# AY23 Cognia Science Alternate Assessment (CSAA)/ Multi-State Alternate Assessment (MSAA) Science Standard-Setting Report

**July 18–20, 2023—Wakefield, Massachusetts**

**Prepared by Cognia for the MSAA Partners**

cognia

# TABLE OF CONTENTS

# Chapter 1. Overview of Standard-Setting Procedures

This report is designed to provide a concise summary of the activities carried out during the standard-setting for the Multi State Alternate Assessment (MSAA) in Science, intended for the MSAA Science Partners. The primary objective of the standard setting process was to identify and define the knowledge, skills, and abilities (KSAs) that students must demonstrate to be categorized into each of the performance levels.

In 2022, a standard setting session was conducted to establish provisional cut scores. Subsequently, in the current year of 2023, another standard setting session was conducted to finalize these provisional cut scores. Specifically, for the science alternate assessment, there are four distinct performance levels, denoted as Level 1, Level 2, Level 3, and Level 4."

The standard-setting process used was the Item-Descriptor (ID) Matching method (Ferrara & Lewis, 2012; Cizek & Bunch, 2007). The ID Matching method was selected because it reduces cognitive burden on panelists (see Appendix A for a list of panelists) as compared to other standard-setting methods that require probability judgments about hypothetical high- and low-performing students, and it most clearly translates content standards into performance categories as compared to other methods of standard setting (Cizek, Bunch, & Koons, 2004).

It is worth noting that the Technical Advisory Committee (TAC) actively engaged in a series of iterative reviews concerning the methodology, ultimately reaching a consensus on the structure and proposed procedures of ID Matching. Furthermore, in light of challenges encountered during the provisional standard-setting meeting, TAC recommended the presence of an external observer to oversee the execution of the proposed procedures.

The standard-setting meeting was held from July 18th through July 20th of 2023. In all, 21 panelists that were recruited from the partners participated in the process and were organized into three groups of 7 panelists each, plus a facilitator provided by Cognia for each grade. Furthermore, all MSAA Science Partners, except for the Virgin Islands, attended to observe the standard-setting process, with Rachel Quenemoen, a Technical Advisory Committee member, serving as the external observer. Panelists were organized according to the grade level in which each panelist had the most professional experience. Table 1-1 illustrates how the report is organized into three major sections, describing tasks completed before, during, and after the standard-setting meeting.

**Table 1-1. Standard-Setting Tasks and the Report Layout**

| Chapter 2 | Chapter 3 | Chapter 4 |
|---|---|---|
| *Tasks Completed Before the Standard-Setting Meeting* | *Tasks Completed During the Standard-Setting Meeting* | *Tasks Completed After the Standard-Setting Meeting* |
| • Creation of Performance Level Descriptors<br>• Preparation of Materials<br>• Preparation of Instructions for Facilitators<br>• Content-based benchmarks<br>• Preparation of Systems and Materials for Use During the Meeting<br>• Selection of Panelists | • Overview of the ID Matching Method<br>• General Orientation and Panelist Training<br>• Review of Assessment Materials<br>• Completion of the Item Map Form<br>• Review of Performance Level Descriptors<br>• Judgment Rounds and Feedback<br>• Vertical Articulation | • Analysis and Review of Panelists' Feedback<br>• Policy Adjustments for Level 2 Cut Scores<br>• Approval of the Cut Scores<br>• Preparation of Standard-Setting Report |

# Chapter 2. Tasks Completed Before Standard Setting

## 2.1 Creation of Performance Level Descriptors

In 2022, Cognia collaborated with MSAA science partners and educators to create performance level descriptors. Cognia Content and Accessibility Specialists collaborated with the MSAA Science Partners to develop Performance Level Descriptors (PLDs) for the Science Alternate Assessment. During the early design tasks, a claim was developed for the assessment. From this claim, Policy PLDs were developed for each grade that served as the defining descriptions for each performance level in grades 5, 8, and high school. The Policy PLDs provide overarching policy level student performance expectations for Level 1, Level 2, Level 3, and Level 4. Range PLDs were developed from these Policy PLDs. Cognia Content and Accessibility Specialists used their expertise in the test design, Extended Performance Expectations (EPEs) structure, and knowledge of the students to develop Range PLDs for each performance level for each grade. The Range PLDs describe the knowledge, skills, and abilities that students must demonstrate to be classified into a performance level (Level 1, Level 2, Level 3, Level 4).

Stakeholders were recruited by Cognia and the MSAA Science Partners to participate in a review of the Policy and Range PLDs to provide input on the expectations outlined in 2022. Feedback was gathered on the clarity and consistency of the knowledge, skills, and abilities outlined in each performance level, as well as the incorporation of the science dimensionality aspects that are part of the EPEs and the Science Alternate Assessment test design. The stakeholders included content and special education experts. In addition, MSAA Science Partners actively participated in the review meeting. Panelists met virtually on May 16, 2022, to review the draft Policy and Range PLDs. Materials included the draft PLDs, the EPEs, and a review checklist. The review meeting started with introductions and a detailed overview of the Science Alternate Assessment, including the assessment design, the EPEs, and how the PLDs were developed. An overview of the materials to be used during the meeting and a detailed walkthrough of the review checklist was also provided. The panelists then began their review of the PLDs using the review checklist for guidance. Feedback was gathered and noted in the PLD document by the facilitator. At the end of the review, meeting panelists were asked to complete a feedback survey. The demographics information of the PLD review panel, along with the feedback survey and its results, are available in Appendix B (PLD Reviewer Information).

Following this PLD review meeting, Cognia incorporated this feedback and provided the MSAA Science Partners time to review and provide additional input. Edits incorporated included clarifying language around the multidimensional aspects of the EPEs, applying consistency in range PLD wording within a level, and formatting the range PLDs to facilitate understanding within and across grade levels. Once the Policy and Range PLDs were set, Cognia Content and Accessibility Specialists developed the Borderline PLDs. The PLDs are provided in Appendix C.

## 2.2 Preparation of Materials

The following materials were assembled for presentation to the panelists at the standard-setting meeting in paper or digital form (as indicated):

- Opening session PowerPoint
- Meeting agendas
- Nondisclosure forms
- Test booklets
- Performance Level Descriptors
- Content standards
- The Cognia Standard-Setting Toolkit

Copies of the meeting materials, including a list of panelists, PLDs, PowerPoint presentations, the Cognia Standard-Setting Toolkit, the readiness surveys, the workshop evaluation and results, the Standard-Setting meeting memo, and a synopsis of the procedural validity evidence are included in Appendices A through K.

## 2.3 Preparation of Instructions for Facilitators

Facilitators attended training sessions led by Cognia before the standard setting. The purpose of the training was to prepare the facilitators for the panel activities and to ensure consistency in the implemented procedures. During this training, facilitators were trained on how to lead the panelist review of the ordered item booklet, to lead the discussion of borderline PLDs, to facilitate panel discussion throughout the standard setting, to collect and review the standard-setting materials, and to control secure materials. Facilitators were expected to ensure that discussion and logistics within each grade panel were conducted fairly and efficiently. Facilitator scripts, which are embedded in grade-specific PowerPoint slides, were created for the facilitators to refer to while working through each step of the standard-setting process. An example of the grade 8 instructions for facilitators slides are included in Appendix D.

## 2.4 Content-Based Benchmarks

In standard setting, benchmarks refer to any content- or policy-based information that comes from an external source and is presented to panelists. The exact way that the benchmarks are used in the standard setting depends upon the methodology. However, the general use is the same: Standard-Setting panelists see and consider information from these external measures as they engage in the Standard-Setting meeting activities.

This Science standard setting used Content-based benchmarks. The procedure for determining the content-based benchmarks was as follows:

- Prior to the Standard-Setting meeting, Cognia content teams reviewed each item in the OIB and matched the items to one of three PLD levels (Level 2-4). Note that the Cognia content specialists did not assign any items to the Level 1 PLD. This is because all MSAA Science

items are written according to level 2 and above, and the Level 1 performance level is described simply as the inconsistency to perform at the Level 2.

- Cognia psychometricians then compiled the content specialists' item-PLD alignments and calculated threshold regions through logistic regression. Specifically, the regions were calculated by combining the item-PLD judgments to derive a set of cut scores with two standard errors added below and above each cut score. See Appendix E for calculation details.

- The above process resulted in content-based benchmark regions for the Level 3 and 4 performance level cuts.

**Special Considerations for the Level 2 Benchmark Region.** As mentioned previously, the Level 1 performance level is described as the inconsistency of Level 2 performance; therefore, items were not written to Level 1 and, by extension, it was not feasible to align items to Level 1. Since there were no Level 1 item-PLD alignments, the above logistic regression method could not be employed to calculate a cut and corresponding region for Level 2.

Thus, to facilitate the Level 2 cut score identification, Cognia psychometricians empirically derived the cut score by constructing a miniature Test Characteristic Curve (TCC) based on items that were aligned to the Level 2 PLD. Cognia interpreted the borderline PLD of 50% to mean that a student placed in the Level 1 performance level should be answering items aligned to the Level 2 PLD correctly 50% of the time, considering chance. Thus, Cognia calculated a theta value that was associated with 50% beyond chance of the expected score of the mini TCC. The '50% beyond chance' criterion is reflected in the performance level descriptor and takes guessing into account. Two OIB pages were added below and above the empirical cut score to create an empirical threshold region for the Level 2 cut.

# 2.5 Preparation of Systems and Materials for Use During the Meeting

This section provides details about the Cognia Standard-Setting Toolkit that panelists used to complete all standard-setting activities during the meeting. In addition, the setup of the grade-specific digital ordered item booklets with their associated target cut scores and benchmark regions is discussed.

The Cognia Standard-Setting Toolkit was developed, tested, and set up by Cognia before the meeting and included the following components:

- Digital ordered item booklet: A booklet specific to each grade in the form of an item list. Items were ordered with the easiest item at the top and the most difficult at the bottom. Items in the benchmark regions were shaded for easy reference. The benchmark regions were presented to panelists at the beginning of Round 3.

- Items: A PDF of each item (along with associated stimuli). The item view for each item also included notes on the specific EPE associated with that item.

- Judgment forms: Integrated within the booklet and item views of the digital tool, the judgment forms provided space for panelists to note (1) the relevant knowledge, skills, and abilities (KSAs) needed to answer the item (2) content-based rationales, (3) item descriptor matches.

- Readiness surveys: Digital readiness surveys that panelists completed before undertaking each judgment round.

- Evaluation form: The final workshop evaluation form that panelists completed after the standard-setting meeting.

Additional details of the Cognia Standard-Setting Toolkit are available in Appendix F.

Within the digital tool, the ordered item booklet contained one item per page, ordered from the easiest item to the most difficult item. The ordered item booklet was created by sorting the items according to their item response theory (IRT)-based difficulty values (RP0.50 was used). A two-parameter logistic IRT model was used to calculate the RP0.50 values for dichotomous items.

## 2.6 Selection of Panelists

As emphasized in Cizek and Bunch (2007), regardless of the method used, the selection of panelists is an important factor in determining standard-setting outcomes and maximizing the validity of the standard-setting process. The guidance provided by the *Standards for Educational and Psychological Testing* (AERA et al., 1999) states that "a sufficiently large and representative group of judges should be involved to provide reasonable assurance that results would not vary greatly if the process were repeated." Consistent with the above guidelines and respecting practical considerations regarding the maximum size of a group that can be successfully managed, the goal was to recruit a standard-setting panel of 8–10 members representing different stakeholder groups to set standards for the science alternate assessment. Additionally, in consideration of the various MSAA Science Partners' locations, an attempt was made to ensure the panels included representation from each Partner. Targets for the size and composition of the panel were also consistent with federal guidelines as described in *Standards and Assessment Peer Review Guidance: Information and examples for meeting requirements of the No Child Left Behind Act of 2001* (U.S. Department of Education, 2009).

MSAA Science Partners selected panelists before the standard-setting meeting. The goal for panel selection was to include participants who were primarily special education and/or general education teachers but also included school administrators, and stakeholders from other interest groups. Moreover, to the extent possible, panelists were selected to reflect a balance of gender, race/ethnicity, and geographic location. Finally, panelists were selected who were familiar not only with the subject matter but also with the grade for which they would be setting standards. A list of the panelists is included in Appendix A.

**Table 2-2. Panelists Demographic Information Summary**

| Panelist Demographics | | N | Percentage |
|---|---|---|---|
| **Gender** | Female | 20 | 95% |
| | Male | 1 | 5% |
| **Ethnicity** | American Indian or Alaska Native | 4 | 19% |
| | Asian or Asian American | 1 | 5% |
| | Native Hawaiian or other Pacific Islander | 7 | 33% |
| | White | 9 | 43% |
| **State Representation** | American Samoa | 3 | 14% |
| | Arizona | 3 | 14% |
| | BIE | 4 | 19% |
| | CNMI | 2 | 10% |
| | Guam | 3 | 14% |
| | Maine | 3 | 14% |
| | Vermont | 3 | 14% |
| **Panelist Teaching Experience** | | | |
| **Grade Band*** | Elementary | 15 | 32% |
| | Middle | 14 | 30% |
| | High | 10 | 21% |
| | All Grades | 8 | 17% |
| **Educational Setting*** | Special Education | 16 | 48% |
| | General Education | 14 | 42% |
| | School Administrator | 3 | 9% |
| **Years of Experience in Education** | 0-5 | 0 | 0% |
| | 5-10 | 5 | 24% |
| | 10-15 | 7 | 33% |
| | More than 15 | 9 | 43% |

*Several Panelists indicated multiple Grade Band and Educational Setting Categories.*

# Chapter 3. Tasks Completed During the Standard-Setting Meeting

## 3.1 Overview of the ID Matching Method

The Item-Descriptor (ID) Matching method is appropriate for setting standards for standards-aligned assessments like the science alternate assessment. Assessment programs around the world have used ID Matching (e.g., Delaware, Massachusetts, Maryland, Mississippi, New Mexico, New York, South Carolina, and West Virginia; the Chicago and Philadelphia Public Schools; and programs in Brazil and Germany).

ID Matching has advantages over Bookmark, Angoff, and other standard-setting methods. Specifically, its cognitive-judgmental task requires that standard-setting panelists, who are typically classroom educators, undertake a judgmental task that they are well suited for—matching item knowledge and skill response demands with knowledge and skill expectations in performance level descriptors (PLDs). The Bookmark and other methods require panelists to make probability judgments—something that people in general do not do well (e.g., Murphy, 2002). In addition, panelists do not need to hold a hypothetical borderline student in mind when they match items to descriptors and recommend cut scores, so the cognitive load and complexity of ID Matching is more manageable.

During standard setting using ID Matching, panelists use PLDs as their guide to match items to performance level descriptors. The structure of the PLDs provides a general characterization of expected student knowledge and skill at each level and examples of the knowledge and skills that students at each performance level can be expected to demonstrate. Panelists identified knowledge and skills required by each item by answering two questions: "What are the knowledge, skills, and abilities a student needs to respond to this item?" The ordering of items by their empirical difficulty facilitates the matching process. By matching test items to specific claims from the borderline Level 3 PLD, for example, panelists identify the evidence in test items that supports the claims in that descriptor. Supporting the claims represented in the borderline Level 3 PLD contributes to the validity of interpretations of student achievement, based on the PLDs, and to the overall validity argument that a student who achieves that level on the assessment has demonstrated adequate understanding of essential concepts with respect to the standards being measured. This logic applies to all cut scores and performance levels.

One limitation of the ID Matching Method is that it required assessment items to be written for each performance level, For MSAA Science assessments, all items were developed according to PLD levels 2-4. In other words, no items were specifically designed for PLD level 1. Consequently, an alternative approach was employed to establish the level 2 cut score. See 2.4 Special Considerations for the Level 2 Benchmark Region for more details on the method.

## 3.2 General Orientation and Panelist Training

Concerning panelist training, Standards for Educational and Psychological Testing (AERA et al., 2014) states the following:

Care must be taken to assure these persons understand what they are to do and that their judgments are as thoughtful and objective as possible. The process must be such that well-qualified participants can apply their knowledge and experience to reach meaningful and relevant judgments that accurately reflect their understandings and intentions. (p. 101)

The training of the panelists began with a general orientation session at the start of the standard-setting meeting which included an overview of assessing students on an alternate assessment and participation criteria. The purpose of the orientation was to ensure that all panelists received the same information about the need for and the goals of standard setting, and about their part in the process.

In a collaborative effort, Science partners, Cognia, and MSAA's technical advisory committee member, Rachel Quenemoen serving as the observer, welcomed the standard-setting participants. The orientation session began with a thorough introduction, provided a comprehensive overview of the MSAA Science assessment, highlighted its key features, and explained its accessibility features. Moreover, the purpose of the standard setting was clearly articulated to set the context for the workshop.

Following the introduction, postsecondary outcomes and their direct correlation to performance levels and cut scores were introduced to participants. The link between these elements underlined the importance of establishing meaningful cut scores that align with students' readiness for inclusive college programs and integrated workforce opportunities.

Next, the Lead Science Content and Accessibility Specialist provided an overview of the Science design, administration, timeline of work leading up to the Standard Setting, an overview of the policy, range, and borderline PLDs, and specific logistical details (e.g., materials review, content security, attendance). Once the general orientation was complete, panelists broke out into grade-specific groups, where they received more detailed training and completed the three rounds of the standard-setting activities in a secure environment.

## 3.3 Review of Assessment Materials

The first step after the opening session was for the panelists to review the test. The purpose of this step was to familiarize the panelists with the assessment and the test-taking activities expected of students during administration. Panelist questions about the assessment materials were answered by the facilitator.

## 3.4 Completion of the Item Map Form

Panelists reviewed their grade-specific ordered item booklets, considering the knowledge, skills, and abilities (KSAs) students needed to answer each item. The ordered item booklet contained one item per page, ordered from the easiest item to the most difficult item. The ordered item booklet was created by sorting the items according to their item response theory (IRT)-based difficulty values (*RP* 0.50 was used). A two-parameter logistic IRT model was used to calculate the *RP* 0.50 values for dichotomous items.

Panelists then completed the item map form using the provided laptop computers. The item map form listed the items in the same order as they were presented in the ordered item booklet. The form included space for the panelists to type in the KSAs required to answer each item correctly.

After working individually, panelists had the opportunity to discuss the item map with members of their group and make necessary additions or adjustments. The purpose of this step was to ensure that panelists became familiar with the ordered item booklet and understood the relationships among the ordered items.

## 3.5 Review of Performance Level Descriptors

Before engaging in the judgment tasks, panelists reviewed the borderline PLDs. This important step was designed to ensure that panelists thoroughly understood the KSAs needed for students to be classified into the four performance levels (Level 1, Level 2, Level 3, Level 4). The borderline PLDs are provided in Appendix C.

## 3.6 Judgment Rounds and Feedback

During the main portion of the standard-setting workshop, panelists completed a practice round followed by three consecutive rounds of judgments. After the completion of each judgment round, Cognia psychometricians calculated a variety of statistics, such as theta scale cut scores, the conditional standard error of measurement (SEM) for each of the scale cut scores, and impact data (i.e., the percentage of students in each performance level. These statistics served as reporting to Cognia and the MSAA Science Partners as intermediate evidence for the impact of panelists' judgments, and as quality control metrics.

To begin, the panelists completed a practice round of judgments. The purpose of the practice round was to familiarize the panelists with all the materials they would be using for the standard-setting process and become facile with the ID Matching judgments. Panelists used the provided laptop computers to access digital copies of the borderline PLDs and standards (aka EPEs). In addition, panelists were provided with credentials to access the Cognia Standard-Setting Toolkit. Within the digital tool, panelists were presented with a practice ordered item booklet, which consisted of 4 items in each grade, representing the range of difficulty on the test, as well as the integrated digital judgment forms.

Within each grade-specific group, the facilitator demonstrated how to navigate within the standard-setting tool and how to use the tool to make their judgments. Then, beginning with the first ordered item and considering the skills and abilities needed to complete it, panelists were instructed to ask themselves two questions: (1) "What are the knowledge, skills, and abilities a student needs to respond to this item?" and (2) "Why is this item more difficult than the previous item?" Panelists considered each ordered item in turn, asking themselves the same two questions and assigning item descriptor matches (i.e., Level 1, Level 2, Level 3, Level 4, or the threshold between two levels) to each item. The facilitator then led the panelists in a readiness discussion, asking panelists to share the reasoning behind their item descriptor matches with the group and assessing each panelist's understanding of the judgment task and borderline PLDs.

At the end of the practice round, panelists completed the Round 1 Readiness Survey (Appendix G). The readiness survey was designed to ascertain whether the panelists were comfortable moving ahead to the judgment task. Once all panelists completed the Round 1 Readiness Survey, Cognia psychometricians reviewed the responses to make sure panelists were ready to undertake the first round of judgments. In the event of any uncertainty (based on the survey responses), the specific information was relayed to the

facilitator so that any questions or issues could be addressed before proceeding to the Round 1 judgments. The facilitator moved on to the next round of judgement when all panelists indicated "Yes" to all survey questions.

## 3.6.1 Round 1 Judgments and Results

In the first round, panelists worked individually with PLDs, the standard-setting tool, and the ordered item booklet (OIB). Beginning with the first ordered item in the grade-specific OIB, described previously, and considering the skills and abilities needed to complete it, panelists considered each ordered item in turn, asking themselves the same two questions and assigning item-descriptor matches (i.e., Level 2, Level 3, Level 4) to each item. They continued in this manner until they had looked at all the items in the OIB.

After the completion of Round 1, Cognia psychometricians calculated a variety of statistics as described previously. Table 3-1 displays Round 1 OIB page numbers, associated raw scores, theta cut score and associated standard error, and impact data (percentage of students in each performance level), respectively. All statistics in Table 3-1 are listed for technical documentation purposes. Figure 1 is the grade 8 example chart that was shown to panelists.

**Figure 1. Round 1 Results for Panelists-Grade 8 as example**
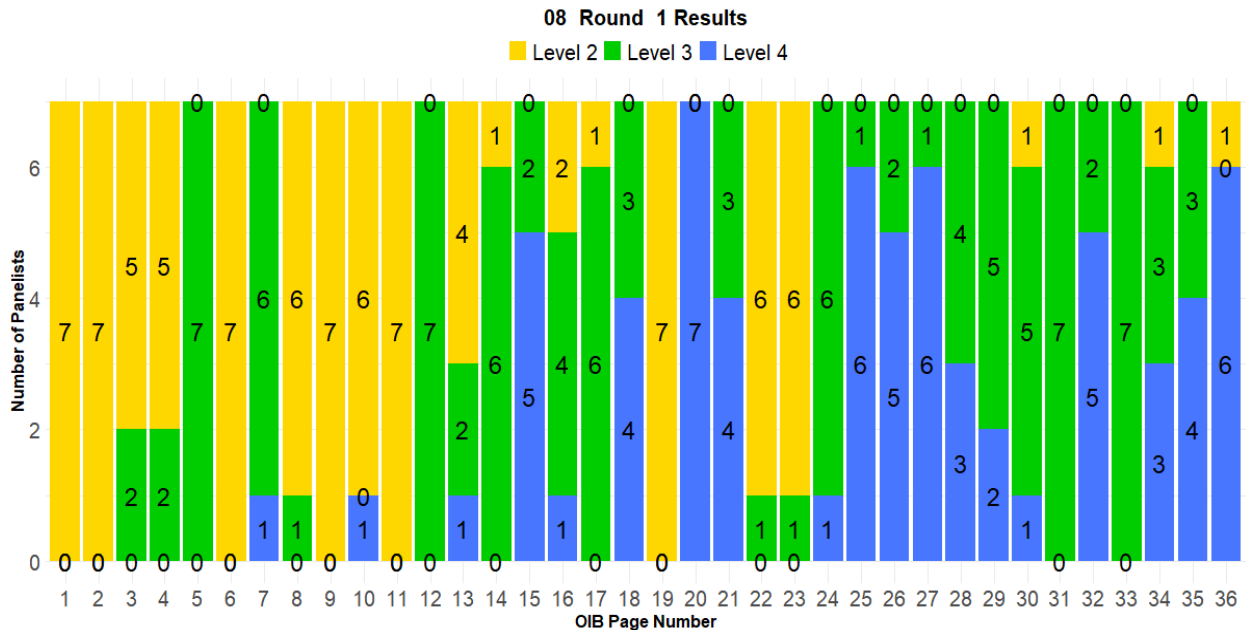
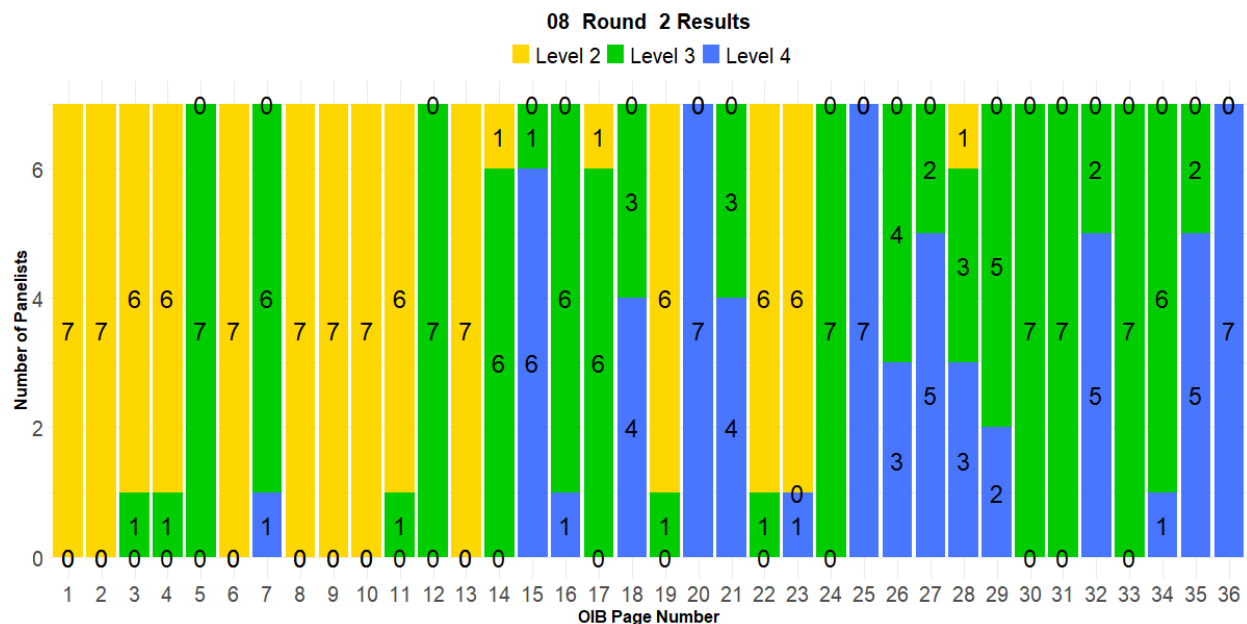**Table 3-1. MSAA Science Standard-Setting Round 1 Results**

| Grade | Level 1 | Level 2 | | | Level 3 | | | Level 4 | | | Level 3 + 4 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Percent | OIB# | Theta | Percent | OIB# | Theta | Percent | OIB# | Theta | Percent | Percent |
| 5 | 46 | 8 | -0.3 | 10 | 14 | -0.1 | 20 | 30 | 0.5 | 23 | 43 |
| 8 | 50 | 8 | -0.4 | 5 | 13 | -0.2 | 30 | 33 | 1.0 | 16 | 46 |
| 11 | 56 | 10 | -0.0 | 7 | 13 | 0.16 | 26 | 34 | 1.0 | 11 | 37 |

## 3.6.2 Round 2 Judgments and Results

Prior to beginning Round 2, the panelists at each table were presented with Figure 1 based on their Round 1 ratings for each cut point in that grade. Within each grade, panelists were then allowed to share their rationales for their judgments in terms of the necessary knowledge and skills for each performance level. Once the discussions were complete, panelists completed the Round 2 Readiness Survey (Appendix G). The readiness survey was designed to ascertain whether the panelists were comfortable moving ahead to the second round of the judgment task.

The purpose of Round 2 was for panelists to discuss their Round 1 placements and, if necessary, to revise their ratings. Once all panelists indicated that they were ready to undertake the next round, they proceeded to Round 2. In Round 2, panelists were allowed to revise their Round 1 ratings on the toolkit. When Round 2 ratings were complete, Cognia staff members calculated the statistics described above and discussed the results with MSAA Science Partners. Table 3-2 displays Round 2 OIB page numbers, theta cut score, and impact data (percentage of students in each performance level), respectively.

**Figure 2. Round 2 Results for Panelists-Grade 8 as an Example**

**Table 3-2. MSAA Science Standard-Setting Round 2 Results**

| Grade | Level 1 Percent | Level 2 OIB# | Level 2 Theta | Level 2 Percent | Level 3 OIB# | Level 3 Theta | Level 3 Percent | Level 4 OIB# | Level 4 Theta | Level 4 Percent | Level 3 + 4 Percent |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 5 | 46 | 7 | -0.3 | 10 | 14 | -0.1 | 17 | 30 | 0.4 | 27 | 44 |
| 8 | 50 | 8 | -0.4 | 11 | 13 | -0.1 | 24 | 33 | 1.0 | 16 | 40 |
| 11 | 47 | 10 | -0.1 | 16 | 13 | 0.1 | 26 | 34 | 1.0 | 11 | 37 |

## 3.6.3 Round 3 Judgments and Results

Prior to Round 3, panelists were presented with their Round 2 judgments for each performance level in that grade (Figure 2). Facilitators presented this information to the group using a projector and laptop and explained how to use it as they completed their discussions.

Within their grade-specific groups, panelists were then allowed to share their rationales for their item-PLD alignment in terms of the necessary knowledge and skills for each classification. Once the discussions were complete, panelists completed the Round 3 Readiness Survey (Appendix G). The readiness survey was designed to ascertain whether the panelists were comfortable moving ahead to the third round of the judgment task.

The purpose of Round 3, the final round, was for panelists to discuss their Round 2 recommendations and, if necessary, to revise their judgments. Once all panelists indicated that they were ready to undertake the next round, they were allowed to discuss and revise their Round 2 judgments within the tool. When Round 3 judgments were complete, Cognia psychometricians calculated the statistics described previously and discussed the results with MSAA Science Partners.

A summary of the results approval by Round 3 judgment is provided in Table 3-3 below. They display final OIB page numbers, associated raw scores, median cut points on the theta scale, and impact data (percentage of students in each performance level), respectively.

After the completion of round 3, all panelists took part in the final workshop evaluation survey, see section 4.1 for more details related to survey results.

