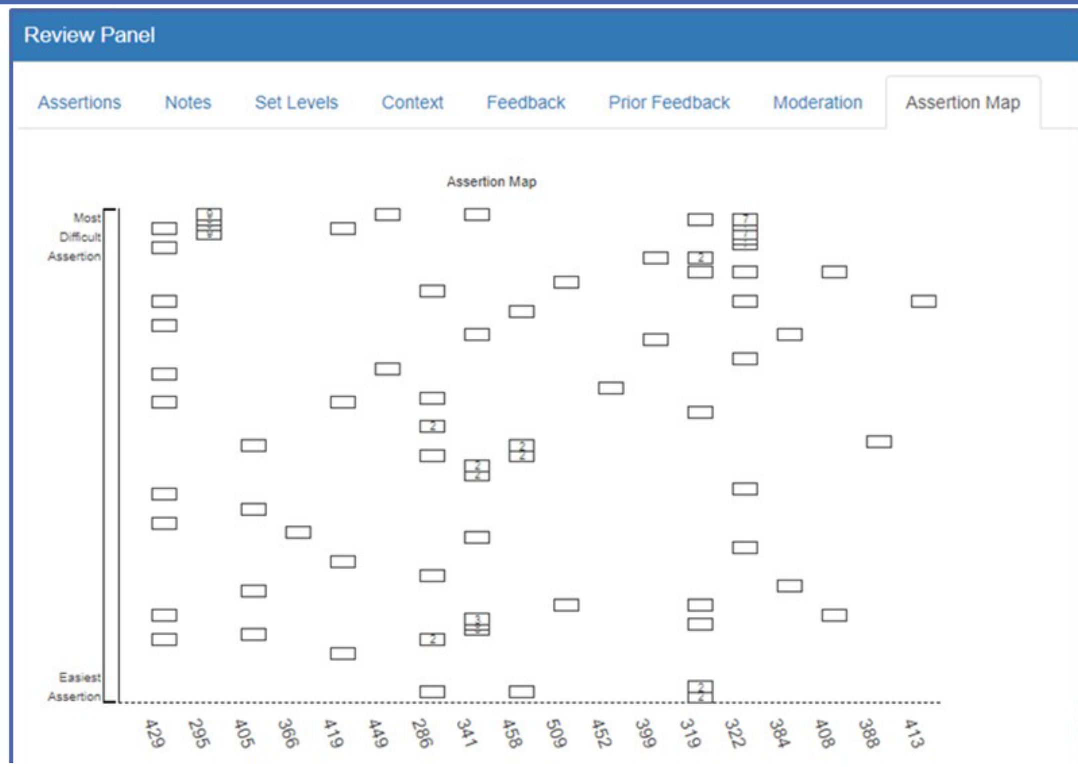
Ordered Scoring Assertion Booklet: Grade 8 OSAB Assertion Map

42



Ordered Scoring Assertion Booklet: Grade 11 OSAB Assertion Map

43



Cambium Assessment

What If an Item Seems Wrong or Unfair?

44

- 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

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Studying the Items and Scoring Assertions

45

- 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 assertion?*
  3. *How does the scoring assertion and the underlying interactions relate to the ALDs?*
- Working as a group
  - Discuss how item interactions support scoring assertions
  - Discuss ordering of scoring assertions
  - Discuss how scoring assertions are related to the ALDs

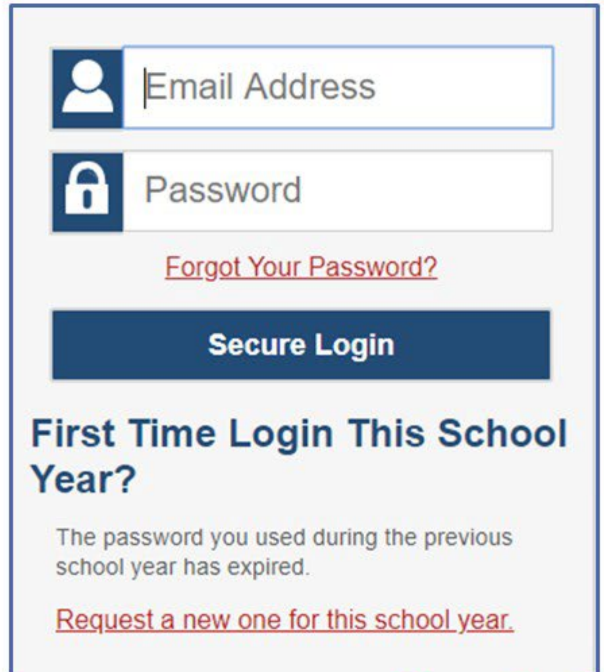
Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Accessing the OSAB

46

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



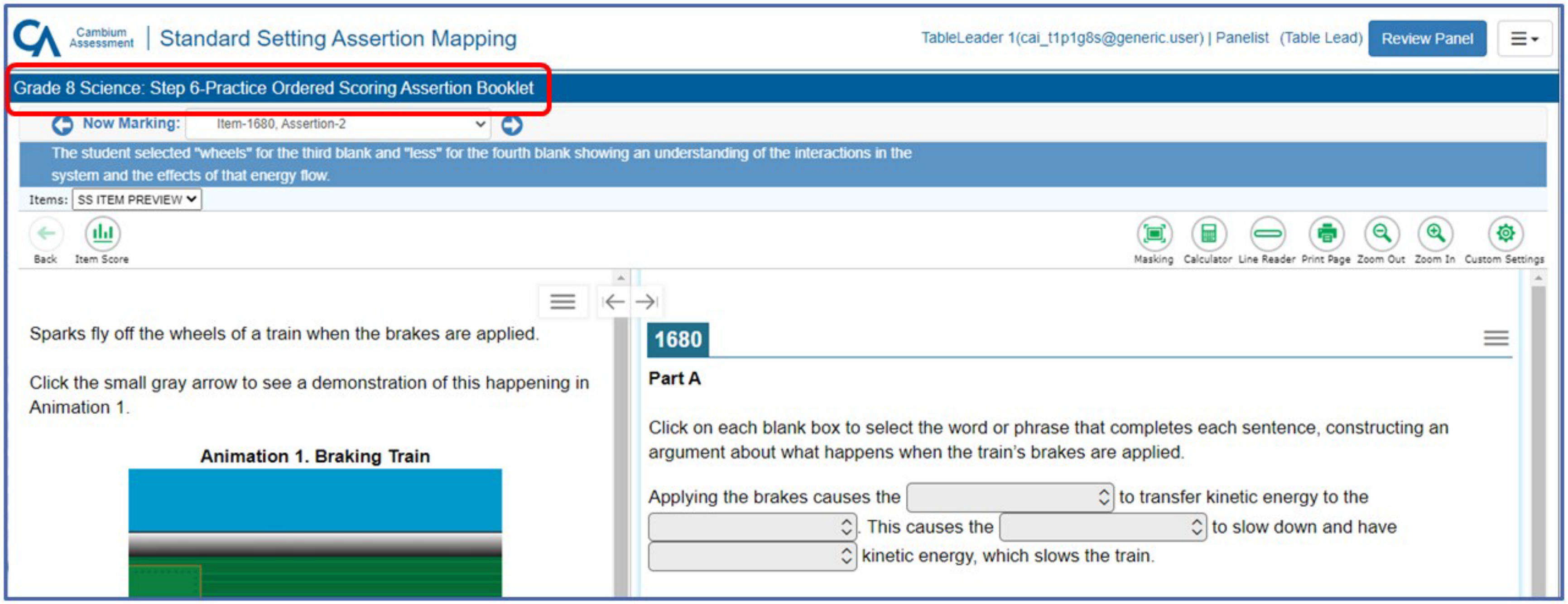
Cambium Assessment

south dakota DEPARTMENT OF EDUCATION
Learning. Leadership. Service.

Navigating the OSAB

47

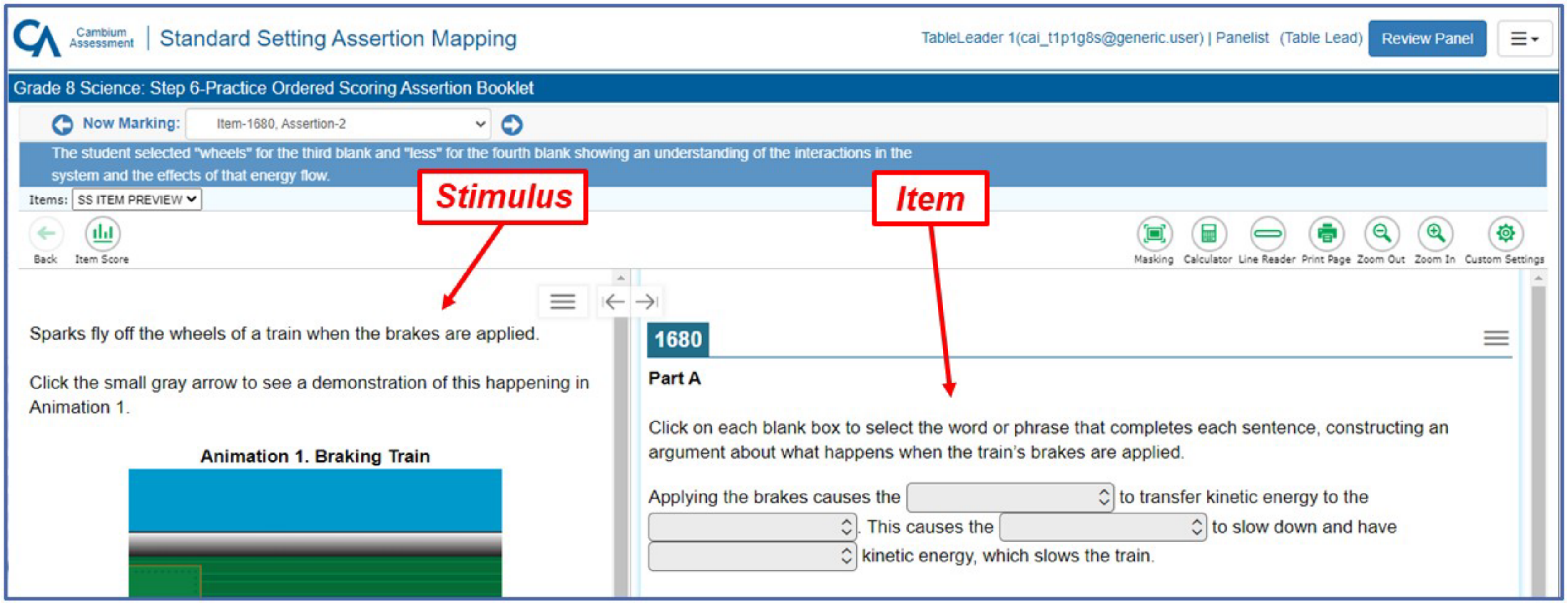
- Test and step we are working on shown at the top of the screen



Navigating the OSAB

48

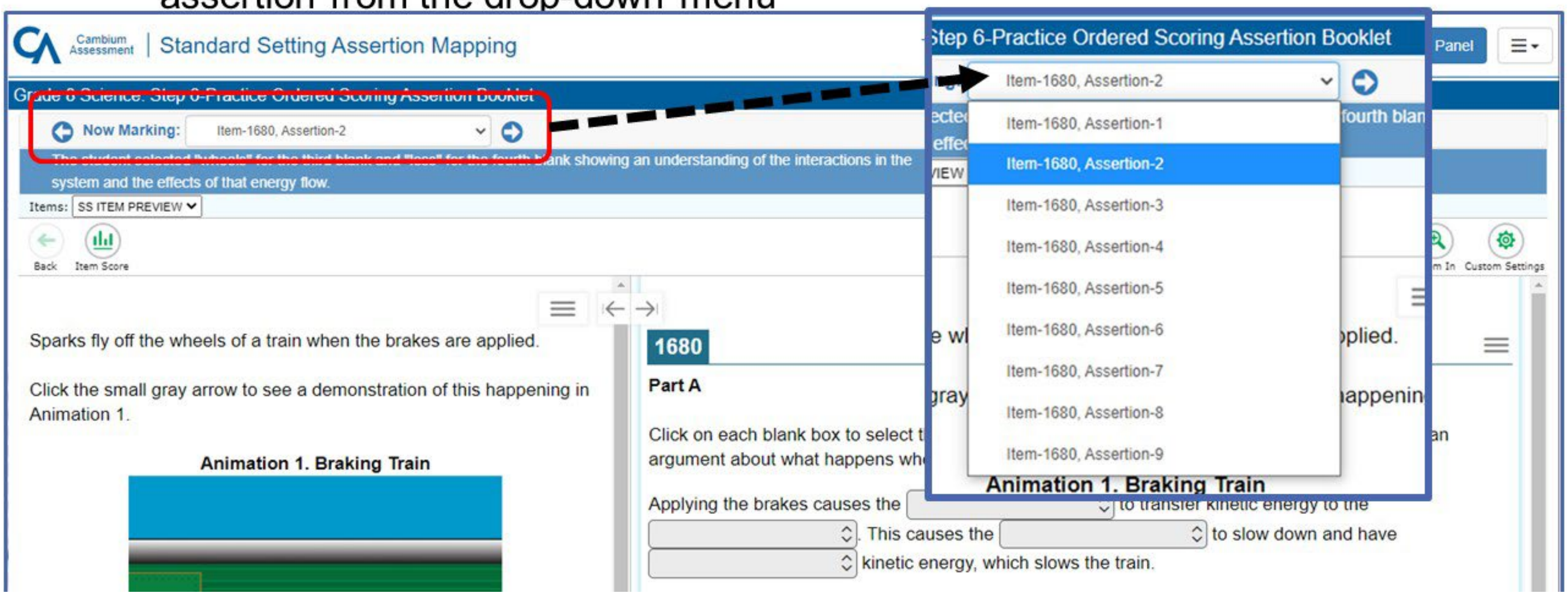
- View the stimulus on the left side of the screen and the item on the right



Navigating the OSAB

49

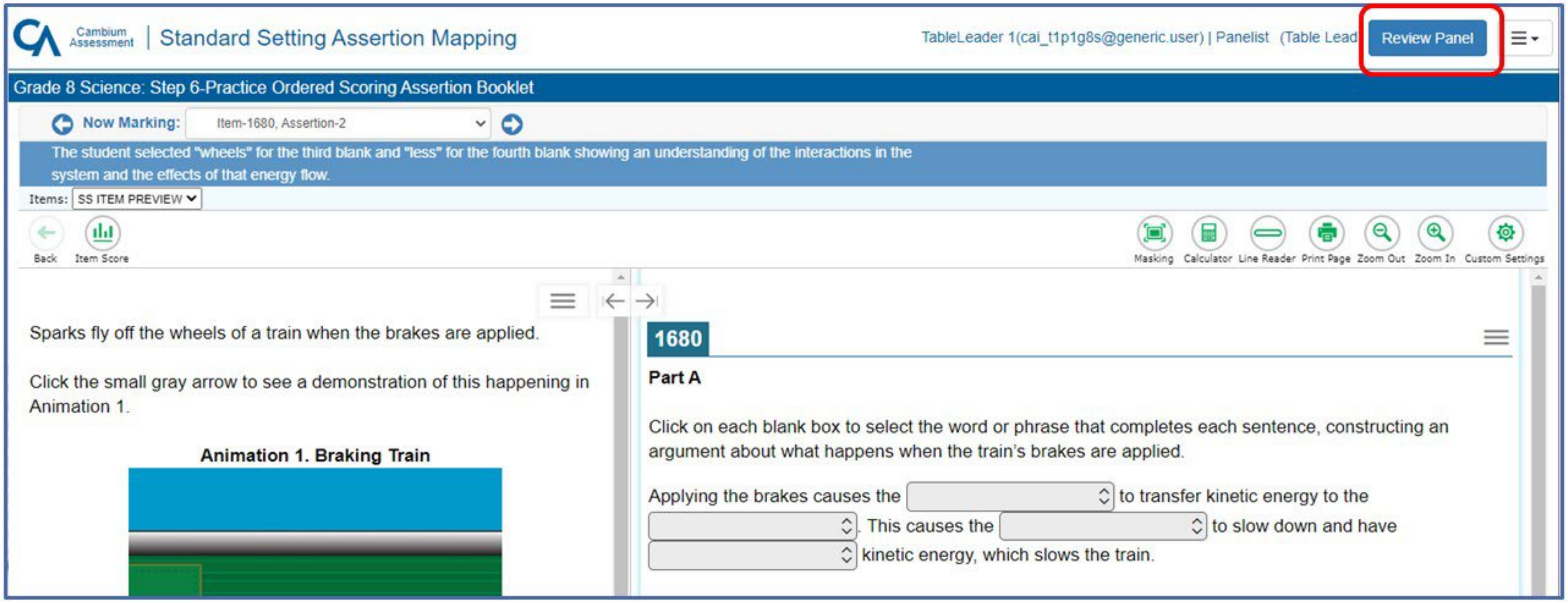
- Move forward in the OSAB using the navigation arrows or select an assertion from the drop-down menu



Navigating the OSAB

50

- Access the Review Panel on the top right of the screen



Review Panel – Assertions

51

Review Panel

Assertions | Notes | Set Levels | Context | Feedback | Prior Feedback | Moderation | Assertion Map

Assertion Rubric Order	Interpretation	Room Selection	Your Selection
1	When asked to indicate when igneous rocks formed, the student selected "400 million years ago," providing some evidence that the student understands how to use evidence provided in the stimulus to explain the age of the igneous rock formation near Fitchburg, MA.	NA	NA
2	When asked to indicate when Fitchburg, MA was covered by ice, the student selected "27,000 to 17,000 years ago," providing some evidence that the student understands how to interpret maps of the North American landscape thousands of years ago.	NA	NA
3	When asked to indicate by what process the boulder was exposed, the student selected "a glacier," providing some evidence that the student understands how to interpret evidence presented in maps in order to construct an explanation for how glaciers moved along Earth's surface and caused erosion of rocks at Earth's surface thousands	NA	NA

More about this item

Content Alignment: SDSS-MS-ESS|ESS2|MS-ESS2-2.

MS-ESS2-2. Construct an explanation based on evidence for how geoscience processes have changed Earth's surface at varying time and spatial scales. (SEP: 6; DCI: ESS2.A, ESS2.C; CCC: Scale/Prop.)

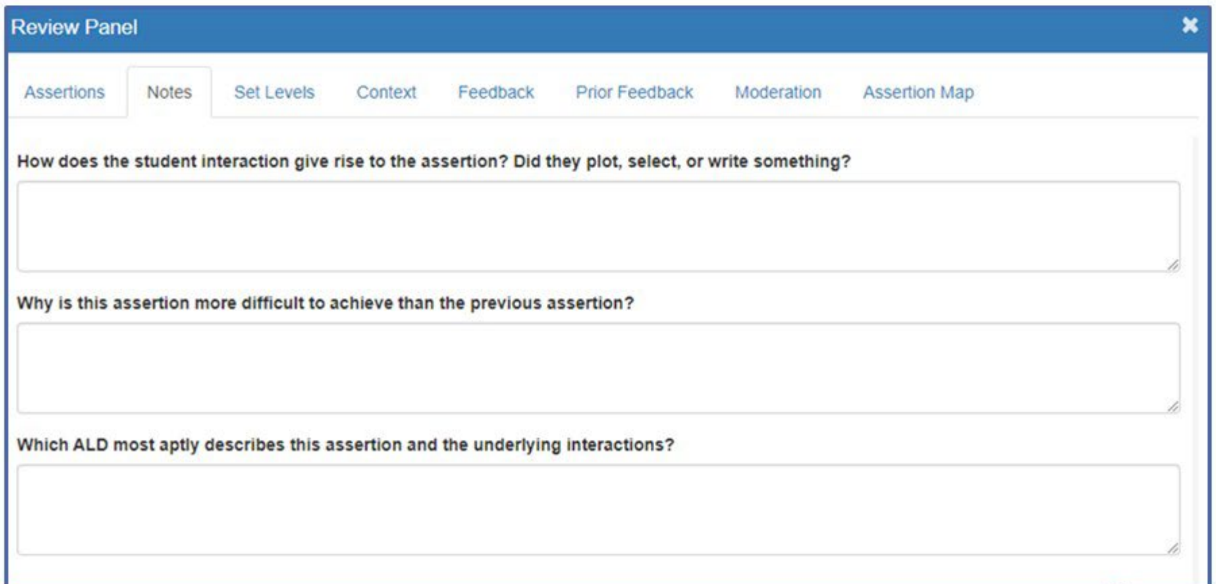
Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Review Panel – Notes

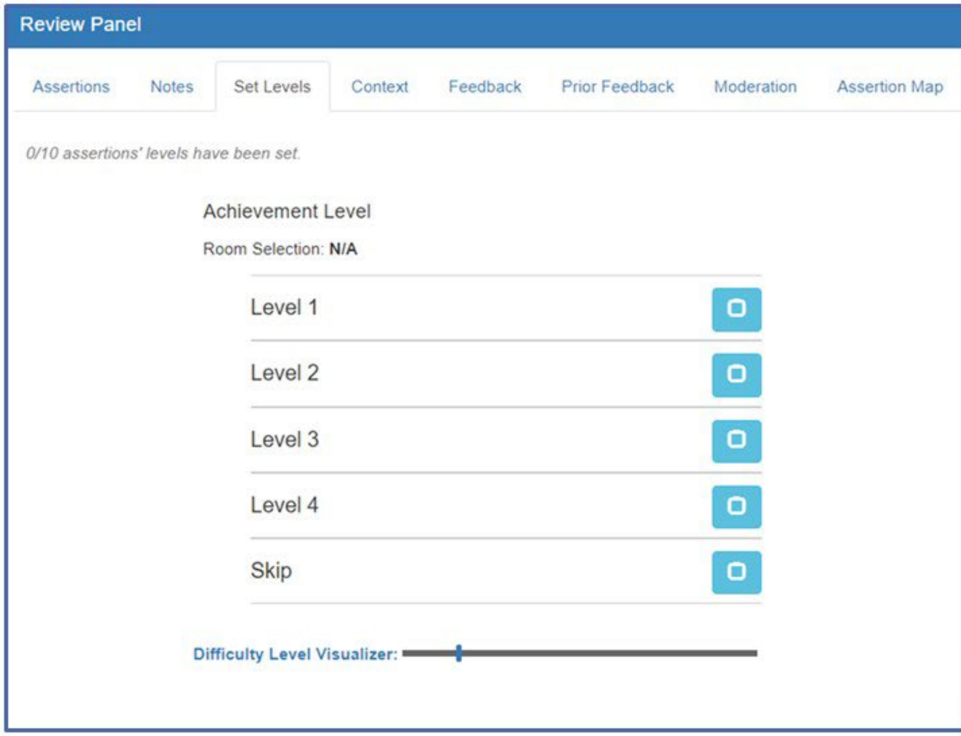
52

- “Notes” tab is for your reference



Review Panel – Set Levels

53



Review of the OSAB

54

- Let's review the items together

Studying the Items and Scoring Assertions

55

- 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 ALDs?*

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

56

Standard-Setting Workshop Day 2

Recommending Achievement Standards for Grade 5 Science

Standard-Setting Workshop Day 2 Agenda

57

- 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

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

58

Training on Assertion-Mapping Task

Assertion-Mapping Key Concepts

59

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

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Assertion-Mapping Procedure (AMP)

60

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

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Important Concepts

61

- "Just barely" meets the achievement level
  - Differentiate students who just barely qualify for entry into an achievement level from those just below
- Assertion mapping
  - Map each scoring assertion to the achievement 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 ALDs*
  - Mapping of assertions to achievement levels should reflect the ordering – *no inversions within an item**
  - Pay attention to the Difficulty Level Visualizer and Assertion Map across items

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Mapping Ordered Assertions to Achievement Levels

62

- You will map each scoring assertion to an achievement level using the following tools:
  - ALDs
  - Difficulty Level Visualizer
  - Assertion Map
  - 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

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Practice Online Assertion-Mapping Task

63

- 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 and one stand-alone item
- Log into the system and review the items and ordered scoring assertions answering the three questions as you go
- Then, map each scoring assertion to an achievement level and click "*confirm*"
- This is meant to help you become familiar with the tool and process

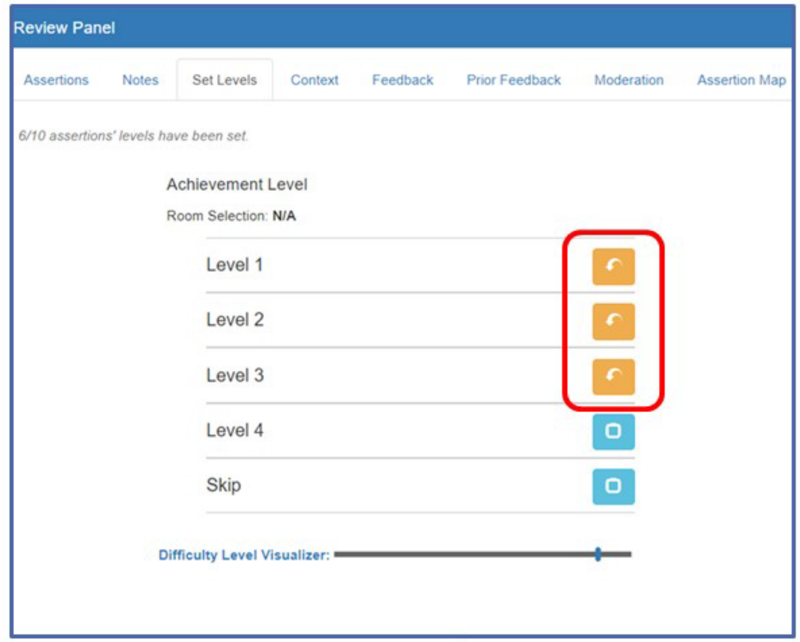
Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Assertion Mapping – Mapping Inversion within Item

64

- The standard-setting tool will not let you map an assertion to a lower achievement level than the previous assertions within the item that you have already mapped



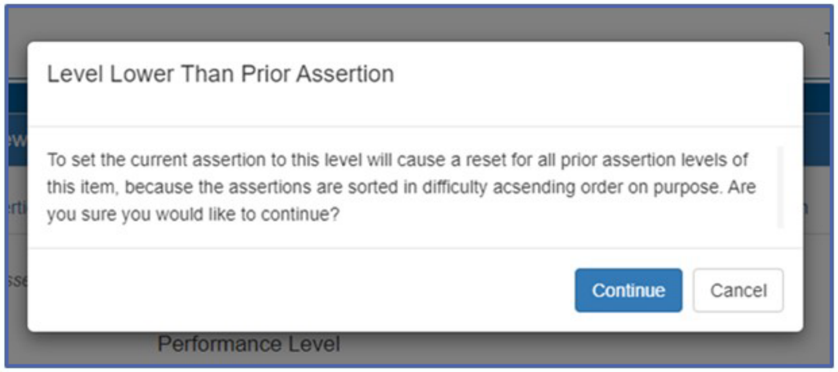
Cambium Assessment

south dakota DEPARTMENT OF EDUCATION
Learning. Leadership. Service.

Assertion Mapping – Mapping Inversion within Item

65

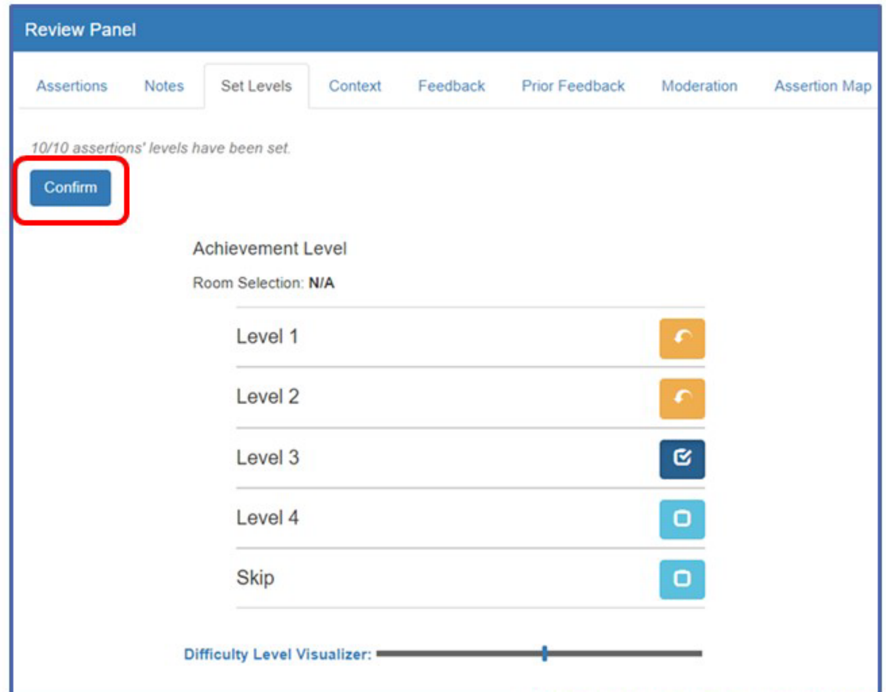
- If you want to map the assertion to a lower achievement level:
  - All previous mappings within the item that are lower will be un-mapped
  - You will need to go back and re-map those assertions



Assertion Mapping – Confirm

66

- Once all assertions are mapped, a "Confirm" button will appear in the "Set Levels" tab



Cambium Assessment

DEPARTMENT OF EDUCATION
Learning. Leadership. Service.

67

Practice Assertion-Mapping Task and Standard-Setting Quiz

Step 6: Practice Ordered Scoring Assertion Booklet

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION
Learning. Leadership. Service.

68

Round 1 Assertion Mapping

Step 8: Round 1 Assertion Mapping

Round 1 Readiness Form

69

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

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION
Learning. Leadership. Service.

Round 1 Assertion Mapping

70

- You will use the next 75 minutes to map each assertion to an achievement 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 achievement 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

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION
Learning. Leadership. Service.

71

Review Panelist Feedback Data and Discuss Round 1 Results

Step 10: Results of Round 1

Group Feedback and Discussion

72

- 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

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION
Learning. Leadership. Service.

Feedback and Impact Data

73

- 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 achievement standard seem reasonable?
  - What are the implications for the achievement standards?
  - All achievement standard recommendations should be based on content rationales

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Feedback Table

74

Form: Grade 5 Science OSAB Step: All Steps

Summary of tentative standards

Room 1 (Grade 5 Science) Step: 8 (Results of Round 1 Assertion Mapping) Form: Grade 5 Science Ordered Scoring Assertion Booklet

Step 8 - Round 1 Assertion Mapping

	Level 2	Level 3	Level 4
Form-Grade 5 Science OSAB	477.0	508.0	526.0
Room-1	477.0	508.0	526.0
Table-1	477.0	508.0	526.0

Step 8 - Round 1 Assertion Mapping Panelist Statistic Data

Form	Room	Table	UserName	Level 2	Level 3	Level 4	Time Confirmed
Grade 5 Science OSAB	1	1	SD_T1P1_G5S@generic.user	461.0	485.0	526.0	9/16/2021 2:14:32 PM
Grade 5 Science OSAB	1	1	SD_T1P2_G5S@generic.user	474.0	508.0	528.0	9/16/2021 1:18:17 PM
Grade 5 Science OSAB	1	1	SD_T1P3_G5S@generic.user	477.0	513.0	542.0	9/16/2021 1:17:58 PM
Grade 5 Science OSAB	1	1	SD_T1P4_G5S@generic.user	477.0	502.0	521.0	9/16/2021 1:20:57 PM

Cambium Assessment

DEPARTMENT OF EDUCATION
Learning. Leadership. Service.

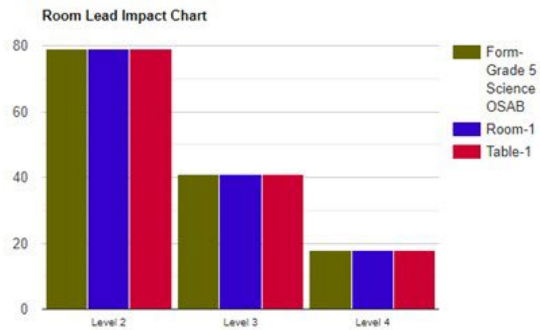
Feedback Chart

75

Sequence Type: Percent At or Above

Step 8 - Round 1 Assertion Mapping, Percent At or Above Data

	Level 2	Level 3	Level 4
Form-Grade 5 Science OSAB	79	41	18
Room-1	79	41	18
Table-1	79	41	18

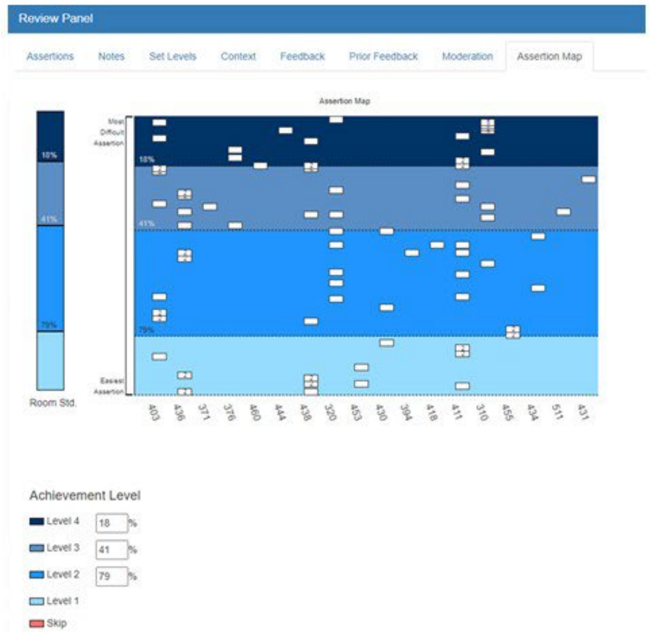


Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Assertion Map

76



Cambium Assessment

south dakota DEPARTMENT OF EDUCATION
Learning. Leadership. Service.

Variance Monitor

77

- 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

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Contextual Information for Round 2 Review Panel – Context

78

- “Context” tab presents student impact data and benchmarking data

Review Panel

Assertions | Notes | Set Levels | Context | Feedback | Prior Feedback | Moderation | Assertion Map

Some facts about the difficulty of this assertion.

Context Category	Value
Overall percent of South Dakota students that perform at or above this level:	51
A cut-score at this assertion is comparable to:	Level 3 on the South Dakota English Language Arts Assessment
A cut-score at this assertion is comparable to:	Level 2 on the South Dakota Math Assessment

Contextual Information for Round 2

79

- Does the percentage of students who performed at or above the specified RP value associated with each assertion seem reasonable?
- What are the implications for the achievement standards?
- All achievement standard recommendations should be based on content rationales

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Contextual Information – Student Impact Data

80

- The impact data for an assertion is defined as the percentage of students who performed at or above the specified RP value associated with the assertion.

Review Panel

Assertions | Notes | Set Levels | Context | Feedback | Prior Feedback | Moderation | Assertion Map

Some facts about the difficulty of this assertion.

Context Category	Value
Overall percent of South Dakota students that perform at or above this level:	51
A cut-score at this assertion is comparable to:	Level 3 on the South Dakota English Language Arts Assessment
A cut-score at this assertion is comparable to:	Level 2 on the South Dakota Math Assessment

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Contextual Information – Benchmarking Data

81

- Are achievement standards nationally competitive and represent on track for college readiness?
  - Smarter Balanced ELA and Mathematics
- Achievement levels for benchmark assessments will provide context about the general neighborhood in which achievement standards likely reside

South Dakota 2021 ELA Assessment Results			
Grade	At or Above		
	Level 2	Level 3	Level 4
5	73	51	20
8	78	52	15
11	86	66	28

South Dakota 2021 Math Assessment Results			
Grade	At or Above		
	Level 2	Level 3	Level 4
5	68	38	17
8	69	40	18
11	68	39	14

Cambium Assessment

DEPARTMENT OF EDUCATION
Learning. Leadership. Service.

Contextual Information – Benchmarking Data

82

Review Panel

Assertions | Notes | Set Levels | Context | Feedback | Prior Feedback | Moderation | Assertion Map

Some facts about the difficulty of this assertion.

Context Category	Value
Overall percent of South Dakota students that perform at or above this level:	51
A cut-score at this assertion is comparable to:	Level 3 on the South Dakota English Language Arts Assessment
A cut-score at this assertion is comparable to:	Level 2 on the South Dakota Math Assessment

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION
Learning. Leadership. Service.

83

Round 2 Assertion Mapping

Step 12: Round 2 Assertion Mapping

Round 2 Readiness Form

84

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

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION
Learning. Leadership. Service.

Round 2 Assertion Mapping

85

- You will use the next 60 minutes to map each scoring assertion to an achievement level
- Use the tools and documents along with your professional judgment, **contextual information – student impact data and benchmarking 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
- Should be a logical progress of achievement levels (within an item) – *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 3:30 pm

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION
Learning. Leadership. Service.

86

Round 2 Results

Step 14: Results of Round 2

Feedback Table

87

Form: Grade 5 Science OSAB Step: 12. Round 2 Assertion M

Summary of tentative standards

Room 1 (Grade 5 Science) Step: 12 (Results of Round 2 Assertion Mapping) Form: Grade 5 Science Ordered Scoring Assertion Booklet

Step 12 - Round 2 Assertion Mapping

	Level 2	Level 3	Level 4
Form-Grade 5 Science OSAB	477.0	508.0	527.0
Room-1	477.0	508.0	527.0
Table-1	477.0	508.0	527.0

Step 12 - Round 2 Assertion Mapping Panelist Statistic Data

Form	Room	Table	UserName	Level 2	Level 3	Level 4	Time Confirmed
Grade 5 Science OSAB	1	1	SD_T1P1_G5S@generic.user	477.0	501.0	526.0	9/16/2021 3:49:27 PM
Grade 5 Science OSAB	1	1	SD_T1P2_G5S@generic.user	474.0	508.0	528.0	9/16/2021 3:34:03 PM
Grade 5 Science OSAB	1	1	SD_T1P3_G5S@generic.user	477.0	508.0	527.0	9/16/2021 3:49:57 PM
Grade 5 Science OSAB	1	1	SD_T1P4_G5S@generic.user	477.0	502.0	521.0	9/16/2021 3:47:38 PM

Cambium Assessment

akota ENT OF EDUCATION Learning. Leadership. Service.

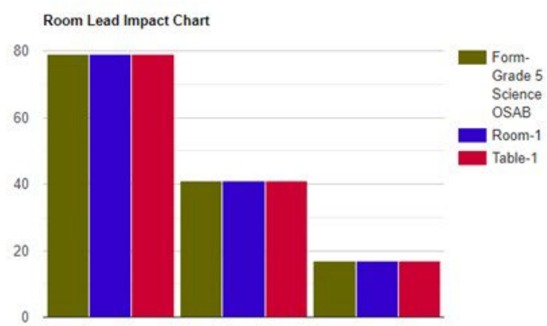
Feedback Chart

88

Sequence Type: Percent At or Above

Step 12 - Round 2 Assertion Mapping, Percent At or Above Data

	Level 2	Level 3	Level 4
Form-Grade 5 Science OSAB	79	41	17
Room-1	79	41	17
Table-1	79	41	17

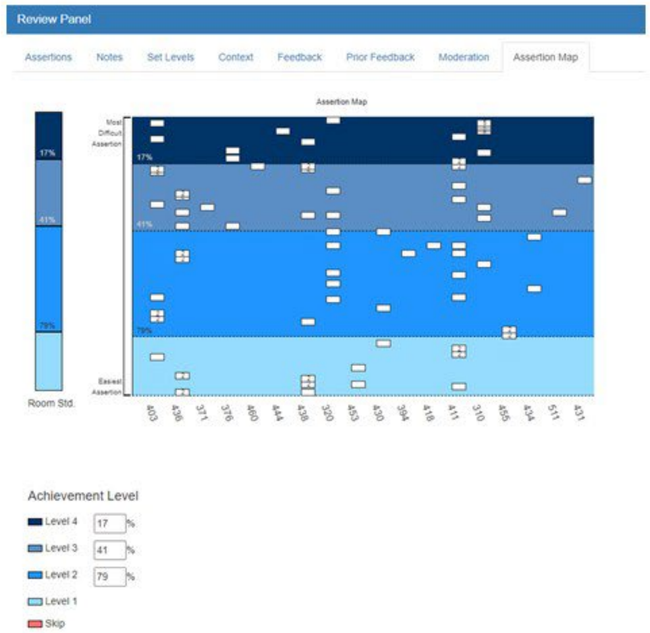


Cambium Assessment

south dakota DEPARTMENT OF EDUCATION
Learning. Leadership. Service.

Assertion Map

89



Cambium Assessment

south dakota DEPARTMENT OF EDUCATION
Learning. Leadership. Service.

90

Workshop Evaluations

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

91

Moderation

Step 16: Moderation

Creating a System of Achievement Standards

92

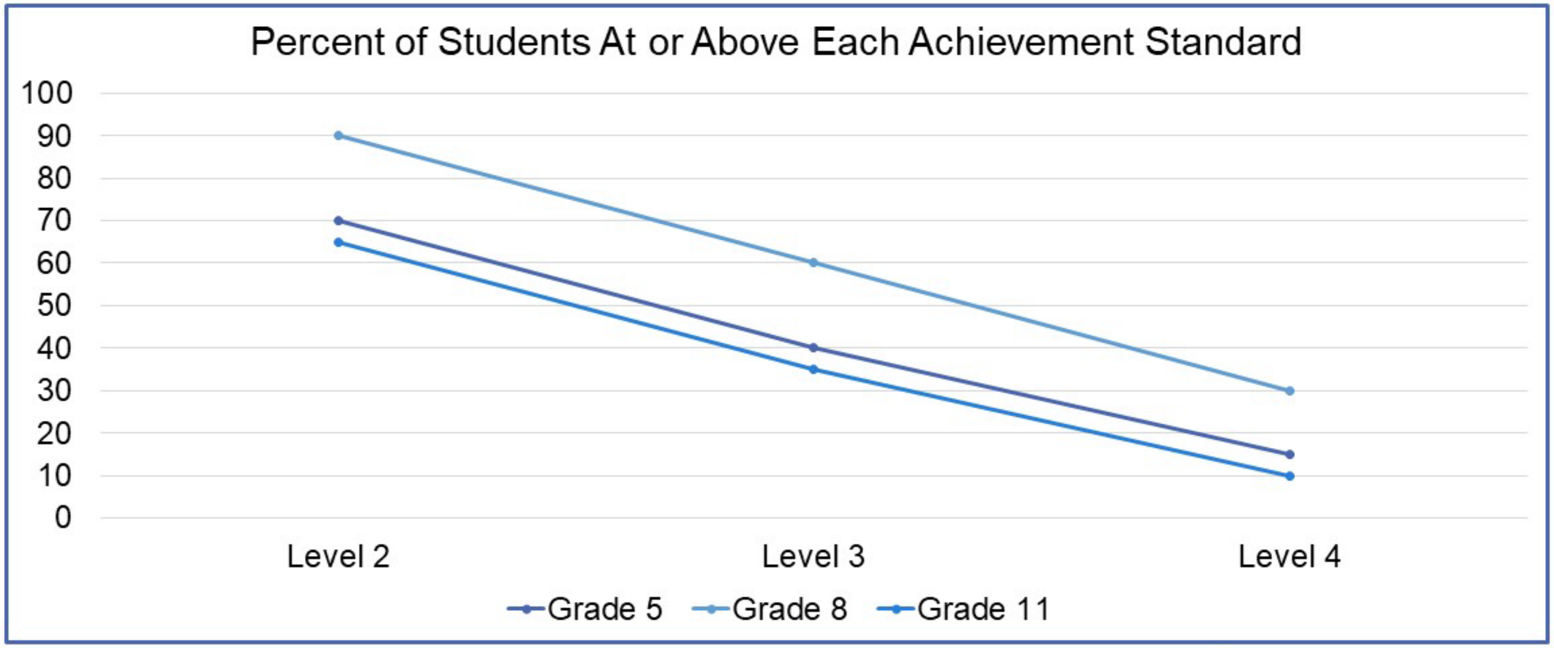
- Achievement standards for a statewide system must be coherent across grades and subjects
  - Articulation
  - Benchmarking
  - Moderation

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Articulation

93



Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Moderation

94

- 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 achievement standards (assuming there is a good content reason for doing so)

Appendix 3-F

Standard-Setting Practice Quiz

Standard-Setting Practice Quiz

Exhibit 3-F-1. Standard-Setting Practice Quiz

2021 South Dakota Science Assessment Standard Setting - Assertion Mapping Practice Quiz

* Required

1. Name: *

2. Panelist ID (e.g., SD_T1P1_G5S): *

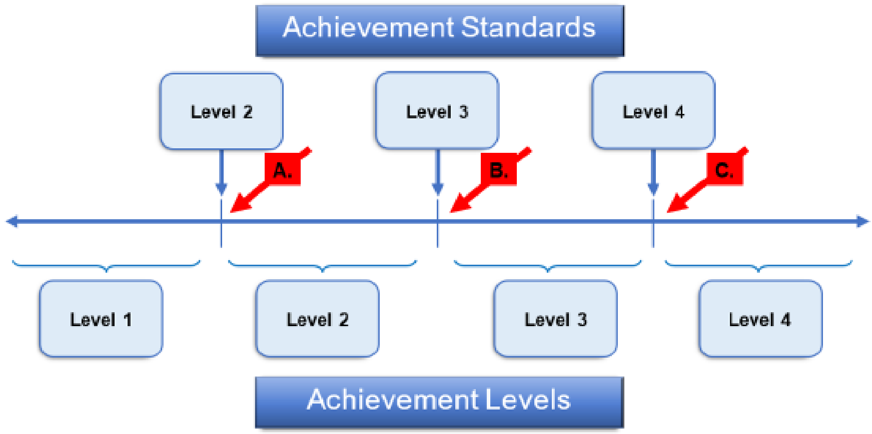
3. Assigned Committee: *

   Mark only one oval.