**Figure 2. Round 3 Results for Panelists-Grade 8 as example**
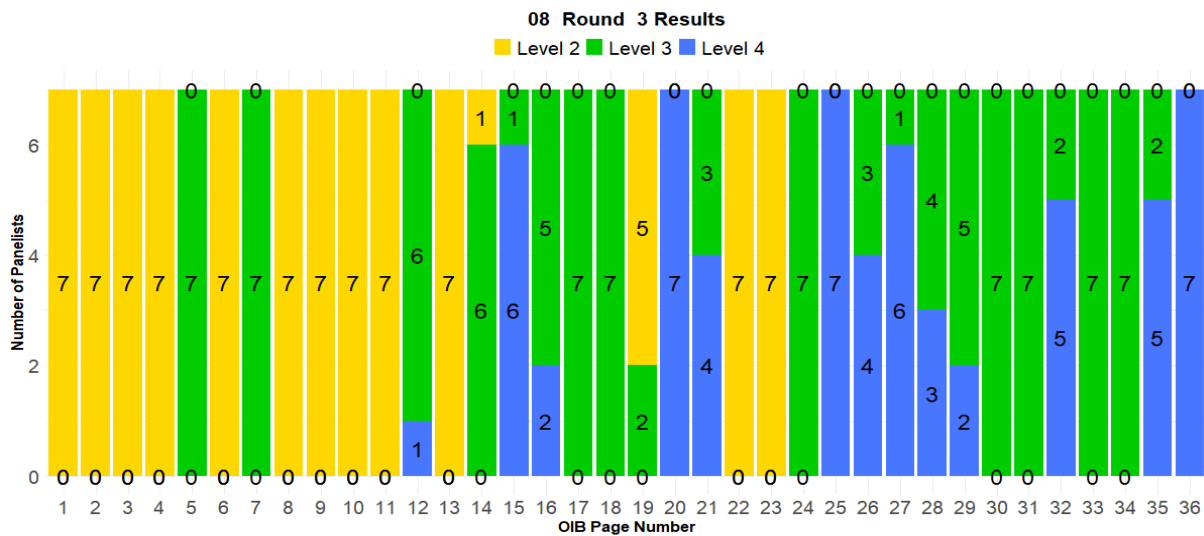


**Table 3-3. MSAA Science Standard-Setting Round 3 Results**

| Grade | Level 1 Percent | Level 2 | | | Level 3 | | | Level 4 | | | Level 3 + 4 Percent |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | OIB# | Theta | Percent | OIB# | Theta | Percent | OIB# | Theta | Percent | |
| 5 | 46 | 7 | -0.3 | 10 | 15 | -0.1 | 20 | 30 | 0.5 | 23 | 43 |
| 8 | 50 | 8 | -0.4 | 11 | 13 | -0.1 | 24 | 33 | 1.0 | 15 | 39 |
| 11 | 47 | 10 | -0.1 | 16 | 12 | 0.1 | 24 | 32 | 1.0 | 13 | 37 |

## 3.6.4 Vertical Articulation

For the vertical articulation panel, two panelists were selected from each of the three grade-specific standard-setting panels. The vertical articulation panel convened during the second half of day 3, and the panel was joined by relevant MSAA partners, MSAA TAC observer, and Cognia staff.

The articulation meeting process started with introductions, an overview of the articulation process, and articulation-specific training. The articulation facilitator explained differences between the standard setting and articulation goals/procedures. Next, panelists were asked to share their experiences related to the standard-setting meeting. Following the feedback session, panelists delved into Performance Level Descriptors (PLDs) across all grades. They examined the expectations for students moving across the grades (5, 8, and high school) and considered the knowledge, skills, and abilities required to transition from a lower to higher PLD level (e.g., from Level 1 to Level 2). In addition, panelists considered how the standards change across grades.

Two panelists from each grade were asked to share their experiences with standard setting. The feedback was positive, participants appreciated the thoughtfulness that went into the assessment development to provide meaningful scores for students and families.

Following the feedback session, panelists delved into Performance Level Descriptors (PLDs) and the discussion revealed general trends in each grade's PLDs. The vertical articulation panel agreed on the following:

- **Progression across performance levels:** Panelists agreed that there were "repeating patterns for verbs across grades" based on performance level. For example, one panelist noted that "level 2 largely uses the verb 'identify' across grades" (Grade 8 panelist) and the rest of the panel agreed. Building on this conversation, other panelists noted that the verbiage generally shifted from 'identify' to 'compare' to 'inference' when moving across Levels 2, 3, and 4 within each grade. Additionally, they agreed that skills required progress in complexity.
- **Progression across grades:** While similar verbs showed up in the different PLDs across grades, the context changed. Ranging from grade 5 through 8 and then high school, the context "moves further away from students' everyday experiences" (Grade 5 panelist). The panel agreed that the context was more concrete in Grade 5 and became less so (i.e., more abstract) with each higher grade. In terms of skill across grades, Grade 5 emphasized lower-level skills like recall, while grade 8 shifted to thinking of evidence and potential outcomes, and high school PLDs involved planning and investigating.
- **Progression through performance level with respect to different grades:** Panelists agreed that Level 2 required a similar amount of effort from students across the different grades, but that the required effort to progress to the next level (i.e., from Level 2 to 3, and from Level 3 to 4) became "steeper" within each higher grade. One panelist summarized the conversation by noting that "elementary [grades] was a gentler gradient, and then each grade becomes steeper as you move from 2 to 3 to 4" (Grade 8 panelist).

Following the PLD discussion, specific items from each grade were reviewed and discussed. The discussion again highlighted similar verbs across grades, with varying content complexity. Panelists agreed to the following specific expectations regarding student achievement at each performance level across grades:

- Level 2: Similar effort required to reach Level 2 across all three grades, thus the expectation is that student achievement would be fairly consistent at this level across the three grades.
- Level 3 and Level 4: Increased effort needed to move from Level 3 to 4 within each grade, but also across grades.

After reaching consensus about the expectations for student achievement, the group reviewed the percentage of students in each performance level across the three grades based on the cut scores that resulted from the standard-setting meeting. Panelists expressed surprise at the higher percentage of students classified in performance level 1 and the lower and more inconsistent percentages in performance Level 2.

Cognia, partners, and observers held a debrief session during which they determined they had sufficient information from the panelists to dismiss them and move forward with conversations to apply the panelist feedback.

# Chapter 4. Tasks Completed After the Standard-Setting Meeting

Upon conclusion of the standard-setting meeting, several important tasks were completed. These tasks centered on the following: presenting the results to the MSAA Science Partners in a memo (Appendix H) and a follow-up meeting; and making any final revisions or adjustments based on policy considerations, under the direction of the MSAA Science Partners.

## 4.1 Analysis and Review of Panelists' Feedback

The standard-setting literature considers the evaluation of the workshop and its results to be another product of the standard-setting process (e.g., Reckase & Chen, 2012), as it provides important validity evidence supporting the cut scores that are obtained. To provide evidence of the participants' views of the standard-setting process, panelists were asked to complete a questionnaire at the end of the meeting.

After the evaluation forms (Appendix I) were completed, the panelists' responses were reviewed. This review did not reveal any anomalies in the standard-setting process nor indicate any reason that a particular panelist's data should not be included when the final cut score calculation. In general, participants felt that the processes were appropriate and that their judgments were based on appropriate information and decision-making.

## 4.2 Policy Adjustments for Level 2 Cut Scores

As a result of incorporating the recommendations of the vertical articulation panel and the consensus reached among MSAA Science partners, a policy decision was made to limit the classification of students at level 1 to approximately 40%, as opposed to the approximately 50% determined solely through content-based judgment across three grades. The Level 2 cut scores for all three grades were lowered, while Level 3 and Level 4 cut scores remained unchanged. The proposed adjustments for Level 2 cuts were informed by several factors:

1. The adjustments were made with the goal of enhancing the alignment of Level 2 cut scores with student performance expectations within the PLDs. These adjustments were informed by qualitative data provided by articulation panelists, which can be found in section 3.6.4 where panelists' feedback on expected progression across performance levels and grades is detailed.

2. The adjusted Level 2 cut scores ensure improved alignment with the provisional cut from 2022, maintaining consistent percentages of students in Level 1 across years as the standard setting workshop aims to solidify the provisional cut scores established in the previous year.

3. Additional psychometric considerations led to lowering the Level 2 cuts by one raw score point. For instance, in grade 5, the cut score was adjusted to align with a raw score of 15 instead of a raw score of 16. Raw-score comparisons are available in Table 4-1 with the adjusted cut scores highlighted for easy reference.

**Table 4-1. Comparison of Raw Scores between Committee Recommended and Proposed Adjustment**

| Performance Level Cut | AY23 Standard-Setting Recommendations | | Proposed Adjustments | |
|---|---|---|---|---|
| | Raw Score | Theta | Raw Score | Theta |
| **Grade 5** | | | | |
| Level 2 | 16 | -0.310 | 15 | -0.398 |
| Level 3 | 18 | -0.060 | 18 | -0.060 |
| Level 4 | 23 | 0.508 | 24 | 0.508 |
| **Grade 8** | | | | |
| Level 2 | 15 | -0.370 | 14 | -0.395 |
| Level 3 | 17 | -0.074 | 17 | -0.074 |
| Level 4 | 24 | 1.060 | 24 | 1.060 |
| **Grade 11** | | | | |
| Level 2 | 14 | -0.060 | 13 | -0.173 |
| Level 3 | 16 | 0.108 | 16 | 0.108 |
| Level 4 | 21 | 0.926 | 21 | 0.926 |

Table 4-2 displays the theta cut scores and associated impact data for each grade based on the provisional cut scores from 2022, the AY23 standard-setting recommended cut scores, and the proposed adjusted cut scores. As discussed above, only the Level 2 cut scores (indicated by highlighted cells) were adjusted while the Level 3 and Level 4 cut scores remain unchanged.

Exhibit 1 gives a visual representation of the impact data across grades based on the proposed adjusted cut scores. Additionally, exhibits 2–4 visually represent comparisons of impact data between the provisional, standard-setting recommendation, and proposed adjusted cut scores for each grade, respectively.

**Table 4-2. Cut Scores and Impact Data based on Provisional, Standard-Setting, and Adjusted Cut Scores**

| Performance Level | Provisional Cuts on AY23 Student Data* | | AY23 Standard-Setting Recommendations | | Proposed Adjustments | |
|---|---|---|---|---|---|---|
| | Cut Score (Theta) | % Students | Cut Score (Theta) | % Students | Cut Score (Theta) | % Students |
| **Grade 5** | | | | | | |
| Level 1 | -- | 39.48 | -- | 46.02 | -- | 39.48 |
| Level 2 | -0.385 | 16.92 | -0.310 | 10.37 | -0.398 | 16.92 |
| Level 3 | -0.017 | 29.00 | -0.060 | 20.34 | -0.060 | 20.34 |
| Level 4 | 1.018 | 14.60 | 0.508 | 23.26 | 0.508 | 23.26 |
| Level 3 + 4 | -- | 43.61 | -- | 43.61 | -- | 43.61 |
| **Grade 8** | | | | | | |
| Level 1 | | 42.98 | -- | 49.95 | -- | 42.98 |
| Level 2 | -0.459 | 22.81 | -0.37 | 10.77 | -0.395 | 17.74 |
| Level 3 | 0.099 | 12.14 | -0.074 | 23.76 | -0.074 | 23.76 |
| Level 4 | 0.705 | 22.07 | 1.060 | 15.52 | 1.060 | 15.52 |
| Level 3 + 4 | -- | 34.21 | -- | 39.28 | -- | 39.28 |
| **High School** | | | | | | |
| Level 1 | -- | 46.83 | -- | 46.83 | -- | 39.28 |
| Level 2 | -0.070 | 16.29 | -0.060 | 16.29 | -0.173 | 23.83 |
| Level 3 | 0.137 | 22.04 | 0.108 | 24.07 | 0.108 | 24.07 |
| Level 4 | 0.799 | 14.85 | 0.926 | 12.81 | 0.926 | 12.81 |
| Level 3 + 4 | -- | 36.89 | -- | 36.89 | -- | 36.89 |

*\* The goal of the AY23 standard setting is to finalize the cut scores initially established in 2022. Applying the provisional cut scores from the 2022 standard setting to AY23 student data enables a direct comparison of student performance between the provisional cut scores and the newly established standard setting cut scores.*
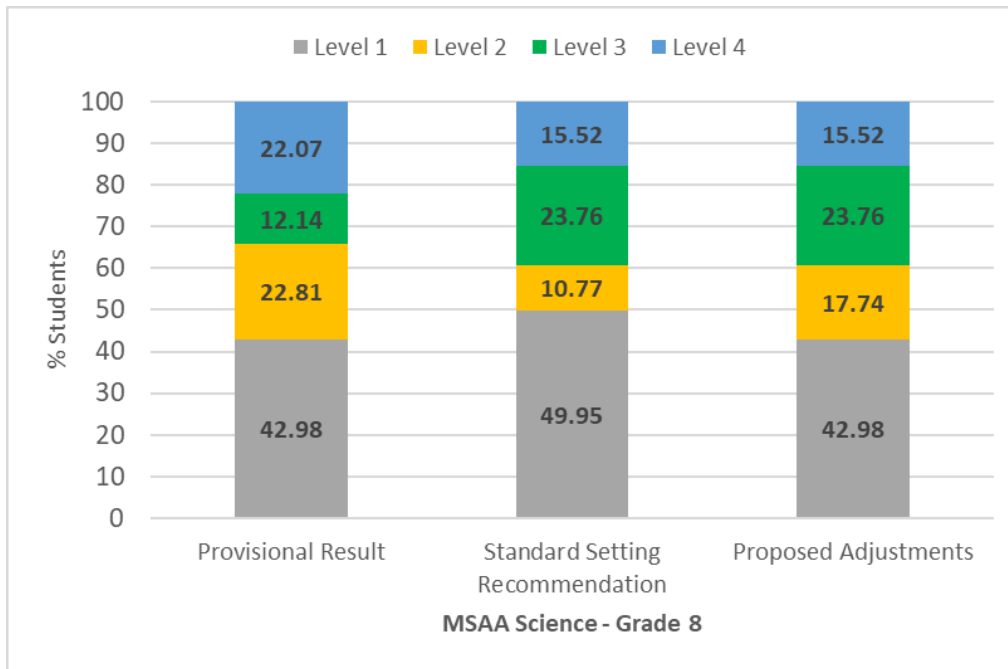
**Exhibit 1. MSAA Science Impact Data across Grades based on Adjusted Cut Scores**
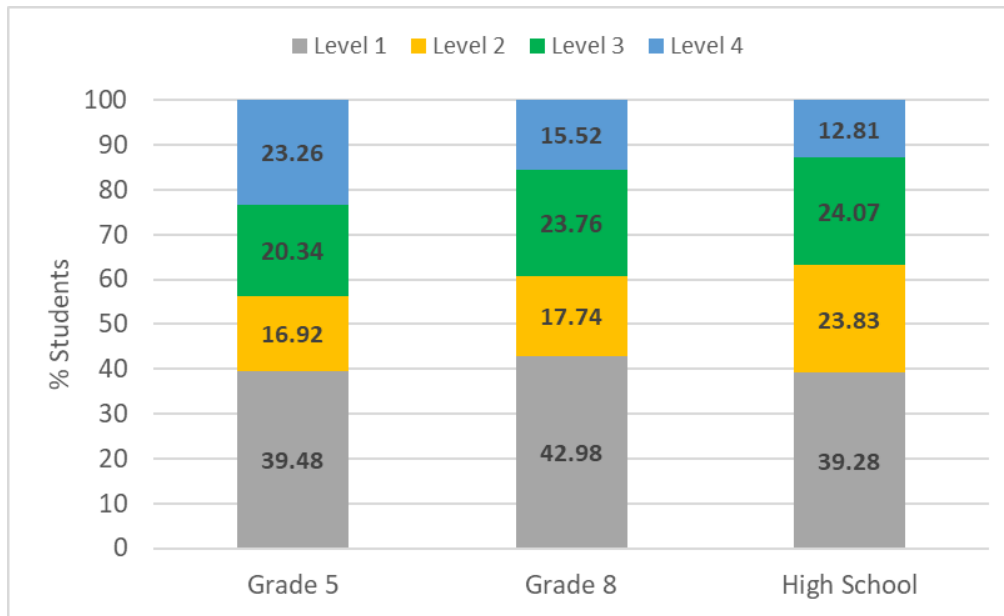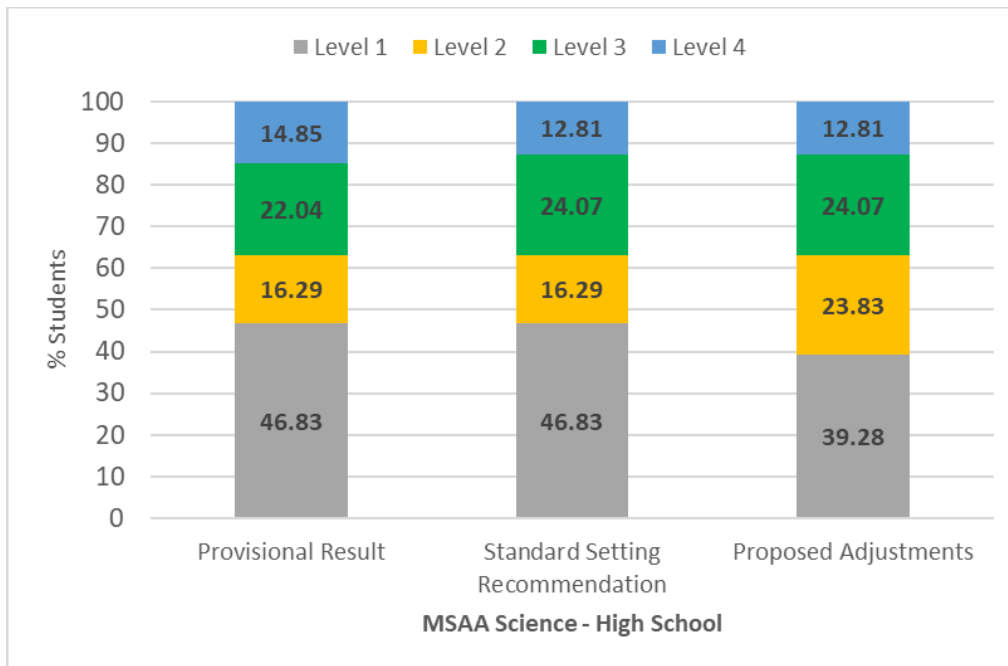


**Exhibit 2. Grade 5 Impact Data based on Provisional, Standard-Setting, and Adjusted Cut Scores**
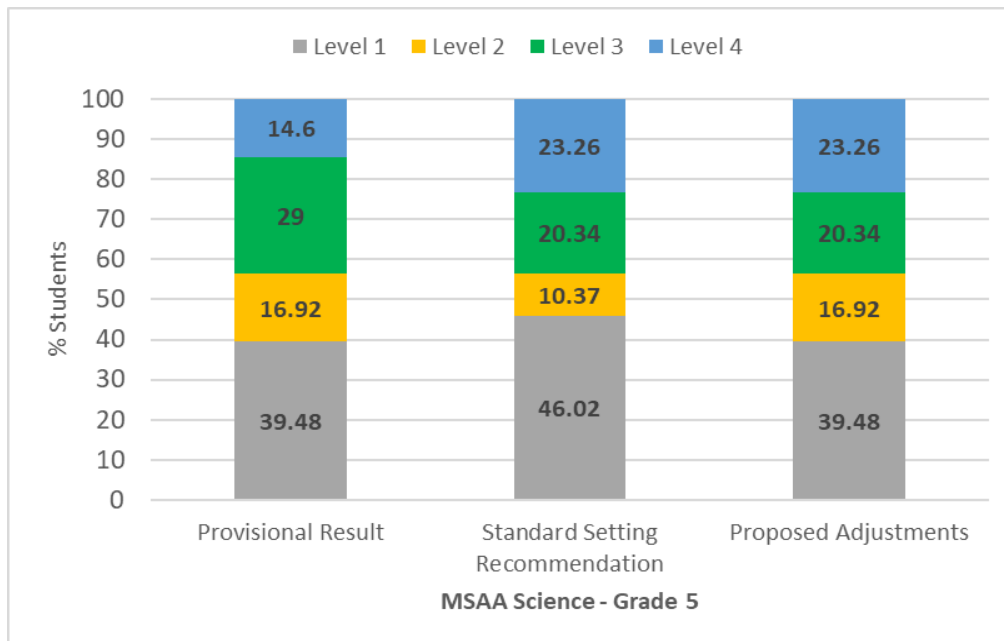
**Exhibit 3. Grade 8 Impact Data based on Provisional, Standard-Setting, and Adjusted Cut Scores**



**Exhibit 4. High School Impact Data based on Provisional, Standard-Setting, and Adjusted Cut Scores**

# 4.3 Approval of the Cut Scores

A summary of the results approved by policy review is provided in Table 4-3 below. This displays final cut points on the theta scale and the resulting impact data (percentage of students in each performance level). The final percentage of students in each performance level on final score reports may exhibit minor variations once all reporting rules have been implemented.

**Table 4-3. MSAA Science Standard-Setting Policy Review Approved Results**

| Performance Level | Cut Score (Theta) | % Students |
|---|---|---|
| Level 1 | -- | 39.48 |
| Level 2 | -0.398 | 16.92 |
| Level 3 | -0.060 | 20.34 |
| Level 4 | 0.508 | 23.26 |
| Level 3 + 4 | -- | 43.61 |
| Level 1 | -- | 42.98 |
| Level 2 | -0.395 | 17.74 |
| Level 3 | -0.074 | 23.76 |
| Level 4 | 1.060 | 15.52 |
| Level 3 + 4 | -- | 39.28 |
| Level 1 | -- | 39.28 |
| Level 2 | -0.173 | 23.83 |
| Level 3 | 0.108 | 24.07 |
| Level 4 | 0.926 | 12.81 |
| Level 3 + 4 | -- | 36.89 |

# 4.4 Preparation of Standard-Setting Report

Following the final compilation of standard-setting results, Cognia prepared this report, which documents the procedures and results of the 2023 standard-setting meeting that was held to establish performance standards for the assessment.

# References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999*). Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

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

Cizek, G. J., & Bunch, M. B. (2007). Standard setting: Establishing and evaluating performance standards on tests. Thousand Oaks, CA: Sage Publications.

Cizek, G. J., Bunch, M. B., & Koons, H. (2004). An NCME Instructional Module on Setting Performance Standards: Contemporary Methods. *Educational Measurement: Issues and Practice, 23*(4), 31–50.

Ferrara, S., & Lewis, D. (2012). The Item-Descriptor (ID) Matching method. In G. J. Cizek (Ed.), *Setting performance standards: Foundations, methods, and innovations* (2nd ed., pp. 255-282). Routledge.

Murphy, G. L. (2002). The big book of concepts. Cambridge, MA: The MIT Press

Reckase, M. D., & Chen, J. (2012). The Role, Format, and Impact of Feedback to Standard Setting Panelists. In *Setting performance standards* (pp. 172-187). Routledge.

U.S. Department of Education. (2009). Standards and assessments peer review guidance: Information and examples for meeting requirements of the No Child Left Behind Act of 2001. Washington, DC: U.S. Department of Education Office of Elementary and Secondary Education. Retrieved June 10, 2010, from the World Wide Web: www2.ed.gov/policy/elsec/guid/saaprguidance.pdf.

# Appendices

# APPENDIX A
# PANELISTS

**Table A-1. Final list of Panelists who Participated in the 2023 CSAA Standard Setting**

|  | Name | State | Email |
|---|---|---|---|
| **Grade 5** | Jessica McFarland | Arizona | mcfarlandjessi@gmail.com |
|  | Helene S. A. Cruz | Guam | hsacruz@gdoe.net |
|  | Tricia T. Taitano | CNMI | tricia.taitano@cnmipss.org |
|  | Tracie Surridge | Vermont | tsurridge@kingdomeast.org |
|  | Kelley Wallace | Maine | k.wallace@msad17.org |
| **Grade 5 & Articulation & Policy Review** | Josephine Naranjo - Montoya | BIE | jmontoya@khapoeducation.org |
|  | Kathryn Rossman | Vermont | Kathryn.rosssman@vermont.gov |
| **Grade 8** | Kimberly Aikins | Arizona | kimberly.aikins@asdb.az.gov |
|  | Christine Hernandez | Guam | cchernandez@gdoe.net |
|  | Karen Felise-Ioka | American Samoa | karen.ioka@doe.as |
|  | Racheal Ahearn | Maine | Rahearn@rsu13.org |
|  | Christine Depatie | Vermont | christine.depatie@mvsdschools.org |
| **Grade 8 & Articulation & Policy Review** | Devon Morrill | Vermont | dmorrill@cvsdvt.org |
|  | Tracy Lynn Del Rosario | CNMI | tracylynn.delrosario@cnmipss.org |
| **HS** | Jason Cowles | Arizona | jcowles@pusd11.net |
|  | Janice J. Almoguera | Guam | jjalmoguera@gdoe.net |
|  | Fiti Sua | American Samoa | fiti.sua@doe.as |
|  | Karen Etsitty | BIE | Karen.Etsitty@bie.edu |
|  | Samantha Cantell | Vermont | scantell@gisu.org |
| **HS & Articulation** | Morgan Brewer | Maine | Brewer@lincolnacademy.org |
|  | Roque Castro Indalecio | CNMI | roque.indalecio@cnmipss.org |

# APPENDIX B
# PLD REVIEW INFORMATION

| Name* | State Entity* | Email* | Phone Number* | Cell Phone | School* | Current Position/ Title* | Grade Levels Currently working with (select all that apply)* | Area of Expertise (select al that apply)* | Content Area* | Years of experience in Education* | Administered the MSAA previously?* | Educational Setting* | Gender | Race/Ethnicity (check all that apply) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Benjamin Altsher | New Hampshire | benaltsher@gmail.com | 603-489-8738 | 603-489-8738 | none | Freelance | Middle All Grades | Special Education | ELA | 5 - 10 | No | Suburban | Man | White or Caucasian |
| Lacey Todd | Maine | ltodd@rsu10.org | (207) 890-6763 | (207) 890-6763 | Mountain Valley Middle School | Grade 5 Science Teacher | Elementary Middle | Special Education General Education | Science | more than 15 | Yes | Rural | Woman | White or Caucasian |
| Agosto Jerusalem | US Virgin Islands | agosto.jerusalem@vide.vi | 340-643-7871 | 340-643-7871 | Ivanna Eudora Kean High School | Special Education Teacher | Elementary Highschool | Special Education General Education Other (Please specify) *Grade 5 Science, Autism, Behavior, Emotional, Transition* | Science | more than 15 | yes | Urban | Man | Asian or Asian American |
| Kristen Nash | Arizona | kristen.nash@husd.org | 480-279-7153 | 480-435-0431 | Higley Unified School District | Assistant Director of Special Education | All Grades | Special Education Administrator | ELA Math Science | 10 - 15 | Yes | Suburban | Woman | White or Caucasian |
| Amanda Simcock | Maine | amanda.simcock@cognia.org | 2077101585 | 2077101585 | N/A | Accessibility Assessment Specialist | All grades | Special Education | N/A | more than 15 | No | Suburban | Woman | White or Caucasian |
| Francine Galko | Texas | f.galko@yahoo.com | 512-906-8480 | 512-906-8480 | No current school affiliation, but was teaching at Lake Travis STEM Academy spring 2021 | STEM Education Consultant | All grades | Special Education General Education Other (please specify) *ESL, ELL, math, science, engineering* | Math Science | more than 15 | No | Suburban | Woman | White or Caucasian |
| Donald Griffin | Bureau of Indian Education | donald.griffin@bie.edu | (703) 282-3316 | (skipped) | BIE Partner - BIE Chief Academic Office | BIE Section 504 Program Coordinator; BIE MSAA Coordinator | All Grades | Special Education General Education Administrator | N/A | more than 15 | No | Rural Urban Suburban | Man | American Indian or Alaska Native |

**Notes**

An Asterisk* indicate that the question was mandatory
this survey was given to the panelists via a SurveyMonkey
link from PM

**2022 Science Achievement Level Descriptors (ALD) Review Meeting Feedback Survey**

\* Overall, the review worked well.

| Strongly Agree | Agree | Neither Agree nor Disagree | Disagree | Strongly Disagree |
|:---:|:---:|:---:|:---:|:---:|
| ○ | ○ | ○ | ○ | ○ |

Comments:

---

* The opening overview group training was helpful.

| Strongly Agree | Agree | Neither Agree nor Disagree | Disagree | Strongly Disagree |
|:---:|:---:|:---:|:---:|:---:|
| ○ | ○ | ○ | ○ | ○ |

Comments:

[text box]

![MSAA - Multi-State Alternate Assessment logo]

2022 Science Achievement Level Descriptors (ALD) Review Meeting Feedback Survey

* The process for individual review, feedback, and recommendations worked well.

| Strongly Agree | Agree | Neither Agree nor Disagree | Disagree | Strongly Disagree |
|:---:|:---:|:---:|:---:|:---:|
| ○ | ○ | ○ | ○ | ○ |

Comments:

[text box]

* I had all the materials necessary to complete the task. (available electronically and displayed by the Facilitator)

| Strongly Agree | Agree | Neither Agree nor Disagree | Disagree | Strongly Disagree |
|:---:|:---:|:---:|:---:|:---:|
| ○ | ○ | ○ | ○ | ○ |

Comments:

2022 Science Achievement Level Descriptors (ALD) Review Meeting
Feedback Survey

Three things I liked best about this experience:

Three things I would change about this experience:

Do you have any other comments, questions or concerns?