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

Achievement Standards and Achievement Levels

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



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

Mark only one oval.

- Box A
- Box B
- Box C

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

Mark only one oval.

- Box A
- Box B
- Box C

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

   Mark only one oval.

   - Box A
   - Box B
   - Box C

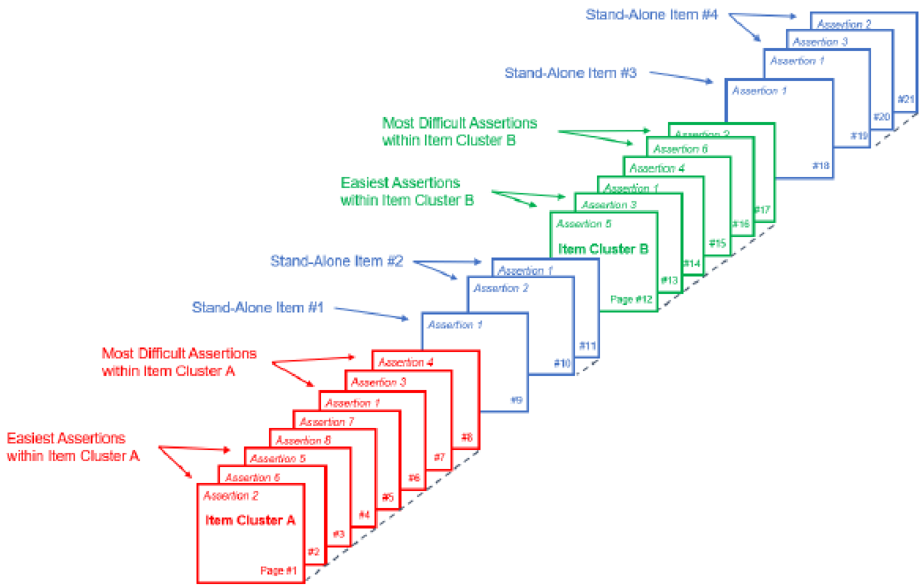
7. Which achievement standard differentiates between the Level 2 achievement level and the Level 3 achievement level? *

   Mark only one oval.

   - Level 2
   - Level 3
   - Level 4

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 within the OSAB, scoring assertions are ordered by difficulty. In the 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

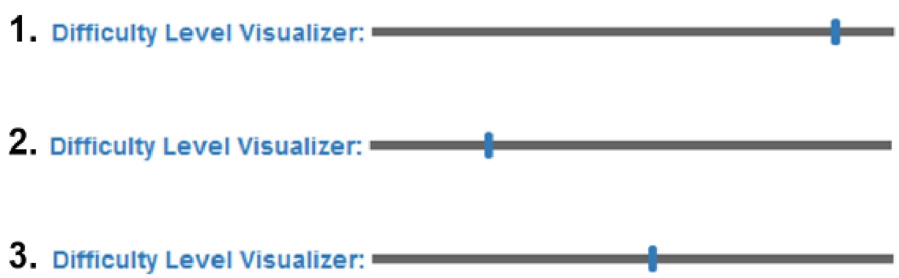
Standard-Setting Assertion Mapping Tool

9. Do you have to assign each scoring assertion to an achievement 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-G

Standard-Setting Readiness Forms

Standard-Setting Readiness Forms

Exhibit 3-G-1. Standard-Setting Round 1 Readiness Form

2021 South Dakota Science Assessment Standard Setting Educator Panel - Readiness Form

Preparation for Round 1 Assertion Mapping

* Required

1. Name: *

2. Panelist ID (e.g., SD_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 Achievement-Level Descriptors (ALDs) and fully explained the concept of threshold ALDs. *

   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 achievement 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 achievement 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 achievement 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 achievement levels. (Please initial below.) *

    Mark only one oval.

    - Yes
    - No
    - Not applicable

14. Initial: *

This content is neither created nor endorsed by Google.

Google Forms

Exhibit 3-G-2. Standard-Setting Round 2 Readiness Form

2021 South Dakota Science Assessment Standard Setting Educator Panel - Readiness Form

Preparation for Round 2 Assertion Mapping

* Required

1. Name: *

2. Panelist ID (e.g., SD_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 achievement 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 achievement 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 achievement 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 achievement levels. (Please initial below.) *

Mark only one oval.

- Yes
- No
- Not applicable

14. Initial: *

Appendix 3-H

Round 1 and Round 2 Standard-Setting Assertion Maps

Round 1 Standard-Setting Assertion Maps

Exhibit 3-H-1. Round 1 Standard-Setting Assertion Map, Science Grade 5

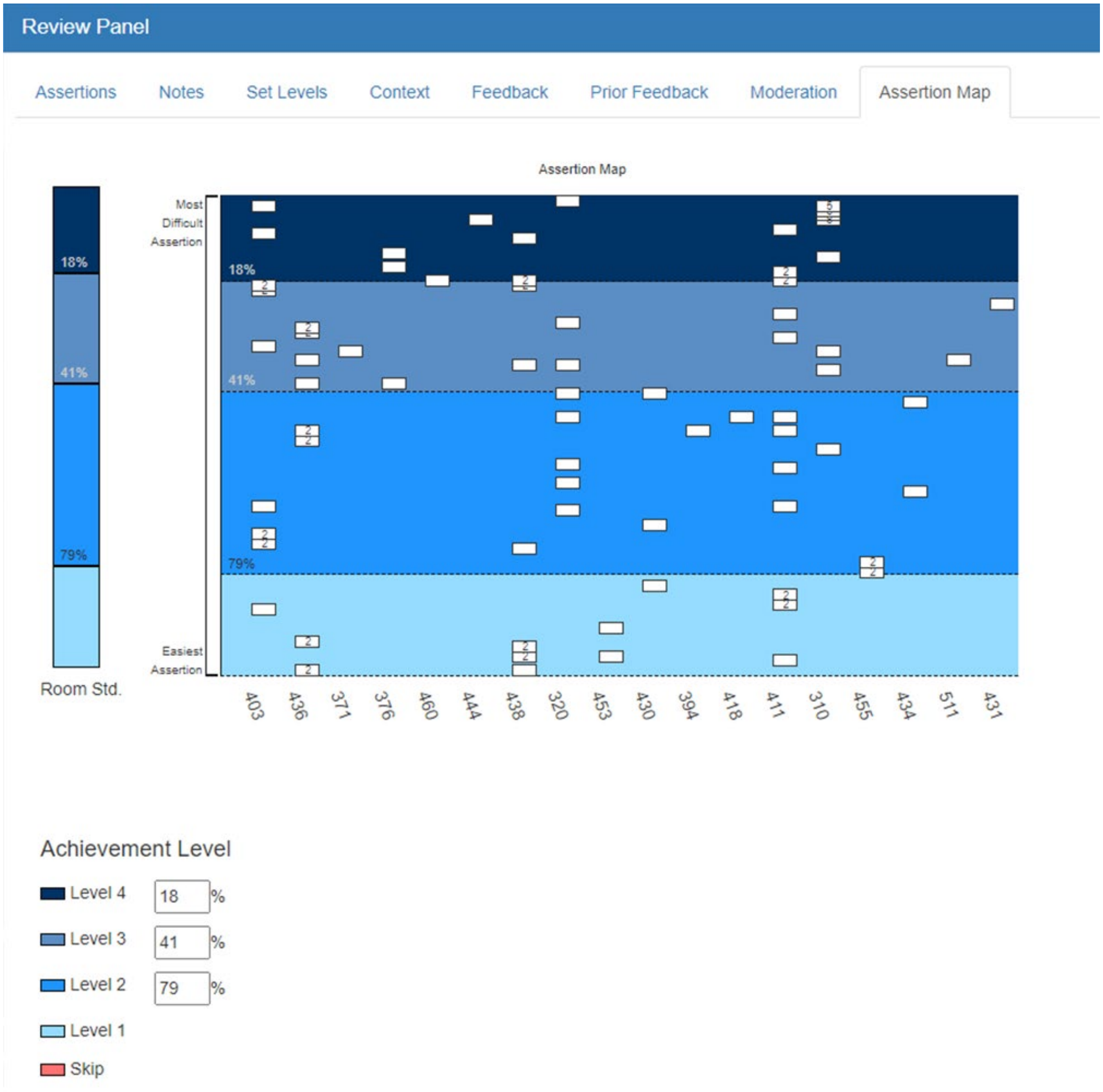


Exhibit 3-H-2. Round 1 Standard-Setting Assertion Map, Science Grade 8

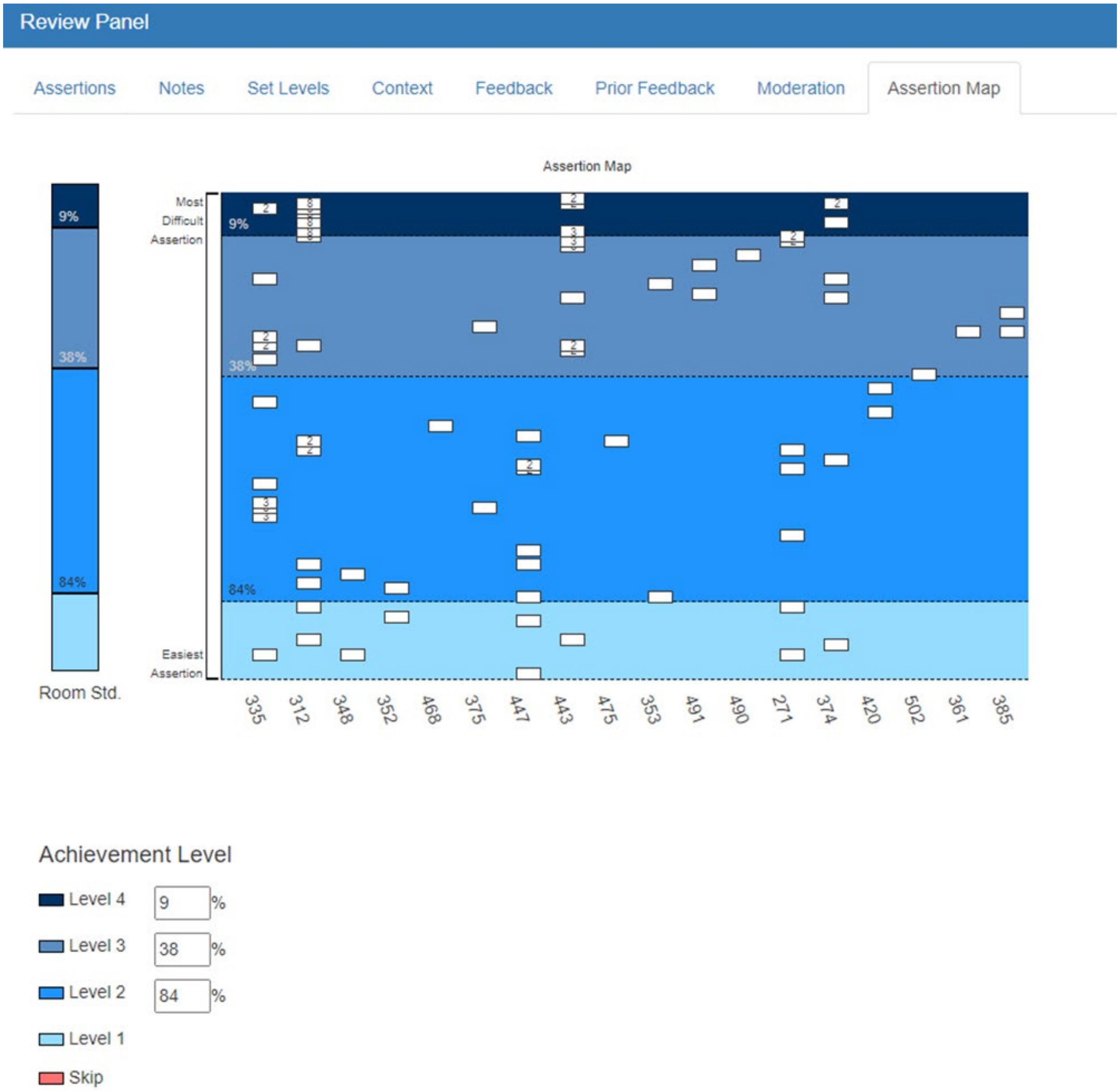
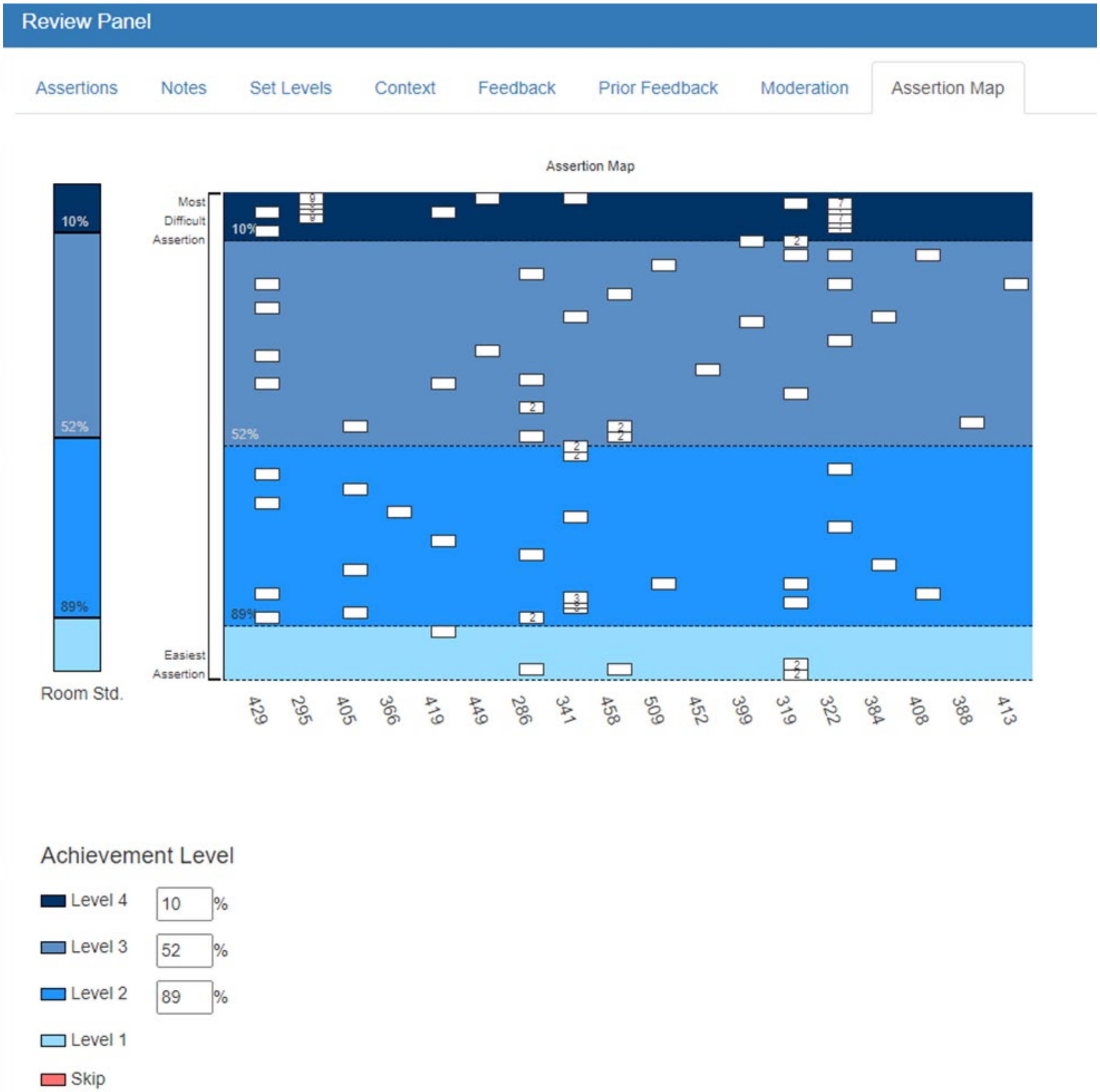


Exhibit 3-H-3. Round 1 Standard-Setting Assertion Map, Science Grade 11



Round 2 Standard-Setting Assertion Maps

Exhibit 3-H-4. Round 2 Standard-Setting Assertion Map, Science Grade 5

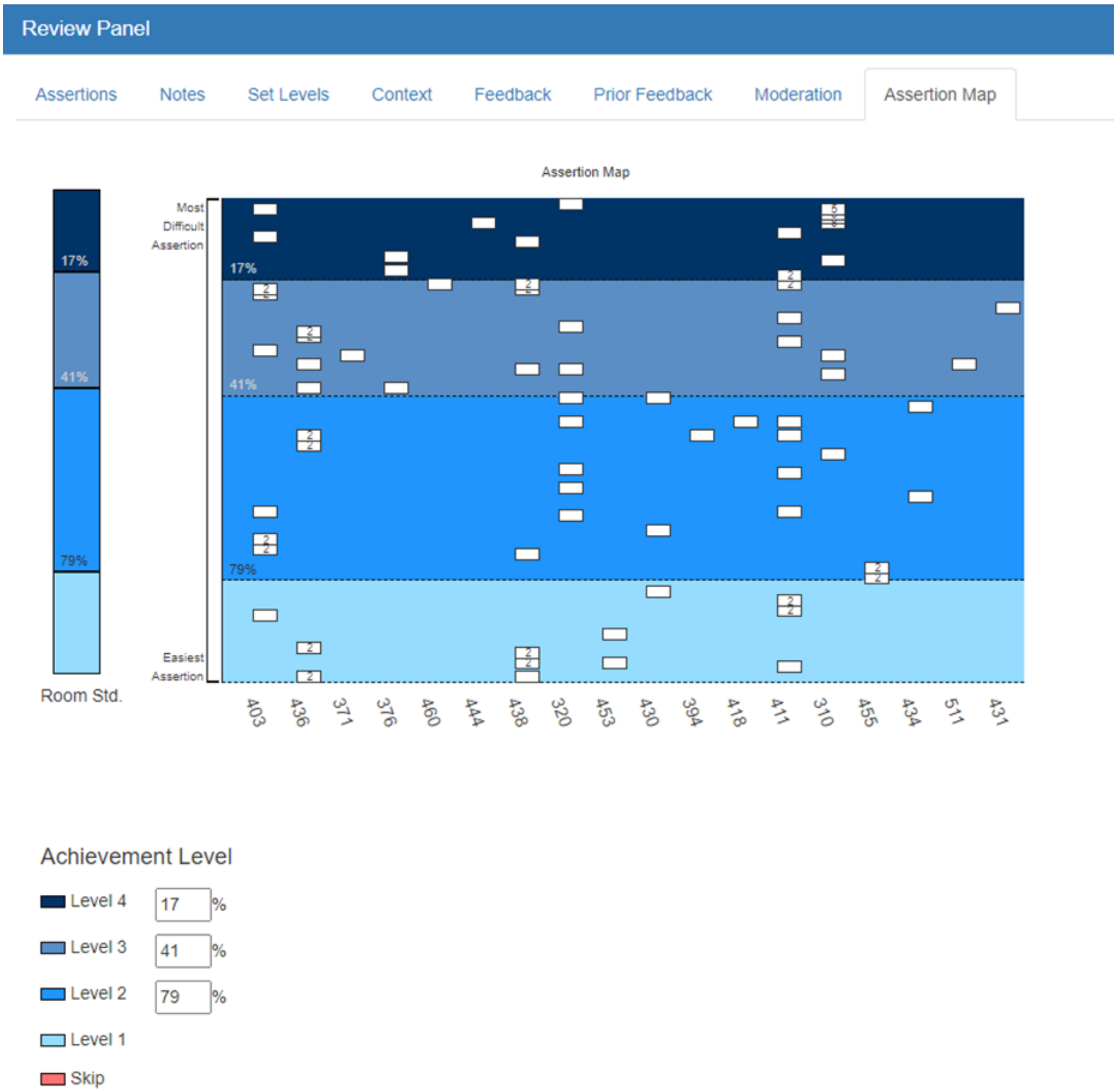


Exhibit 3-H-5. Round 2 Standard-Setting Assertion Map, Science Grade 8

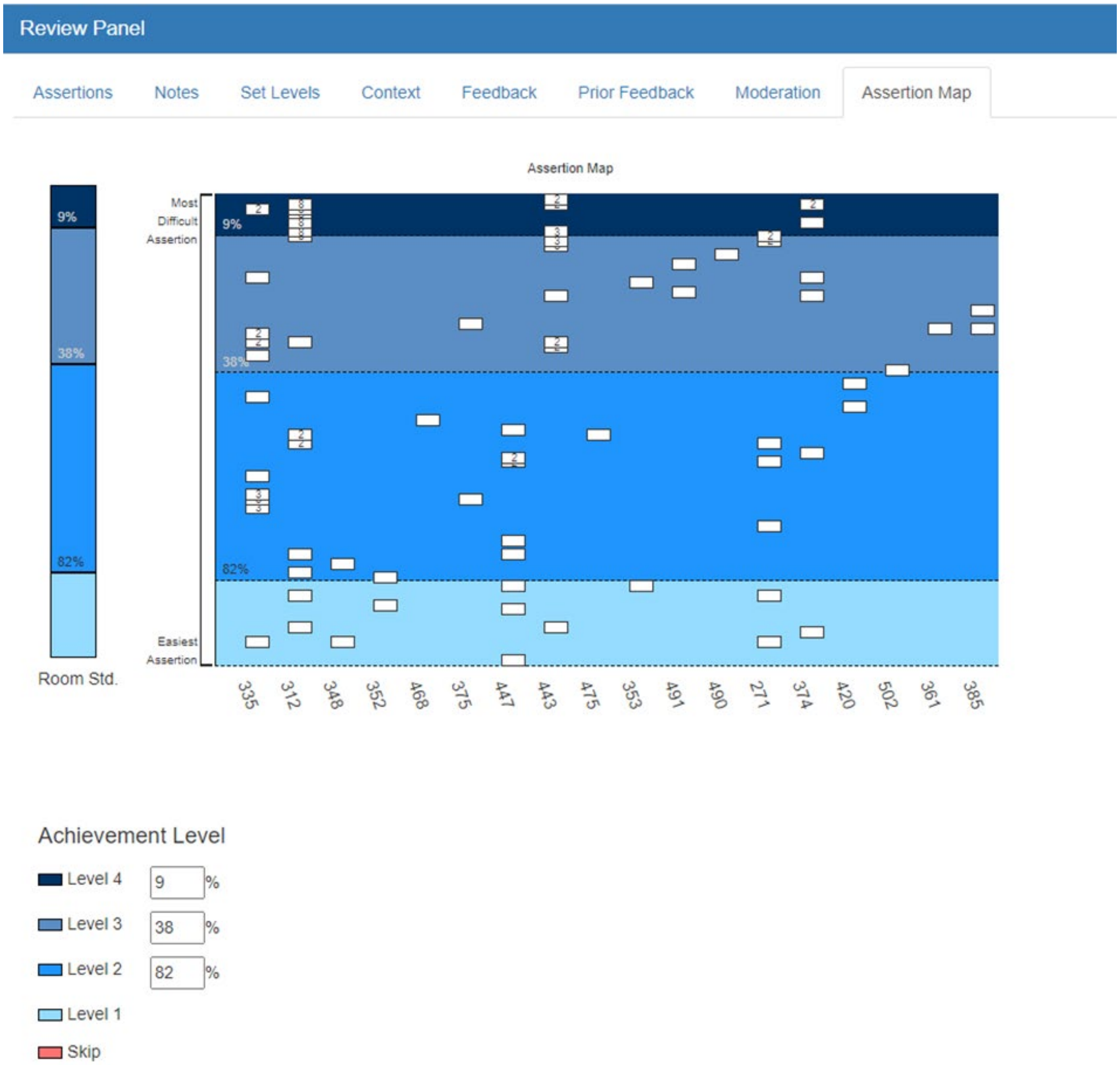
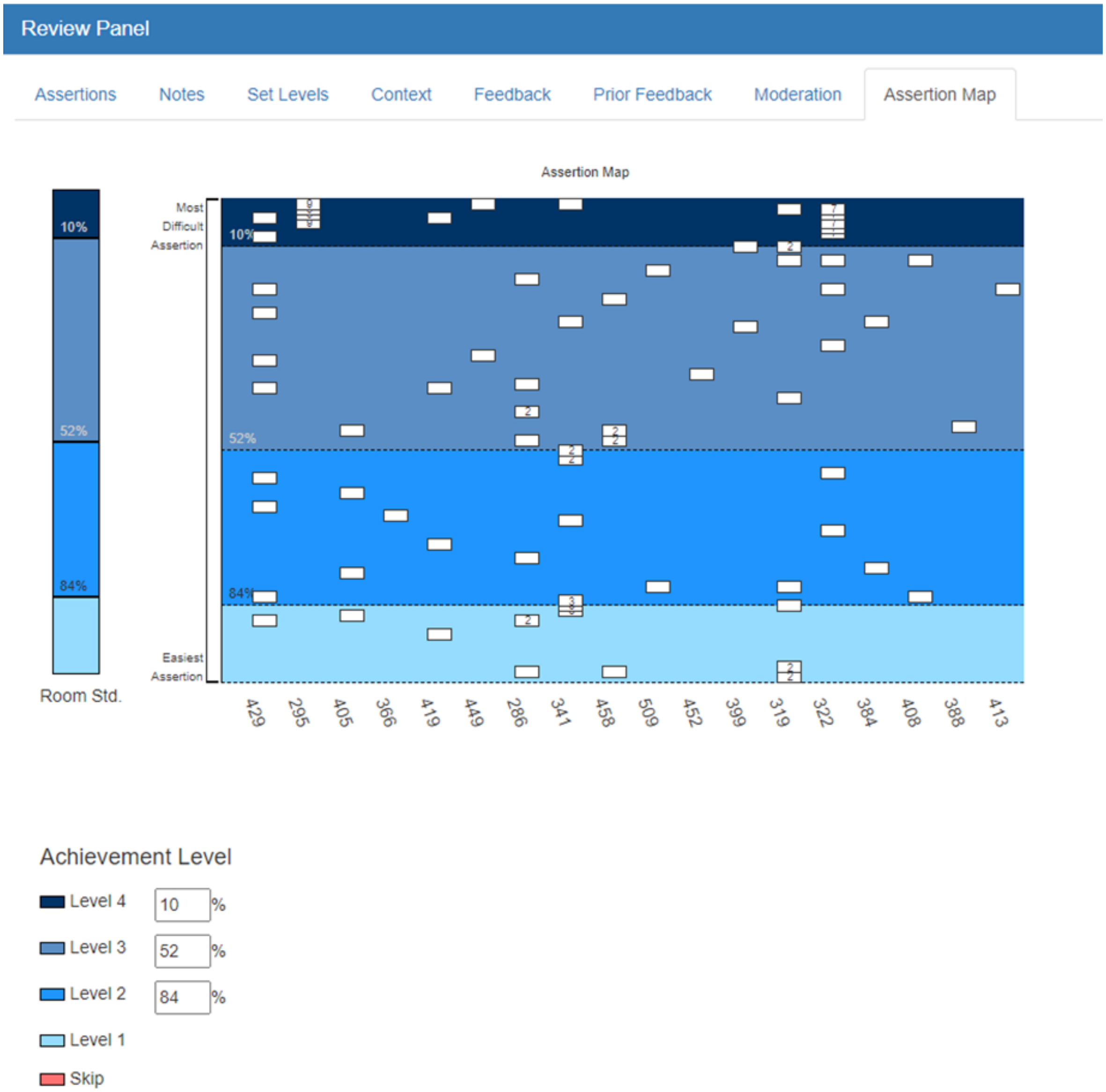


Exhibit 3-H-6. Round 2 Standard-Setting Assertion Map, Science Grade 11



South Dakota
Science Assessment

2023–2024

Addendum to Volume 3:
Confirming Achievement Standards

south dakota
DEPARTMENT OF EDUCATION
Learning. Leadership. Service.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

1. EXECUTIVE SUMMARY

In May 2015, the South Dakota State Board of Education adopted the new South Dakota Science Standards. The new standards employ a three-dimensional conceptualization of science understanding, including science and engineering practices, crosscutting concepts, and disciplinary core ideas. The South Dakota Science Assessment (SDSA) was developed to measure student achievement relative to those standards. Under contract to the South Dakota Department of Education (SDDOE), Cambium Assessment, Inc. (CAI) conducted a standard-setting workshop to recommend a system of achievement standards for the SDSA in grades 5, 8, and 11. The workshop was conducted remotely from September 15 to September 16, 2021, after the first administration of SDSA in spring 2021.

Because the standards were set in a year disrupted by the COVID-19 pandemic, SDDOE sought to verify the location of SDSA achievement standards after the 2022 administration. In response to the request from SDDOE, CAI conducted a standards confirmation workshop remotely from July 28 to July 29, 2022.

South Dakota science educators, serving as standard-setting panelists, followed a rigorous standardized procedure to confirm achievement standards demarcating each achievement level. Per grade, panelists familiarized themselves with the online SDSA testing environment, reviewed grade-band appropriate Achievement-Level Descriptors (range ALDs), and discussed grade-level threshold ALDs. Subsequently panelists acquired knowledge of scoring assertions and the assertion map, reviewed item clusters and stand-alone items in the ordered scoring assertion booklet, and evaluated scoring assertions. Panelists were provided with English language arts (ELA) and mathematics benchmark information as well as science impact data based on 2021 and 2022 operational administrations. Finally, panelists engaged in a group discussion about the defensibility of the location of the achievement standards, followed by a private vote to agree or disagree on the existing location of the achievement standards. An evaluation of the standards confirmation workshop concluded the two-day meetings.

For the standards confirmation workshop, 17 South Dakota science educators were selected to serve as science standard-setting panelists: six for the grade 5 panel, six for the grade 8 panel, and five for the grade 11 panel. Among them, four participated in the standard-setting workshop in September 2021 (one in the grade 5 panel, two in the grade 8 panel, and one in the grade 11 panel). Across the 2021 standard-setting and the 2022 standards confirmation workshops, 26 South Dakota science educators [1] formed the three grade-level panels, with nine, eight, and nine participants for grades 5, 8, and 11, respectively. 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 South Dakota Science Standards contributed to the standard setting and standards confirmation processes.

For the SDSA, each grade-level panel discussed and evaluated three achievement standards used to assign student performances into four achievement levels: Level 1 (Not Met), Level 2 (Nearly

[1] See Section 3.1.4, Educator Participants for more information on the panelists.

Met), Level 3 (Met), and Level 4 (Exceeded). Table 1 summarized the results of the standards confirmation workshop, in terms of the vote on the defensibility of the existing achievement standards. In grade 5, all six panelists voted that the Level 2, Level 3, and Level 4 standards were defensible. In grade 8, five panelists voted that the Level 2 standard was defensible, and one panelist voted that the location of the standard was not defensible; all six panelists in grade 8 voted that the Level 3 and Level 4 standards were defensible. In grade 11, all five panelists voted that the Level 2, Level 3, and Level 4 standards were defensible.

Table 1. Agree/Disagree Vote Counts of Defensibility Statement

	Defensibility Statement of Achievement Standards					
	The existing Level 2 achievement standard is defensible.		*The existing Level 3 achievement standard is defensible.*		*The existing Level 4 achievement standard is defensible.*	
Grade	**Agree**	**Disagree**	**Agree**	**Disagree**	**Agree**	**Disagree**
5	6	0	6	0	6	0
8	5	1	6	0	6	0
11	5	0	5	0	5	0

2. Introduction

South Dakota adopted the 2014 South Dakota Science Standards on May 18, 2015. The SDDOE and its assessment vendor, CAI, developed and administered a new assessment to measure the new standards. In spring 2021, the SDSA, which is aligned to the South Dakota Science Standards, was first administered to all grade 5, 8, and 11 students in South Dakota. To meet legislatively mandated score reporting requirements, SDDOE worked with CAI to conduct a standard-setting workshop for the SDSA after the spring 2021 test administration.

Achievement standards (also referred to as cut scores) were recommended for grades 5, 8, and 11 at the conclusion of the September 2021 standard-setting workshop. SDDOE sought to confirm the standards after the spring 2022 administration for the following reasons.

Although the AMP method used to recommend achievement standards in summer 2021 is primarily a content-based standard-setting procedure, panelists were provided with impact data from the 2021 SDSA administration. The spring 2021 administration was the first state test administration following the pandemic. Due to pandemic related disruptions to instruction, achievement was likely impacted and lower than achievement that would have been observed in a "regular" academic year. In addition, student participation rates, especially for some subgroups, were not consistent with pre-pandemic participation rates. In particular, there was a reduced participation of Native Americans: Native Americans had participation rates of 74.34% in grade 5, 68.95% in grade 8, and 68.64% in grade 11 in 2021 while the participation rates for this group of students were higher than 90% in previous years. Hence, impact data were not based on the entire student population, and the students that did participate were not fully representative of the entire population. Besides, SDDOE was not able to recruit the recommended number of panelists per grade band, which potentially threatens the validity of the standards.

In response to the concerns and request from SDDOE, CAI conducted a workshop designed to evaluate the appropriateness of the achievement standards. The remote standards confirmation workshop from July 28 to 29, 2022 followed a similar design and agenda as the standard-setting workshop conducted in September 2021. The main difference and key event of this standards confirmation workshop relied in the judgment task which was suggested by the Technical Advisory Committee (TAC). Rather than mapping the scoring assertions to achievement levels in two rounds, panelists was asked whether the location of the current achievement standards classify students into each of the achievement levels in a defensible way.

The purpose of this report is to document the activities in the standards confirmation workshop and to summarize the workshop results. Descriptions of the South Dakota Science Standards, introduction of the SDSA, and the September 2021 standard setting including its methodology, processes, results, and evaluations are documented in the South Dakota Science Assessment 2020–2021 Technical Reports—“Volume 3: Setting Achievement Standards.” With Volume 3, the current report—“Addendum to Volume 3: Confirming Achievement Standards” completes the development and validation of the achievement standards recommended for the SDSA.

3. CONFIRMATION STANDARD SETTING

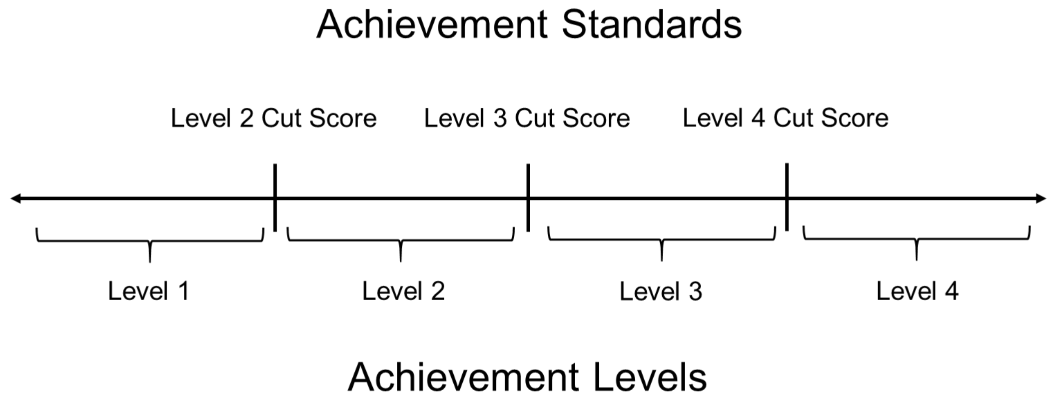
After the 2022 administration of the SDSA, 17 science educators from South Dakota convened remotely from July 28 to 29, 2022, to confirm the location of the achievement standards recommended by the standard-setting panel in 2021. Table 2 displays the achievement standards for each grade.

Achievement standards, or *cut scores*, define achievement levels that specify how many of the performance expectations students must know and be able to do in order to meet the minimum for each achievement level. As illustrated in Figure 1, three achievement standards are needed to define four achievement levels for the SDSA: Level 1 (Not Met), Level 2 (Nearly Met), Level 3 (Met), and Level 4 (Exceeded).

Table 2. Achievement Standards Recommended for Science

Grade	Level 2	Level 3	Level 4
5	477	508	527
8	773	810	836
11	1073	1102	1134

Figure 1. Three Achievement Standards Defining Four Achievement Levels for the South Dakota Science Assessment



3.1 PARTICIPANTS AND ROLES

3.1.1 South Dakota Department of Education Staff

Staff from the SDDOE were present throughout the process and provided overall policy context and answered any policy questions that arose.

From SDDOE, attendees included

- Matthew Gill, Administrator of the Office of Assessment & Accountability;
- Christina Booth, Program Specialist for General Assessment Support, Science Assessment, and Science ALT Assessment and MSAA; and
- Beth Schiltz, Program Specialist for Special Education.

3.1.2 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.

- Sandi Hendrick, Jennifer Chou, Mackenzie Worn, and Kimberly David, Program Management Team, managed processes and logistics throughout the meeting.
- Floyd Helm and Mark Palomo, System Support Agents, troubleshot technology during the workshop.

3.1.3 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. They had led standard-setting processes before, and they could answer any questions about the workshop, 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 as follows:

- Kevin Dwyer, Hibbah Haddam, and Olivia Francois facilitated the science grade 5 panel.
- Vanessa Johnson and Mark Warner facilitated the science grade 8 panel.
- Matthew Davis and Jared Taylor facilitated the grade 11 panel.

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

3.1.4 Educator Participants

To verify achievement standards, SDDOE recruited participants from across the state. Panelists included science teachers from general or special education, administrators, and representatives from other stakeholder groups (e.g., coaches) to ensure that a range of perspectives contributed to the standard-setting process and the product. In recruiting panelists, SDDOE targeted the recruitment of participants to be representative of the gender and geographic representation of South Dakota's teacher population. All participants also had to be familiar with the South Dakota Science Standards content and tests.

SDDOE selected Seventeen educators from the resulting potential panelist pool and invited them to participate in the standards confirmation workshop. Six educators were assigned to the grade 5 panel, six to the grade 8 panel, and five to the grade 11 panel. Among them, one in the grade 5 panel, two in the grade 8 panel, and one in the grade 11 panel already participated in the September 2021 standard-setting workshop.

For the standards confirmation workshop, panelists were 24% male and 0% non-white. Represented stakeholder groups included Administrators, Coaches, General Education Teachers, and Special Education Teachers, with General Education Teachers comprising 71% of the panels overall. The majority of panelists taught in the grades to which they were assigned to verify standards. Overall, 41% of panelists taught elementary school and the others taught some combination of grades in middle and high schools. Panelists worked in schools (59%), schools and districts (29%), one was an assistive technology integrationist in the district (6%), and one was a retired middle and high school teacher (6%). School district areas included rural (47%), suburban (24%), and urban (29%), and were small (53%), medium (35%), and large (12%). Table 3 summarizes the characteristics of the panels.

Table 3. Panelist Characteristics

	Percentage (%) of Panelists, by Panel			
	Science Grade 5	*Science Grade 8*	*Science Grade 11*	*Overall*
Characteristics				
Male	0	17	60	24
Non-White	0	0	0	0
Stakeholder Groups[a]				
Coach	0	50	20	24
General Education Teacher	50	83	80	71
Special Education Teacher	17	17	20	18
Other[b]	33	0	0	18
Current Position				
School	50	50	80	59
District	17	0	0	6
School and District	17	50	20	29
Other[c]	17	0	0	6
School District Area Size				
Large	50	17	100	53
Medium	33	67	0	35
Small	17	17	0	12
School District Area Urbanicity				
Rural	50	83	0	47
Suburban	17	0	60	24
Urban	33	17	40	29
Primary Grades Taught				
ES (grades 1–5)	17	0	0	6
MS (grades 6–8)	33	83	0	41
ES and MS (Preschool, Kindergarten, grades 1–8)	17	0	0	6
MS and HS (grades 6–12)	17	17	0	12
HS (grades 9–12)	17	0	100	35

Note. [a]The total sums to over 100% for "Stakeholder Groups" as participants had multiple roles in local education systems. [b]Others included an English Language (EL) teacher, an assistive technology integrationist, and a retiree. [c]One panelist was a retired science teacher after over 20 years of services.

For the results of any judgment-based method to be valid, the judgments must be made by individuals who are qualified to perform judgment tasks. Participants in the SDSA standards confirm workshop were highly qualified. They brought a variety of experience and expertise. Overall, 47% of panelists had earned a master's degree; 94% had taught in their assigned panel's grade. The average time teaching the South Dakota Science Standards was over six years. Many had experience teaching special populations: 88% taught students eligible to receive free or reduced-price lunch, 76% taught English learners (ELs), and 88% taught students on an Individual Education Plan (IEP). Table 4 summarizes the qualifications of the panels. Appendix A, Standards Confirmation Workshop Panelist Characteristics, provides additional information about the individuals participating in the standards confirmation workshop.