# Q1 Overall, the review worked well.

Answered: 8     Skipped: 0
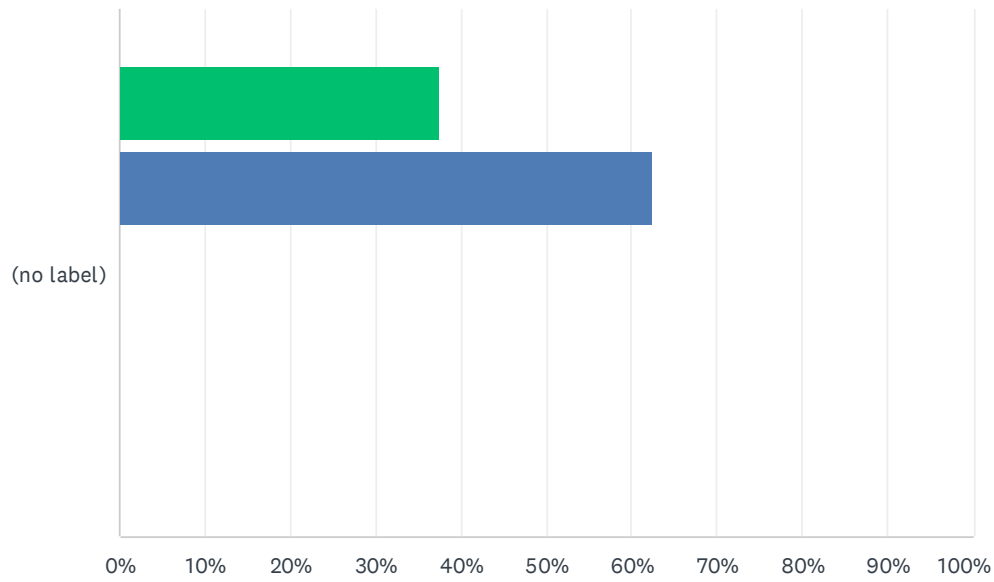


Strongly Ag...   Agree   Neither Agr...   Disagree
Strongly Di...

|  | STRONGLY AGREE | AGREE | NEITHER AGREE NOR DISAGREE | DISAGREE | STRONGLY DISAGREE | TOTAL | WEIGHTED AVERAGE |
|---|---|---|---|---|---|---|---|
| (no label) | 37.50% 3 | 62.50% 5 | 0.00% 0 | 0.00% 0 | 0.00% 0 | 8 | 1.63 |

# Q2 Comments:

Answered: 6    Skipped: 2

| # | RESPONSES | DATE |
|---|-----------|------|
| 1 | The process was very efficient. | 5/18/2022 7:57 AM |
| 2 | I think a little more time could have been spent reviewing the process involved for creating the draft ALD's. The training seemed slightly rushed. | 5/17/2022 8:07 AM |
| 3 | Everyone was open to comments, friendly, polite, and worked to elicit, understand, and carefully capture feedback. | 5/16/2022 8:15 PM |
| 4 | All the participants joined the discussion. | 5/16/2022 7:40 PM |
| 5 | The relationship between ESS and ALD was scrutinized and discussed. The level and classification of the progress of skills were also studied and discussed. I am glad to be part of this activity. | 5/16/2022 7:37 PM |
| 6 | lauren testing survey | 5/16/2022 1:33 PM |

# Q3 The opening overview group training was helpful.

Answered: 7    Skipped: 1
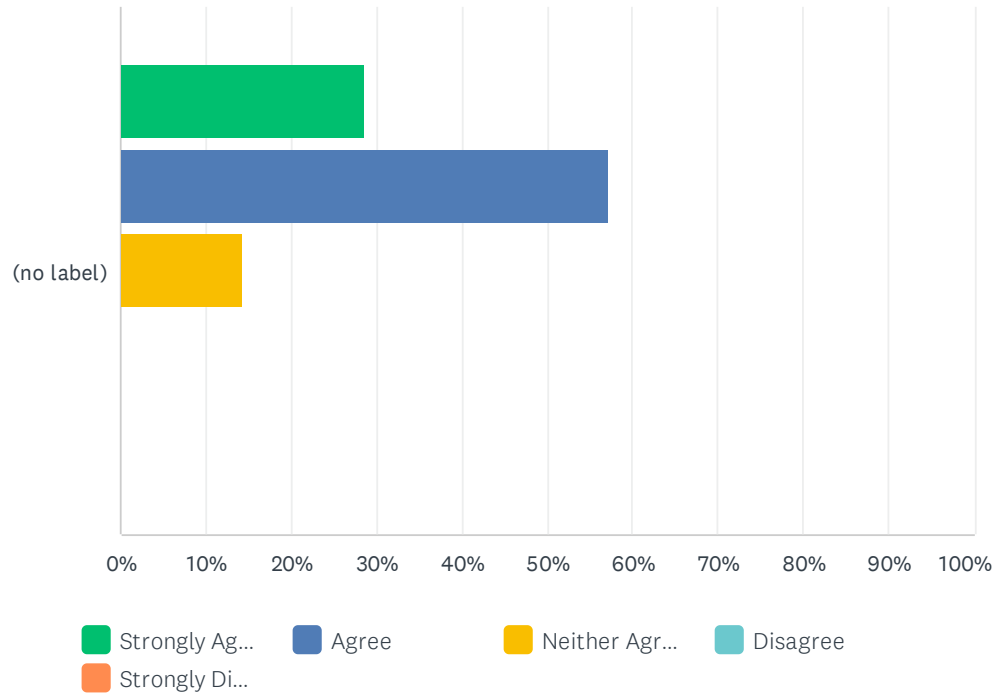


|  | STRONGLY AGREE | AGREE | NEITHER AGREE NOR DISAGREE | DISAGREE | STRONGLY DISAGREE | TOTAL | WEIGHTED AVERAGE |
|---|---|---|---|---|---|---|---|
| (no label) | 28.57%<br>2 | 57.14%<br>4 | 14.29%<br>1 | 0.00%<br>0 | 0.00%<br>0 | 7 | 1.86 |

# Q4 Comments:

Answered: 3   Skipped: 5

| # | RESPONSES | DATE |
|---|-----------|------|
| 1 | The intro was clear and concise. There are a lot of acronyms! It wasn't clear until I dug into the materials that I needed to flip back and forth. It would have been more helpful to do one row together to practice going through the materials and better understand the task and how the materials were to be used. | 5/16/2022 8:17 PM |
| 2 | It's helpful and important, but my gadget was uncooperative in the start. | 5/16/2022 7:42 PM |
| 3 | fdf | 5/16/2022 1:33 PM |

# Q5 The process for individual review, feedback, and recommendations worked well.

Answered: 7    Skipped: 1
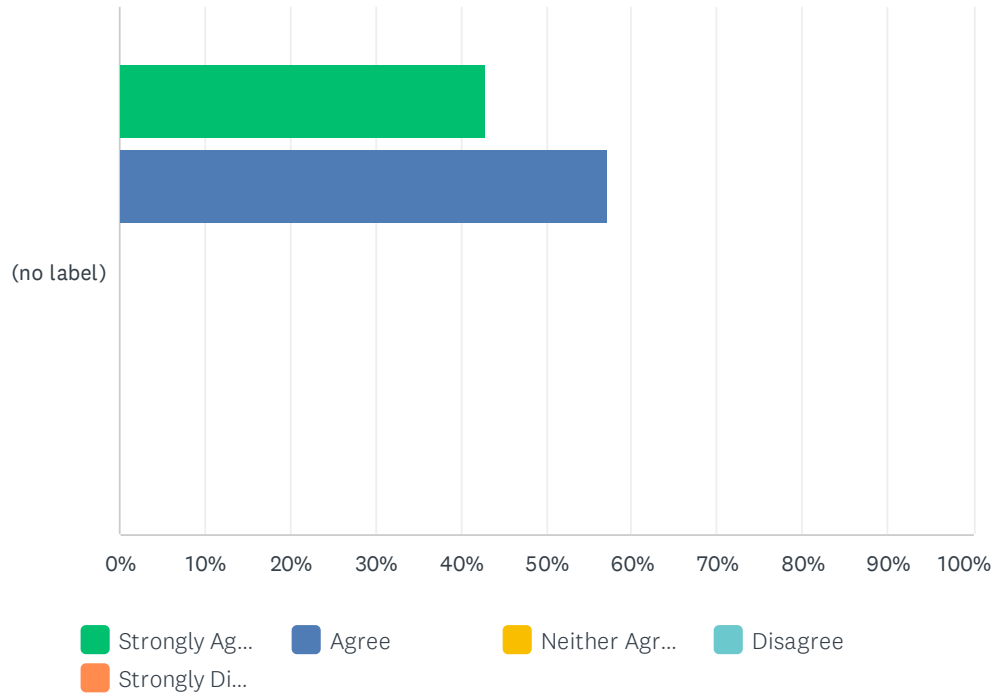


| | STRONGLY AGREE | AGREE | NEITHER AGREE NOR DISAGREE | DISAGREE | STRONGLY DISAGREE | TOTAL | WEIGHTED AVERAGE |
|---|---|---|---|---|---|---|---|
| (no label) | 42.86% 3 | 57.14% 4 | 0.00% 0 | 0.00% 0 | 0.00% 0 | 7 | 1.57 |

# Q6 Comments:

Answered: 3    Skipped: 5

| # | RESPONSES | DATE |
|---|-----------|------|
| 1 | I think it worked well overall, although it was hard to get through the materials in the allotted time. | 5/17/2022 8:09 AM |
| 2 | The process was organized. The facilitators were engaged and helpful. The part of the process that was hard was finding the EPEs because they were in a separate document and not in the same order as the spreadsheet. It would have been easier if the EPEs were copied over to the spreadsheet to streamline and speed up the process. I question whether all the reviewers were actually comparing the two documents because there was not enough time to do so. | 5/16/2022 8:19 PM |
| 3 | It's true that all participants shared their thoughts. | 5/16/2022 7:43 PM |

# Q7 I had all the materials necessary to complete the task. (available electronically and displayed by the Facilitator)

Answered: 7    Skipped: 1
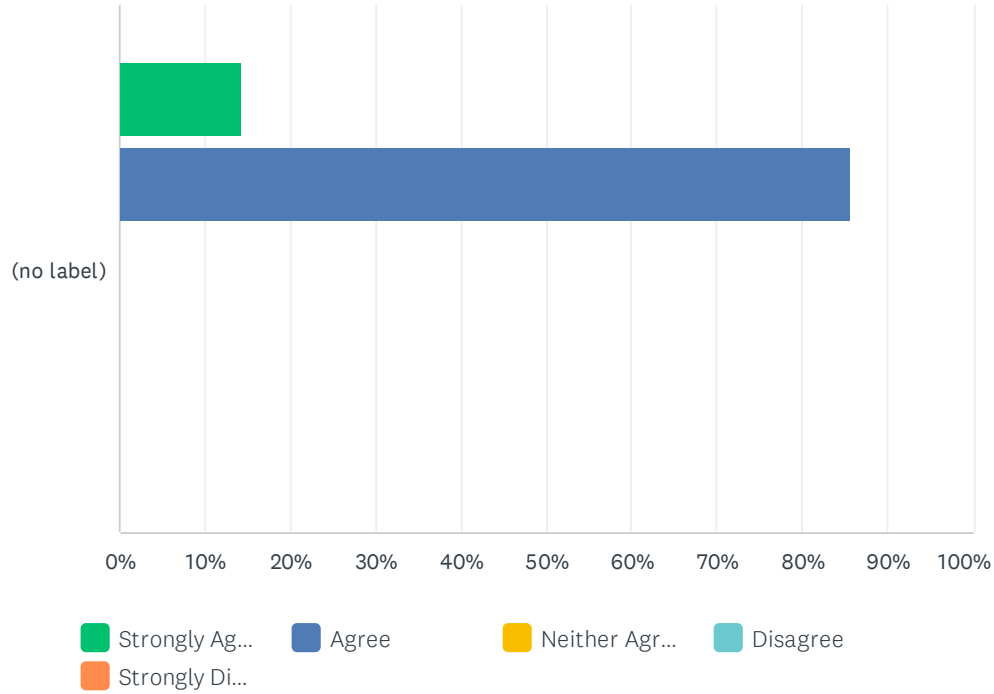


Strongly Ag...    Agree    Neither Agr...    Disagree
Strongly Di...

| | STRONGLY AGREE | AGREE | NEITHER AGREE NOR DISAGREE | DISAGREE | STRONGLY DISAGREE | TOTAL | WEIGHTED AVERAGE |
|---|---|---|---|---|---|---|---|
| (no label) | 14.29% 1 | 85.71% 6 | 0.00% 0 | 0.00% 0 | 0.00% 0 | 7 | 1.86 |

# Q8 Comments:

Answered: 5     Skipped: 3

| # | RESPONSES | DATE |
|---|-----------|------|
| 1 | It would have been helpful if we could have received the materials the night before so that we could print them. It was difficult staring at the computer screen all day. | 5/18/2022 7:58 AM |
| 2 | The EPE's could have been better organized for this meeting. It took a lot of time to find the EPE that aligned to the ALD. | 5/17/2022 8:10 AM |
| 3 | All of the materials were available and Mary-Alice did a good job of reading from the specs as needed. However, the materials could have been better organized to make the task easier and faster. | 5/16/2022 8:20 PM |
| 4 | I agreed, but in the beginning, my main computer got into trouble. The other computer was not checked to open the required files. The iPad was late in opening my mic. The facilitator was well prepared, not me. | 5/16/2022 7:47 PM |
| 5 | fdsf | 5/16/2022 1:33 PM |

# Q9 Three things I liked best about this experience:

Answered: 7    Skipped: 1

| # | RESPONSES | DATE |
|---|-----------|------|
| 1 | 1. Seeing the learning progression 2. Hearing everyone's perspectives 3. The efficiency of the process | 5/18/2022 2:10 PM |
| 2 | Collaborative work of the review. The materials were organized well. The facilitator did a great job of keeping participants engaged and on-track. | 5/17/2022 10:25 AM |
| 3 | Great group | 5/17/2022 8:11 AM |
| 4 | It was a chance to see how the test is put together. Feedback is greatly encouraged and all thoughts are welcome. | 5/17/2022 8:10 AM |
| 5 | (1) The facilitators!! They had such a patient, friendly, helpful, engaged attitude. Jill has the best tone and both Jill and Mary-Alice kept us on pace and reassured us they "got" the comments. (2) The open-mindedness of everyone. (3) That we had time before and after the process to get other work done--it did not take up the full day and it was online so easier for people to participate. | 5/16/2022 8:25 PM |
| 6 | 1. I learned that the ESS and EPE must be coherent. 2. I was able to observe, or participate in how this program was developed. 3. The framework of CSAA ALD was jointly discussed by panelists of color. | 5/16/2022 8:08 PM |
| 7 | testing survey | 5/16/2022 1:34 PM |

# Q10 Three things I would change about this experience:

Answered: 6     Skipped: 2

| # | RESPONSES | DATE |
|---|-----------|------|
| 1 | 1. Provide materials earlier to give enough time to print. I can't think of anything else I would change. | 5/18/2022 2:10 PM |
| 2 | More time to review More organized materials More thorough explanation of development process | 5/17/2022 8:11 AM |
| 3 | A little more clarity on what the ALD is used for. Maybe slightly longer breaks although the administrators were not sticklers on time | 5/17/2022 8:10 AM |
| 4 | Streamline the materials (put the EPEs side-by-side with the ADLs), better explain the purpose of the ADLs vs EPEs, do the first 1-2 reviews together as a group to ensure that everyone is using the materials as intended and thinking about the aspects that are most useful to you. | 5/16/2022 8:25 PM |
| 5 | I must prepare my computer set up. I must have 3 computers working. | 5/16/2022 8:08 PM |
| 6 | testing survey | 5/16/2022 1:34 PM |

# Q11 Do you have any other comments, questions or concerns?

Answered: 4    Skipped: 4

| # | RESPONSES | DATE |
|---|-----------|------|
| 1 | No | 5/18/2022 2:10 PM |
| 2 | I liked how positive everyone was and thank you for not letting us get too bogged down in the details. | 5/16/2022 8:25 PM |
| 3 | I think the discussions were so broad and we needed more specific examples to elaborate on each piece of information. | 5/16/2022 8:08 PM |
| 4 | testing survey | 5/16/2022 1:34 PM |

# APPENDIX C
# PERFORMANCE LEVEL DESCRIPTORS

**cognia**™

# Cognia Science Alternate Assessment Achievement Level Descriptors

## Policy Achievement Level Descriptors (ALDs)

Policy ALDs define the knowledge and skill level expectations for grades 5, 8, and high school for the Cognia Science Alternate Assessment (CSAA).

### Level 4 (Exceeding Expectations)

Students at Level 4 can be expected to demonstrate understanding and skills of the K-12 science framework Extended Performance Expectations.

Students can be expected to demonstrate understanding and skills in the disciplinary core ideas, practices, and crosscutting concepts from the K-12 science framework Extended Performance Expectations in more sophisticated ways than students in Level 3 to address science phenomena and problems that are complex, more abstract, and/or multi-factorial. Students are expected to describe, explain, and/or respond to phenomena and problems using reasonably complex evidence, analysis, and inference in Level 4.

At Level 4, students are expected to have the knowledge and skills described in Level 3.

### Level 3 (Meeting Expectations)

Students at Level 3 can be expected to demonstrate knowledge and skills of the K-12 science framework Extended Performance Expectations.

Students can be expected to demonstrate knowledge and skills in the majority of disciplinary core ideas, practices, and crosscutting concepts from the K-12 science framework Extended Performance Expectations to address moderately complex science phenomena and problems, some concrete and some abstract in Level 3.

At Level 3, students are expected to have the knowledge and skills of Level 2 and may be able to demonstrate some of the knowledge and skills described in Level 4.

## Level 2 (Approaching Expectations)

Students at Level 2 can be expected to demonstrate developing knowledge and skills of the K-12 science framework Extended Performance Expectations.

Students can be expected to demonstrate developing knowledge and skills in some disciplinary core ideas together with some aspects of the practices and crosscutting concepts from the K-12 science framework Extended Performance Expectations to address primarily basic and concrete science phenomena and problems in Level 2.

At Level 2, students are expected to have the knowledge and skills of Level 1 and may be able to demonstrate some of the knowledge and skills described in Level 3.

## Level 1 (Beginning – In need of additional support)

Students at Level 1 are beginning to access the science content and can be expected to need additional support to demonstrate knowledge and skills of the K-12 science framework Extended Performance Expectations.

Students attempt to perform basic science tasks but will require additional support in order to demonstrate knowledge and skills of the K-12 science framework Extended Performance Expectations by using disciplinary core ideas, practices, and/or crosscutting concepts to address more basic and concrete science phenomena and problems in Level 1.

# Borderline ALDs

*Borderline ALDs describe the knowledge and skills that students just barely within each proficiency level are expected to be able to demonstrate. In line with the nature of the Next Generation Science Standards (NGSS), the statements combine science and engineering practices (SEPs), disciplinary core ideas (DCIs), and crosscutting concepts (CCCs) that students are expected to integrate and demonstrate.*

## Grade 5, Level 4 (Exceeding Expectations)

Students performing at the borderline of Level 4 Exceeding Expectations can be expected at least 50% of the time to demonstrate knowledge and skills such as those listed below as evidence of just barely exceeding expectations. This includes understanding and use of the Extended Performance Expectations in more sophisticated ways than students in Level 3 to address science phenomena and problems that are complex, more abstract, and/or multi-factorial. Students may also sometimes demonstrate the skills and understandings at the Meeting Expectations level rather than the skills and understandings of the Exceeding Expectations level.

- Show that the weight of matter does not change when substances are heated, cooled, or mixed by measuring, graphing, or using mathematical relationships.

- Determine predictable patterns in the motion of an object by using observations or measurements from an investigation.

- Support the claim that Earth's gravity pulls objects downward (toward Earth's center) by describing evidence (observations, data, or a model).

- Identify which design or improvement will maximize energy transfer from one form to another by designing or modifying a device.

- Describe how the energy animals obtain from food comes from the Sun by using a model.

- Describe evidence to support a claim that parts of plants and/or animals have specific functions that help them survive, grow, or reproduce by using evidence from data and/or a model.

- Describe how patterns in trait variation between groups of organisms (e.g., parents and their offspring, siblings, populations of similar organisms) provide evidence of inheritance between parents and their offspring and that there are differences in these traits by analyzing and interpreting data.

- Describe the type of environment in which plants and/or animals lived on Earth long ago by using observations of fossils and/or data.

- Predict or infer patterns concerning the rotation of Earth, Earth's orbit around the Sun, or the Moon's orbit around Earth by analyzing data (e.g., length and direction of shadows, day and night, seasonal appearance of stars) or a model.

- Predict weather conditions for a particular season by analyzing patterns in weather data.

- Represent the interaction between two Earth systems (e.g., geosphere, hydrosphere, atmosphere, biosphere) by developing a model.

- Describe how humans are using science to protect Earth's resources and/or the environment by using data.

## Grade 5, Level 3 (Meeting Expectations)

Students performing at the borderline of Level 3 Meeting Expectations can be expected at least 50% of the time to demonstrate knowledge and skills such as those listed below as evidence of just barely meeting expectations. This includes satisfactory knowledge and skills in most disciplinary core ideas, practices, and crosscutting concepts from the K-12 science framework Extended Performance Expectations to address moderately complex science phenomena and problems, some concrete and some abstract. Students may also sometimes demonstrate the skills and understandings of the Approaching Expectations level rather than the skills and understandings of the Meeting Expectations level.

- Compare the weight of matter before and after heating, cooling, or mixing by using data.
- Predict the future motion of an object by using observations or data.
- Show the direction objects move when released on Earth (downward toward Earth's center) by identifying or developing a model.
- Describe the various ways that energy transfer can occur between everyday objects or devices.
- Describe the direction of energy transfer between two organisms (e.g., plant to animal, animal to animal) or between the Sun and a plant by using a model.
- Describe how parts of plants or animals have specific functions that help them survive, grow, or reproduce by using data and/or a model.
- Describe patterns in trait variation between groups of organisms (e.g., parents and their offspring, siblings, populations of similar organisms) by using data or observations.
- Describe how modern-day plants or animals compare to their ancestors by using observations of fossils and/or data.
- Identify patterns concerning the rotation of Earth, Earth's orbit around the Sun, or the Moon's orbit around Earth by analyzing data (e.g., length and direction of shadows, day and night, seasonal appearance of stars) or a model.
- Describe patterns of weather conditions for a particular season by analyzing weather data.
- Describe the interaction between two Earth systems (e.g., geosphere, hydrosphere, atmosphere, biosphere) by using a model.
- Describe an effect (positive or negative) of human activities on the environment by using data.

## Grade 5, Level 2 (Approaching Expectations)

Students performing at the borderline of Level 2 Approaching Expectations can be expected at least 50% of the time to demonstrate knowledge and skills such as those listed below as evidence of just barely approaching expectations. This includes some disciplinary core ideas together with some aspects of the practices and crosscutting concepts from the Extended Performance Expectations to address primarily basic and concrete science phenomena and problems. Students may also sometimes demonstrate the skills and understandings of the Beginning level rather than the skills and understandings of the Approaching Expectations level.

- Identify the appropriate tools or units of measurement (for weight, time, temperature, or volume) for a scientific task.

- Identify patterns in the motion of an object by using observations or data.

- Identify patterns in the motion of falling objects on Earth by using observations.

- Identify the various forms of energy present in a system.

- Identify that the Sun is a source of energy for ecosystems by using patterns in food chains or drawings of ecosystems.

- Identify the parts of plants or animals that have specific functions by using evidence from data and/or a model.

- Identify patterns in trait variations between parents and their baby/babies by using data or observations.

- Identify that plants and/or animals lived on Earth long ago by using information about fossils and/or data.

- Identify the positions of the Sun, the Moon, and Earth in the solar system by using data or a model.

- Describe weather conditions by using data of weather observations.

- Identify parts of an Earth system (e.g., geosphere, hydrosphere, atmosphere, biosphere) by using data or a model.

- Identify a natural or human impact on the environment by using data.

# Grade 5, Level 1 (Beginning - In need of additional support)

Students attempt to perform basic science tasks but will require additional support to demonstrate knowledge and skills of the K-12 science framework Extended Performance Expectations by using disciplinary core ideas, practices, and/or crosscutting concepts to address more basic and concrete science phenomena and problems in Level 1.

- Attempt to identify the appropriate tools or units of measurement (for weight, time, temperature, or volume) for a scientific task.
- Attempt to identify patterns in the motion of an object by using observations or data.
- Attempt to identify patterns in the motion of falling objects on Earth by using observations.
- Attempt to identify various forms of energy present in a system.
- Attempt to identify that the Sun is a source of energy for ecosystems by using patterns in food chains or drawings of ecosystems.
- Attempt to identify the parts of plants or animals that have specific functions by using evidence from data and/or a model.
- Attempt to identify patterns in trait variations between parents and their baby/babies by using data or observations.
- Attempt to recognize that there was life on Earth long ago by using information about fossils and/or data.
- Attempt to identify the positions of the Sun, the Moon, and Earth in the solar system by using data or a model.
- Attempt to describe weather conditions by using data of weather observations.
- Attempt to identify parts of an Earth system (e.g., geosphere, hydrosphere, atmosphere, biosphere) by using data or a model.
- Attempt to identify a natural or human impact on the environment by using data.

# Cognia Science Alternate Assessment Achievement Level Descriptors

## Policy Achievement Level Descriptors (ALDs)

Policy ALDs define the knowledge and skill level expectations for grades 5, 8, and high school for the Cognia Science Alternate Assessment (CSAA).

### Level 4 (Exceeding Expectations)

Students at Level 4 can be expected to demonstrate understanding and skills of the K-12 science framework Extended Performance Expectations.

Students can be expected to demonstrate understanding and skills in the disciplinary core ideas, practices, and crosscutting concepts from the K-12 science framework Extended Performance Expectations in more sophisticated ways than students in Level 3 to address science phenomena and problems that are complex, more abstract, and/or multi-factorial. Students are expected to describe, explain, and/or respond to phenomena and problems using reasonably complex evidence, analysis, and inference in Level 4.

At Level 4, students are expected to have the knowledge and skills described in Level 3.

### Level 3 (Meeting Expectations)

Students at Level 3 can be expected to demonstrate knowledge and skills of the K-12 science framework Extended Performance Expectations.

Students can be expected to demonstrate knowledge and skills in the majority of disciplinary core ideas, practices, and crosscutting concepts from the K-12 science framework Extended Performance Expectations to address moderately complex science phenomena and problems, some concrete and some abstract in Level 3.

At Level 3, students are expected to have the knowledge and skills of Level 2 and may be able to demonstrate some of the knowledge and skills described in Level 4.

## Level 2 (Approaching Expectations)

Students at Level 2 can be expected to demonstrate developing knowledge and skills of the K-12 science framework Extended Performance Expectations.

Students can be expected to demonstrate developing knowledge and skills in some disciplinary core ideas together with some aspects of the practices and crosscutting concepts from the K-12 science framework Extended Performance Expectations to address primarily basic and concrete science phenomena and problems in Level 2.

At Level 2, students are expected to have the knowledge and skills of Level 1 and may be able to demonstrate some of the knowledge and skills described in Level 3.

## Level 1 (Beginning – In need of additional support)

Students at Level 1 are beginning to access the science content and can be expected to need additional support to demonstrate knowledge and skills of the K-12 science framework Extended Performance Expectations.

Students attempt to perform basic science tasks but will require additional support in order to demonstrate knowledge and skills of the K-12 science framework Extended Performance Expectations by using disciplinary core ideas, practices, and/or crosscutting concepts to address more basic and concrete science phenomena and problems in Level 1.

# Borderline ALDs

*Borderline ALDs describe the knowledge and skills that students just barely within each proficiency level are expected to be able to demonstrate. In line with the nature of the Next Generation Science Standards (NGSS), the statements combine science and engineering practices (SEPs), disciplinary core ideas (DCIs), and crosscutting concepts (CCCs) that students are expected to integrate and demonstrate.*

## Grade 8, Level 4 (Exceeding Expectations)

Students performing at the borderline of Level 4 Exceeding Expectations can be expected at least 50% of the time to demonstrate knowledge and skills such as those listed below as evidence of just barely exceeding expectations. This includes understanding and use of the Extended Performance Expectations in more sophisticated ways than students in Level 3 to address science phenomena and problems that are complex, more abstract, and/or multi-factorial. Students may also sometimes demonstrate the skills and understandings at the Meeting Expectations level rather than the skills and understandings of the Exceeding Expectations level.

- Determine whether a chemical reaction occurred by using data or observations on the properties of substances before and after an interaction.
- Describe how the mass of an object or the force on an object will change the motion of the object by using data from an investigation.
- Make or support a claim that a transfer of energy occurs when the kinetic energy of an object changes by using data as evidence.
- Represent what happens to waves when they are reflected, absorbed, or transmitted through different materials by developing a model.
- Make a claim about two body systems (e.g., circulatory, respiratory, muscular, digestive, nervous, excretory) working together to carry out various functions by using evidence.
- Explain how the growth of organisms is influenced by various environmental and/or genetic factors by using data.
- Identify evidence of a cause-and-effect relationship between resource availability and growth of organisms and/or populations of organisms by analyzing data.
- Describe how energy is transferred or how matter is cycled among living and nonliving parts of ecosystems by developing a model.
- Compare or show patterns in seasons, lunar phases, or eclipses by using or developing a model of the Earth-Sun-Moon system.
- Explain how geological processes on Earth have caused changes to Earth's surface at various time or spatial scales by using evidence to support an explanation.
- Describe how the Sun's energy or the force of gravity moves water through the water cycle by developing a model.
- Select or evaluate a design for a method that is intended to minimize a human impact on the environment by using data.

# Grade 8, Level 3 (Meeting Expectations)

Students performing at the borderline of Level 3 Meeting Expectations can be expected at least 50% of the time to demonstrate knowledge and skills such as those listed below as evidence of just barely meeting expectations. This includes satisfactory knowledge and skills in most disciplinary core ideas, practices, and crosscutting concepts from the K-12 science framework Extended Performance Expectations to address moderately complex science phenomena and problems, some concrete and some abstract. Students may also sometimes demonstrate the skills and understandings of the Approaching Expectations level rather than the skills and understandings of the Meeting Expectations level.

- Determine the identities of substances by using data or observations on the properties of substances.

- Identify the change in an object's motion when the mass of the object or the force on the object is changed by using data from an investigation.

- Identify the forms of energy that increase or decrease when the kinetic energy of an object changes by using data as evidence.

- Describe the path of a wave that is reflected, absorbed, or transmitted through different materials by using a model.

- Identify the parts that belong to a particular body system and the organization of those parts by using a model.

- Determine whether a particular factor is affecting the growth of organisms by analyzing data.

- Describe the effects of resource availability on organisms and/or populations of organisms by using data or observations.

- Identify how energy is transferred or that matter is cycled from one specific part of an ecosystem to another specific part by using a model.

- Describe or compare the positions of the Sun, the Moon, and Earth or the amount or path of light in the cyclic patterns of seasons, lunar phases, or eclipses by using a model.

- Identify whether a geological process or event on Earth was small/large scale and/or whether a process or event happened gradually/rapidly by using information in charts, diagrams, or graphic organizers.

- Describe the state of water or how water changes state in various parts of the water cycle by using a model.

- Make a claim about how a particular method would work to reduce a human impact on the environment by using data.

## Grade 8, Level 2 (Approaching Expectations)

Students performing at the borderline of Level 2 Approaching Expectations can be expected at least 50% of the time to demonstrate knowledge and skills such as those listed below as evidence of just barely approaching expectations. This includes some disciplinary core ideas together with some aspects of the practices and crosscutting concepts from the Extended Performance Expectations to address primarily basic and concrete science phenomena and problems. Students may also sometimes demonstrate the skills and understandings of the Beginning level rather than the skills and understandings of the Approaching Expectations level.