Table 4. Panelist Qualifications

	Percentage (%) of Panelists, by Panel			
	Science Grade 5	*Science Grade 8*	*Science Grade 11*	*Overall*
Highest Degree				
Bachelor	50	50	60	53
Master	50	50	40	47
Years of Teaching Experience				
1–5 years	33	17	60	35
6–10 years	0	0	20	6
11–15 years	17	33	0	18
16–20 years	17	33	0	18
More than 20 years	33	17	20	24
Years of Teaching Experience in Assigned Grade				
None	17	0	0	6
1–5 years	33	17	60	35
6–10 years	33	33	20	29
11–15 years	17	50	0	24
More than 20 years	0	0	20	6
Subject Areas Currently Teaching[a]				
English Language Arts (ELA)	33	33	0	24
Mathematics	33	17	0	18
Social Studies	0	17	0	6
Science	67	100	100	88
Other[b]	17	17	20	12
Years of Professional Experience in Education Other Than Classroom Teaching[c]				
None	67	83	80	76
1–5 years	17	17	20	18
6–10 years	17	0	0	6
Experience Teaching Special Student Populations				
Students eligible to receive free/reduced price lunch	100	83	100	88
English Learners (ELs)	83	67	100	76

	Percentage (%) of Panelists, by Panel			
	Science Grade 5	*Science Grade 8*	*Science Grade 11*	*Overall*
Students on an Individual Education Plan (IEP)	100	83	100	88
Years of Experience with the South Dakoda Science Standards[d]				
Average years teaching the South Dakota Science Standards	7.9	4.7	6.6	6.4

Note. [a]The total sums to over 100% for "Subject Areas Currently Teaching" as many participants taught multiple subjects/areas. [b]Other Areas Currently Teaching included behavior, vocational skills, and special education. [c]Years of Professional Experience in Education Other Than Classroom Teaching included administrative positions in schools, specialists, etc. The current panels had an educator as an instructional coach and another educator. [d]For experience over 20 years and less than a year, we used 20.5 and 0.5 years, respectively, in aggregation.

3.2 MATERIALS

3.2.1 Achievement-Level Descriptors

Achievement-Level Descriptors (ALDs) provide definitions in terms of content-area knowledge, skills, and processes that students at each achievement level can demonstrate. While ALDs link the content standards to the achievement standards, there are four types: policy ALDs, range ALDs, threshold ALDs, and reporting ALDs. Introduction to ALDs was included in the SDSA 2020–2021 Technical Reports—Volume 3: Setting Achievement Standards. The Policy-Level Descriptors (i.e., policy ALDs) of the four achievement levels used in South Dakota were also addressed.

During the July 2022 standards confirmation workshop, range ALDs and threshold ALDs were profoundly relied on in different activities. Range ALDs are detailed grade- and content-area-specific descriptions that communicate exactly what students performing at each level know and can do. Threshold ALDs were created during and used at both the standard-setting and standards confirmation workshops, as they describe what a student just barely scoring into each achievement level knows and can do.

Prior to the September 2021 standard-setting workshop, SDDOE drafted range ALDs that describe observable evidence for what student performance looks like in science at each achievement level and grade. SDDOE and CAI reviewed the draft range ALDs to ensure that the language accurately represented the goals and policies of the state; revisions were made wherever necessary at this stage. Next, the range ALDs were reviewed, revised, and approved by the group of South Dakota educators selected to be table leaders at the 2021 workshop. The final range ALDs were then used to guide panelists to transform content standards to achievement standards on the reporting score scale. The final range ALDs for the SDSA are documented in Appendix B of the SDSA 2020–2021 Technical Reports—Volume 3: Setting Achievement Standards—South Dakota Science Assessment Range Achievement-Level Descriptors.

3.2.2 Ordered Scoring Assertion Booklets

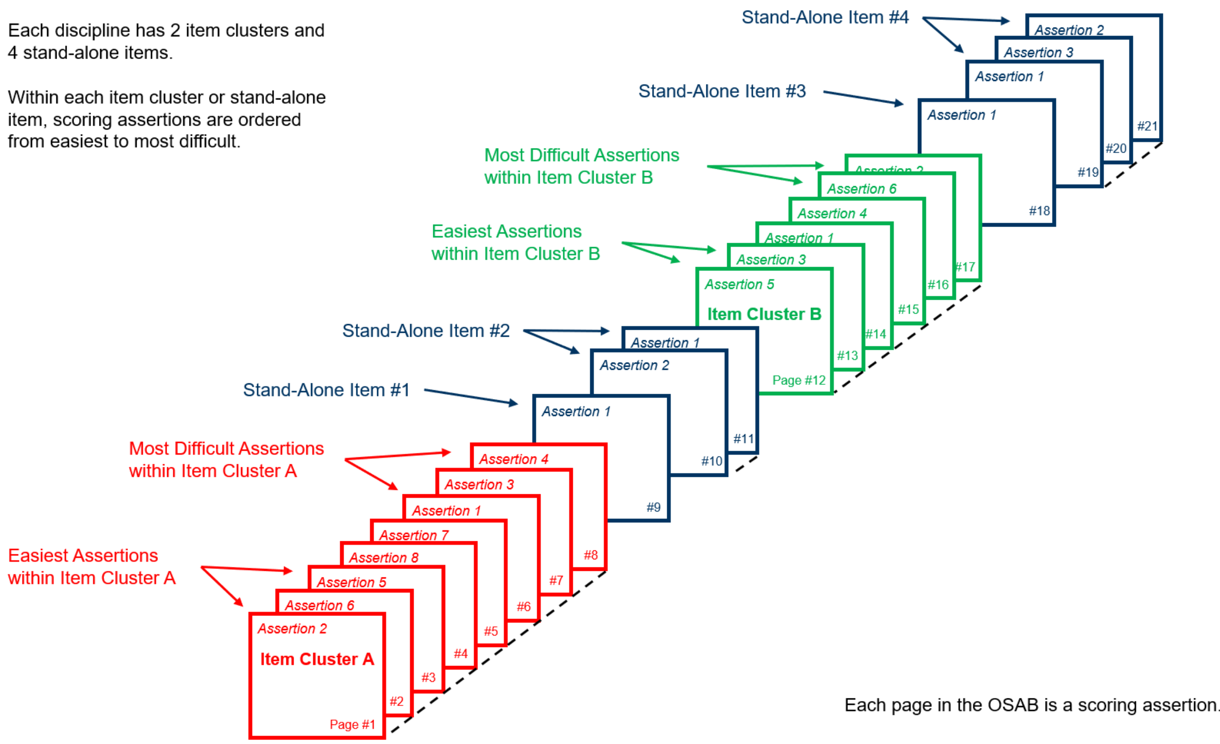
The test-centered Assertion-Mapping Procedure (AMP; Rijmen, Cohen, Butcher, & Farley, 2018) is the method for establishing achievement standards for the SDSA. The AMP method uses booklets of ordered test materials—scoring assertions—to present content-balanced and psychometrically-sound testing instance to panelists. The booklets are called ordered scoring assertion booklets (OSABs).

Prior to the September 2021 standard-setting workshop, CAI psychometricians created three grade-specific OSABs based on the operational test blueprint and the item pool of the SDSA. With psychometricians, CAI content experts reviewed the OSABs before SDDOE reviewed and approved them. The items and scoring assertions selected to construct the OSAB ensured the test blueprint of the SDSA were met and the reported score gaps were minimal. To minimize cognitive load for panelists when implementing the AMP method, the total number of scoring assertions was not allowed to exceed 85.

The creation and evaluation of the items and their assertions on the OSABs are documented in the SDSA 2020–2021 Technical Reports. In the July 2022 standards confirmation workshop, the OSABs consisted of the same items and scoring assertions used in the September 2021 workshop. Each OSAB contained three disciplines and 18 items, including six item clusters and 12 stand-alone items. The grade 5 OSAB contained 75 assertions, the grade 8 OSAB contained 77 assertions, and the grade 11 OSAB contained 83 assertions. The OSABs were presented to the panelists using CAI's online standard-setting tool and enabled the display of complex item clusters and stand-alone items, as well as contextual information, and item attributes such as performance expectation alignment.

Figure 2 shows an example of the structure of the OSAB. Within each item cluster or stand-alone item, scoring assertions are ordered by difficulty. Easier assertions are those that most students are able to demonstrate, and difficult assertions are those that the fewest students are able to demonstrate. Note that assertions are ordered by difficulty within items only. Across all items, this is generally not the case; for example, the most difficult assertion of an item presented early on in the OSAB is typically more difficult than the easiest assertion of the next item in the OSAB. That is, the order of assertions in Figure 2 represents the order of presentation to the panelists, but assertions are not ordered by overall difficulty across all items.

Figure 2. Ordered Scoring Assertion Booklet (OSAB)



3.2.3 Benchmark Information and Impact Data

To be adoptable, achievement standards for a statewide system must be coherent across grades and subjects. To that end, panelists are provided with benchmark data from other assessments as well as impact data during the standard-setting workshop. Analogous to the September 2021 standard-setting workshop, the South Dakota English language arts (ELA) and mathematics assessments were provided as benchmarks in the July 2022 standard confirmation workshop.

For purposes of confirming the achievement standards, panelists were provided with the location of the current SDSA achievement standards. CAI calculated and provided the impact data from the 2022 administration for the current SDSA achievement standards. In addition, for each assertion, panelists were provided with the percentage of students that performed at or better than the level of achievement associated with the specified response probability value of the assertion.

Having been provided with the benchmark information and impact data, panelists were instructed to use them as part of their OSAB review during the standards confirmation workshop. Panelists received instruction on the benchmark information and impact data before reviewing the OSAB.

3.2.4 Assertion Maps

Assertion maps accompanied the OSABs and provided a visual representation of the assertion difficulties within and across items. They were provided to panelists to help reduce the cognitive load of the AMP during their OSAB review.

The 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 helped panelists easily identify more- or less-difficult assertions and compare the difficulty of assertions across items. The assertion maps provided during the standards confirmation workshop also displayed the achievement standards recommended by the September 2021 standard-setting workshop.

3.3 Workshop Technology

In the July 2022 standards confirmation workshop, panelists used CAI's online application to experience testing environments and items, to review OSABs, and to read assertion maps. 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 provided his or her judgments to evaluate the defensibility of the location of the achievement standards.

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

3.4 Events

The standards confirmation workshop occurred over two days. Table 5 summarizes the daily events, followed by more detailed descriptions of each event. Appendix B, Standards Confirmation Workshop Agenda, provides the full workshop agenda.

Table 5. Summary of the Standards Confirmation Workshop Agenda

Day 1: Thursday, July 28, 2022
• Participate in Large-Group Orientation • Review and Take the Operational Test • Review Range ALDs • Discuss Threshold ALDs • Review OSAB
Day 2: Friday, July 29, 2022
• Continue Reviewing OSAB • Review Achievement Standard Judgment Task • Discuss Defensibility of Current Achievement Standards • Cast Private Vote • Evaluate Standards Confirmation Workshop

3.4.1 Practice Technical Check and Participant Login

Panelists were required to attend a technical check prior to the standards confirmation workshop to ensure they had access to the required sites needed for 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.

3.4.2 Attend Large-Group Introductory Training

Matthew Gill and Christina Booth, SDDOE, welcomed panelists to the workshop and provided context and background for the SDSA. Matthew Gill outlined the roles and responsibilities of the participants at the workshop: panelists, CAI staff, and SDDOE personnel. Dr. Rijmen from CAI then oriented participants to the workshop by describing the purpose and objectives of the meeting, explaining the processes to be implemented to meet those objectives, and outlining the events that would occur each day. He explained that panelists were selected because they were experts. He continued that the processes to be implemented over the two days were designed to elicit and apply the panelists' expertise to evaluate whether the achievement standards adopted for the SDSA were defensible. Finally, he described what the judgment task would require, how individual, private vote would be done, and what would happen once all the votes were collected. Appendix C, Standards Confirmation Workshop Training Slides, provides the slides used during the large-group training.

3.4.3 Comply with 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 operational SDSA pool, requiring confidentiality to maintain their security. Participants were forbidden to do the following either during, or after, the workshop:

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

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

3.4.4 Take the Test

Following the large-group orientation, panelists broke out into their separate grade-level virtual meeting rooms. As their introduction to the standards confirmation, panelists took a form of the test that students took in 2021, 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 as students take provided 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.

3.4.5 Review Range Achievement-Level Descriptors and Discuss Threshold Achievement-Level Descriptors

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 setting achievement standards. The ALDs were used as a reference for evaluating student performance and for developing a representation of students who just barely qualify for entry into each of the achievement levels based on the ALDs. Thus, it was important for panelists to understand the critical role of ALDs to evaluate the appropriateness of the locations of the achievement standards.

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 that students scoring just below the achievement level do not have. Characterizing just barely qualified students is not an intuitive judgment, and panelists spent time working to identify the minimum characteristics of student achievement for entry into each achievement level. Each panel produced a set of threshold/just barely ALDs to help guide their discussions and evaluation of the achievement standard locations. To develop a common understanding among panelists, each panel was asked to

1. review and parse range ALDs;
2. discuss characteristics of students classified near thresholds of achievement standards
3. identify the characteristics that distinguish students just above the achievement standard from those just below;
4. determine what evidence was necessary to conclude that a student possessed the minimum knowledge and skills needed to meet the achievement standard; and
5. summarize knowledge and skills of students who just barely meet each achievement standard, or are just barely described by each ALD.

These discussions were intended to yield common descriptions of students just barely characterized by each ALD within each panel room. The purpose of the threshold ALD discussion was to enhance the panelists' understanding of the differences between ALD levels by paying attention to the transition areas between achievement levels. Threshold ALDs were used to evaluate the appropriateness of the location of the achievement standards.

3.4.6 Review Ordered Scoring Assertion Booklet

After completing the review and discussion of the ALDs, panelists began reviewing the OSAB, including the item interactions that gave rise to the scoring assertions, as well as the associated contextual information for each scoring assertion within CAI's standard-setting tool. For each scoring assertion, panelists were instructed to consider the item interactions and the scoring assertions derived from the interactions. For each assertion, panelists were asked to consider the following:

- How does the student interaction give rise to the assertion? Did they plot, select, or write something?
- Why is this assertion more difficult to achieve than the previous one (within the item)?
- Are the knowledge and skill requirements of this assertion consistent with the achievement-level classification?

The first two questions were the same questions that panelists consider in the standard AMP method. The third question was intended to calibrate panelists' judgments about the achievement levels indicated by each assertion in the OSAB.

3.4.7 Perform Judgment Task and Cast Private Vote

After reviewing the ALDs, developing representations of students who just barely qualify for entry into each of the achievement-level classifications, and reviewing the OSAB, panelists prepared to perform the achievement standard judgment task. The question of the judgment task was whether the location of the current SDSA achievement standards was defensible.

Panelists were oriented to the location in the assertion map of the current achievement standards. They engaged themselves in a group discussion about whether the scoring assertions located near each of the current achievement standards accurately differentiate students who just barely qualify for entry into the achievement level from students who do not yet qualify.

The workshop facilitator guided the panelists through the group discussion. After confirming their readiness to vote using Appendix D, Standards Confirmation Readiness Form, via secure Google forms, panelists independently cast an individual vote in response to the defensibility statement for each achievement level. Panelists chose to "Agree" or "Disagree" with each of the following statements:

- *The existing Level 2 achievement standard is defensible.*
- *The existing Level 3 achievement standard is defensible.*
- *The existing Level 4 achievement standard is defensible.*

Panelist votes were collected immediately. CAI summarized the vote results and sent them to SDDOE for final decisions.

3.5 PRIVATE VOTE RESULTS

The vote counts of "Agree" and "Disagree" with the defensibility statement were tallied. In grade 5, all six panelists voted that the Level 2, Level 3, and Level 4 standards were defensible. In grade 8, five panelists voted that the Level 2 standard was defensible, and one panelist voted that the location of the standard was not defensible; all six panelists in grade 8 voted that the Level 3 and Level 4 standards were defensible. In grade 11, all five panelists voted that the Level 2, Level 3, and Level 4 standards were defensible. The CAI psychometric team collected the results from each grade panel, summarized the findings, and delivered them to SDDOE.

The grade 8 panelist who disagreed with the current location of the Level 2 achievement standard indicated that for some items, their Level 2 assertions did not line up with the Level 2 ALDs but the Level 3 ALDs, which was the reason for the disagreement.

3.6 Student Performance

Given that the standards were deemed defensible by all panelists except for one panelist for the Level 2 cut score of grade 8, the 2022 student performance is summarized with respect to the achievements standards that were recommended in 2021. Figure 3 displays the percentage of students that reached or exceeded each of the recommended achievement standards in 2022.

Figure 3. Percentage of Students Reaching or Exceeding Each Recommended Science Achievement Standard in 2022

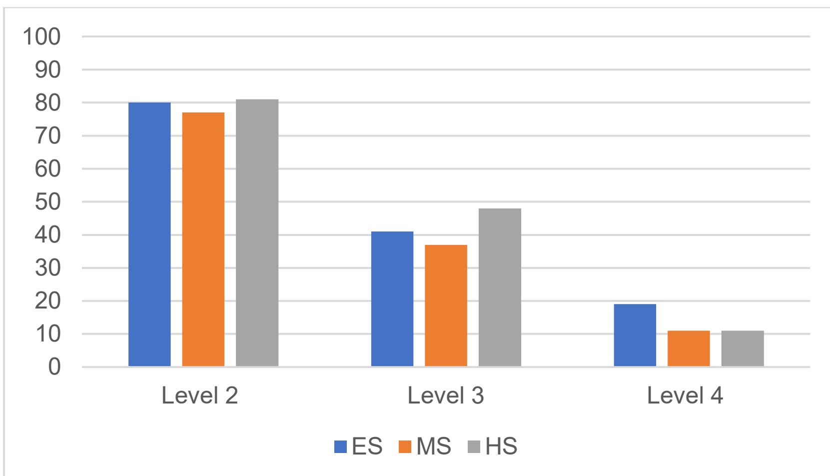
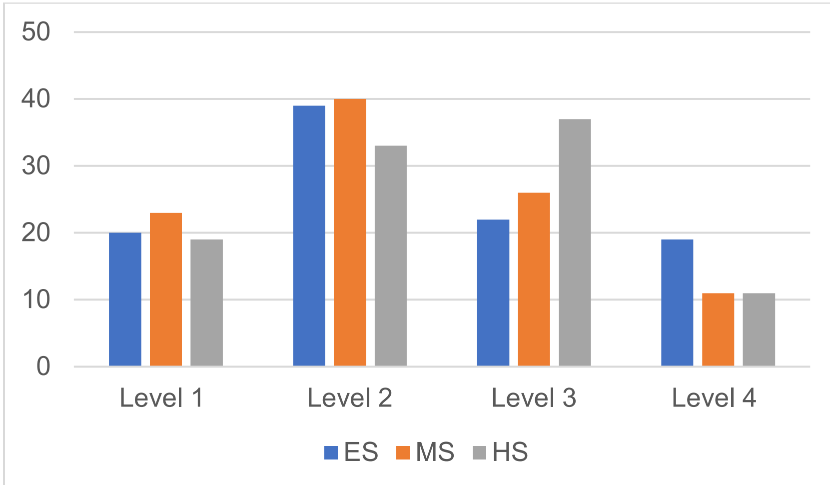


Table 6 shows the percentage of students classified within each of the achievement levels in 2022. The values are displayed graphically in Figure 4.

Table 6. Percentage of Students Classified Within Each Science Achievement Level in 2022

Grade	Level 1	Level 2	Level 3	Level 4
5	20	39	22	19
8	23	40	26	11
11	19	33	37	11

Figure 4. Percentage of Students Classified Within Each Science Achievement Level in 2022



Compared to the percentage of students per achievement level in 2021[2], across two years it was found for grades 5 and 11, the percentages were similar within a level, with a difference of 3% or less. For grade 8, there were a higher percentage of students at Level 1 (23% in 2022 vs. 18% in 2021), a lower percentage at Level 2 (40% in 2022 vs. 44% in 2021), and a slightly higher percentage of students at Level 4 (11% in 2022 vs. 9% in 2021). An investigation on the scale score distributions revealed that the grade 8 score distribution of 2022 scores had a larger variance than the score distribution of 2021 scores, which explains to discrepancies in the percentage of students in achievement levels addressed here.

3.7 Workshop Evaluations

After finishing all activities, panelists completed online workshop evaluations independently, in which they described and evaluated their experience taking part in the standards confirmation workshop. Table 7, Table 8, Table 9, and Table 10 summarize the results of the evaluations.

Panelists' understanding of the standards confirmation workshop was evaluated by self-reported level of agreement on statements that described workshop processes and tasks. In general, panelists had good understanding of the workshop processes (see Table 7). Of the panelists who did not agree,

- one grade 8 panelist disagreed with the statement regarding being able to understand just barely ALDs;
- another grade 8 panelist disagreed that the ALDs provided clear expectations and disagreed that the assertion map, the benchmark data, and the impact data were helpful when judging the appropriateness of the achievement standards; and
- one grade 11 panelist disagreed with the statement regarding comfortability expressing their opinions throughout the workshop.

[2] See Table 12 and Figure 13 in the SDSA 2020–2021 Technical Reports—Volume 3: Setting Achievement Standards.

Table 7. 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"**			
	Science Grade 5	*Science Grade 8*	*Science Grade 11*	*Overall*
I understood the purpose of this standards confirmation workshop.	100	100	100	100
The procedures used to complete the achievement standard judgment task were fair and unbiased.	100	100	100	100
The training provided me with the information I needed to complete the subsequent tasks.	100	100	100	100
Taking the online assessment helped me to better understand what students need to know and be able to do to receive credit for each assertion.	100	100	100	100
The Achievement-Level Descriptors (ALDs; description 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	83	100	94
I was able to develop an understanding of the knowledge and skills demonstrated by students who are just barely described by the ALDs.	100	83	100	94
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 receive credit for each assertion.	100	100	100	100
I found the assertion map helpful when judging the appropriateness of the achievement standards.	100	83	100	94
I found the benchmark data and discussions helpful when judging the appropriateness of the achievement standards.	100	83	100	94
I found the impact data (percentage of students that would achieve at the level indicated by the assertion difficulty) helpful when judging the appropriateness of the achievement standards.	100	83	100	94
I felt comfortable expressing my opinions throughout the workshop.	100	100	80	94
Everyone was given the opportunity to express his or her opinions throughout the workshop.	100	100	100	100

Note. Number of responses = 17 (grade 5 responses = 6, grade 8 responses = 6, and grade 11 responses = 5). Evaluation response options included "Strongly Disagree," "Disagree," "Agree," and "Strongly Agree."

Clarity of the workshop materials and process was evaluated. Overall, panelists overwhelmingly indicated clarity in the instructions, materials, data, and process (see Table 8).

Table 8. 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"**			
	Science Grade 5	*Science Grade 8*	*Science Grade 11*	*Overall*
Instructions provided by the workshop leader	100	100	100	100
Achievement-Level Descriptors (ALDs)	100	100	100	100
Ordered Scoring Assertion Booklet (OSAB)	100	100	100	100
Assertion Map	100	100	100	100
Impact Data (percentage of students that would achieve at the level indicated by the assertion difficulty)	100	100	100	100

Note. Number of responses = 17 (grade 5 responses = 6, grade 8 responses = 6, & grade 11 responses = 5). Evaluation response options included "Very Unclear," "Somewhat Unclear," "Somewhat Clear," and "Very Clear."

Appropriateness of the standards confirmation process was evaluated in terms of the amount of time given to workshop panelists. Panelists felt that the time allocated to various workshop tasks may be adjusted (see Table 9). Of the panelists who did not think the amount of time for the task was about right

- two grade 5 panelists indicated all the tasks were too long;
- two grade 8 panelists and one grade 11 panelist indicated having too much time for taking the tests;
- three grade 8 panelists and one grade 11 panelist reported having too much time to review the ALDs;
- one grade 8 panelist and one grade 11 panelist reported having too little time to discuss the skills demonstrated by students who are just barely described by each ALD, and four from the grade 8 panel reported having too much time for the discussion;
- three grade 8 panelists and one grade 11 panelist indicated having too much time to review the OSAB while one grade 11 panelist reported having too little time for this task; and
- two grade 8 panelists indicated having too much time for the achievement judgment task.

Table 9. Evaluation Results: Appropriateness of Process

How appropriate was the amount of time you were given to complete the following components of the standards confirmation process?	**Percentage (%) Indicating "About Right"**			
	Science Grade 5	*Science Grade 8*	*Science Grade 11*	*Overall*
Large-group orientation	67	67	60	65
Experiencing the online assessment	67	67	80	71
Reviewing the Achievement-Level Descriptors (ALDs)	67	50	80	65
Discussion of the skills demonstrated by students who are just barely described by each ALD	67	17	80	53
Reviewing the Ordered Scoring Assertion Booklet (OSAB)	67	50	60	59
Achievement standard judgment task	67	67	100	76

Note. Number of responses = 17 (grade 5 responses = 6, grade 8 responses = 6, & grade 11 responses = 5). Evaluation response options included "Too Little," "Too Much," and "About Right."

Importance of the materials during standards confirmation was accessed. Participants appreciated the importance of the multiple factors contributing to achievement standard judgment task, with all but two grade 8 panelist rating just barely ALDs not important. One of the two panelists also indicated ALDs and impact data were not important (see Table 10).

Table 10. 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"**			
	Science Grade 5	*Science Grade 8*	*Science Grade 11*	*Overall*
Achievement-Level Descriptors (ALDs)	100	83	100	94
Just barely ALDs	100	67	100	88
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	100	100	100	100
External benchmark data	100	100	100	100
Impact data (percentage of students that would achieve at the level indicated by the assertion difficulty)	100	83	100	94

Note. Number of responses = 17 (grade 5 responses = 6, grade 8 responses = 6, and grade 11 responses = 5). Evaluation response options included "Not Important," "Somewhat Important," and "Very Important."

3.7.1 Workshop Participant Feedback

Finally, panelists responded to two open-ended questions: "What suggestions do you have to improve the training or standards confirmation process?" and "Do you have any additional comments? Please be specific."

Eleven panelists responded to the first question, and 16 responded to the second. Most responses were positive comments expressing appreciation for the workshop and the opportunity to participate. Panelists also indicated the training was effective and the process was clear.

Participants provided minor suggestions, such as shortening or lengthening the time allocated for some tasks. Many appreciated the SDDOE and testing vendor, well-prepared materials, technology, and technical support, and many panelists complimented the professionalism and expertise of the facilitators.

A returning panelist commented:

"I participated in the standard setting in September '21, so it was interesting to come back and see the work we had done and talk with others about it. It was reaffirming to see the work that we had done resulted in pretty consistent outcomes between the 2021 and 2022 testing. I appreciate the work of everyone involved as this is a big undertaking."

A panelist actively participating in standard-setting workshops for South Dakota commented:

"I liked it the way it was. I was glad I participated in setting cut scores with the ALT standards. It really helped me understand the whole process better."

Additional panelist comments included:

"Our group leader Matt was excellent and really made the sessions balanced, the tasks understandable, and on track."

"I liked how the standards were combined under their specific content, it made the process of going through ALDs much easier."

"Thank you for allowing me to be a part of this process. It was very interesting for me, especially because I often teach these standards at a basic level."

"Thank you for a great experience. Cambium tech team were amazing and patient."

4. VALIDITY EVIDENCE

The achievement standards of the SDSA are a crucial component to determining whether students have met the learning objectives defined by the South Dakota Science Standards. Evidence to support the development, use, and interpretation of the achievement standards was collected and documented in the SDSA 2020–2021 Technical Reports—Volume 3: Setting Achievement Standards. The standards confirmation workshop was conducted to further validate the achievement standards.

Internally, panelists achieved consensus to agree on the defensibility of the achievement standards for each grade. Impact data from the 2021 and 2022 administrations showed consistent patterns for percentage of students reaching or exceeding each recommended achievement standards and for percentage of students classified within each achievement level. To address the concern regarding low participation rates among Native American students, additional analysis showed that their participation rates in 2022 were 98%, 95%, and 100% for grades 5, 8, and 11, respectively. These were much higher than the participation rates in 2021 (74% for grade 5, 69% for grade 8, and 69% for grade 11). Participation rates of other race/ethnicity groups, including African American, Asian, Hispanic, Multi-Racial, Pacific Islander, and White, were also computed from the 2022 administration and found to be fairly high across groups (ranging from 90% to 100%). High student participation in the 2022 administration ensured that the impact data used in the standards confirmation workshop were an accurate reflection of the achievement of the 2022 student populations.

Procedurally, the purpose and processes of the workshop were articulated and well perceived by panelists. The judgment task for panelists was straightforward. Panelists reported confidence in the workshop processes and outcomes, providing evidence to support the validity of the standards confirmation workshop process and procedures. The current report served as comprehensive documentation to feature the standards confirmation process.

5. References

Rijmen, F., Cohen, J., Butcher, T., & Farley, D. (2018, June 28). Scoring and reporting for assessments developed for the new science standards [Symposium]. National Conference on Student Assessment, San Diego, CA, United States.

South Dakota Department of Education. (2021). *The South Dakota Science Assessment 2021–2022 Technical Reports—Volume 3: Setting Achievement Standards*. Pierre, SD: Author.

Appendix 3-A

Standards Confirmation Workshop Panelist Characteristics

Standards Confirmation Workshop Panelist Characteristics

Table 3-A-1. Standards Confirmation Workshop 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 South Dakota Science Standards	School District Size	School District Area Urbanicity
General Education Teacher	School	Female	White	Master's degree (e.g., M.A., M.S.)	1 to 5	None	4	Medium	Rural
Retired science teacher	home	Female	White	Master's degree	More than 20	None	20+	Medium	Rural
Assistive Technology Integrationist Birth-21	District	Female	White	Master's degree	16 to 20	6 to 10	5	Large	Urban
Special Education Teacher	School	Female	White	Bachelor's degree (e.g., B.A., B.S.)	More than 20	None	1	Small	Rural
General Education Teacher	School	Female	White	Bachelor's degree	1 to 5	1 to 5	3	Large	Suburban
General Education Teacher	School & District	Female	White	Bachelor's degree	11 to 15	None	14	Large	Urban

Table 3-A-2. Standards Confirmation Workshop 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 South Dakota Science Standards	School District Size	School District Area Urbanicity
Coach	School & District	Female	White	Master's degree	16 to 20	1 to 5	19	Large	Urban
General & Special Education Teacher	School & District	Female	White	Master's degree	More than 20	None	2	Medium	Rural
General Education Teacher	School	Female	White	Master's degree	1 to 5	None	4	Medium	Rural
General Education Teacher, Coach & Parent	School	Female	White	Bachelor's degree	11 to 15	None	2	Medium	Rural
General Education Teacher	School	Female	White	Master's degree	16 to 20	None	1	Medium	Rural
General Education Teacher, Coach & Parent	School & District	Male	White	Bachelor's degree	11 to 15	None	This is the first year I have been part of the process	Small	Rural

Table 3-A-3. Standards Confirmation Workshop 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 South Dakota Science Standards	School District Size	School District Area Urbanicity
Special Education Teacher	School	Female	White	Master's degree	6 to 10	None	7	Large	Urban
General Education Teacher, Coach, ELL Teacher	School	Female	White	Bachelor's degree	1 to 5	None	5	Large	Urban
General Education Teacher	School, District	Male	White	Master's degree	1 to 5	None	3	Large	Suburban
General Education Teacher	School	Male	White	Bachelor's degree	1 to 5	None	6	Large	Suburban
General Education Teacher	School	Male	White	Master's degree	More than 20	1 to 5	12	Large	Suburban

Appendix 3-B

Standards Confirmation Workshop Agenda

Standards Confirmation Workshop Agenda

Exhibit 3-B-1. Day 1 Standards Confirmation Workshop Agenda

south dakota
DEPARTMENT OF EDUCATION
Learning. Leadership. Service.

2022 Standards Confirmation for the SDSA

SCIENCE PANEL AGENDA

July 28 – 29, 2022

Standards Confirmation Workshop Day 1 – Thursday, July 28, 2022	
8:30 – 9:00 a.m.	**Participant Login**
9:00 – 9:15 a.m.	**Welcome and Introductions from the South Dakota Department of Education (SDDOE)**
9:15 – 10:00 a.m.	**Large-Group Orientation and Introductory Training** Welcome and introductions Purpose of standards confirmation workshop General overview of standards confirmation procedures and key concepts • Achievement-level descriptors (ALDs) • Item clusters and stand-alone items Item interactions Scoring assertions • Item cluster review • Contextual information – benchmark and impact data • Individual Vote
10:00 – 10:15 a.m.	**Break, and Separate into Small Virtual Group Rooms**
10:15 – 11:45 a.m.	**Panelists Experience Online Operational Assessment and Test Environment**
11:45 – 12:45 p.m.	**Review Range ALDs and Discuss Threshold ALDs** Parse range ALDs to identify specific claims within achievement levels Identify knowledge and skills differentiating student achievement between levels
12:45 – 1:30 p.m.	**Lunch (on your own)**
1:30 – 2:30 p.m.	**Continue Discussions of ALDs**
2:30 – 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? • Are the knowledge and skill requirements of this assertion consistent with the achievement level classification? 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 3-B-2. Day 2 Standards Confirmation Workshop Agenda

2022 Standards Confirmation for the SDSA

SCIENCE PANEL AGENDA

July 28 – 29, 2022

Standards Confirmation Workshop Day 2 – Friday, July 29, 2022	
9:00 – 9:15 a.m.	**Reconvene Panels**
9:15 – 12:00 p.m.	**Continue Review of OSAB Items**
12:00 – 1:00 p.m.	**Lunch (on your own)**
1:00 – 3:00 p.m.	**Achievement Standard Judgment Task, Group Discussion, Private Vote** Review of Achievement Standard Judgment Task key concepts Group discussion of whether the current achievement standard locations are defensible Private panelist vote
3:00 – 3:30 p.m.	**Workshop Evaluations**
3:30 p.m.	**Adjourn**

Appendix 3-C

Standards Confirmation Training Slides

Standards Confirmation Training Slides

Exhibit 3-C-1. Large-Group Orientation Slides

Cambium Assessment

south dakota
DEPARTMENT OF EDUCATION
Learning. Leadership. Service.

Standards Confirmation: Science

July 28 – 29, 2022
South Dakota Science Assessment (SDSA)

2

Welcome and Introductions

South Dakota Department of Education

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION
Learning. Leadership. Service.

State Education Representatives

3

South Dakota Department of Education (SDDOE)

- Matthew Gill, Office of Assessment Administrator
- Chris Booth, Program Specialist
- Beth Schiltz, Special Education Program Specialist

4

Large-Group Orientation

Cambium Assessment, Inc.

Workshop Leaders

5

Cambium Assessment, Inc.

- Psychometrics
  - Frank Rijmen
  - Yi-Fang Wu
  - Jiajun Xu
- Room Facilitators
  - Grade 5: Kevin Dwyer, Hibbah Haddam, and Olivia Francois
  - Grade 8: Vanessa Johnson and Mark Warner
  - Grade 11: Matt Davis and Jared Taylor

Purpose of the Standards Confirmation Workshop

6

- South Dakota Science Assessment (SDSA) Standard-Setting Workshop in September 2021
  - After first administration of SDSA in Spring 2021
- Panels of educators convened to recommend achievement standards for the new SDSA
- Effects of pandemic on standard setting
  - Student performance could have been affected
  - Challenges in recruiting a sufficient number of workshop panelists
- Given the historical context, SDDOE seeks to verify the location of SDSA achievement standards

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION
Learning. Leadership. Service.

Main Workshop Activities

7

- Large-Group Orientation
- Panel Training
  - Take the Online Operational Assessment
  - Review Range ALDs
  - Discuss Just Barely ALDs
  - Review the Ordered Scoring Assertion Booklet
- Evaluate Achievement Standards Locations
  - Determine whether the location of the current achievement standards classify students into each of the achievement levels in a defensible way
  - Individually vote on the appropriateness of the achievement standard locations
- Workshop Evaluation

Importance of Security

8

- Cameras are required for participants
- 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

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Description of the Science Test Design

9

- Grades 5, 8, and 11 tests assess students' understanding of the South Dakota Science Standards
- The SDSA at grades 5, 8, and 11 includes 6 item clusters and 12 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

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Scoring Assertions

10

- 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

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

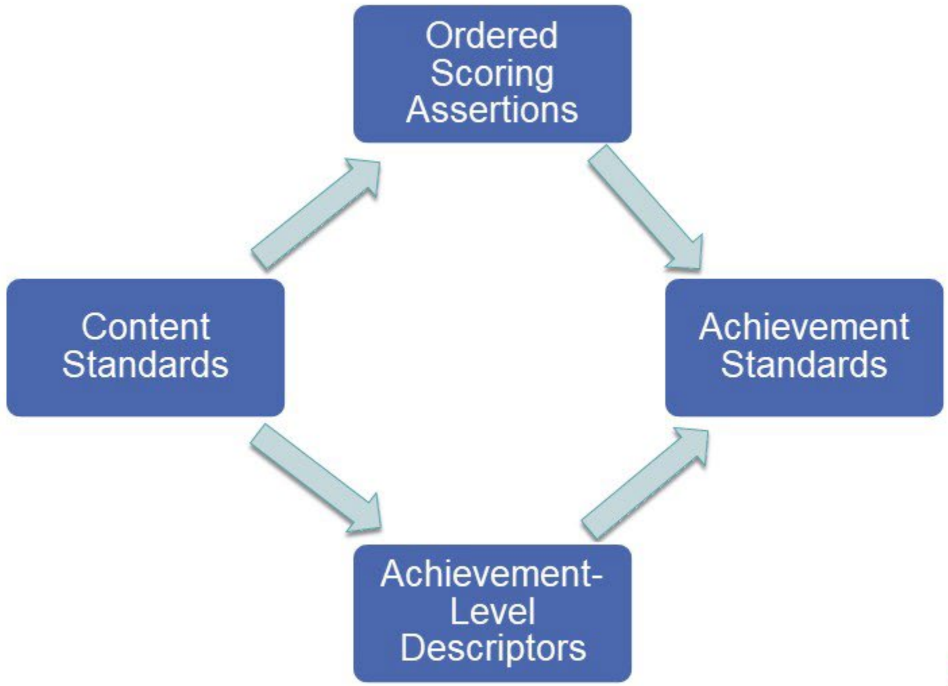
Standard Setting

11

- 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 achievement level on the test

From Content Standards to Achievement Standards

12

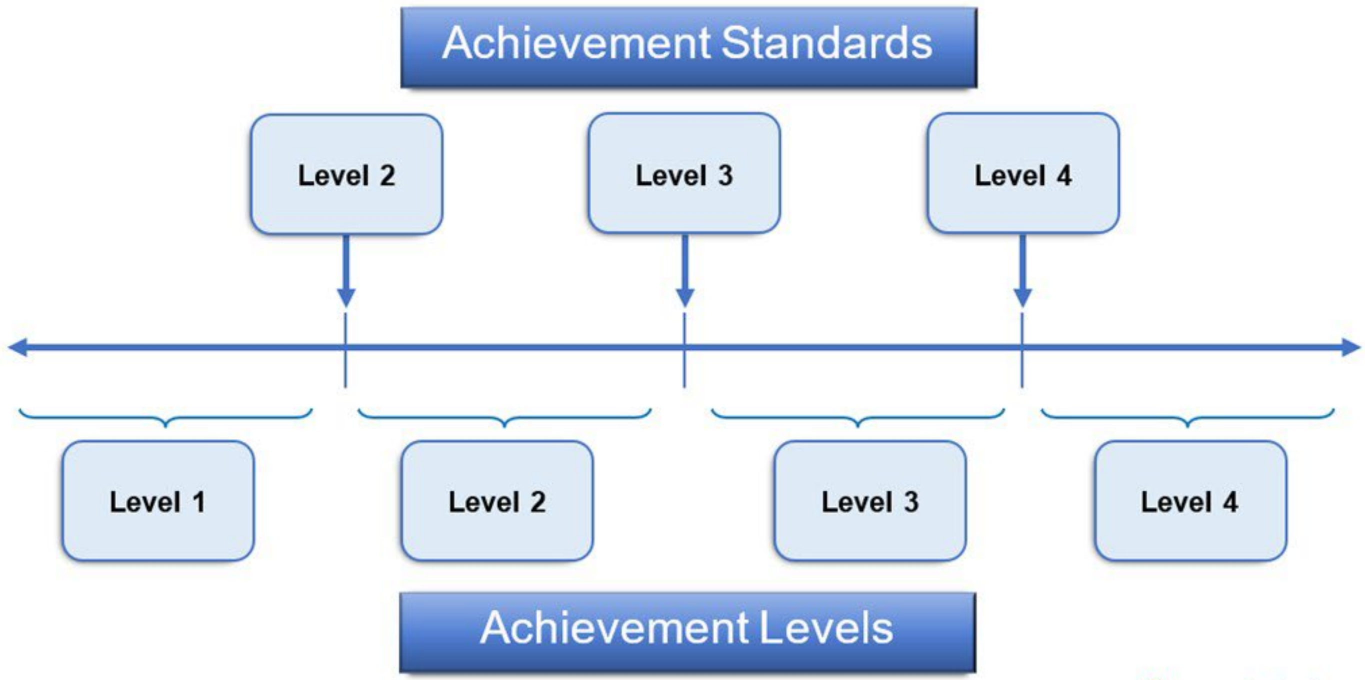


Cambium Assessment

south dakota DEPARTMENT OF EDUCATION
Learning. Leadership. Service.

Achievement Standards and Achievement Levels

13



Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Assertion-Mapping Procedure (AMP)

14

- Procedure used to recommend the SDSA achievement standards in the 2021 standard-setting workshop
- Test-centered procedure
- Employs an ordered item procedure adapted to accommodate new multiple interaction item types
- Map ordered scoring assertions to achievement levels
- Is being employed to recommend achievement standards in multiple states assessing three-dimensional science standards

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Key Elements of the Standards Confirmation Procedure

15

- Achievement-Level Descriptors (ALDs)
  - Range ALDs
  - Threshold ALDs (just barely meets)
- Ordered scoring assertions
- Assertion map
- Evaluate SDSA achievement standards
  - Student impact and benchmarking data

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Achievement-Level Descriptors (ALDs)

16

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

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Grade 8 Range ALDs – Level 3

17

Physical Sciences

- **MS-PS1:** Analyze patterns in graphical displays of data and develop and/or use a model to explain the conservation of mass when two substances react…
- **MS-PS2:** Ask questions, plan and conduct an investigation, and analyze and interpret data to make and support a claim regarding the relationships between mass, force, and motion, and the attractive and repulsive forces that act at a distance (electric, magnetic, and gravitational forces).
- **MS-PS3:** Develop and/or use a model or investigation to construct an argument to support a claim about how kinetic and potential energy interact, transform, or transfer to another object…
- **MS-PS4:** Develop and/or use mathematical representations in a model to describe the patterns observed between wave characteristics and wave energy…

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Grade 8 Range ALDs Across Achievement Levels

18

MS-PS2 Motion and Stability: Forces and Interactions

- **Level 1: Identify components of an investigation, and identify data** regarding the relationships between mass, force, and motion, and the attractive and repulsive forces that act at a distance (electric, magnetic, and gravitational forces) that could be used to **support a claim.**
- **Level 2: Identify questions, conduct an investigation, and organize and use data to make a claim** regarding the relationships between mass, force, and motion, and the attractive and repulsive forces that act at a distance (electric, magnetic, and gravitational forces).
- **Level 3: Ask questions, plan and conduct an investigation, and analyze and interpret data to make and support a claim** regarding the relationships between mass, force, and motion, and the attractive and repulsive forces that act at a distance (electric, magnetic, and gravitational forces).
- **Level 4: Ask questions to conduct, evaluate, and revise an investigation; and analyze and evaluate data to predict and support a claim** regarding the relationships between mass, force, and motion, and the attractive and repulsive forces that act at a distance (electric, magnetic, and gravitational forces).