- Identify properties of a substance by using data or observations.
- Identify the effects of pushes and pulls on objects by using data from an investigation.
- Determine whether energy is being transferred in a system by asking questions or by using data.
- Identify whether a wave is being reflected, absorbed, or transmitted through a material by using data or a model.
- Identify structures that are part of human body systems and those that are not by using charts, diagrams, or graphic organizers.
- Identify factors that could be affecting the growth of an organism by asking questions.
- Identify resources (e.g., food, water, nutrients, space) that are necessary for the growth or survival of organisms or populations of organisms by using data.
- Identify the role of organisms (e.g., producer, consumer, decomposer) or nonliving things (e.g., the Sun, water, minerals, air) in cycling energy or matter in an ecosystem by using a model.
- Show the positions of Earth (with its tilt), the Sun, and the Moon as Earth orbits the Sun and the Moon orbits Earth in the solar system by identifying a model.
- Identify the process or agent that causes a particular change to Earth's surface by using observations as evidence.
- Trace the path of water through Earth's systems by using a model.
- Identify an environmental problem caused by human activities/impact by using data.

# Grade 8, Level 1 (Beginning - In need of additional support)

Students attempt to perform basic science tasks but will require additional support to demonstrate knowledge and skills of the K-12 science framework Extended Performance Expectations by using disciplinary core ideas, practices, and/or crosscutting concepts to address more basic and concrete science phenomena and problems in Level 1.

- Attempt to identify properties of a substance by using data or observations.

- Attempt to identify the effects of pushes and pulls on objects by using data from an investigation.

- Attempt to determine whether energy is being transferred in a system by asking questions or by using data.

- Attempt to identify whether a wave is being reflected, absorbed, or transmitted through a material by using data or a model.

- Attempt to identify structures that are part of human body systems and those that are not by using charts, diagrams, or graphic organizers.

- Attempt to identify factors that could be affecting the growth of an organism by asking questions.

- Attempt to identify resources (e.g., food, water, nutrients, space) that are necessary for the growth or survival of organisms or populations of organisms by using data.

- Attempt to identify the role of organisms (e.g., producer, consumer, decomposer) or nonliving things (e.g., the Sun, water, minerals, air) in cycling energy or matter in an ecosystem by using a model.

- Attempt to show the positions of Earth (with its tilt), the Sun, and the Moon as Earth orbits the Sun and the Moon orbits Earth in the solar system by identifying a model.

- Attempt to identify the process or agent that causes a particular change to Earth's surface by using observations as evidence.

- Attempt to trace the path of water through Earth's systems by using a model.

- Attempt to identify an environmental problem caused by human activities/impact by using data.

# Cognia Science Alternate Assessment Achievement Level Descriptors

## Policy Achievement Level Descriptors (ALDs)

Policy ALDs define the knowledge and skill level expectations for grades 5, 8, and high school for the Cognia Science Alternate Assessment (CSAA).

### Level 4 (Exceeding Expectations)

Students at Level 4 can be expected to demonstrate understanding and skills of the K-12 science framework Extended Performance Expectations.

Students can be expected to demonstrate understanding and skills in the disciplinary core ideas, practices, and crosscutting concepts from the K-12 science framework Extended Performance Expectations in more sophisticated ways than students in Level 3 to address science phenomena and problems that are complex, more abstract, and/or multi-factorial. Students are expected to describe, explain, and/or respond to phenomena and problems using reasonably complex evidence, analysis, and inference in Level 4.

At Level 4, students are expected to have the knowledge and skills described in Level 3.

### Level 3 (Meeting Expectations)

Students at Level 3 can be expected to demonstrate knowledge and skills of the K-12 science framework Extended Performance Expectations.

Students can be expected to demonstrate knowledge and skills in the majority of disciplinary core ideas, practices, and crosscutting concepts from the K-12 science framework Extended Performance Expectations to address moderately complex science phenomena and problems, some concrete and some abstract in Level 3.

At Level 3, students are expected to have the knowledge and skills of Level 2 and may be able to demonstrate some of the knowledge and skills described in Level 4.

## Level 2 (Approaching Expectations)

Students at Level 2 can be expected to demonstrate developing knowledge and skills of the K-12 science framework Extended Performance Expectations.

Students can be expected to demonstrate developing knowledge and skills in some disciplinary core ideas together with some aspects of the practices and crosscutting concepts from the K-12 science framework Extended Performance Expectations to address primarily basic and concrete science phenomena and problems in Level 2.

At Level 2, students are expected to have the knowledge and skills of Level 1 and may be able to demonstrate some of the knowledge and skills described in Level 3.

## Level 1 (Beginning – In need of additional support)

Students at Level 1 are beginning to access the science content and can be expected to need additional support to demonstrate knowledge and skills of the K-12 science framework Extended Performance Expectations.

Students attempt to perform basic science tasks but will require additional support in order to demonstrate knowledge and skills of the K-12 science framework Extended Performance Expectations by using disciplinary core ideas, practices, and/or crosscutting concepts to address more basic and concrete science phenomena and problems in Level 1.

# Borderline ALDs

*Borderline ALDs describe the knowledge and skills that students just barely within each proficiency level are expected to be able to demonstrate. In line with the nature of the Next Generation Science Standards (NGSS), the statements combine science and engineering practices (SEPs), disciplinary core ideas (DCIs), and crosscutting concepts (CCCs) that students are expected to integrate and demonstrate.*

## High School, Level 4 (Exceeding Expectations)

Students performing at the borderline of Level 4 Exceeding Expectations can be expected at least 50% of the time to demonstrate knowledge and skills such as those listed below as evidence of just barely exceeding expectations. This includes understanding and use of the Extended Performance Expectations in more sophisticated ways than students in Level 3 to address science phenomena and problems that are complex, more abstract, and/or multi-factorial. Students may also sometimes demonstrate the skills and understandings at the Meeting Expectations level rather than the skills and understandings of the Exceeding Expectations level.

- Construct an explanation for why specific chemical reactions occur by using a periodic table.

- Select, evaluate, or revise the design of a familiar device that minimizes the forces on a macroscopic object during a collision.

- Plan or conduct an investigation to determine cause-and-effect relationships between magnetic fields and electric current.

- Describe how energy is conserved at the macroscopic or particle level when kinetic and/or potential energy are transferred or converted from one form to another in a system by developing or using models.

- Explain how a factor affects population size or biodiversity in an ecosystem at different scales (e.g., habitat size compared to population size) by using mathematical representations of data.

- Ask questions that will provide information about the cause-and-effect relationships among DNA/chromosomes and/or traits that are inherited from parents to offspring.

- Describe how comparing patterns in data (e.g., DNA sequences, amino acid sequences, structures found in organisms, embryos, fossils) provide evidence for evolution and common ancestry of living things.

- Demonstrate that organisms with helpful traits increase in proportion to organisms lacking those traits by using data as evidence.

- Explain Earth's formation and early history by using data about ancient Earth materials, meteorites, or other planetary surfaces.

- Predict or draw conclusions about how various factors (e.g., large volcanic eruptions, human activity, solar output, changes to Earth's orbit and axis, changes to atmospheric composition) affect Earth's climate (measured as changes in surface temperatures, precipitation patterns, glacial ice volumes, sea levels, biosphere distribution) by using models.

- Plan or conduct an investigation of the properties of water and its effects on Earth materials and surface processes (e.g., stream transportation and deposition by using a stream table, frost wedging by the expansion of water as it freezes, chemical weathering and recrystallization by testing the solubility of different materials).

- Evaluate or refine the design of a local technological solution that reduces the negative impact of human activities on natural systems.

## High School, Level 3 (Meeting Expectations)

Students performing at the borderline of Level 3 Meeting Expectations can be expected at least 50% of the time to demonstrate knowledge and skills such as those listed below as evidence of just barely meeting expectations. This includes satisfactory knowledge and skills in most disciplinary core ideas, practices, and crosscutting concepts from the K-12 science framework Extended Performance Expectations to address moderately complex science phenomena and problems, some concrete and some abstract. Students may also sometimes demonstrate the skills and understandings of the Approaching Expectations level rather than the skills and understandings of the Meeting Expectations level.

- Identify or classify elements that will react similarly in chemical reactions by using a periodic table.
- Construct a claim for how a familiar device functions to minimize the forces on a macroscopic object during a collision.
- Predict or draw conclusions about how a change to a system affects how electric current produces magnetic fields or how magnetic fields produce electric current by using data.
- Show how kinetic and potential energy change in a system when an object's position changes or when the particles making up an object change their motion by using a model.
- Describe how a factor affects population size or biodiversity in an ecosystem by interpreting data.
- Describe how genes and traits are inherited from parents to offspring by using a model.
- Draw conclusions about patterns of relatedness among organisms by using data (e.g., DNA sequences, amino acid sequences, structures found in organisms, embryos, fossils).
- Describe changes in the distribution of physical traits that can vary in a population by using data.
- Describe Earth's formation and early history by asking questions about ancient Earth materials, meteorites, and other planetary surfaces.
- Describe how energy from the Sun drives Earth's climate system by using a model.
- Draw conclusions about how water affects Earth's materials and surface processes by using data or observations.
- Construct a claim about how a local technological solution reduces the negative impact of human activities on natural systems.

## High School, Level 2 (Approaching Expectations)

Students performing at the borderline of Level 2 Approaching Expectations can be expected at least 50% of the time to demonstrate knowledge and skills such as those listed below as evidence of just barely approaching expectations. This includes some disciplinary core ideas together with some aspects of the practices and crosscutting concepts from the Extended Performance Expectations to address primarily basic and concrete science phenomena and problems. Students may also sometimes demonstrate the skills and understandings of the Beginning level rather than the skills and understandings of the Approaching Expectations level.

- Show how substances react by using provided information to complete an incomplete chemical reaction model.
- Identify how forces are acting on a macroscopic object during a collision in a model.
- Identify examples of electric current producing magnetic fields or magnetic fields producing electric current by using data or observations.
- Identify questions that would determine whether an object's kinetic or potential energy is changing in a system.
- Identify factors that affect population size or biodiversity by using provided information.
- Identify the function of DNA or chromosomes by using provided information.
- Identify how organisms have changed over time by using provided information.
- Identify physical traits that can vary in an organism by using provided information.
- Identify patterns in data about ancient Earth materials, meteorites, or other planetary surfaces by using data.
- Identify how energy flows between two Earth systems by using a model.
- Identify testable questions about how water affects Earth's materials and surface processes.
- Identify the positive or negative impacts of local human activities on natural systems by using data.

# High School, Level 1 (Beginning - In need of additional support)

Students attempt to perform basic science tasks but will require additional support to demonstrate knowledge and skills of the K-12 science framework Extended Performance Expectations by using disciplinary core ideas, practices, and/or crosscutting concepts to address more basic and concrete science phenomena and problems in Level 1.

- Attempt to show how substances react in a chemical reaction by using provided information to complete an incomplete chemical reaction model.

- Attempt to identify how forces are acting on a macroscopic object during a collision in a model.

- Attempt to identify examples of electric current producing magnetic fields or magnetic fields producing electric current by using data or observations.

- Attempt to identify questions that would determine whether an object's kinetic or potential energy is changing in a system.

- Attempt to identify factors that affect population size or biodiversity by using provided information.

- Attempt to identify the function of DNA or chromosomes by using provided information.

- Attempt to identify how organisms have changed over time by using provided information.

- Attempt to identify physical traits that can vary in an organism by using provided information.

- Attempt to identify patterns in data about ancient Earth materials, meteorites, or other planetary surfaces by using data.

- Attempt to identify how energy flows between two Earth systems by using a model.

- Attempt to identify testable questions about how water affects Earth's materials and surface processes.

- Attempt to identify the positive or negative impacts of local human activities on natural systems by using data.

# APPENDIX D
# POWERPOINT PRESENTATIONS

# MSAA Science

Standard Setting Orientation

July 2023

# Welcome!

Multi-State Alternate Assessment Standard Setting Meeting for Grades 5, 8, and High School

# Orientation Session - Agenda

- Introduction of the Standard Setting Team
- Standard Setting Goals and Outcomes
- Overview of the MSAA Science Assessments
- Post-secondary outcomes for MSAA students
- Overview of Standard Setting
- Overview of Performance Level Descriptors
- Overview of Key Concepts and Procedures
- Review Meeting Agenda

# Standard Setting Team

- MSAA Science Partners
  - Kim Pilitati (American Samoa)
  - Anju Kuriakose (Arizona)
  - Bethany Spangenberg (Arizona)
  - Don Griffin (BIE)
  - Fasefulu Tigilau (CNMI)
  - Jodi Bossio-Smith (Maine)
  - Nicole Richardson (Maine)
  - Emma Rose McCadden (Vermont)
  - Kathryn Rossman (Vermont)

- Cognia Staff
  - Jason Brodeur (Program Manager)
  - David Harrison (HS Facilitator)
  - Jill Stepanek (G8 Facilitator)
  - Karen Whisler (G5 Facilitator)
  - Thurman Munn (Content Specialist)
  - Jami Nelson (Content Specialist)
  - Frank Padellaro (Psychometrician)
  - Qi Qin (Psychometrician)
  - Sandra Sweeney (Psychometrician)

- MSAA Technical Advisory Committee Member
  - Rachel Quenemoen (observer)
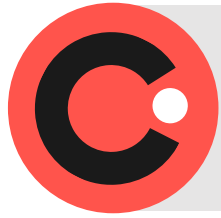
# Standard Setting Goals

- Our shared goals
  - Collect your recommendations on performance standards for MSAA science assessments that provide meaningful and actionable information

- Your goals as panelists
  - Learn concepts and procedures following the Item-Descriptor (ID) Matching method
  - Follow the procedures to complete the standard setting activities
  - Rely on your expertise about the content standards, student learning, and this population of students throughout the process

# Expectations of All Panelists

Listen and Collaborate

High Expectations

Follow the Guided Standard Setting Process
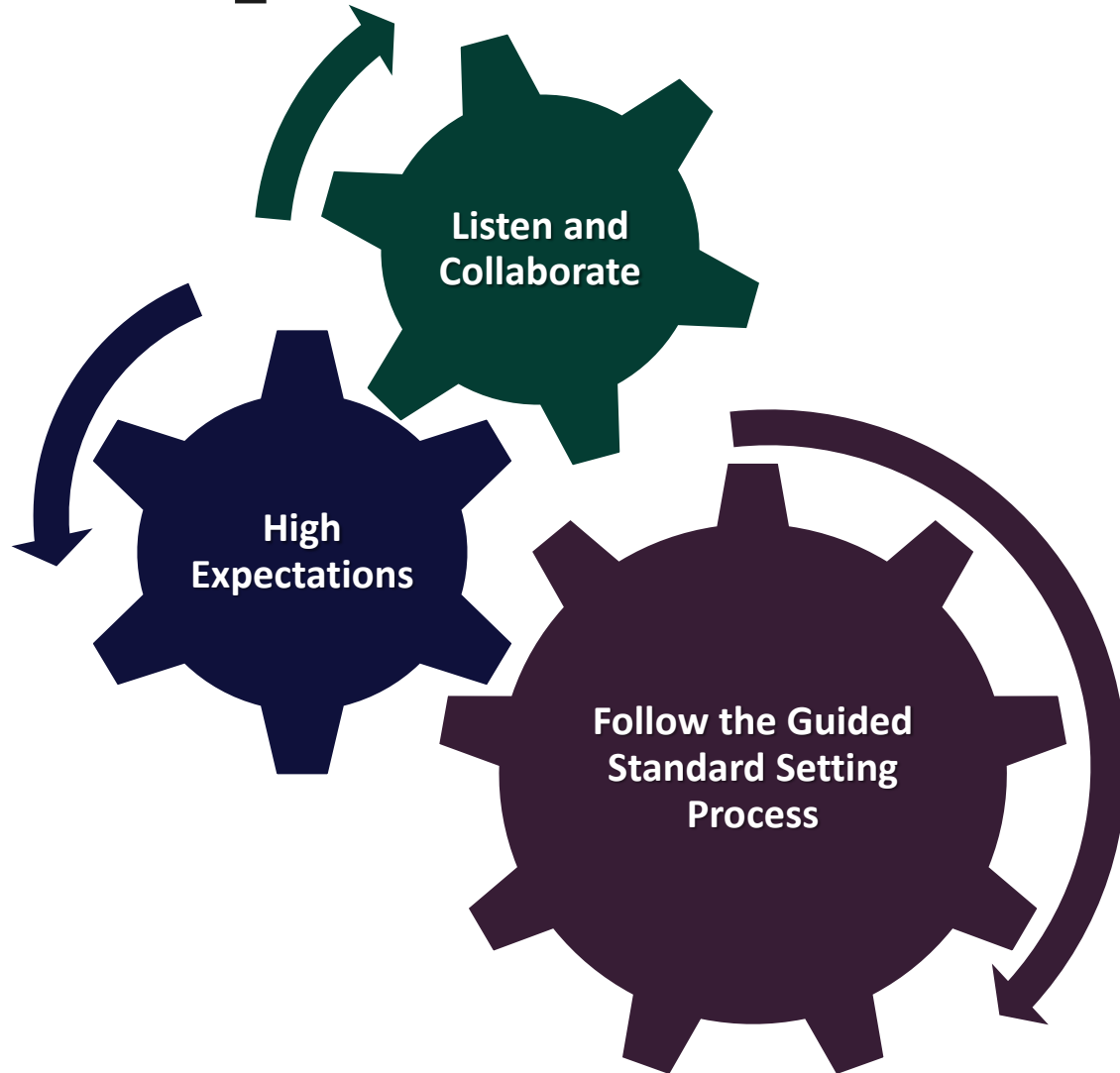
- Security is of the utmost importance
  - You can discuss the process in general terms
- You may NOT
  - Share details about the items or specific details about the process (e.g., results or cut scores)
  - Use your phones or personal devices while in the room
  - Use the Chromebooks for anything other than standard setting activities

# Assessment overview topics

- Assessing students on an alternate assessment
- Postsecondary outcomes for students
- The three dimensions of science learning
- MSAA Science Test Design overview
- Assessment features and accommodations

# What is MSAA Science?
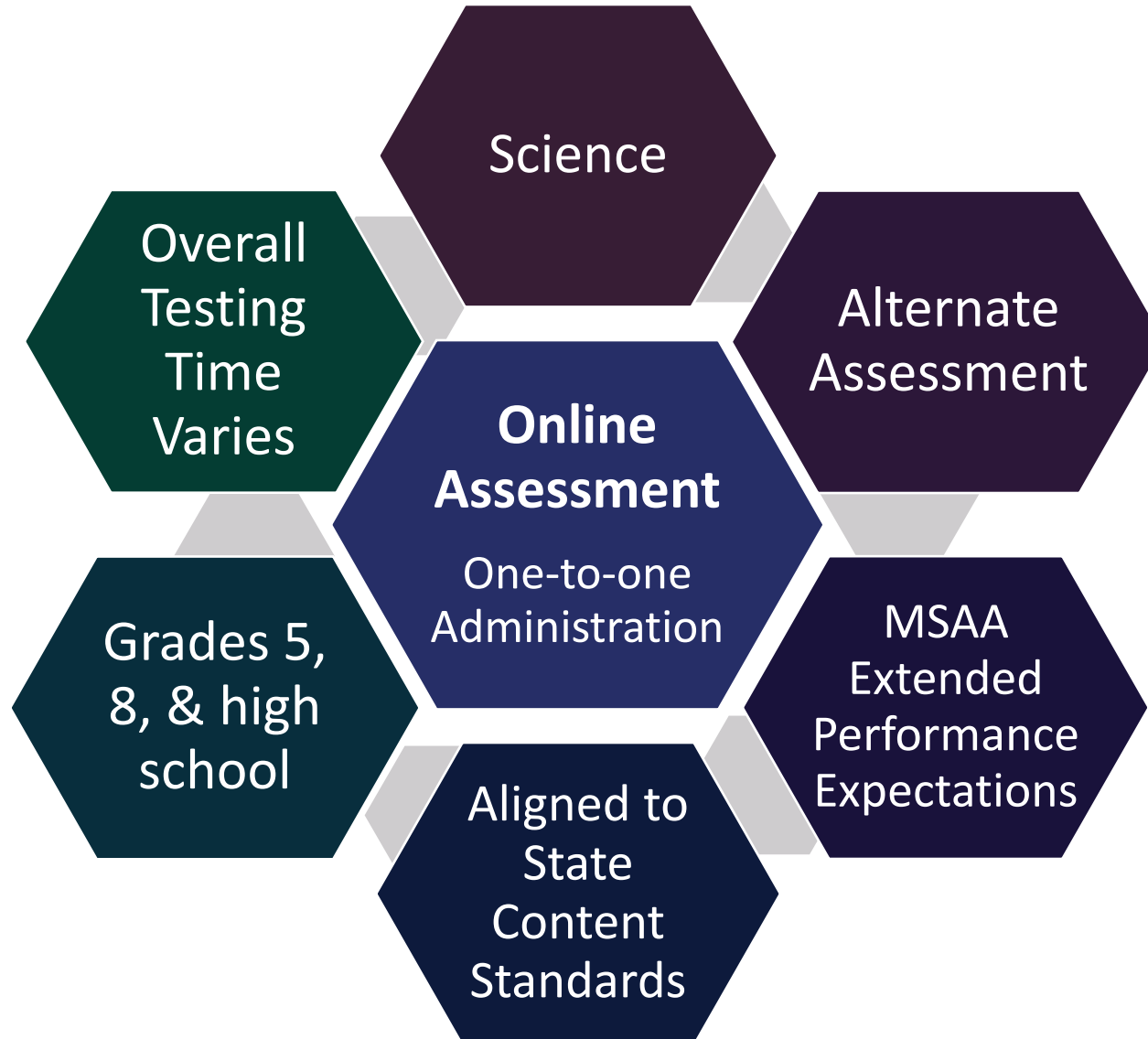
# Learner characteristics

Participation criteria

Learning characteristics

# MSAA Science purpose

| | |
|---|---|
| **Participation** | Grade-level content aligned with the Content Standards |
| | Assesses students in grades 5, 8, and high school |
| **Achievement** | Higher academic outcomes |
| | Prepared for post-secondary options |
| **Accountability** | Every Student Succeeds Act (ESSA) |
| | Individuals with Disabilities Education Act (IDEA) |

# MSAA Development & Administration Process
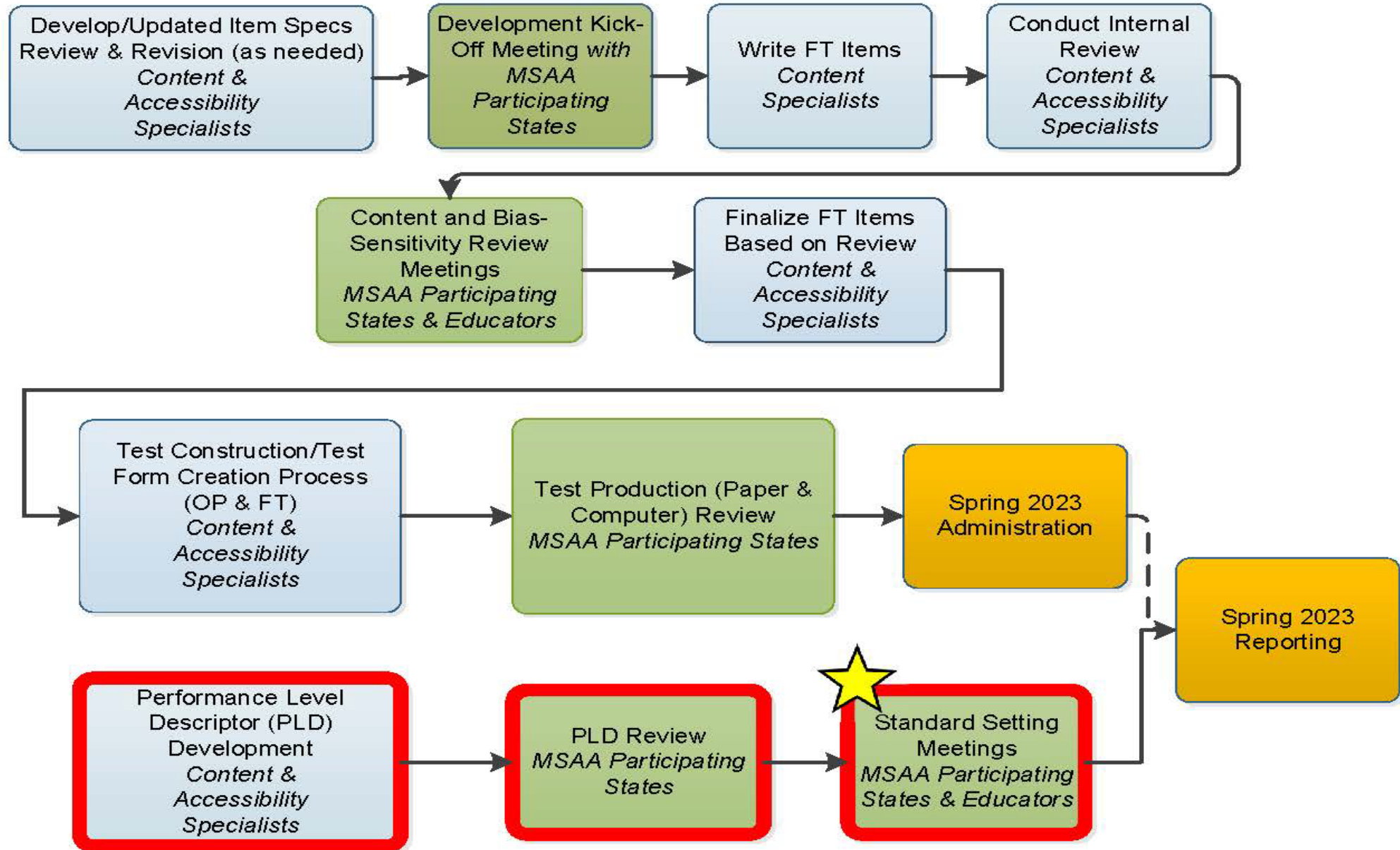
# Postsecondary Outcomes for Students: WIOA expectations

Rachel Quenemoen,
MSAA Technical Advisory Committee
Member Think College National Coordinating Center Accreditation Workgroup,
2016-2021, Chair of the Student Assessment and Learning Outcomes Committee

# What should we expect?

- Requirements for state assessments under the Every Student Succeeds Act (ESSA, 2015) related to achievement standards/expectations

- Overview of two big ideas and two target outcomes

- Existence proofs? Success stories.

- Reflection: What should we expect?

- Your task

# Background information

Quenemoen, R. F., & Thurlow, M. L. (2019). *Students with disabilities in educational policy, practice, and professional judgment: What should we expect?* (NCEO Report 413). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Think College National Coordinating Center Accreditation Workgroup (2021). *Report on Model Accreditation Standards for Higher Education Programs for Students with Intellectual Disability: Progress on the Path to Education, Employment, and Community Living*. Boston, MA: University of Massachusetts Boston, Institute for Community Inclusion.

Thurlow, M. L., Nye-Lengerman, K., and Lazarus, S. S. (2019, January). *Suggestions for aligning alternate achievement standards with WIOA*. (NCEO Brief #16). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

# ESSA Assessment Regulations*

With respect to alternate assessments for students with the most significant cognitive disabilities, measure student performance based on alternate academic achievement standards defined by the State consistent with section 1111(b)(1)(E) of the Act that reflect professional judgment as to the highest possible standards achievable by such students to ensure that a **student who meets the alternate academic achievement standards is on track to pursue postsecondary education or competitive integrated employment**, consistent with the purposes of the Rehabilitation Act of 1973, as amended by the Workforce Innovation and Opportunity Act, as in effect on July 22, 2014.

 * Section 200.2(b)(3)(ii)(B)(2), State Responsibilities for Assessment

# Two big ideas and two big goals

- States are required to coordinate educational services with the requirements of WIOA.

- States must demonstrate for the federal assessment peer review that students with disabilities who meet the state's alternate academic achievement standards are on track to:
  - **pursue postsecondary education or**
  - **competitive integrated employment.**

# Report to Congress on higher education participation, May 2021



www.thinkcollege.net

Report on Model Accreditation Standards for Higher Education Programs for Students with Intellectual Disability: Progress on the Path to Education, Employment, and Community Living

REPORT TO:
United States Secretary of Education
United States Senate Committee on Health, Education, Labor and Pensions
United States House of Representatives Committee on Education and Labor
National Advisory Committee on Institutional Quality and Integrity

The National Coordinating Center
Accreditation Workgroup
May 2021

ICI

ThinkCollege
NATIONAL COORDINATING CENTER

# Postsecondary education

- More than 250 college & university programs, in 49 states, support students with intellectual disabilities.

- Over 5,500 students with intellectual disabilities are attending college in the United States.

- Those engaging in inclusive higher education saw a 135% increase in employment, compared to peers.



**ThinkCollege**
INSTITUTE FOR COMMUNITY INCLUSION, UMASS BOSTON

The inclusive higher education programs in the U.S.

NUMBER OF PROGRAMS
0   1-3   4-6   7-9   10+
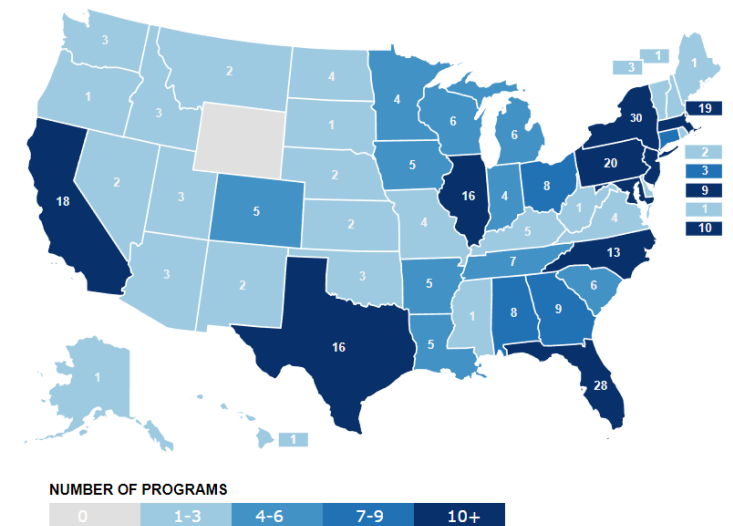
# Success stories

- Robert Lewis, Graduate of Clemson LIFE Program at Clemson University: "What I liked best about college was making new friends, going to football games, managing the Clemson men's basketball team, being in a service fraternity, and working at the Clemson Outdoor Lab and at McAlister's deli. College has changed my life because I learned how to live independently in an apartment with friends. Now that I graduated, I am sharing an apartment with my best friends from college, and I have kept my job as a cashier at McAlister's deli."

- Do you know of students who have participated in Alternate Assessment who are now in postsecondary education settings or have completed them? What knowledge and skills, including science content knowledge and skills, do you think helped them achieve this?