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

"Just Barely" Meets the Achievement Standard

19

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

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Threshold "Just Barely" ALDs

20

Ordered Scoring Assertions

21

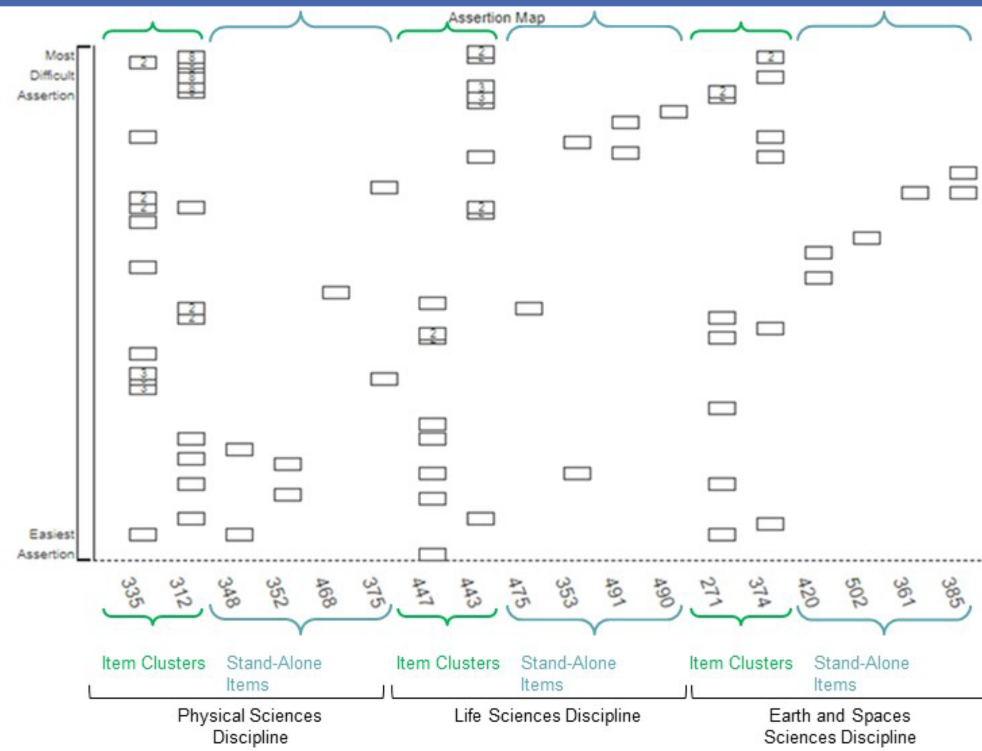
- The ordered scoring assertion booklet (OSAB) constitutes a test administration:
  - A test form that meets test blueprint specifications
  - Spring 2021 OSAB
- It is important to evaluate scoring assertions as they relate to the item interactions
- Assertions within items are ordered by difficulty
  - Assertions within an item may not represent all ALDs

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Assertion Map

22



Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Studying the Items and Scoring Assertions

23

- For each scoring assertion ask yourself:
  1. How does the student interaction give rise to the assertion? Did they plot, select, or write something?
  2. Why is this assertion more difficult to achieve than the previous one (within the item)?
  3. Are the knowledge and skill requirements of this assertion consistent with the achievement level classification?
- Working as a group
  - Discuss how item interactions support scoring assertions
  - Discuss ordering of scoring assertions
  - Discuss how scoring assertions are related to the ALDs

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

What If an Assertion Seems Out of Order?

24

- 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

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

What If an Item Seems Wrong or Unfair?

25

- 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

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Evaluate the SDSA Achievement Standards

26

- Judgment Task
  - Appropriateness of the location of the current SDSA achievement standards that were set in the 2021 standard-setting workshop
- Group Discussion
  - Are the recommended achievement standards defensible?
  - Is the location of the achievement standard in the OSAB consistent with the Threshold ALDs?
- Following discussion, panelists will engage in a private vote regarding whether the current achievement standard locations for their grade band are defensible.
- Panelist votes will be private and will be sent to the SDDOE for final decisions.

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Student Impact and Benchmarking Data

27

- Impact data: percentage of students performing at or above the achievement standards set in 2021
  - Presented for both 2021 and 2022 test administrations
- In addition, impact data will be presented for each assertion in the OSAB

Student Impact and Benchmarking Data

28

- Benchmark data: percentage of students performing at or above the achievement standards for ELA and Mathematics
  - Presented for both 2021 and 2022 test administrations
- In addition, benchmarking data will be presented for each assertion in the OSAB

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

We are Not Recommending Achievement Standards

29

- Standard-setting panelists recommended achievement standards in two rounds
  - Panelists discussed their assertion mappings after round one, but did not, and were not asked to come to agreement
- In some instances, panel recommendations were modified to achieve vertical articulation
- Thus, the current achievement standard locations are likely not the locations that you would have recommended

Break Into Groups

30

Panel	Facilitators
Grade 5 Science	Kevin Dwyer Hibbah Haddam Olivia Francois
Grade 8 Science	Vanessa Johnson Mark Warner
Grade 11 Science	Matt Davis Jared Taylor

Exhibit 3-C-2. Breakout Room Slides

Standards Confirmation: Science

July 28 – 29, 2022
South Dakota Science Assessment (SDSA)

2

Breakout Room

Cambium Assessment, Inc.

Introductions

3

- Room Facilitators
  - Grade 5: Kevin Dwyer, Hibbah Haddam, and Olivia Francois

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Purpose of the Standards Confirmation Workshop

4

- South Dakota Science Assessment (SDSA) Standard-Setting Workshop in September 2021
  - After first administration of SDSA in Spring 2021
- Panels of educators convened to recommend achievement standards for the new SDSA
- Effects of pandemic on standard setting
  - Student performance could have been affected
  - Challenges in recruiting a sufficient number of workshop panelists
- Given the historical context, SDDOE seeks to verify the location of SDSA achievement standards

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION
Learning. Leadership. Service.

Main Workshop Activities

5

- Large-Group Orientation
- Panel Training
  - Take the Online Operational Assessment
  - Review Range ALDs
  - Discuss Just Barely ALDs
  - Review the Ordered Scoring Assertion Booklet
- Evaluate Achievement Standards Locations
  - Determine whether the location of the current achievement standards classify students into each of the achievement levels in a defensible way
  - Individually vote on the appropriateness of the achievement standard locations
- Workshop Evaluation

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Importance of Security

6

- Cameras are required for participants
- 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

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Description of the Science Test Design

7

- Grades 5, 8, and 11 tests assess students' understanding of the South Dakota Science Standards
- The SDSA at grades 5, 8, and 11 includes 6 item clusters and 12 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

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Review of 3D Science Standards

8

- 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).

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Review of Items – 3D Composition

9

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

uth dakota
PARTMENT OF EDUCATION
Learning. Leadership. Service.

Review of Items – 3D Composition

10

- 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

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Item Clusters and Stand-Alone Items

11

- 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

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Structure of Item Clusters

12

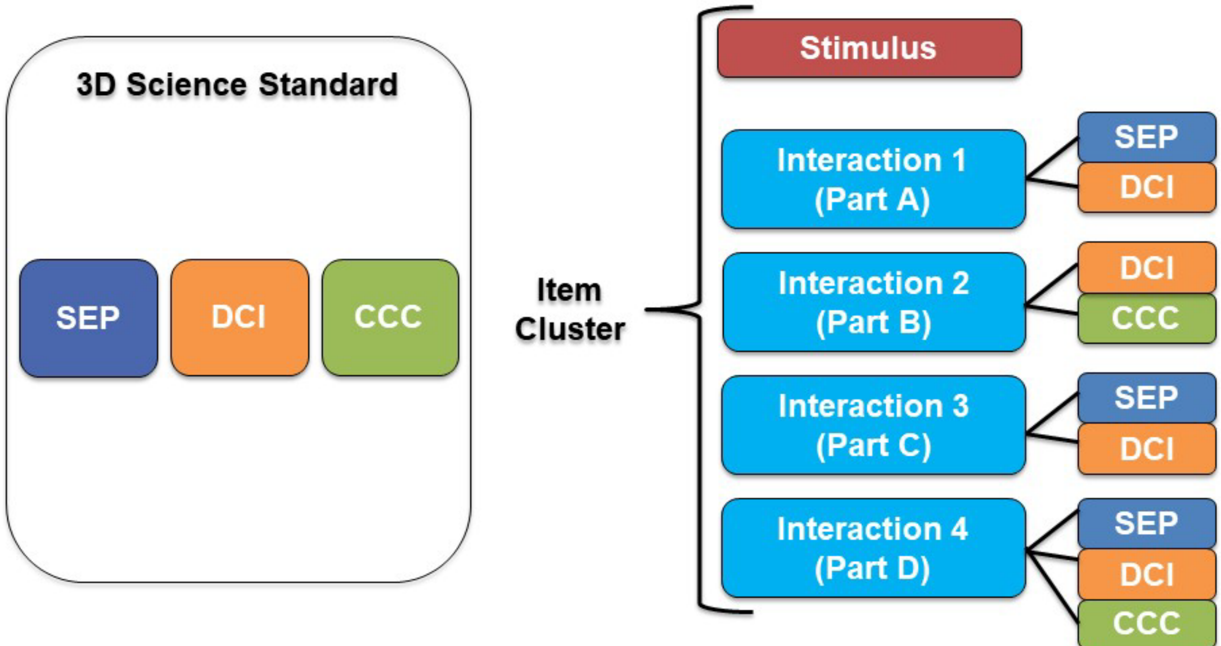
- Each item cluster begins with a ***phenomenon***, which is the observation about the natural world which anchors the entire item cluster. The interactions within the item cluster all address the phenomenon.
- Each item cluster engages the student in a grade-appropriate, meaningful ***scientific activity*** 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

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Review of Item Clusters – Composition

13



Review of Items – Composition Example

14

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

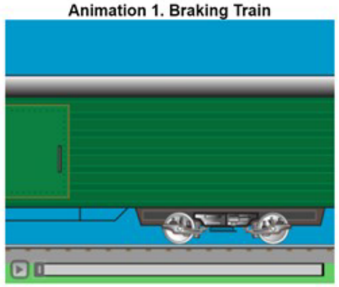


Table 1 explains some properties of the train and its surroundings as energy flows throughout the system.

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?

Ⓐ The surroundings gain energy.

Ⓑ The surroundings lose energy.

Ⓒ The surroundings do not gain or lose energy.

Ⓓ 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

15

- 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

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Review of Items – Scoring Assertions

16

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

outh dakota
EPARTMENT OF EDUCATION
rning. Leadership. Service.

Experience the Online Assessment

17

- 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)

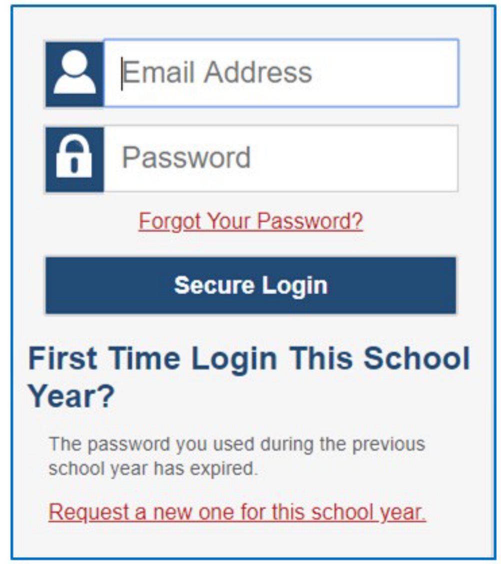
Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Accessing the Online Assessment

18

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



Cambium Assessment

south dakota DEPARTMENT OF EDUCATION
Learning. Leadership. Service.

19

Experience Online Operational Assessment

Step 2: Take the Operational Test

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION
Learning. Leadership. Service.

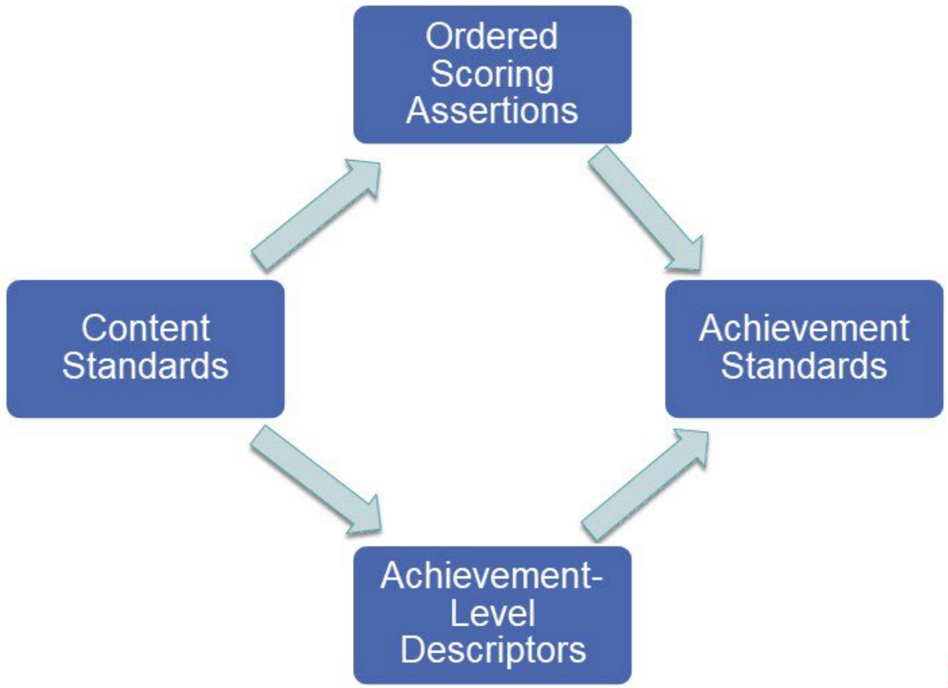
Standard Setting

20

- 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 achievement level on the test

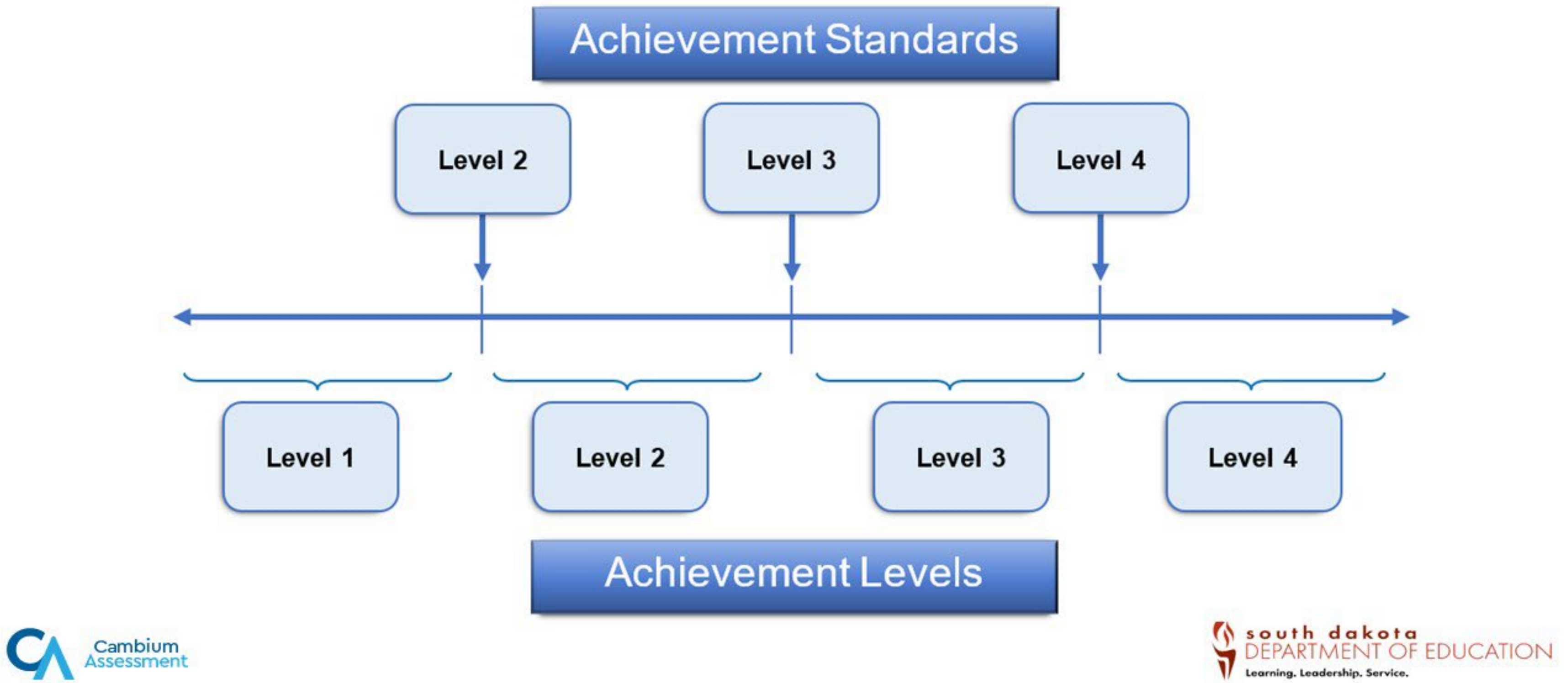
From Content Standards to Achievement Standards

21



Achievement Standards and Achievement Levels

22



Assertion-Mapping Procedure (AMP)

23

- Procedure used to recommend the SDSA achievement standards in the 2021 standard-setting workshop
- Test-centered procedure
- Employs an ordered item procedure adapted to accommodate new multiple interaction item types
- Map ordered scoring assertions to achievement levels
- Is being employed to recommend achievement standards in multiple states assessing three-dimensional science standards

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Key Elements of the Standards Confirmation Procedure

24

- Achievement-Level Descriptors (ALDs)
  - Range ALDs
  - Threshold ALDs (just barely meets)
- Ordered scoring assertions
- Assertion map
- Evaluate SDSA achievement standards
  - Student impact and benchmarking data

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Achievement-Level Descriptors (ALDs)

25

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

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Grade 8 Range ALDs – Level 3

26

Physical Sciences

- **MS-PS1:** Analyze patterns in graphical displays of data and develop and/or use a model to explain the conservation of mass when two substances react…
- **MS-PS2:** Ask questions, plan and conduct an investigation, and analyze and interpret data to make and support a claim regarding the relationships between mass, force, and motion, and the attractive and repulsive forces that act at a distance (electric, magnetic, and gravitational forces).
- **MS-PS3:** Develop and/or use a model or investigation to construct an argument to support a claim about how kinetic and potential energy interact, transform, or transfer to another object…
- **MS-PS4:** Develop and/or use mathematical representations in a model to describe the patterns observed between wave characteristics and wave energy…

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION
Learning. Leadership. Service.

Grade 8 Range ALDs Across Achievement Levels

27

MS-PS2 Motion and Stability: Forces and Interactions

- **Level 1: Identify components of an investigation, and identify data** regarding the relationships between mass, force, and motion, and the attractive and repulsive forces that act at a distance (electric, magnetic, and gravitational forces) that could be used to **support a claim.**
- **Level 2: Identify questions, conduct an investigation, and organize and use data to make a claim** regarding the relationships between mass, force, and motion, and the attractive and repulsive forces that act at a distance (electric, magnetic, and gravitational forces).
- **Level 3: Ask questions, plan and conduct an investigation, and analyze and interpret data to make and support a claim** regarding the relationships between mass, force, and motion, and the attractive and repulsive forces that act at a distance (electric, magnetic, and gravitational forces).
- **Level 4: Ask questions to conduct, evaluate, and revise an investigation; and analyze and evaluate data to predict and support a claim** regarding the relationships between mass, force, and motion, and the attractive and repulsive forces that act at a distance (electric, magnetic, and gravitational forces).

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

“Just Barely” Meets the Achievement Standard

28

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

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Parse and Review the PLDs

29

- 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

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Threshold "Just Barely" ALDs

30

Purpose of Just Barely Discussion

31

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

32

Review of Ordered Scoring Assertion Booklet

Step 4: Review of Ordered Scoring Assertion Booklet

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION
Learning. Leadership. Service.