# What does competitive integrated employment mean? How do we prepare students?

- Competitive, integrated employment: ...full-time or part time work at minimum wage or higher, with wages and benefits similar to those without disabilities performing the same work, and fully integrated with co-workers without disabilities. (34 CFR §§361.5(c)(9)(ii) and 361.5(c) (32)(ii))

- Provisions for supported employment up to age 25, but advocates are seeking clarification for future reauthorization or regulations

- Do you know of students who have participated in Alternate Assessment who are now in competitive, integrated employment or in supported employment? What knowledge and skills, including science content knowledge and skills, do you think helped them achieve this employment?

# Reflection

- How do we prepare ALL students K-12 with essential skills and knowledge for postsecondary opportunities and post school employment?

- How are the skills and knowledge for students who participate in Alternate Assessment the same or different? What about specifically in science?

- What should we expect?

# YOUR TASK THIS WEEK:

- DEFINE STUDENT PERFORMANCE EXPECTATIONS BASED ON ALTERNATE ACADEMIC ACHIEVEMENT STANDARDS THAT … REFLECT PROFESSIONAL JUDGMENT AS TO **THE HIGHEST POSSIBLE STANDARDS ACHIEVABLE BY SUCH STUDENTS**…

# Thank you for your commitment to our children and students!

# The three dimensions of science learning

1. Science and Engineering Practices (SEPs)
   - What students are expected to do

2. Disciplinary Core Ideas (DCIs)
   - What students are expected to know

3. Crosscutting Concepts (CCCs)
   - How students think and connect ideas

# Integrating dimensions

- The standards, called Performance Expectations, integrate all three dimensions (SEPs, DCIs, and CCCs) in each standard.
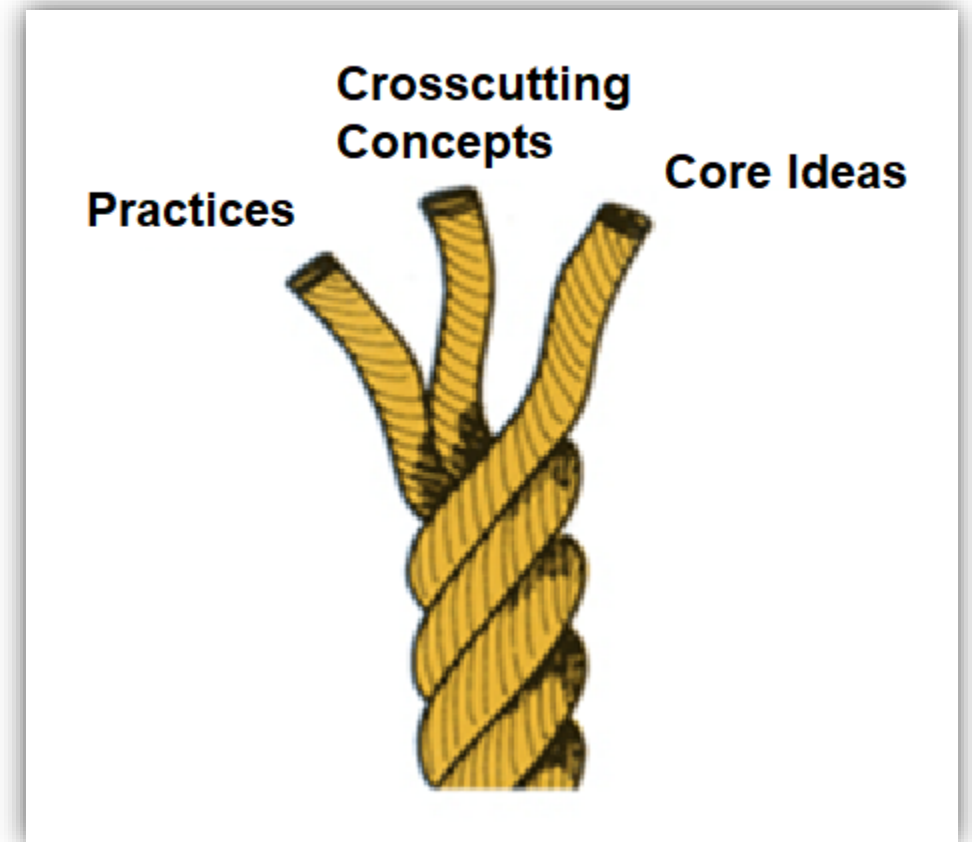
# Extended Performance Expectation (EPE)

- Aligned with *A Framework for K-12 Science Education* and State Content Standards

- Consists of three leveled statements describing three different access points

- Core Ideas represent all content domains.
  - Physical, Life, and Earth and Space, plus at least one Engineering-based PE for each grade band.

- Commitment to maintain multi-dimensional standards expectation

- Each grade level (5, 8, high school) contains 12 priority EPEs, each with its three leveled access points.

# Example: 3-ESS2-1 Item Set

**Standards**

| NGSS Performance Expectation 3-ESS2-1 | | |
|---|---|---|
| **3-ESS2-1** Represent data in tables and graphical displays to describe typical weather conditions expected during a particular season. [Clarification Statement: Examples of data could include average temperature, precipitation, and wind direction.] [Assessment Boundary: Assessment of graphical displays is limited to pictographs and bar graphs. Assessment does not include climate change.] | | |
| **Science and Engineering Practices (SEP)** | **Analyzing and Interpreting Data**<br>• Represent data in tables and various graphical displays (bar graphs and pictographs) to reveal patterns that indicate relationships. | |
| **Disciplinary Core Ideas (DCI)** | **ESS2.D: Weather and Climate**<br>• Scientists record patterns of the weather across different times and areas so that they can make predictions about what kind of weather might happen next. | |
| **Crosscutting Concepts (CCC)** | **Patterns**<br>• Patterns of change can be used to make predictions. | |

| Extended Performance Expectation 3-ESS2-1 | | |
|---|---|---|
| *Level 1* | *Level 2* | *Level 3* |
| *Less Complex* ⬅⋯ ⬅⋯ ⬅⋯ ⬅⋯ ⬅⋯⋯ ⋯⋯➡ ⋯➡ ⋯➡ ⋯➡ | | *More Complex* |
| **3-ESS2-1.1**<br>Use observations to describe weather conditions. | **3-ESS2-1.2**<br>Use tables or graphical displays of data to describe patterns of typical weather conditions in a particular season. | **3-ESS2-1.3**<br>Use tables and/or graphical displays of data to predict patterns of typical weather conditions for a particular season. |
| **Science and Engineering Practices (SEP)** | **Analyzing and Interpreting Data**<br>• Represent data in tables and various graphical displays (bar graphs and pictographs) to reveal patterns that indicate relationships.<br>**Supporting:**<br>**Planning and Carrying Out Investigations** | |
| **Disciplinary Core Ideas (DCI)** | **ESS2.D: Weather and Climate**<br>• Scientists record patterns of the weather across different times and areas so that they can make predictions about what kind of weather might happen next. | |
| **Crosscutting Concepts (CCC)** | **Patterns**<br>• Patterns of change can be used to make predictions. | |

**Extended Performance Expectation with the three access points and their codes, increasing in complexity from left to right**

**SEPs, DCIs, and CCCs that comprise the Extended Performance Expectations (Target and Supporting, as appropriate).**

# Example: Low-level item (3-ESS2-1 set)

Sofia observes a tree bending as air blows on it outside.



| **Low Level** | **3-ESS2-1.1**: Use observations to describe weather conditions. |
|---|---|

Which word can Sofia use to describe the weather?

○ foggy      ○ windy      ○ sleepy

# Example: Medium-level item (3-ESS2-1 set)

This data table shows winter weather conditions in New York in 2015.

**Winter Weather Conditions in New York**

| Weather Condition | Data |
|---|---|
| Average snowfall | 43 cm |
| Average temperature | 31°F |

| | |
|---|---|
| **Medium Level** | **3-ESS2-1.2**: Use tables or graphical displays of data to describe patterns of typical weather conditions in a particular season. |

Based on the data table, which sentence **best** describes winter in New York?

○ It has few windy days.

○ It is warm with lots of rain.

○ It is cold with lots of snow.

# Example: High-level item (3-ESS2-1 set)

This data table shows the high temperatures for a town in Alaska in the spring and fall. Data for June, July, and August are missing.

**High Temperatures for a Town in Alaska**

| Month | High Temperature (°F) |
|---|---|
| April | 46 |
| May | 52 |
| June | ? |
| July | ? |
| August | ? |
| September | 50 |
| October | 47 |
| November | 38 |
| December | 35 |

| | |
|---|---|
| **High Level** | **3-ESS2-1.3**: Use tables and/or graphical displays of data to predict patterns of typical weather conditions for a particular season. |

Based on the data table, which range of high temperatures is expected for the summer?
○ between 20° and 29°F      ○ between 40° and 49°F      ○ between 60° and 69°F

# Overview - Item Types

- Selected-response (multiple choice) items written at three levels with increasing cognitive demand
- Presented in a standardized format
  - Item stimulus
    - Phenomenon or context presented as the SAY section within the Test Administrator Booklet (aka, DTA)
  - Item prompt or question
    - Presented as the ASK section within the Test Administrator Booklet (aka, DTA)
  - Response options

# Example Science Item

# Example Science DTA

**Text to be read aloud is in boldfaced black (if TA is reading the assessment).**

**Any actions the TA should perform are in italics.**

| | Teacher Script |
|---|---|
| SAY | **This data table shows winter weather conditions in New York in 2015.**<br><br>*Indicate and read the data table to the student.*<br><br>**The data table is titled "Winter Weather Conditions in New York." It lists data for weather conditions. Average snowfall is forty-three centimeters. Average temperature is thirty-one degrees Fahrenheit.** |
| ASK | **Based on the data table, which sentence <u>best</u> describes winter in New York?**<br><br>*Indicate and read each response option to the student.*<br><br>**It has few windy days.**<br>**It is warm with lots of rain.**<br>**It is cold with lots of snow.** |

| | Student Response |
|---|---|
| RECORD | *Fill in the circle for the student's response.*<br><br>○ A. It has few windy days.<br>○ B. It is warm with lots of rain.<br>○ C. ***It is cold with lots of snow.***<br>○ D. No Response |

# Overview - Item Set Structure

- Items administered in sets of three
  - Thirteen operational item sets (thirty-nine items total)
- Each item set has three levels of difficulty.
  - Aligned to the three levels of the EPE

Low-level item → Medium-level item → High-level item

- Items in a set are independent of each other.
  - Do not share a stimulus
  - Exception: specialized bundle of item sets referred to as clusters

# Overview - Item Set Structure
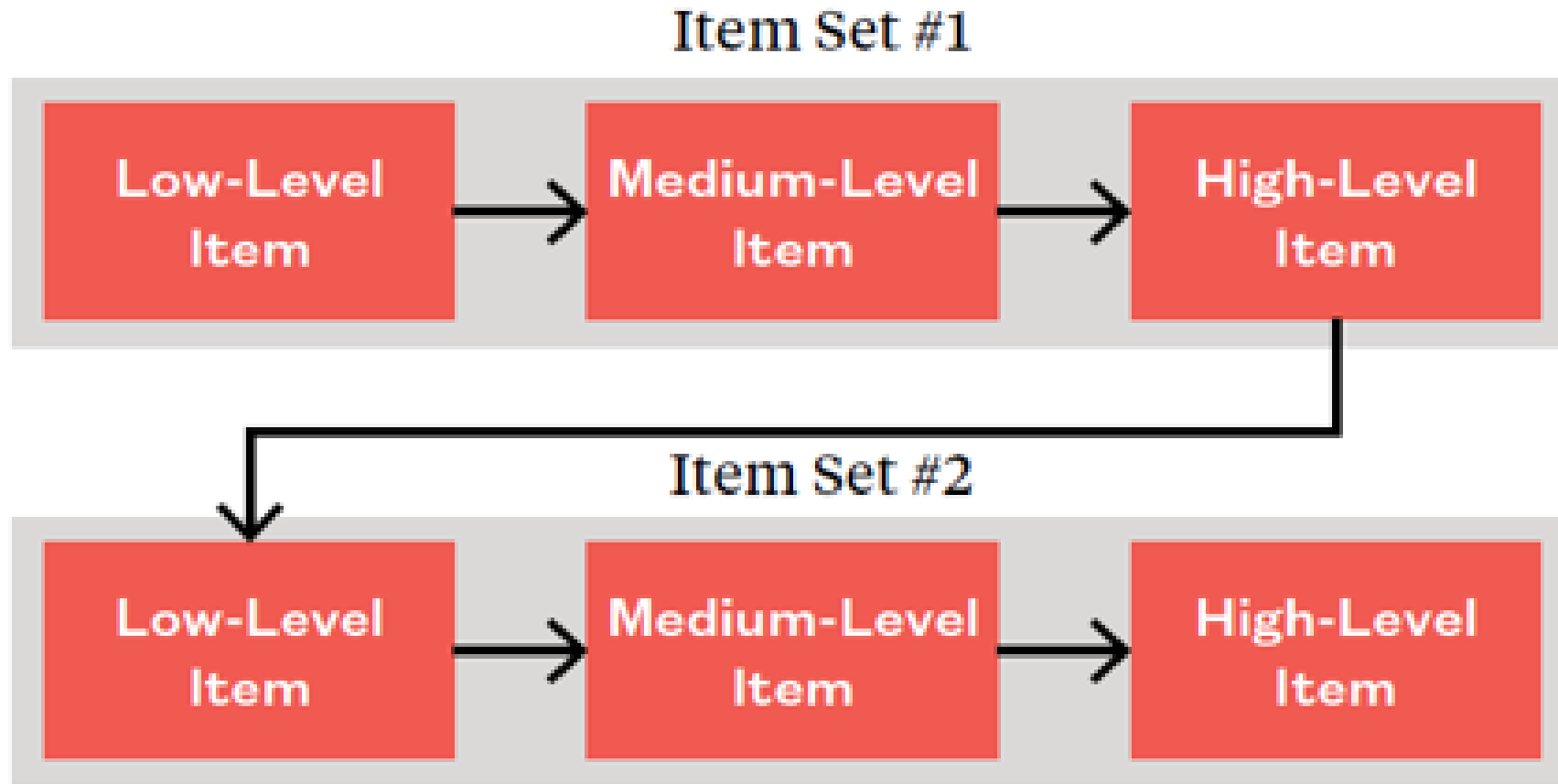
# Overview - Clusters

- Combination of two item sets (six items total)
  - Written to two related state science standards
    - First set aligns to the EPEs derived from the first science standard listed
    - Second set aligns to the EPEs derived from the second science standard listed
- Begin with stimulus containing text and graphics sufficient to support the six items
  - Items within the cluster are independent.

Assessment Features and Accommodations

# Assessment Features
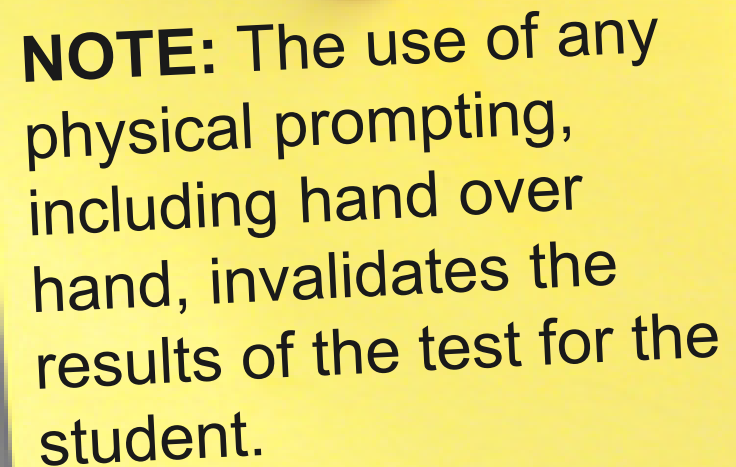
- Computer-Based Assessment Features
    - Answer masking
    - Audio player
    - Alternate color themes
    - Zoom
    - Increase volume
    - Line reader tool
    - Read aloud/reread item directions, response options, passages

# Assessment Features (cont.)

- Assessment Features Provided by Test Administrator
  - Reading aloud as many times as necessary
  - Alternative text
  - Answer masking
  - Magnification
  - Use of manipulatives
  - Line reader
  - Object replacement
  - Tactile graphics or symbols
  - Transcription

# Accommodations

- Assistive Technology (AT) for viewing, responding, or interacting with test items
- Paper version/large print (downloaded from platform)
- Scribe
- Sign language

**NOTE:** The use of any physical prompting, including hand over hand, invalidates the results of the test for the student.

# Standard Setting Overview

# Purpose of Standard Setting

- Allows MSAA Science partners to have educator expertise inform performance standards for the MSAA Science grade 5, grade 8, and high school assessments:
  - Teachers, administrators, and higher-ed and vocational specialists
- Opportunity for educator input on cut scores used to define performance levels
- To ensure recommendations are consistent with expectations stated in the Performance Level Descriptors

# What are performance levels?

- Performance Levels reflect the specific knowledge and skills that a student should be able to demonstrate based on their performance on the test.

| Performance Level 1 | Performance Level 2 | Performance Level 3 | Performance Level 4 |
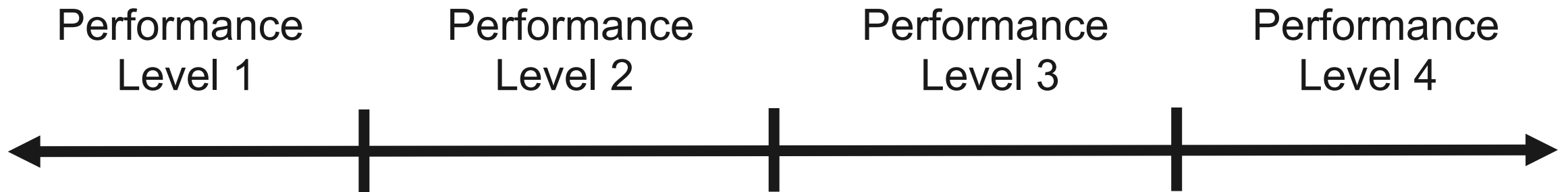
# What are Cut Scores?

- A cut score is the minimum test score a student must earn to be considered at a specific performance level.

- Three cut scores result in four levels of performance.

Cut Score 1        Cut Score 2        Cut Score 3

Level 1        Level 2        Level 3        Level 4
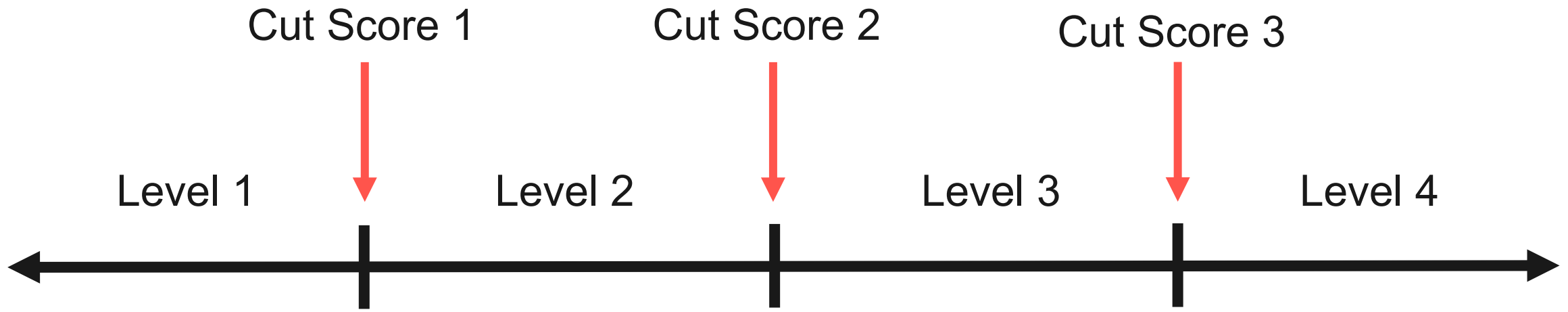
# How do we consider cut scores?

- We use content-based judgments.
  - Content links assessment items, PLDs, and Expected Performance Standards.
  - Content lets you consider MSAA's objective for students.

Performance Level Descriptors

# Performance Level Descriptors (PLDs)

- PLD describes the specific knowledge and skills that a student at a given performance level should be able to demonstrate.

- **Policy PLDs**
  - High-level descriptors that define the knowledge and skill level expectations.
- **Range PLDs**
  - Content-specific descriptors that link back to the EPEs.
- **Borderline PLDs**
  - Specifically define what it takes for a student to attain each performance level, just barely.

**Policy PLDs**

**Range PLDs**

**Borderline PLDs**
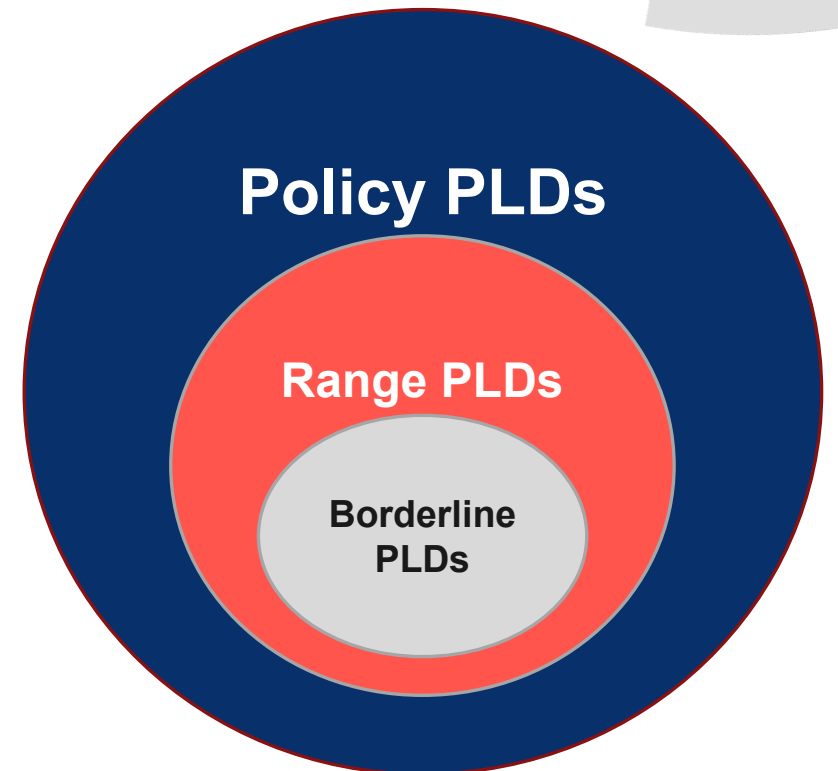
# Performance Level Descriptor Development and Review Process

- What story do we want to tell about students?
- Students are able to use the majority[1] of the disciplinary core ideas, practices, and crosscutting concepts to address moderately complex science phenomena and problems, some concrete and some abstract

[1] majority intended as at least half or more of the science content

Initial PLD drafts developed by Cognia Content & Accessibility Specialists

PLD drafts reviewed by various stakeholders

PLDs updated based on stakeholder feedback

PLDs you will be referencing are final

# Policy PLDs

- How do the levels qualitatively differ from one another?

| Level* | Description |
|--------|-------------|
| Level 1 | Students **attempt** to perform basic tasks but will **require additional support** in order to demonstrate knowledge and skills of the K-12 science framework Extended Performance Expectations by using disciplinary core ideas, practices, and/or crosscutting concepts to address more basic and concrete science phenomena and problems in Level 1. |
| Level 2 | Students can be expected to **demonstrate** developing knowledge and skills in **some** disciplinary core ideas together with some aspects of the practices and crosscutting concepts from the K-12 science framework Extended Performance Expectations to address primarily basic and concrete science phenomena and problems at Level 2. |
| Level 3 | Students can be expected to **demonstrate** knowledge and skills in **the majority of** disciplinary core ideas, practices, and crosscutting concepts from the K-12 science framework Extended Performance Expectations to address moderately complex science phenomena and problems, some concrete and some abstract at Level 3. |
| Level 4 | Students can be expected to **demonstrate understanding** and skills in the disciplinary core ideas, practices, and crosscutting concepts from the K-12 science framework Extended Performance Expectations **in more sophisticated ways** than students in Level 3 to address science phenomena and problems that are complex, more abstract, and/or multi-factorial. Students are expected to describe, explain, and/or respond to phenomena and problems using reasonably complex evidence, analysis, and inference at Level 4. |

*Each level builds on one another, students are expected to be able to demonstrate skills in a lower level

# Range PLDs

- There is a range of performance within each performance level.

- Range PLDs describe the expected knowledge and skills of students achieving in each performance level and correspond to the policy PLD for that level.

- You will have access to, and become more familiar with, the Range PLDs in the **breakout sessions**.

# How the range PLDs are structured

- PLDs are arranged by the EPE content categories: PS-3, LS-2, etc.

| |
|---|
| **PS-3 Energy**<br>•HS-PS3-2<br>**SEP**<br>•Developing and Using Models<br>•Asking Questions and Defining Problems (Supporting)<br>**CCC**<br>•Energy and Matter |
| **LS-2 Ecosystems: Interactions, Energy, and Dynamics**<br>•HS-LS2-2<br>**SEP**<br>•Using Mathematics and Computational Thinking<br>•Obtaining, Evaluating, and Communicating Information (Supporting)<br>**CCC**<br>•Scale, Proportion, and Quantity<br>•Cause and Effect (Supporting) |

# How the range PLDs are structured

- Each PLD descriptor is derived from the EPE language and includes words that refer to the SEPs, DCIs, and CCCs associated with each EPE
  - At Level 1: wording intentionally goes primarily to the EPE DCI, which may overlap with the CCC depending on the EPE
  - At Levels 2 and 3, wording may include language from supporting SEPs or CCCs depending on the EPE
  - At Level 4, wording intentionally goes to the EPE target DCI, SEP, and CCC

| Standard | Level 4 |
|---|---|
| **PS-3 Energy**<br>•HS-PS3-2<br>**SEP**<br>•Developing and Using Models<br>•Asking Questions and Defining Problems(Supporting)<br>**CCC**<br>•Energy and Matter | Show how energy is conserved at the macroscopic or particle level when kinetic and potential energy change in a system by developing or using a model. |

# Borderline PLDs

- Borderline PLDs refer to the minimal competency point separating any two performance levels.

- Example: for Level 3 Borderline PLD
    - All students performing **below the borderline** would be classified as **Level 2.**
    - All students performing **at or above the borderline** would be classified as **Level 3.**
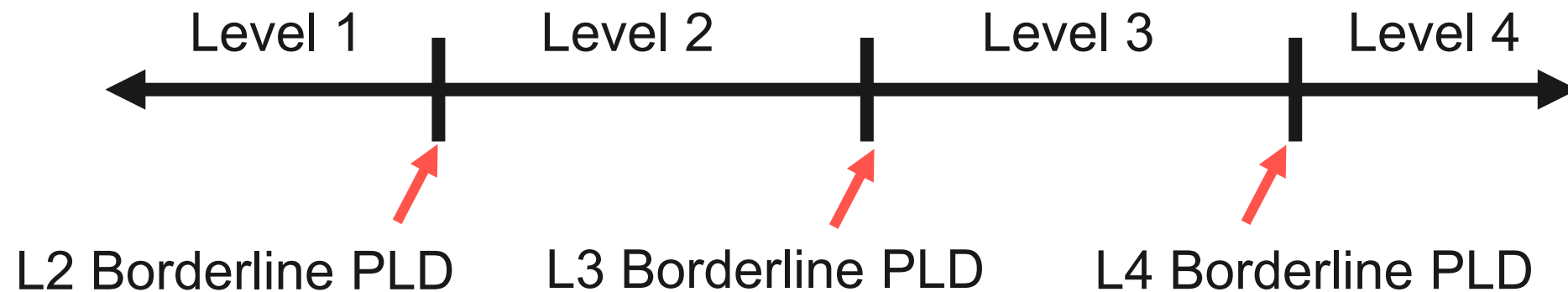


Level 1     Level 2     Level 3     Level 4

L2 Borderline PLD     L3 Borderline PLD     L4 Borderline PLD

# Key Concepts and Procedures Overview

# Item-Descriptor (ID) Matching Method

Item-centered Method

Content-based Judgment

Iterative Process

# Ordered Item Booklet* (OIB)

- One item per page
- Easiest item first
- Items ascend by difficulty
- Hardest item last

*The order of the OIB items is based on their empirical difficulties and not the order in which they appear for students during the test.

**Most Difficult Item**

Item 37

Item ...

Item ...

Item ...

Item 12

Item 11

Item ...

Item 5

Item 4

Item 3

Item 2

Item 1

**Least Difficult Item**

Empirical difficulty based on student data
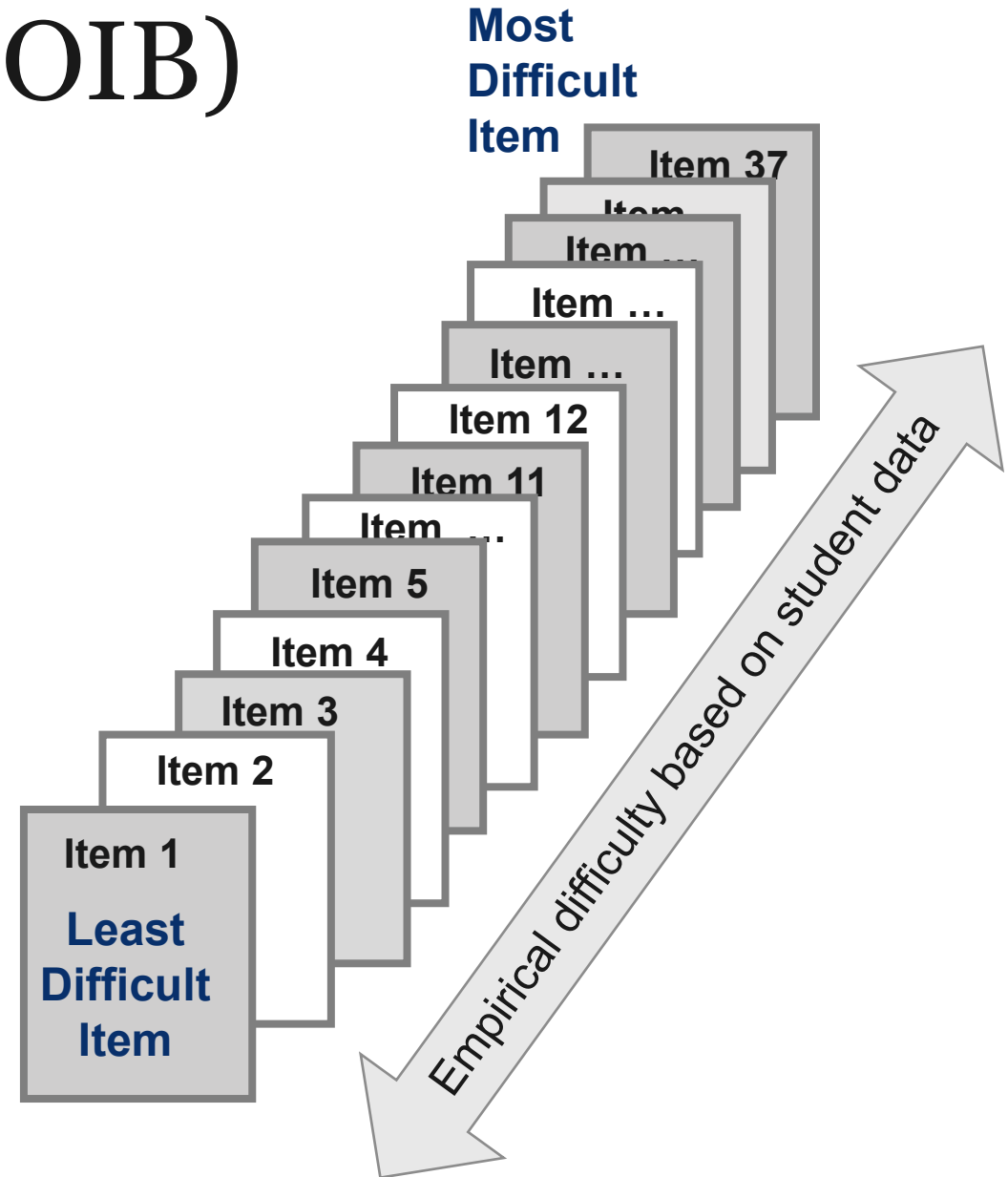
# Overview of ID-Matching Method

↻ Panelists review each item in the OIB.

- Identify the knowledge, skills, and abilities (KSAs) required to answer the item correctly.

↻ Make the following judgment for each item:

- Match the knowledge, skills, and abilities (KSAs) required by the item with the EPEs described in either the Level 2, Level 3, or Level 4 performance level descriptor (PLD).

↻ Judgements are made independently.

# Overview of ID-Matching Method (Cont.)

⟳ Panelists complete 3 rounds of the ID-Matching judgments.

- Round 1: Identify KSAs and align items to PLDs, discuss items.
- Round 2: Introduce benchmarks, align items to PLDs, discuss items.
- Round 3: Align items to PLDs and write group-level content-based rationales if necessary.

# Content-based Benchmarks – Overview

- Benchmarks based on Cognia content team judgements
  - Benchmark regions for Level 3 & 4 will be presented to you at the beginning of Round 2.
  - The benchmark regions represent a transition area between two performance levels.
- Benchmarks serve as a guide.
- You may consider the benchmark as you engage in Round 2 judgment.

# Workshop agenda – General Overview

| Tuesday 7/18 | Wednesday 7/19 | Thursday 7/20 |
|---|---|---|
| Welcome & Overview | Debrief Day 1 | Debrief Day 2 |
| Review Assessments & PLDs | Complete Round 1 | Prepare for Round 3 |
| Training & Practice | Prepare for Round 2 | Complete Round 3 |
| Round 1 Judgements | Complete Round 2 | Vertical Articulation |

# Standard Setting Panels

| | | |
|---|---|---|
| **Grade 5** | Facilitator: Karen Whisler | |
| | Room: Melrose | |
| **Grade 8** | Facilitator: Jill Stepanek | |
| | Room: Montrose | |
| **High School** | Facilitator: David Harrison | |
| | Room: Salem | |

cognia®

# MSAA Science Standard Setting – Breakout Session

Grade 8

Facilitator: Jill Stepanek
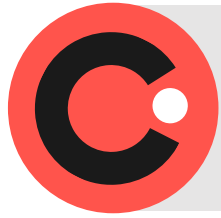
# Breakout Session – Agenda (Day 1)

- Welcome and Introductions – Panelists
- Experience the Test
- Lunch
- Access to the Cognia Standard Setting Toolkit
- Familiarization with PLDs and Content Standards
- Training on the Item-Descriptor (ID) Matching Method
- Modeling and Practice
- Begin Round 1

# Welcome & Introductions - Panelists

- Introduce yourself
  - Your name, school district, what you teach
  - A little selected background information
- Show of hands
  - Who's been involved in standard setting before?
  - Which method(s)?

# Meeting Norms

- All conversations are confidential
  - What happens here stays here.
  - When you return to your state, please do talk about the process we undertake, but do not disclose the specifics.
- Please DO NOT
  - Discuss item specifics outside of the panels or after standard setting.
  - Use personal devices in the room - you may step out at any time if needed.
  - Use the Chromebooks for anything other than the standard setting activities

# Experience the Test

- You will experience the MSAA Science test in a format that is similar to student experience.
  - Briefly examine the tests in the test platform.
  - Try not to linger on any one item; this session is scheduled for a duration of **45 mins**.

- Purpose:
  - Get familiar with the items as they appeared to students.
  - Science items sets appear together in the testing platform but do not appear together in the OIB.
  - You will see most of the items from the testing platform in the OIB.

Lunch
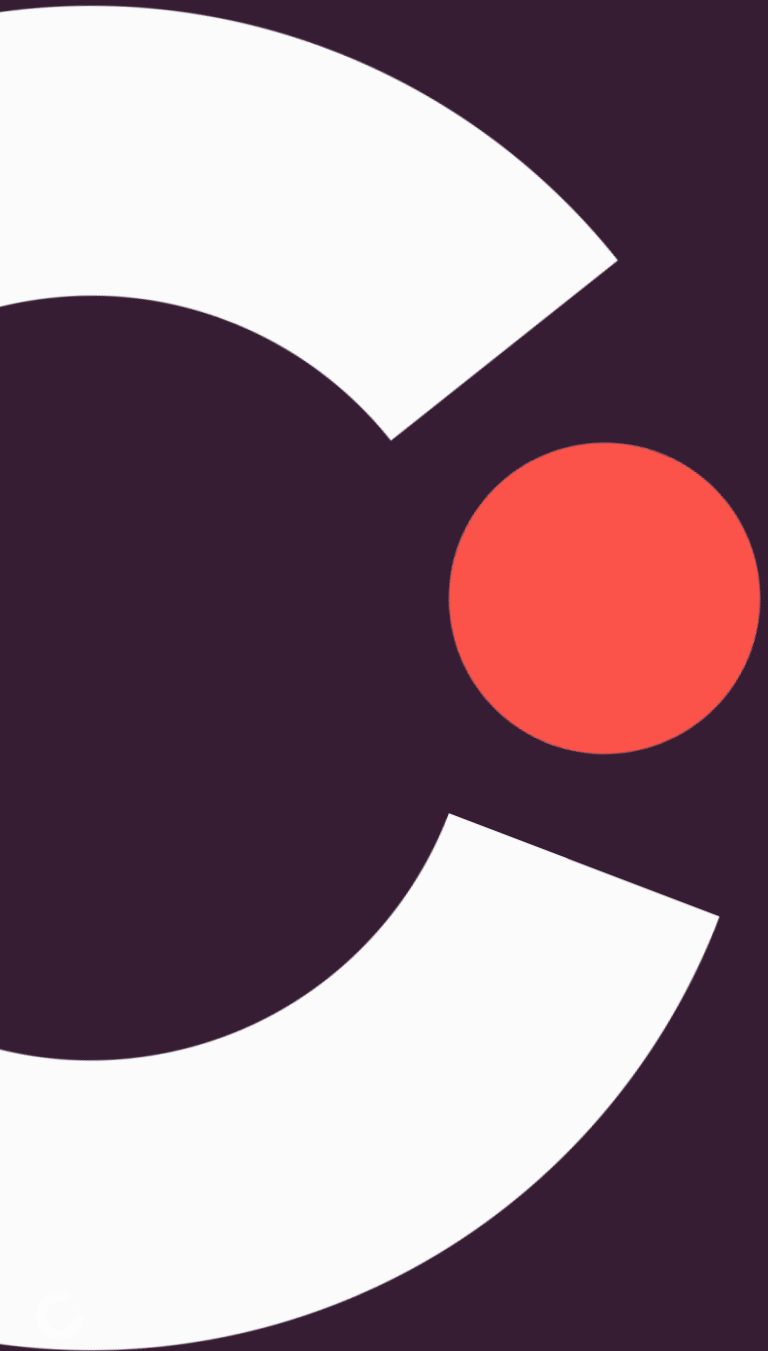
Access to Cognia Standard Setting Toolkit

# Cognia Standard Setting Toolkit

- Use your email and initial password to log on to the platform

- Email: Your own email that was used to register for this meeting

- Password: Everyone has the same initial password
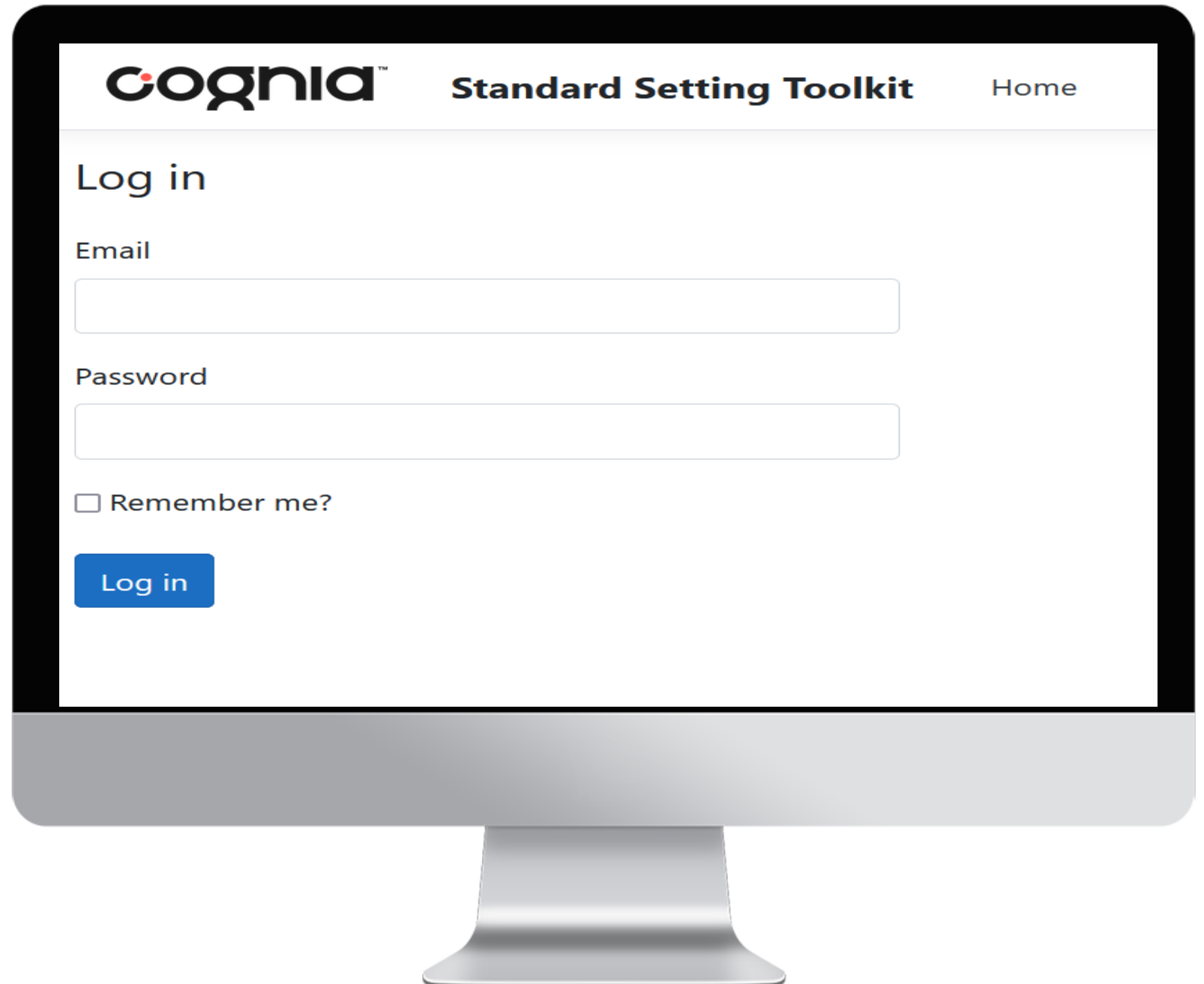
# Cognia Toolkit: Change your password

- Click on your email in the top right corner

- This will bring you to a profile page

- Click on the "Password" tab shown to the left
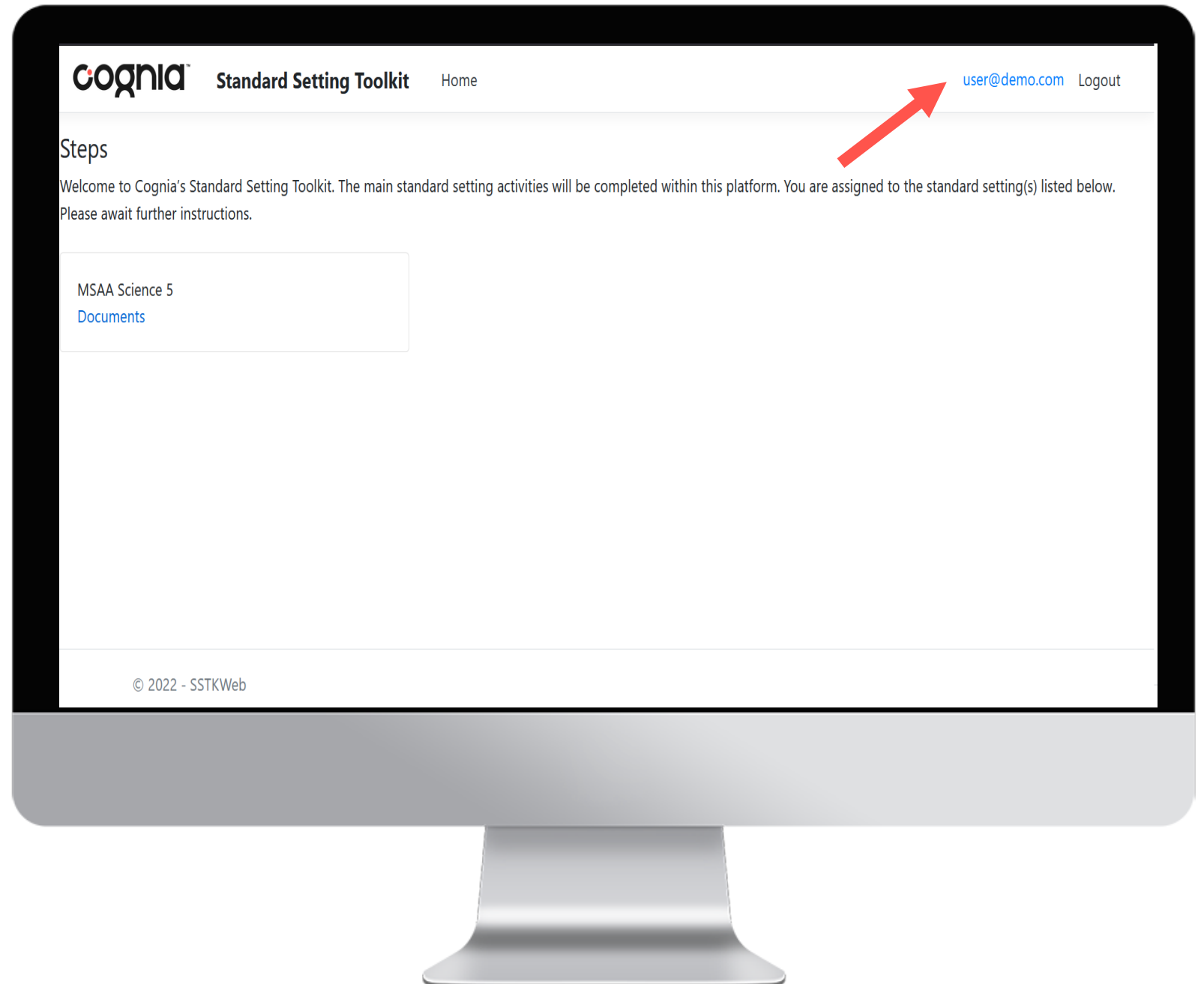
# Cognia toolkit: Change your password

- Click on the "Password" tab shown to the left

- Enter the initial password

- Enter New Password:
  - Upper Case Letter
  - Lower Case Letter
  - Number
  - At least 6 Characters

- Click "Update Password"

- Log out and Log back in with your new password.

You should now be back on the following Screen:
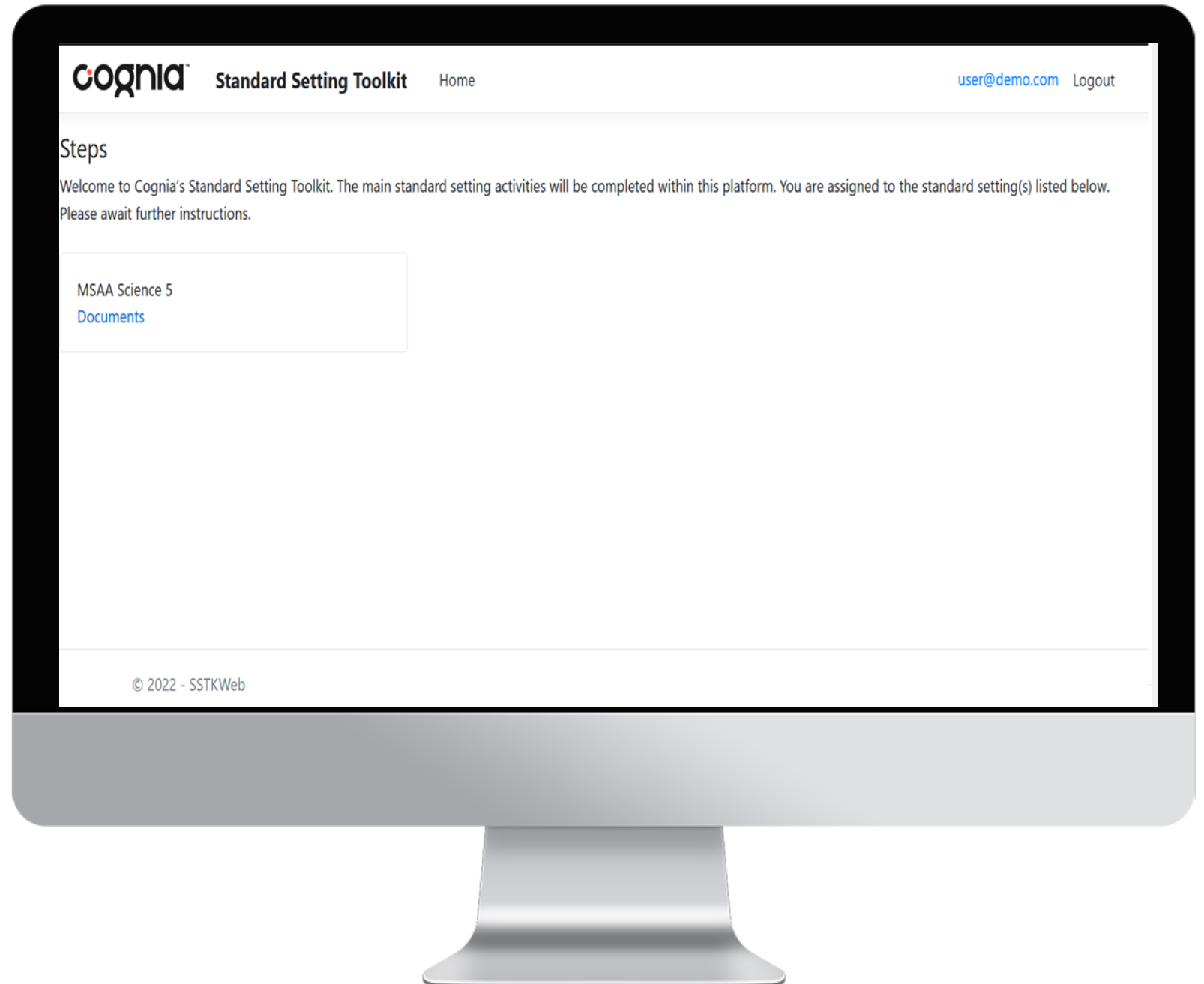
# Familiarization with MSAA Science

Content Standards (aka, EPEs) and Performance Level Descriptors (PLDs)

# Review PLDs

- Brief Background on PLD development
- Obtain an understanding of PLDs in relation to Extended Performance Expectations (EPEs).
  - This activity is critical because you will make judgements based on your understanding of PLDs.
  - The PLD documents will be used throughout the workshop to make item-PLD alignment.
- Individually review PLDs within the Cognia Standard Setting Toolkit.

# Where to find the PLDs and standards

- In the Toolkit: Blue "documents" link

- Document links for the Standards, Range, and Borderline PLDs appear

- Paper copies of the PLDs also distributed for easy reference
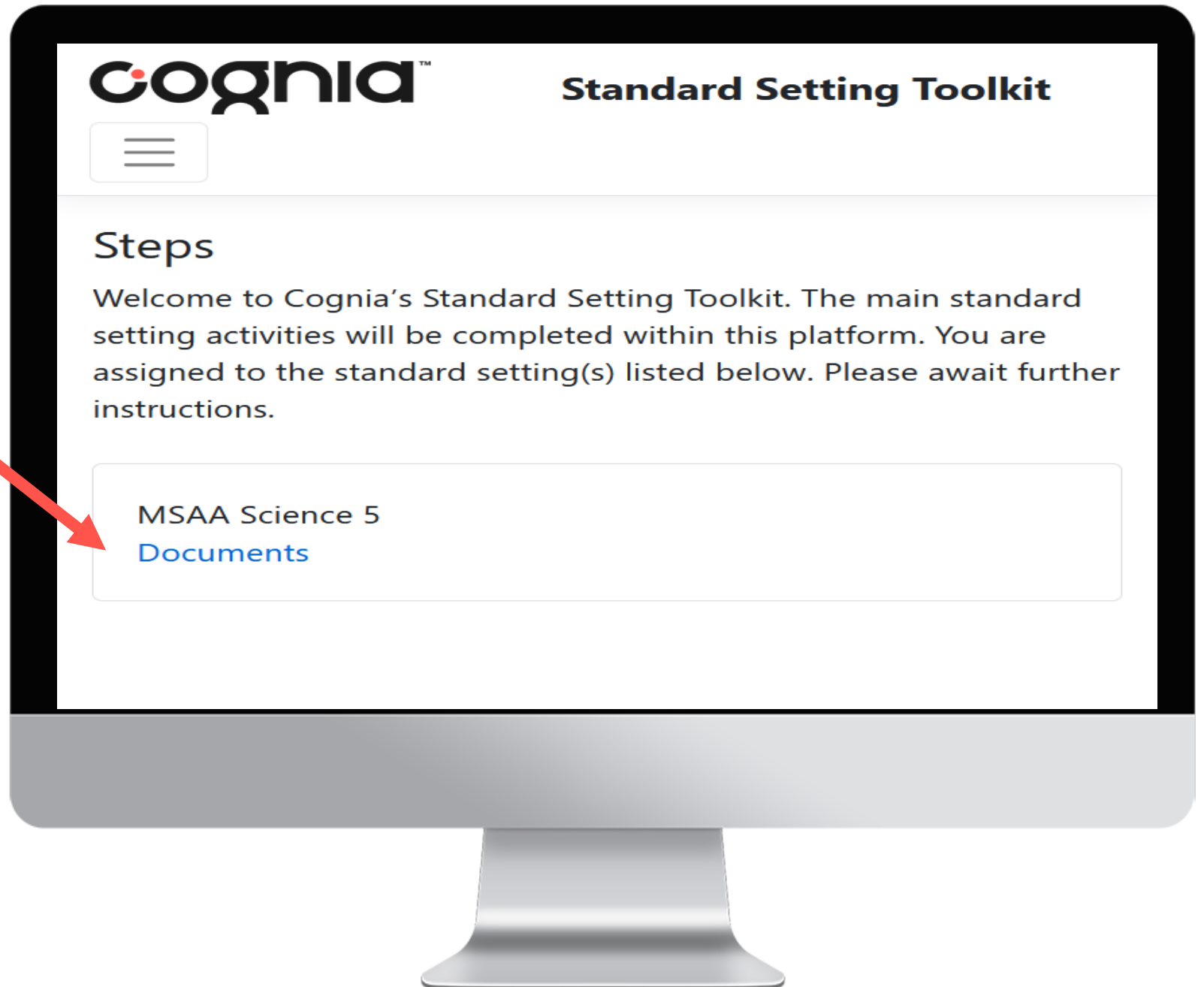
# Discuss range PLDs

- Collegial discussion to clarify questions
- Reach common understanding of what it means to be in each performance level.
  - Start with Level 2 PLD, then Level 3 and Level 4.
  - Focus on how the levels differ in content and dimensions, cognitive complexity, type of phenomenon and accessibility (e.g., concrete vs. abstract).
  - Discuss Level 1 PLD as an extension of Level 2 PLD.

# Discuss range PLDs – MS example

| Standard | Level 2 |
|---|---|
| **MS-LS1-5**<br><br>**SEP**<br>• Constructing Explanations and Designing Solutions<br>• Analyzing and Interpreting Data (Supporting)<br>• Asking Questions and Defining Problems (Supporting)<br><br>**CCC**<br>• Cause and Effect | **Identify** factors that could be affecting the growth of an organism by asking questions. |
| | **Level 3** |
| | **Determine** whether a particular factor is affecting the growth of organisms by analyzing data. |
| | **Level 4** |
| | **Explain** how the growth of organisms is influenced by various environmental and/or genetic factors by using data. |

# Discuss range PLDs – MS example

| Standard: ESS-1 Earth's Place in the Universe<br>•MS-ESS1-1<br><br>Level 3 PLD | Standard: ESS-2 Earth's Systems<br>•MS-ESS2-4<br><br>Level 4 PLD |
|---|---|
| **Describe** or compare the positions of the Sun, the Moon, and Earth or the amount or path of light in the cyclic patterns of seasons, lunar phases, or eclipses by using a model. | **Describe** how the Sun's energy or the force of gravity moves water through the water cycle by developing a model. |

# Discuss Borderline PLDs

- Borderline PLDs describe the characteristics expected of students who just barely meet each performance level.
  - A borderline determination needs to consider whether a student just beginning to demonstrate the KSA in that level would more often than not be able to demonstrate the skills in the particular item. Consideration would typically take into account the complexity and abstractness of the phenomenon, and the science dimensions.
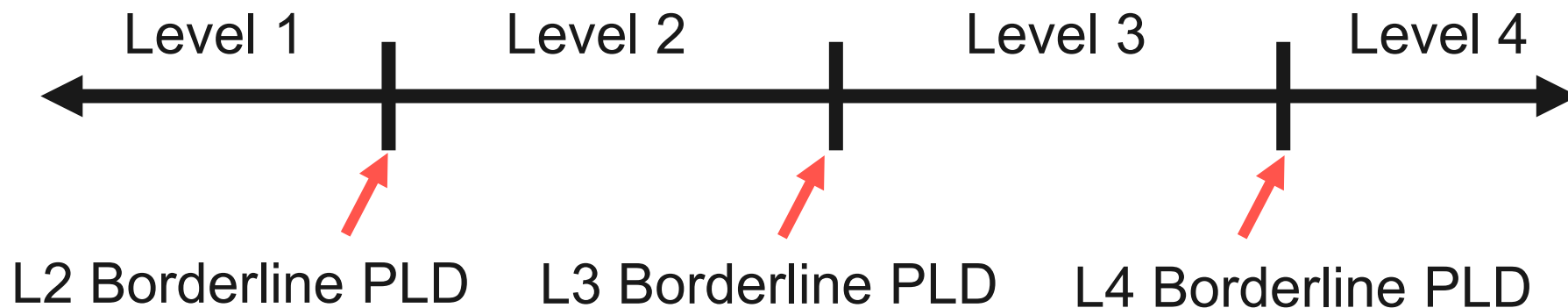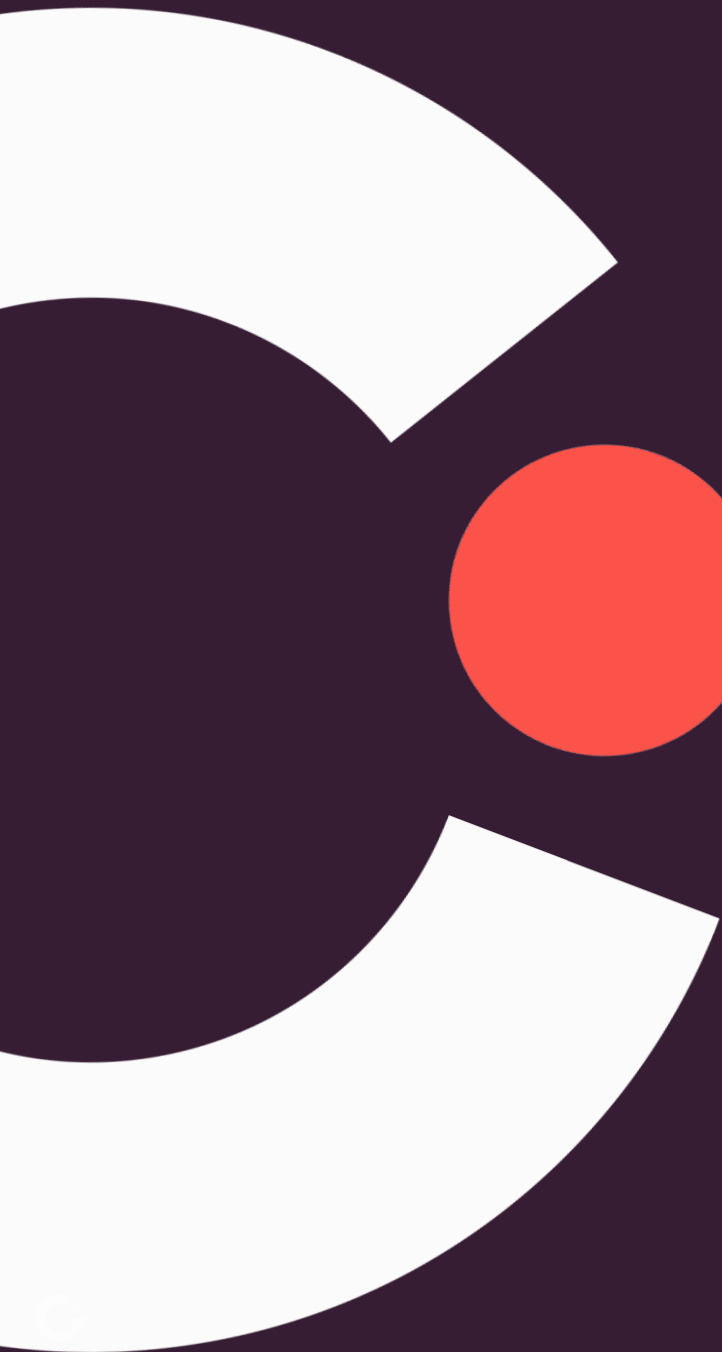
Level 1     Level 2     Level 3     Level 4

L2 Borderline PLD     L3 Borderline PLD     L4 Borderline PLD

Break

# Key Concepts & Process

Ordered Item Booklet (OIB) and Item-Descriptor Matching (ID-Matching)

# Ordered Item Booklet (OIB)

- OIB contains test items ordered by difficulty.