Ordered Scoring Assertions

33

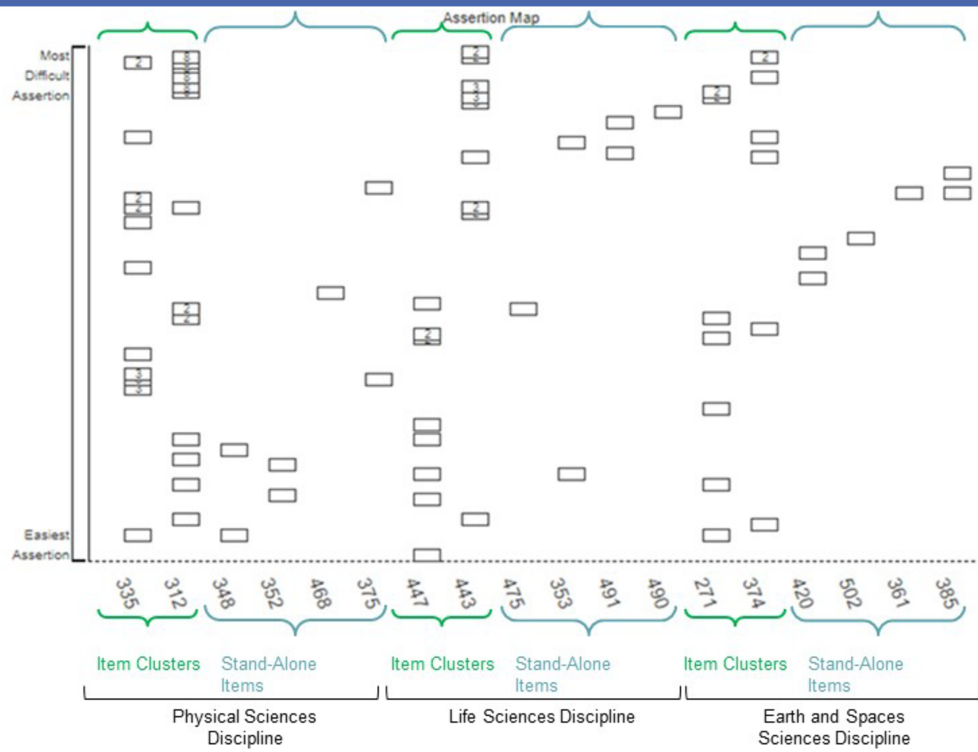
- The ordered scoring assertion booklet (OSAB) constitutes a test administration:
  - A test form that meets test blueprint specifications
  - Spring 2021 OSAB
- It is important to evaluate scoring assertions as they relate to the item interactions
- Assertions within items are ordered by difficulty
  - Assertions within an item may not represent all ALDs

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Assertion Map

34



Cambium Assessment

south dakota DEPARTMENT OF EDUCATION
Learning. Leadership. Service.

What If an Assertion Seems Out of Order?

35

- 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

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

What If an Item Seems Wrong or Unfair?

36

- 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

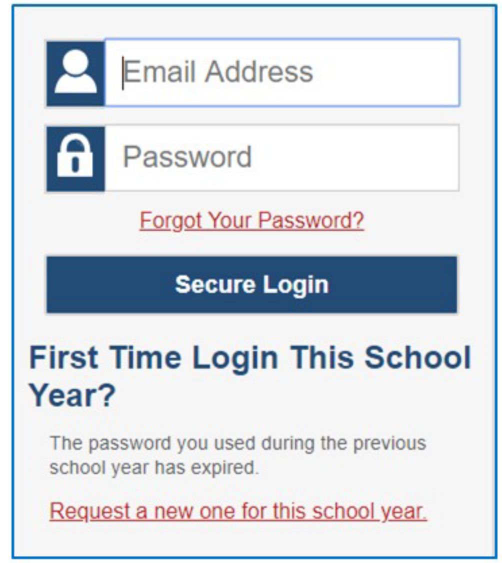
Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Accessing the OSAB

37

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



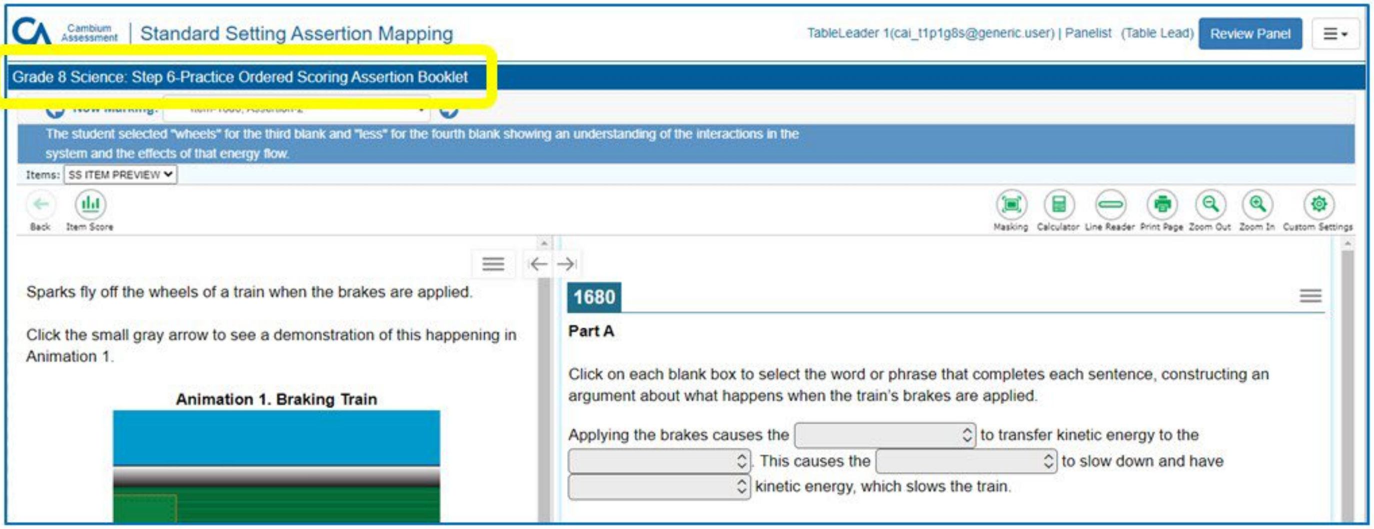
Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Navigating the OSAB

38

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



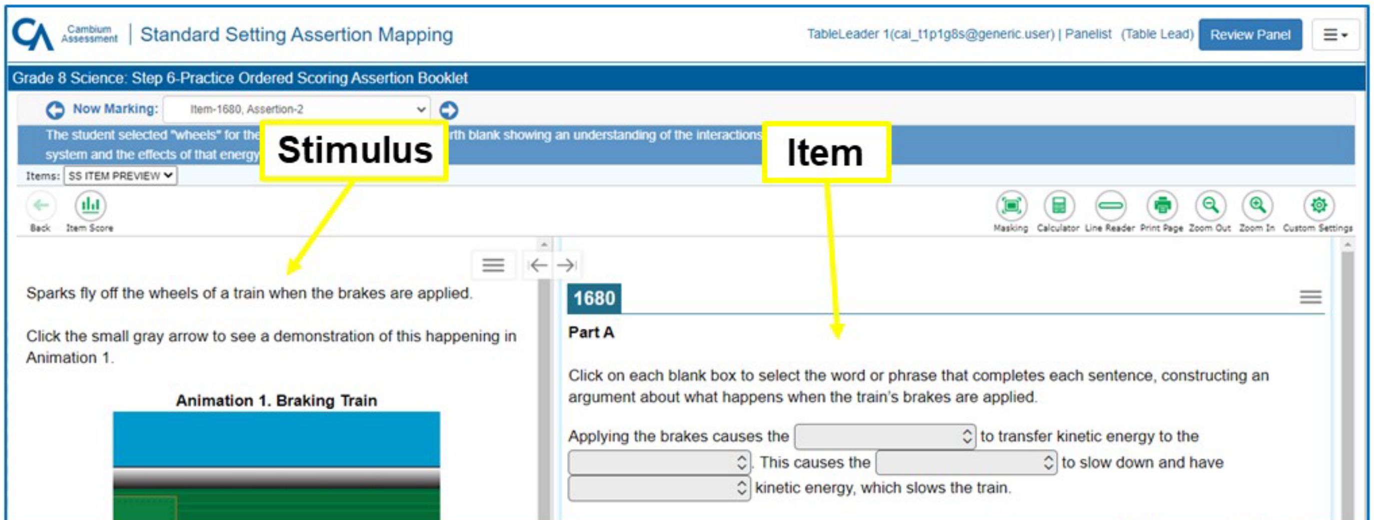
Cambium Assessment

south dakota DEPARTMENT OF EDUCATION
Learning. Leadership. Service.

Navigating the OSAB

39

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



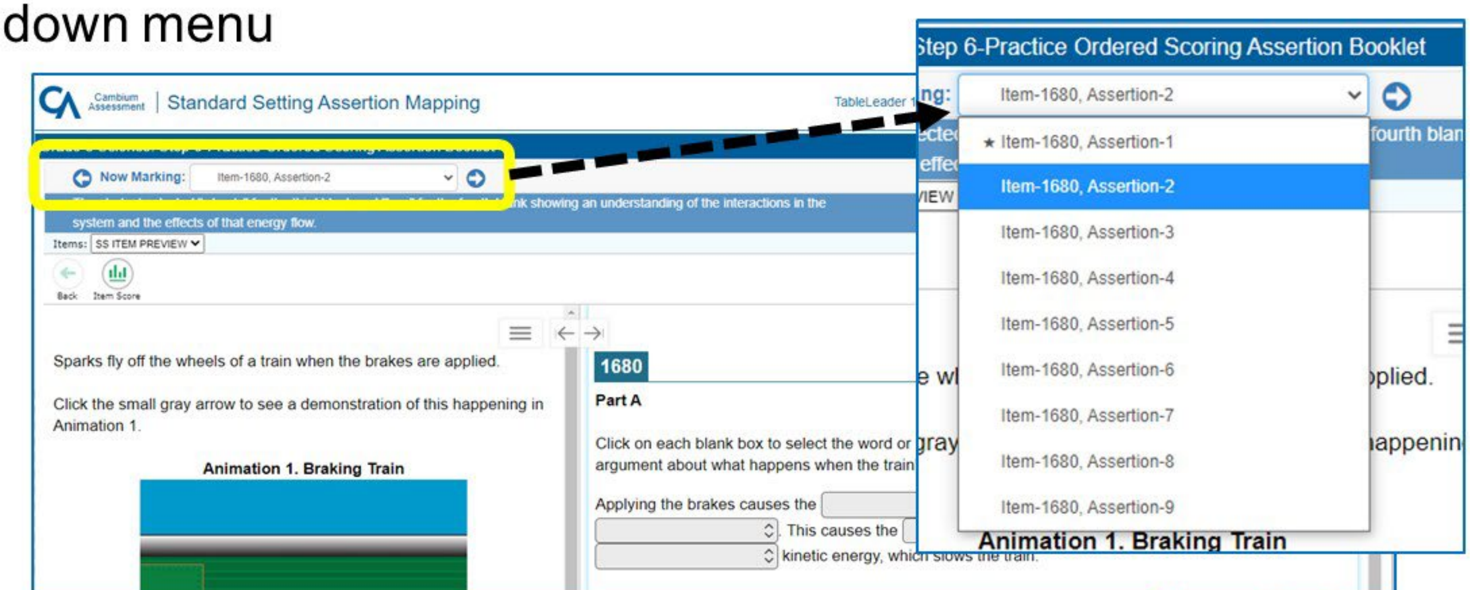
Cambium Assessment

south dakota DEPARTMENT OF EDUCATION
Learning. Leadership. Service.

Navigating the OSAB

40

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



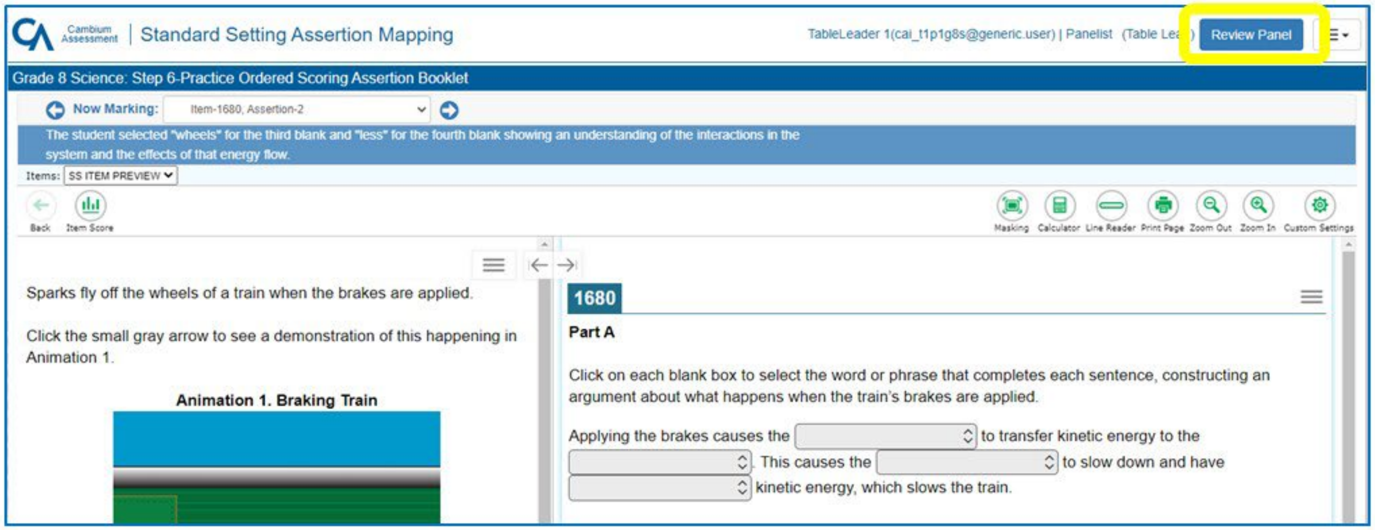
Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Navigating the OSAB

41

- Access the Review Panel



Cambium Assessment

south dakota DEPARTMENT OF EDUCATION
Learning. Leadership. Service.

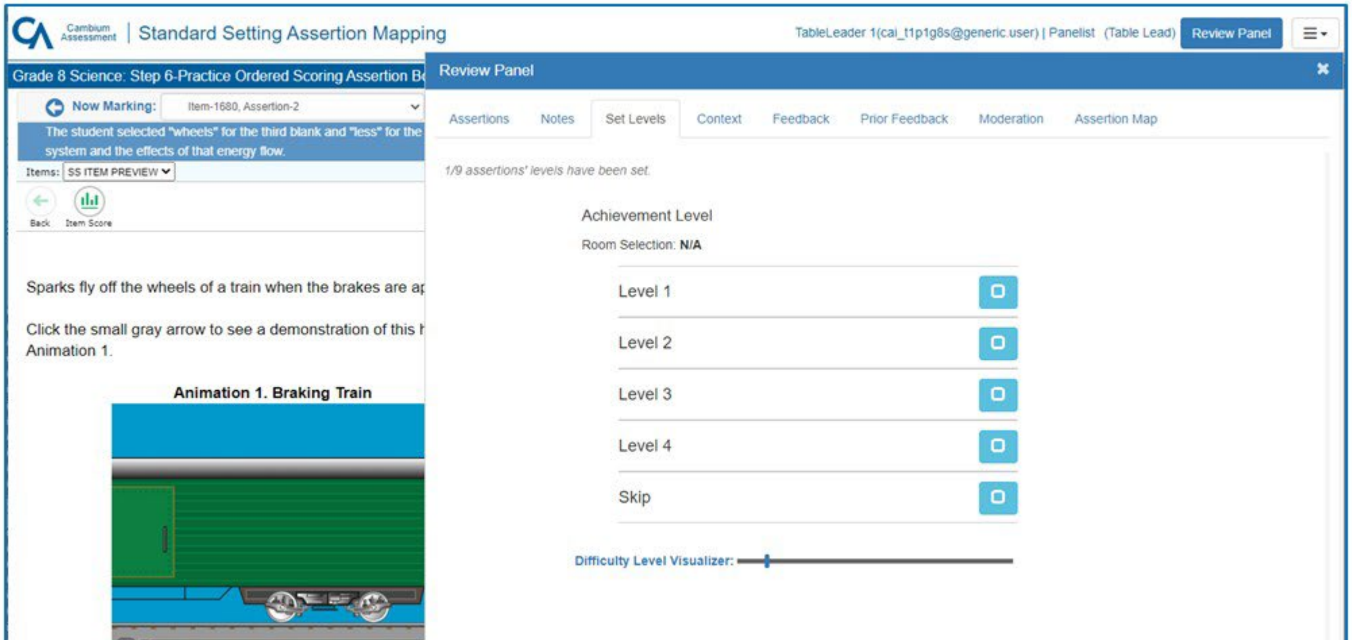
Navigating the OSAB – Review Panel

42

EDUCATION
Learning. Leadership. Service.

Navigating the OSAB – Review Panel

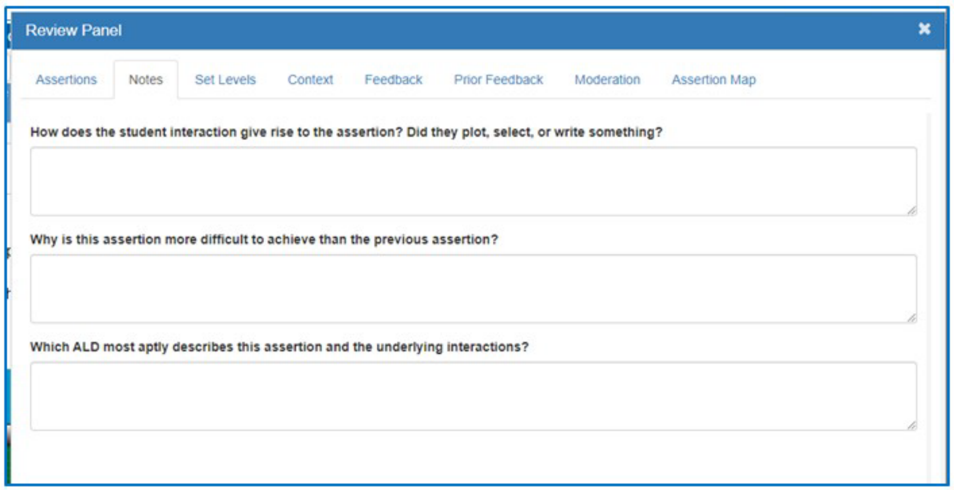
43



Navigating the OSAB – Review Panel

44

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



Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Assertion Map with Benchmark Information

45

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION
Learning. Leadership. Service.

Studying the Items and Scoring Assertions

46

- 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

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

47

Standard Setting Day 2

Recommending Performance Standards for Grade 8 Science

Standard Setting Day 2 Agenda

48

- Continued review of OSAB
- Evaluate the SDSA Achievement Standards

Continue review of OSAB

49

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION
Learning. Leadership. Service.

Evaluate the SDSA Achievement Standards

50

- Judgment Task
  - Appropriateness of the location of the current SDSA achievement standards that were set in the 2021 standard-setting workshop
- Group Discussion
  - Are the recommended achievement standards defensible?
  - Is the location of the achievement standard in the OSAB consistent with the Threshold ALDs?
- Following discussion, panelists will engage in a private vote regarding whether the current achievement standard locations for their grade band are defensible.
- Panelist votes will be private and will be sent to the SDDOE for final decisions.

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Student Impact and Benchmarking Data

51

- Impact data: percentage of students performing at or above the achievement standards set in 2021
  - Presented for both 2021 and 2022 test administrations
- In addition, impact data will be presented for each assertion in the OSAB

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Student Impact and Benchmarking Data

52

- Benchmark data: percentage of students performing at or above the achievement standards for ELA and Mathematics
  - Presented for both 2021 and 2022 test administrations
- In addition, benchmarking data will be presented for each assertion in the OSAB

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Contextual Information – Benchmark Data

53

- Smarter Balanced English Language Arts (ELA)

South Dakota 2021 & 2022 ELA Assessment Results

Grade	Year	At or above		
		Level 2	Level 3	Level 4
5	2021	73	51	20
	2022	71	50	20
8	2021	78	52	15
	2022	76	49	14
11	2021	86	66	28
	2022	84	63	28

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION
Learning. Leadership. Service.

Contextual Information – Benchmark Data

54

- Smarter Balanced Mathematics

South Dakota 2021 & 2022 Mathematics Assessment Results				
Grade	Year	At or above		
		Level 2	Level 3	Level 4
5	2021	68	38	17
	2022	68	39	19
8	2021	69	40	18
	2022	65	37	18
11	2021	68	39	14
	2022	65	37	13

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Contextual Information – Benchmark Data

55

- South Dakota Science Assessment (SDSA)

SDSA 2021 & 2022 Assessment Results				
Grade	Year	At or above		
		Level 2	Level 3	Level 4
5	2021	79	41	17
	2022	80	41	19
8	2021	82	38	9
	2022	77	37	11
11	2021	84	48	10
	2022	81	48	11

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

We are Not Recommending Achievement Standards

56

- Standard-setting panelists recommended achievement standards in two rounds
  - Panelists discussed their assertion mappings after round one, but did not, and were not asked to come to agreement
- In some instances, panel recommendations were modified to achieve vertical articulation
- Thus, the current achievement standard locations are likely not the locations that you would have recommended

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Evaluate the SDSA Achievement Standards

57

- Judgment Task
  - Appropriateness of the location of the current SDSA achievement standards that were set in the 2021 standard-setting workshop
- Group Discussion
  - Are the recommended achievement standards defensible?
  - Is the location of the achievement standard in the OSAB consistent with the Threshold ALDs?
- Following discussion, panelists will engage in a private vote regarding whether the current achievement standard locations for their grade band are defensible.
- Panelist votes will be private and will be sent to the SDDOE for final decisions.

Cambium Assessment

south dakota DEPARTMENT OF EDUCATION Learning. Leadership. Service.

Appendix 3-D

Standards Confirmation Readiness Form

Standards Confirmation Readiness Form

Exhibit 3-D. Standards Confirmation Readiness Form

2022 South Dakota Science Assessment Standards Confirmation Educator Panel - Readiness Form

Preparation for the Achievement Standards Judgment Task

* Required

1. Full Name: *

2. Panelist ID (e.g., SD_T1P1G5S): *

3. Assigned Committee: *

   Mark only one oval.

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

Preparation for the Achievement Standards Judgment Task

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

   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 achievement standard judgment task. *

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 evaluating whether the location of the current SDSA achievement standards are defensible. *

Mark only one oval.

- Yes
- No

9. I have answered "Yes" to the above questions and I understand what I need to do to perform the achievement standard judgment task. (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 perform the achievement standard judgment task. (Please initial below.) *

Mark only one oval.

- Yes
- No
- Not applicable

14. Initial: *