- Each OIB page represents an item.

- The difference in difficulty is not exactly the same between each pair of neighboring items.

- Difficulty is based on data from the AY23 MSAA students who took the test

**Most Difficult Item**

Item 37

Item ...

Item ...

Item ...

Item ...

Item 12

Item 11

Item ...

Item 5

Item 4

Item 3

Item 2

Item 1

**Least Difficult Item**

*Empirical difficulty based on student data*

# OIB in the Standard Setting Toolkit

# ID Matching and Your Judgmental Task

For each item in the OIB:

- Step (a) Review the item and identify KSAs
  - Identify the knowledge, skills, and abilities (KSAs) required to answer the item correctly

What does a student need to know or be able to do to respond to this item?

- Step (b) Make item-PLD alignment judgment
  - Match the KSAs required by the item with the EPEs described in either the Level 2, Level 3, or Level 4 performance level descriptor (PLD).

Which PLD most closely matches the knowledge, skills, and abilities (KSAs) required by the item?

# ID Matching and Your Judgmental Task

As you engage in your judgmental task:

- Use the PLDs to make judgments about items in the OIB.

- Consider the EPEs described in the PLDs.

- Consider the knowledge and skills demand of an item.

- Write brief content-based reasons for your item-PLD matches
  - If an item seems to be aligned between two PLDs, select the PLD that most closely matches the KSAs AND write notes about the item to later inform discussions

- Work independently

- Trust your expertise

# ID Matching Process over 3 rounds:

| | | |
|---|---|---|
| Modeling and Practice | Prepare for Round 1 | Round 1 Judgements |
| R1 Feedback and Discussion | Prepare for Round 2 | Round 2 Judgements |
| R2 Feedback and Discussion | Prepare for Round 3 | Round 3 Judgements |
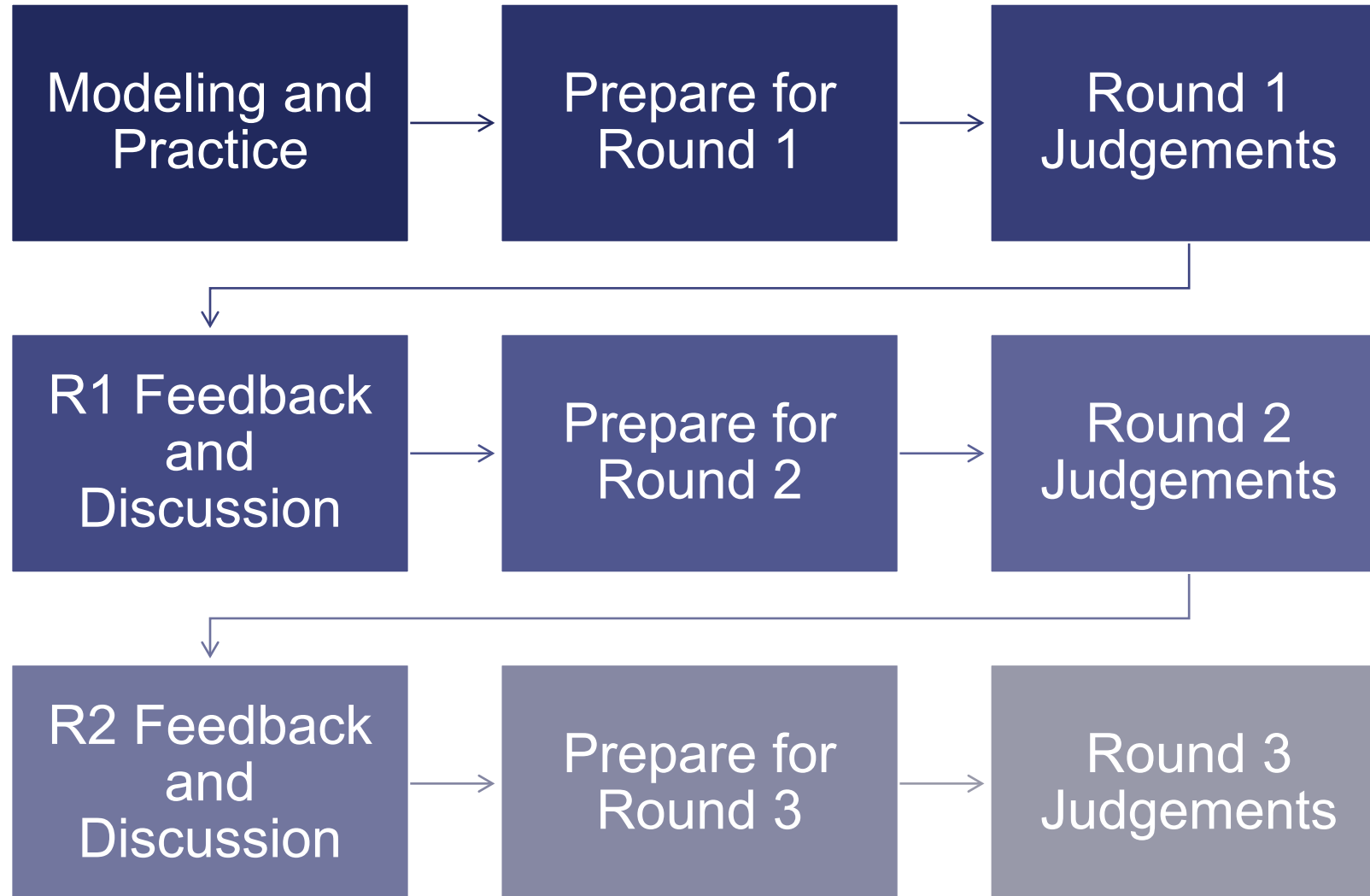
# Overview: From Judgements to Cut Scores

Panelists are presented with Items that are ordered from least to most difficult in OIB based on student data

Panelists will proceed through the items in order of difficulty and make the following judgment:

**Match knowledge, skills, and abilities required by an item with a Performance Level Descriptor**

Particular attention should be given to items within the transition region where PLD alignments transition from one level to the next.

All panelists' judgments will be grouped together to calculate a grade-specific threshold region.
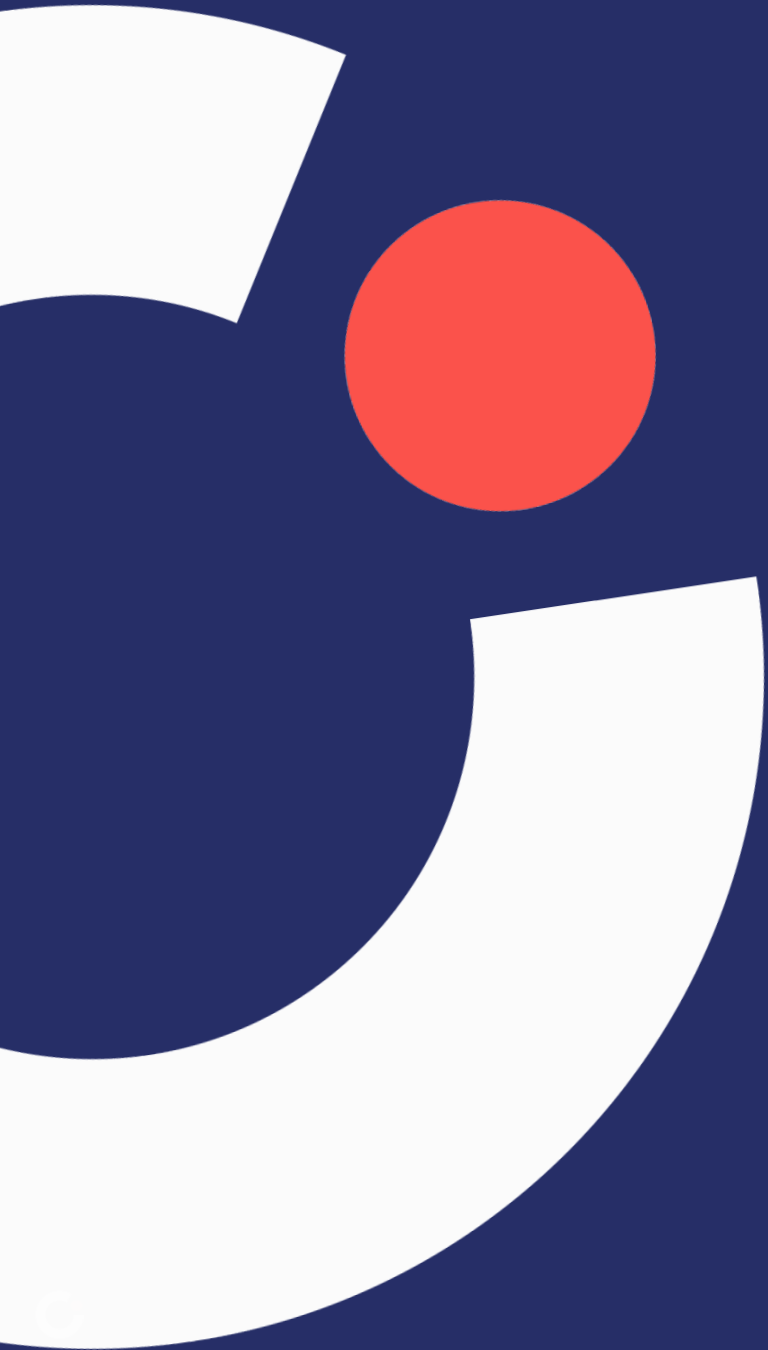
| OIB | PLD Alignment |
|---------|---------------|
| Item 15 | 2 |
| Item 16 | 2 |
| Item 17 | 2 |
| Item 18 | 3 |
| Item 19 | 2 |
| Item 20 | 3 |
| Item 21 | 3 |
| Item 22 | 2 |
| Item 23 | 3 |
| Item 24 | 3 |
| Item 25 | 3 |
| Item 26 | 3 |
| Item 27 | 3 |

Panelist item-PLD Alignment **Transition Region**

Summarize across all panelists using statistics.

| OIB |
|---------|
| Item 15 |
| Item 16 |
| Item 17 |
| Item 18 |
| Item 19 |
| Item 20 |
| Item 21 |
| Item 22 |
| Item 23 |
| Item 24 |
| Item 25 |
| Item 26 |
| Item 27 |

Calculated Grade-specific item-PLD Alignment **Threshold Region**

Modeling and Practice

# Modeling & Practice of the ID-Matching Judgmental Task

We will look at 3 sample items. For each item:

- Step (a) Review the item and identify KSAs
  - Identify the knowledge, skills, and abilities (KSAs) required to answer the item correctly

- Step (b) Make item-PLD alignment judgment
  - Match the KSAs required by the item with the EPEs described in either the Level 2, Level 3, or Level 4 performance level descriptor (PLD).

What does a student need to know or be able to do to respond to this item?

Which PLD most closely matches the knowledge, skills, and abilities (KSAs) required by the item?

# A Reminder: Content-based Judgment

## Good

- Based on Content
- Links items to PLDs
- Refers to specific knowledge, skills, and abilities (KSAs)

## Bad

- Based on something other than the content (e.g., student cognitive disabilities)
- Too general
- Based on a specific student or class

# Content-based Judgment

- A good example:
  - The item requires students to understand that population size is being affected and to be able to read and understand a graph to respond how the population is changing; "describe how a factor affects population size" and "interpreting data" are described in the Level 3 PLD and not in the Level 2 PLD.


- A bad example:
  - The item matches the Level 3 PLD and does not match the Level 2 PLD.

# Practice round process

Complete item-PLD alignment task for 3 sample items.

Discuss matches
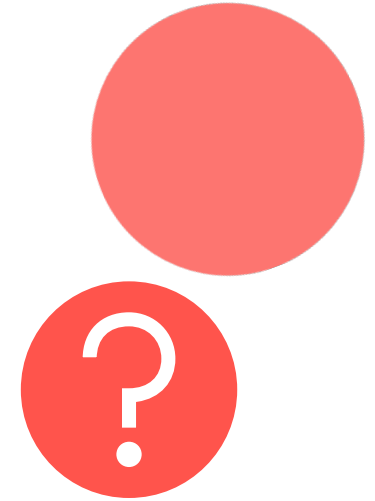
Discuss and clarify range and borderline PLDs.

# Practice Round

- In the Toolkit you will automatically be redirected to the practice round
- You will see 3 practice items

Jeanie has a baby hamster. She wants to be sure it grows up healthy. She feeds it every day.

MS-LS1-5

Which question should Jeanie ask to find out how to help the hamster grow?

   A. Is the hamster a boy or a girl?

   B. What is the best name for the hamster?

   C. Does the hamster get enough food and water?

Jeanie has a baby hamster. She wants to be sure it grows up healthy. She feeds it every day.

MS-LS1-5



Which question should Jeanie ask to find out how to help the hamster grow?

   A. Is the hamster a boy or a girl?

   B. What is the best name for the hamster?

   C. Does the hamster get enough food and water?
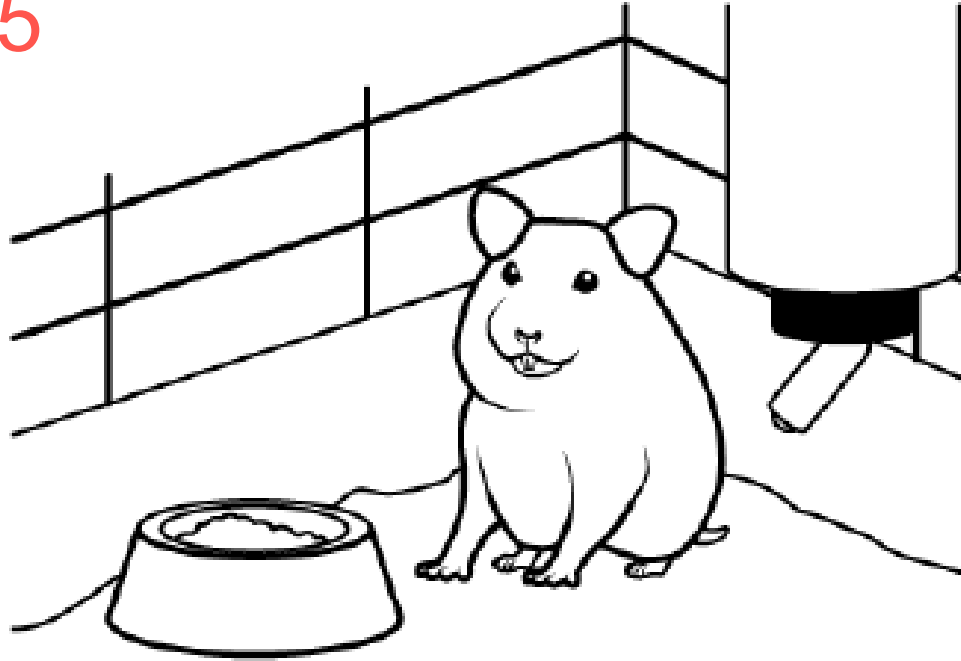
**KSAs**: how organisms grow, interpret diagram (food and water), text says food everyday

**Notes** : PLD L2 "ID factors affecting growth of organism by asking questions", only DCI. Not PLD L3 b/c not "analyzing data".

A scientist measured the growth of small plants growing on the ocean surface. Then, she added fertilizer to the plants and measured their growth again. Her data are shown in the data table.

**Effect of Adding Fertilizer to Ocean Plants**

|  | Before Fertilizer | After Fertilizer |
|---|---|---|
| Plant Growth Rate | 0.25 | 0.65 |

According to the data table, which factor affected the growth of the ocean plants?

A. cloudy skies

B. ocean temperature

C. presence of fertilizer

MS-LS1-5

**KSAs**: The student should identify the factor affecting growth.

**Notes**: PLD L3 because there is a data table.

A scientist measured the growth of small plants growing on the ocean surface. Then, she added fertilizer to the plants and measured their growth again. Her data are shown in the data table.

**Effect of Adding Fertilizer to Ocean Plants**

| | Before Fertilizer | After Fertilizer |
|---|---|---|
| Plant Growth Rate | 0.25 | 0.65 |

According to the data table, which factor affected the growth of the ocean plants?

A. cloudy skies

B. ocean temperature

C. presence of fertilizer

MS-LS1-5

Scientists conducted an investigation to see how different foods available in spring affect geese. They captured geese and weighed them. Then, they weighed how much grass was in the stomach of each goose. The scientists released the geese and then captured them a second time, repeating their measurements. The data table lists their data.

MS-LS1-5

**How Does Eating Grass Affect Goose Weight?**

| Date | Average Weight of Grass Eaten (grams) | Average Goose Weight (kilograms) |
|---|---|---|
| March 1st | 22.7 | 4.65 |
| March 12th | 26.2 | 5.27 |

Based on the data table, how does eating grass affect goose weight?

  A. Goose weight increases when geese eat more grass.

  B. Goose weight decreases when geese eat more grass.

  C. Goose weight stays the same when geese eat more grass.

**KSAs**: The student should describe how goose weight is affected by eating grass.

**Notes**: PLD L4 because the answer options all look similar.

Scientists conducted an investigation to see how different foods available in spring affect geese. They captured geese and weighed them. Then, they weighed how much grass was in the stomach of each goose. The scientists released the geese and then captured them a second time, repeating their measurements. The data table lists their data.

MS-LS1-5

**How Does Eating Grass Affect Goose Weight?**

| Date | Average Weight of Grass Eaten (grams) | Average Goose Weight (kilograms) |
|---|---|---|
| March 1st | 22.7 | 4.65 |
| March 12th | 26.2 | 5.27 |

Based on the data table, how does eating grass affect goose weight?

A. Goose weight increases when geese eat more grass.

B. Goose weight decreases when geese eat more grass.

C. Goose weight stays the same when geese eat more grass.

# Practice Round - Review

- Reviewed three sample items and for each one:
  - Reviewed the item
  - Considered the Knowledge, Skills, and Abilities (KSAs) required by the item
  - Matched the item to either the Level 2, Level 3, or Level 4 PLD
- Borderline considerations
  - Some items seem to be in the border between two adjacent PLDs
  - Select the PLD that most closely matches the item
  - Make notes for yourself next to these items to inform discussions later
- Remaining questions or concerns?
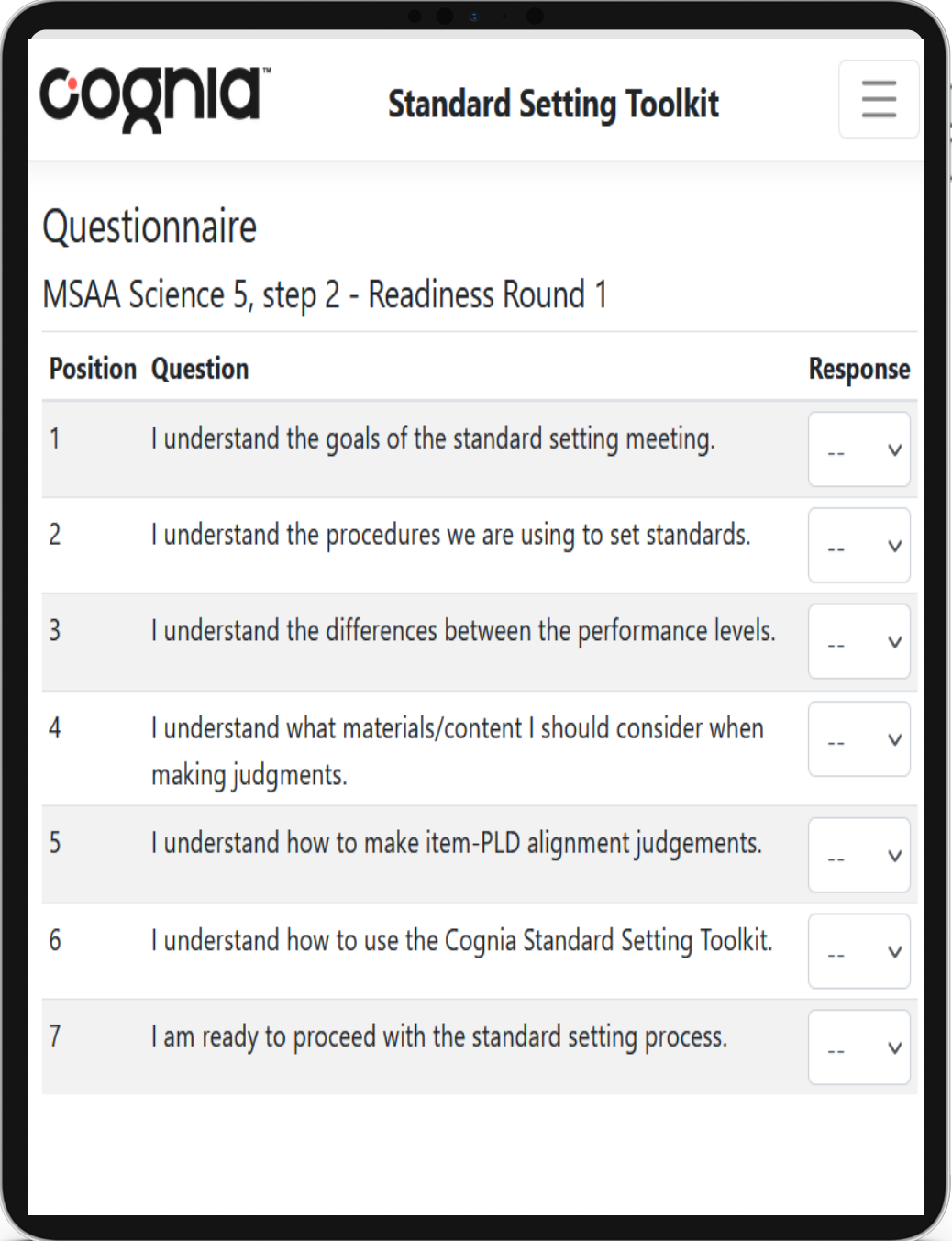
Prepare for Round 1

# Round 1 – Readiness survey

- In a moment, you will be redirected in the Toolkit to a short survey

- Goal: Determine if everyone understands the task at hand and is ready to proceed

- Read each question and answer yes/no

- Once everyone has completed the survey, we will review responses and proceed accordingly.

# Round 1 Judgments

- You will now be redirected to Round 1
  - In the toolkit you will see the full list of items
- Reminder: Your task
  - Review each item
  - Consider the KSAs and match the item to one of the PLDs
  - Write content-based reasons in the "KSAs" box as you go
  - Use the "Notes" box for additional notes (for example: when an item seems to be in-between two PLDs)
- Item-PLD alignment is an individual activity – please do not discuss your work with your colleagues at this time.

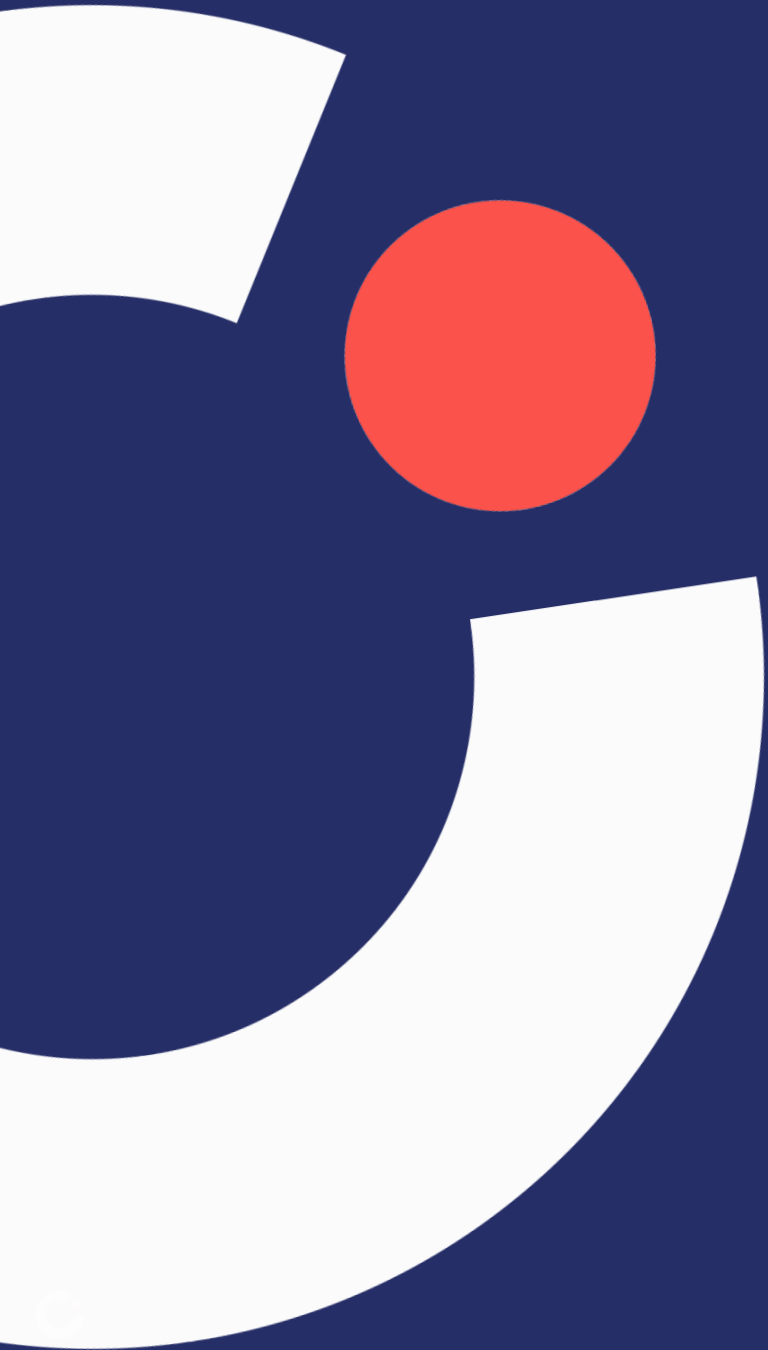# Round 1 Judgements

For each item:

- Step (a) Review the item and identify KSAs
  - Identify the knowledge, skills, and abilities (KSAs) required to answer the item correctly

- Step (b) Make item-PLD alignment judgment
  - Match the KSAs required by the item with the EPEs described in the Level 2, 3, or 4 PLD.

- Work independently
- Trust your expertise

What does a student need to know or be able to do to respond to this item?

Which PLD most closely matches the knowledge, skills, and abilities (KSAs) required by the item?

Adjourn

cognia®

# MSAA Science Standard Setting – Breakout Session Day 2

Grade X

Facilitator: Name here
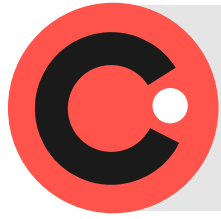
# Breakout Session – Agenda (day 2)

↻ Debrief Day 1

↻ Complete Round 1 Judgments

↻ Feedback and Discussion of Round 1 Results

↻ Prepare for Round 2

↻ Complete Round 2 Judgments

# Swimming pool analogy - Feedback



Relaxing in the shallow end | Approaching deeper water | In the deep end and not in a fun way

3.5 ft.

9 ft.
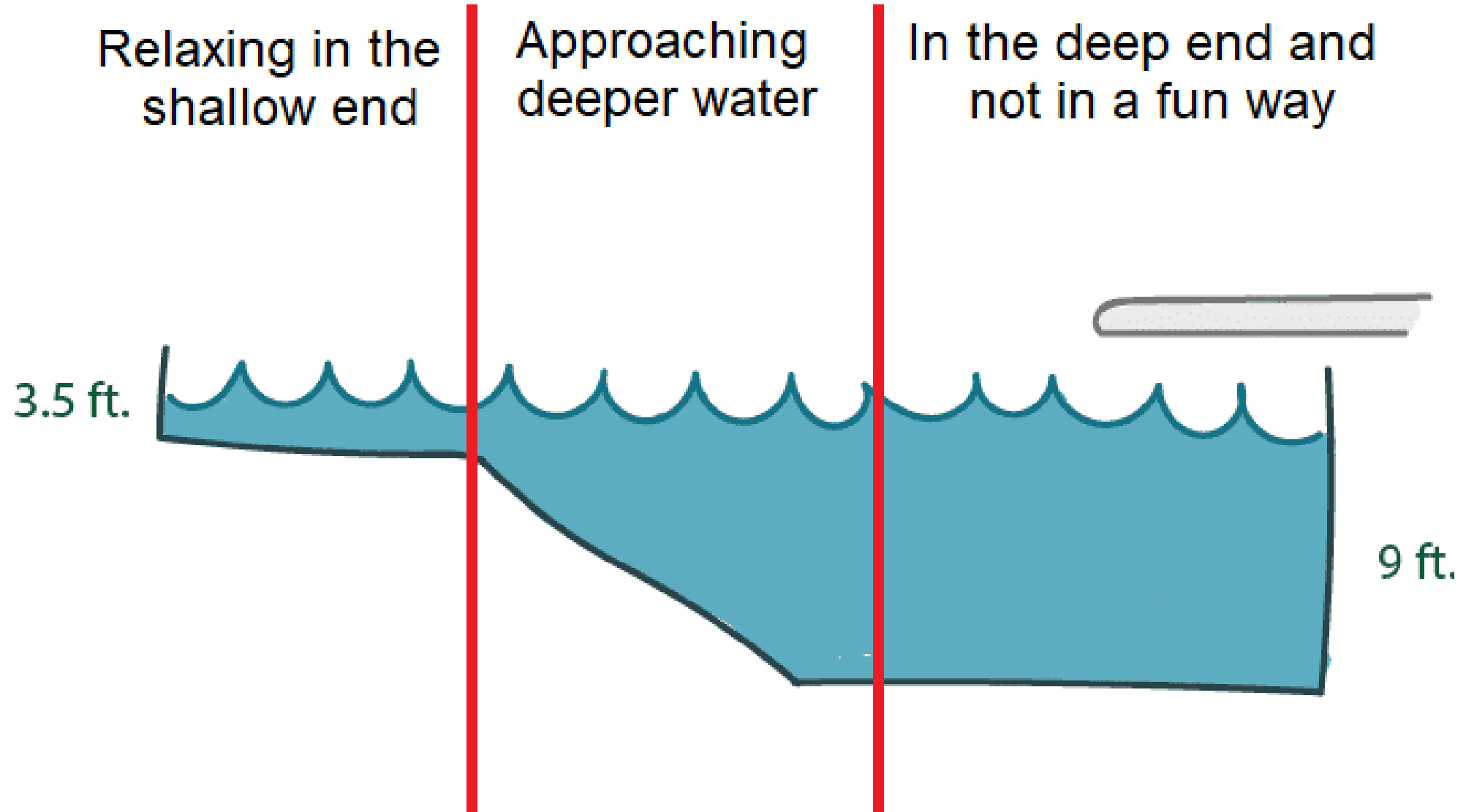
# Feedback – Post its

- Write a positive takeaway from yesterdays standard setting activities – slap on window wall.
- Write an opportunity for growth or improvement that the group can discuss – slap on opposite wall.
- After everyone has slapped, pick one post it from each wall other than your own to share w/the group.

# Swimming pool analogy - Feedback



Relaxing in the shallow end

Approaching deeper water

In the deep end and not in a fun way

3.5 ft.

9 ft.
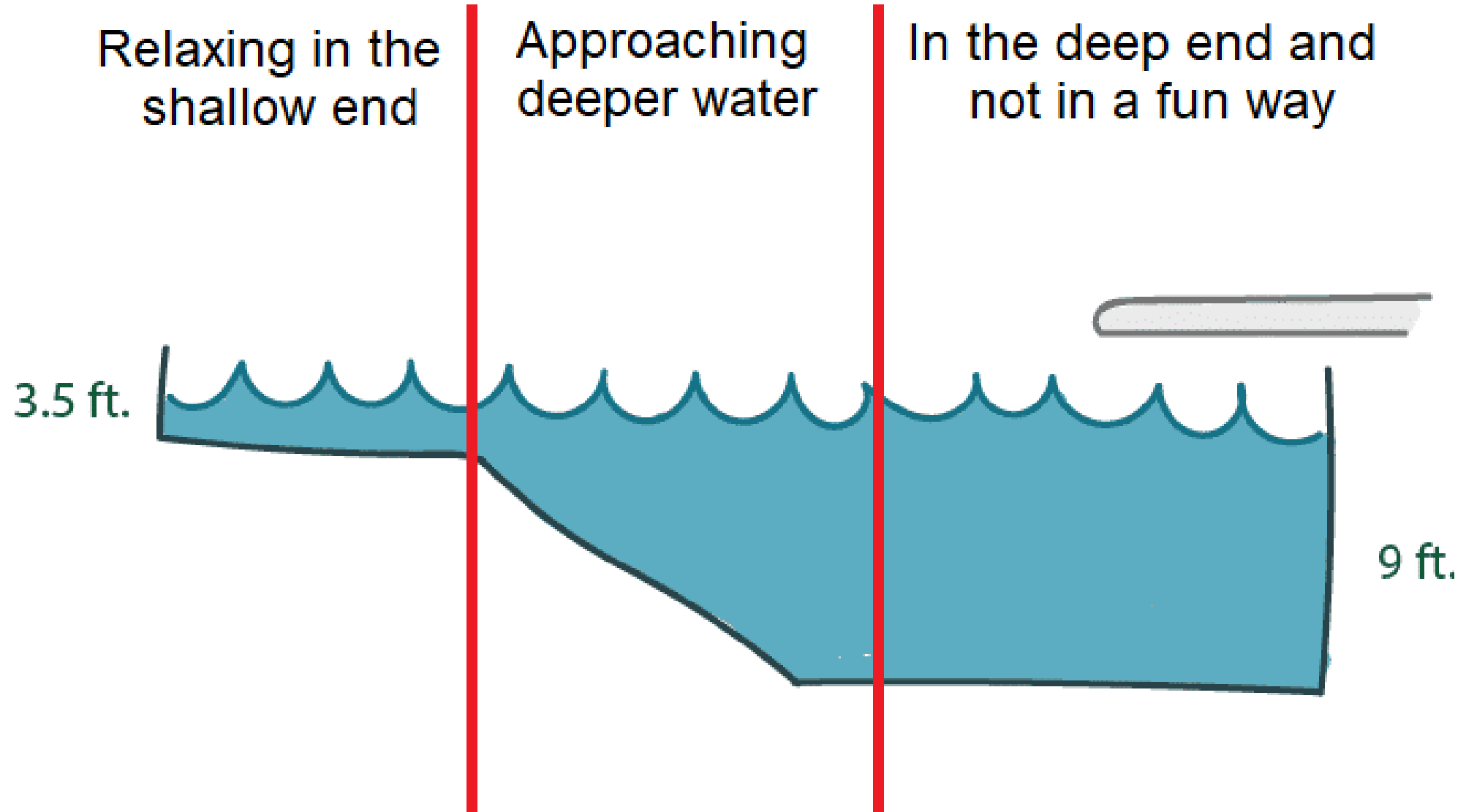
# Round 1 Judgements - Continue

For each item:

- Step (a) Review the item and identify KSAs
  - Identify the knowledge, skills, and abilities (KSAs) required to answer the item correctly

- Step (b) Make item-PLD alignment judgment
  - Match the KSAs required by the item with the EPEs described in the Level 2, 3, or 4 PLD.

- Work independently
- Trust your expertise

What does a student need to know or be able to do to respond to this item?

Which PLD most closely matches the knowledge, skills, and abilities (KSAs) required by the item?

# Lunch
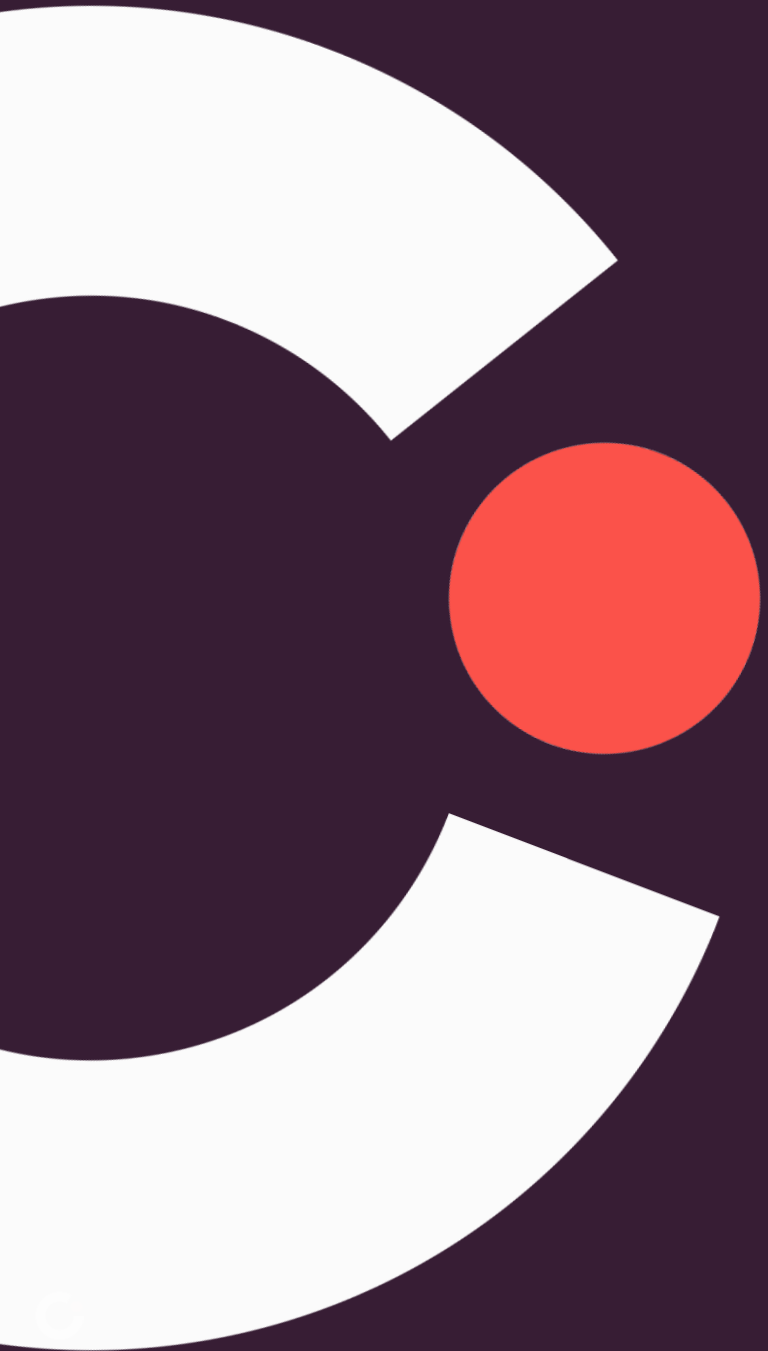
Feedback & Discussion of Round 1 Results

# Round 1 PLD judgements



08  Round  1 Results
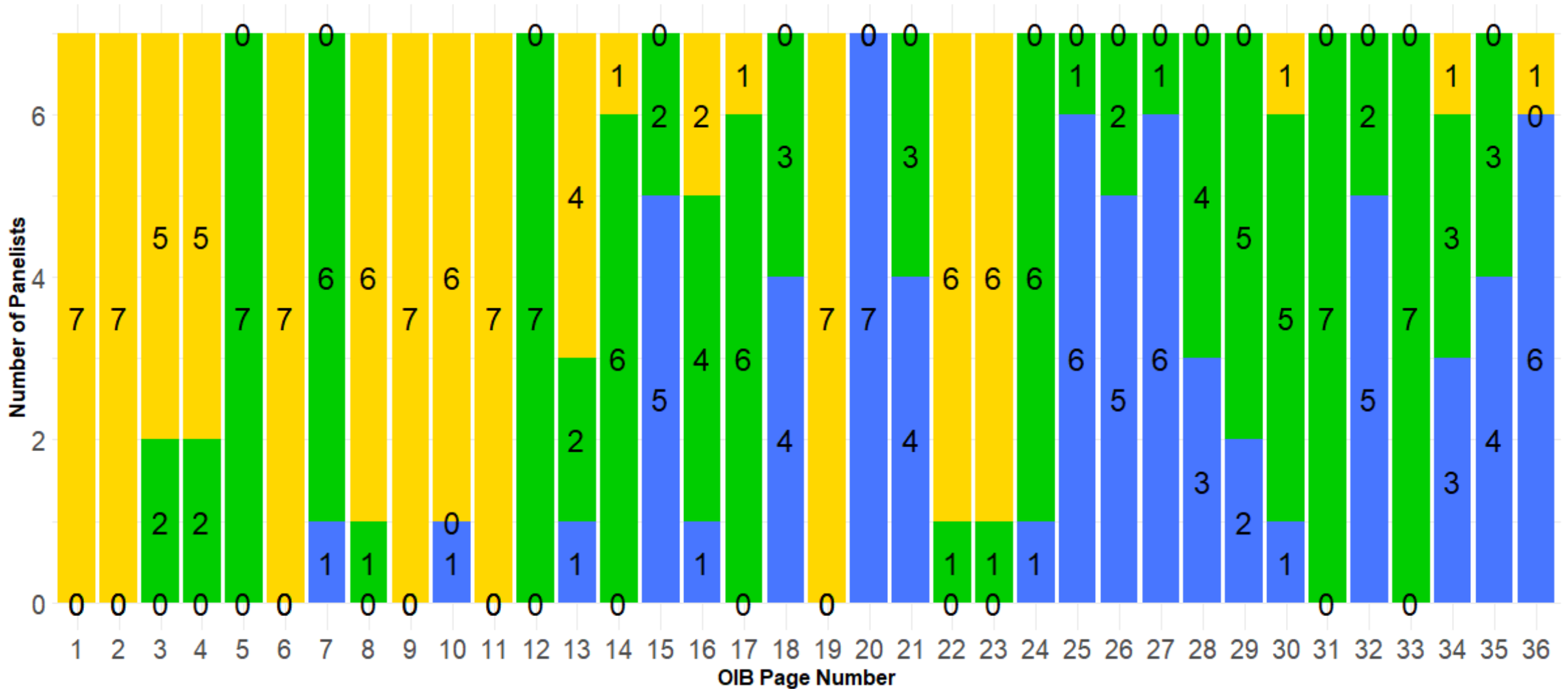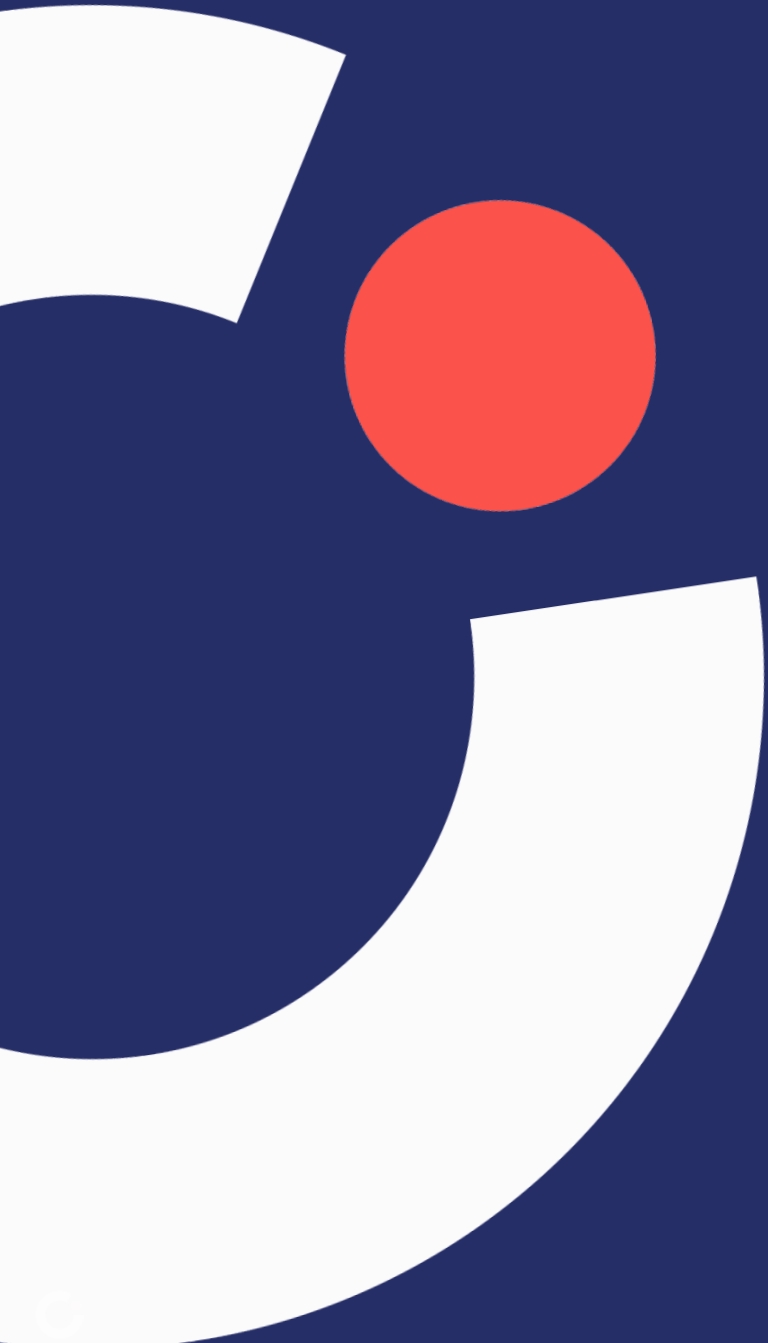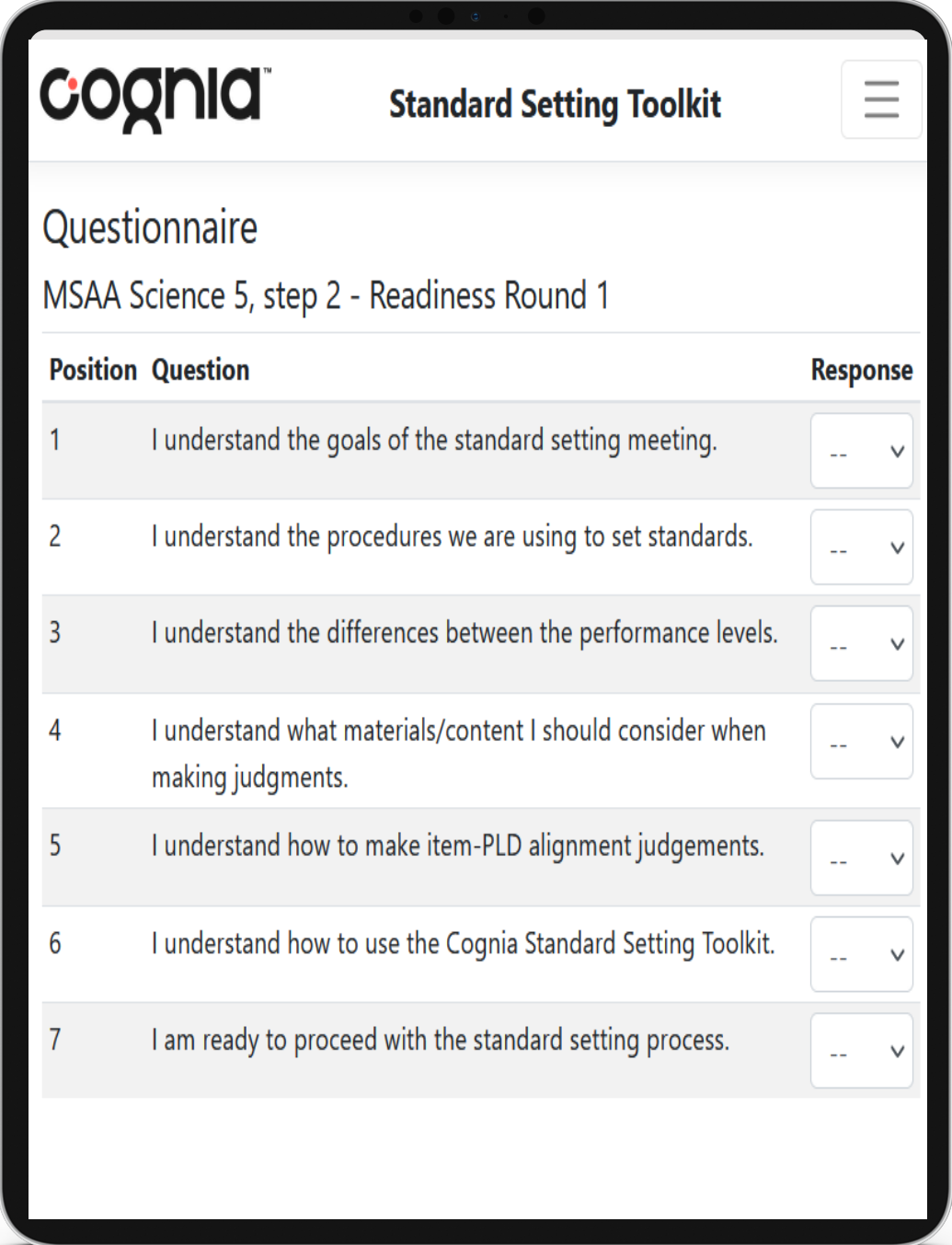
# Preparation for Round 2

# Round 2 – Readiness survey

- In a moment, you will be redirected in the Toolkit to a short survey

- Goal: Determine if everyone understands the task at hand and is ready to proceed

- Read each question and answer yes/no

- Once everyone has completed the survey, we will review responses and proceed accordingly.

# Round 2 judgments

- You will now be redirected to Round 2
  - In the toolkit you will see the same full list of items with your work from round 1 (notes and judgments)
- Reminder: Your task
  - Review items you were previously unsure about
  - Consider the KSAs and decide to keep or change your initial PLD Match
- Item-PLD alignment is an individual activity – please do not discuss your work with your colleagues at this time.
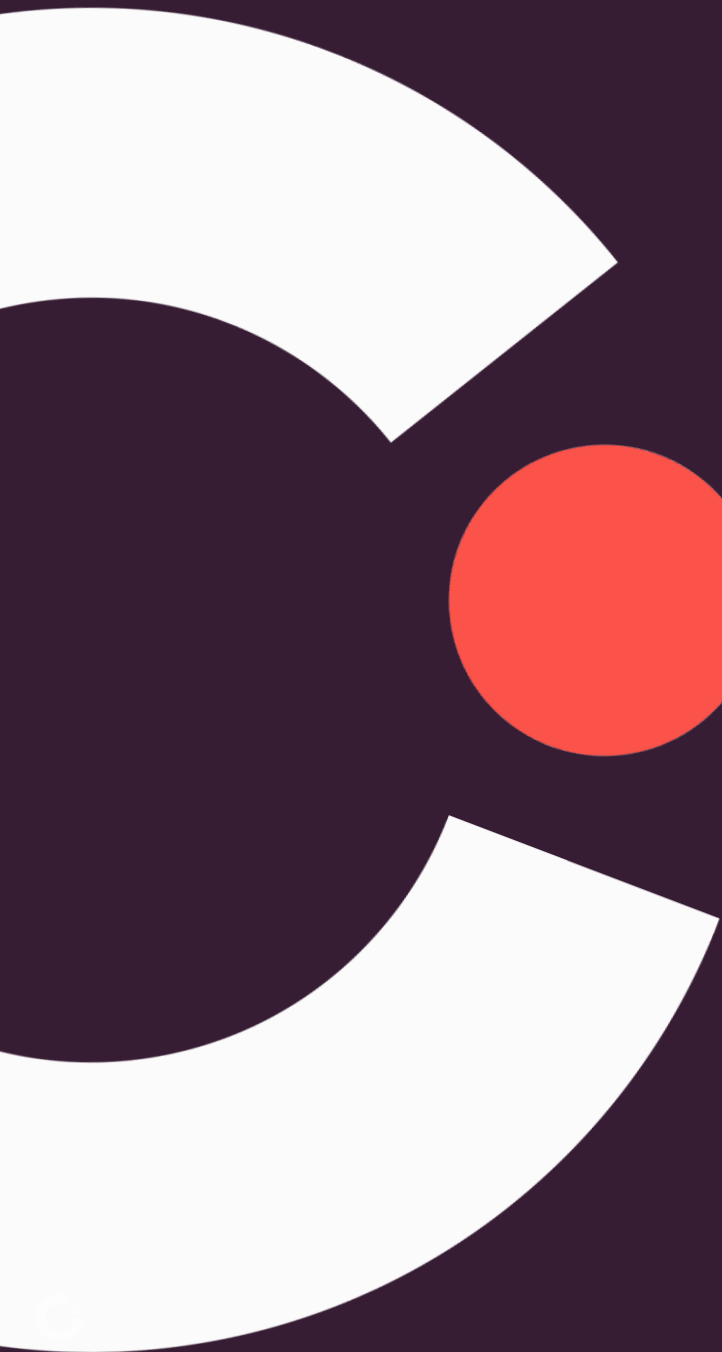
cognia®

# Round 2 Judgements

- Decide to retain/adjust your judgements:
  - Review items in the benchmark regions and items you were previously unsure about
  - Consider the KSAs and decide to keep or change your initial PLD Match

- Reminder:
  - Step (a) Review the item and identify KSAs
  - Step (b) Make item-PLD alignment judgment

- Work independently
- Trust your expertise

What does a student need to know or be able to do to respond to this item?

Which PLD most closely matches the knowledge, skills, and abilities (KSAs) required by the item?
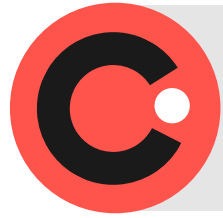
Break

cognia®

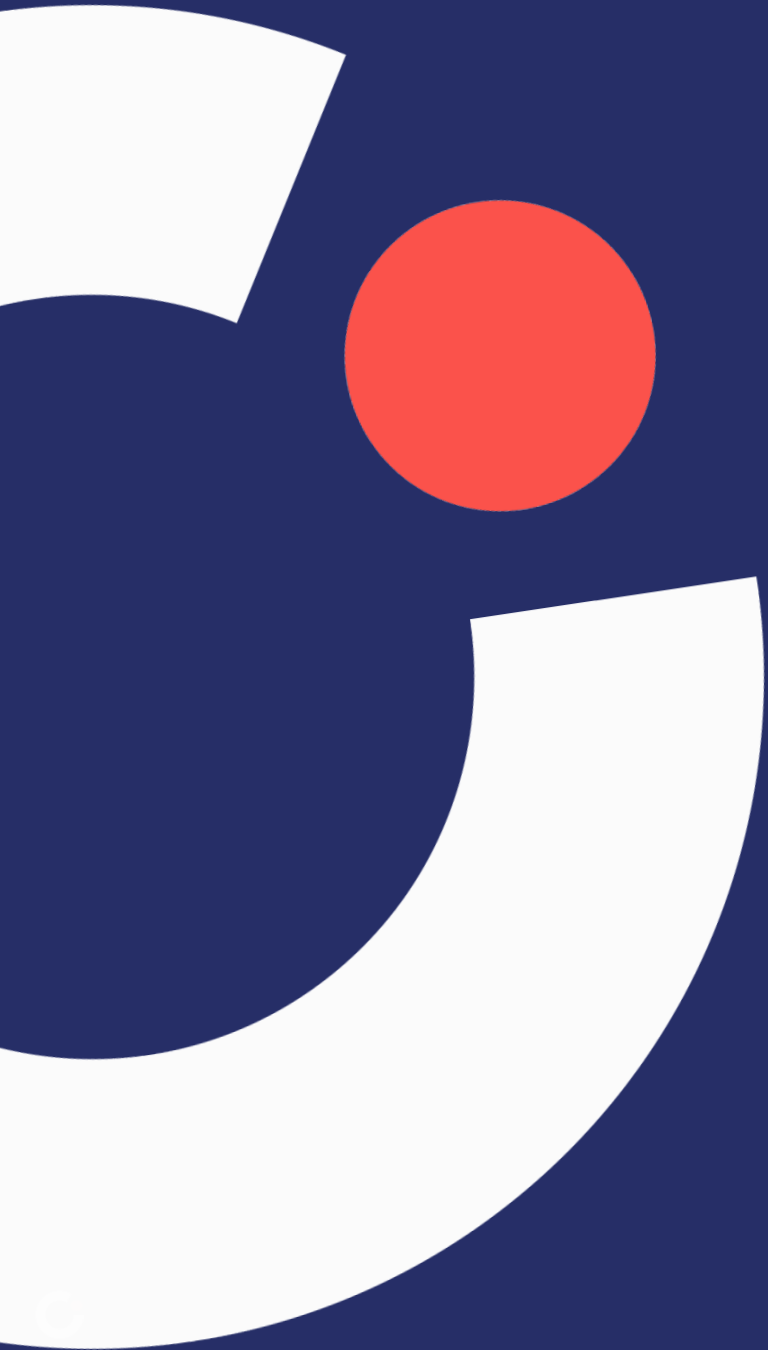# MSAA Science Standard Setting – Breakout Session Day 3

Grade X

Facilitator: Name here

# Breakout Session – Agenda (day 3)

- Debrief Day 2
- Feedback and Discussion of Round 2 Results
- Prepare for Round 3
- Complete Round 3 Judgments
- Write group-level content-based rationales for cut scores (if needed)
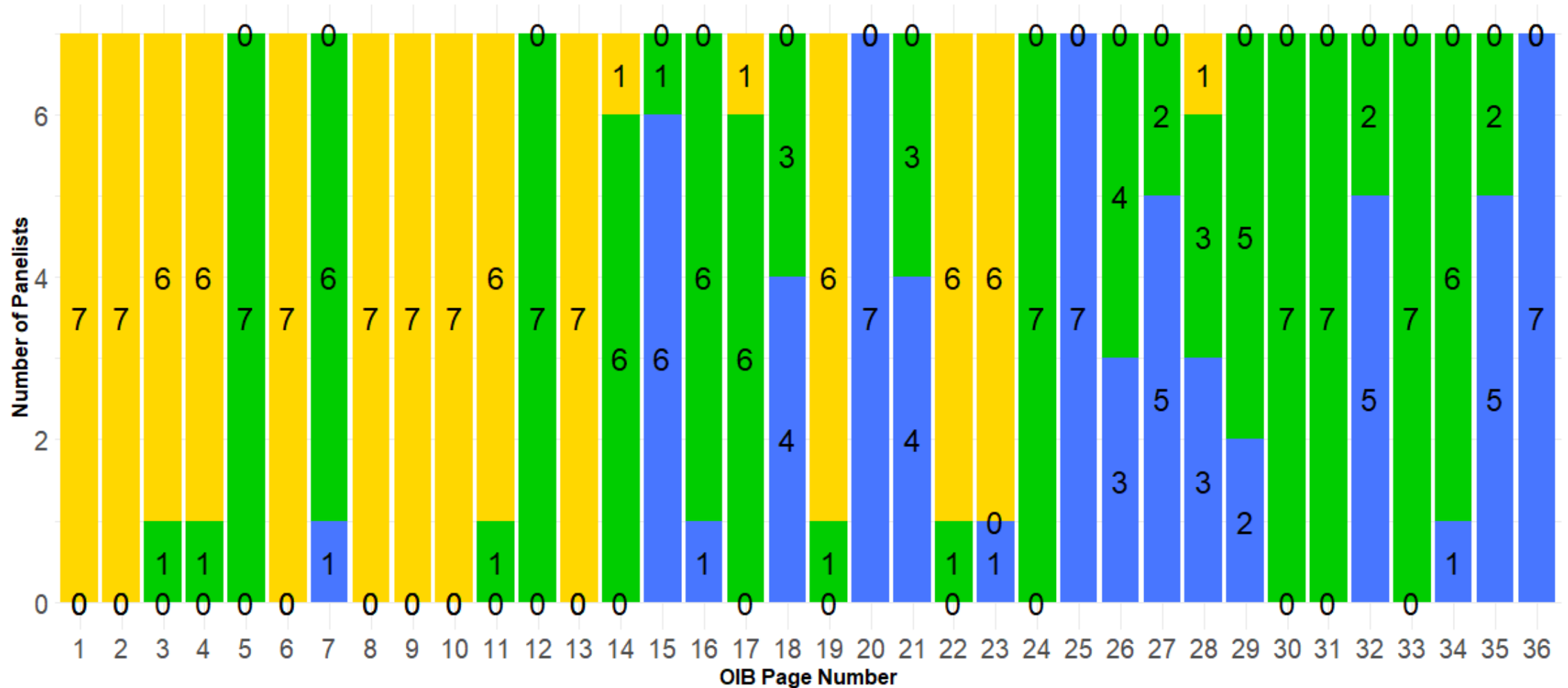- Final Workshop Evaluation Survey
- Lunch/Adjourn
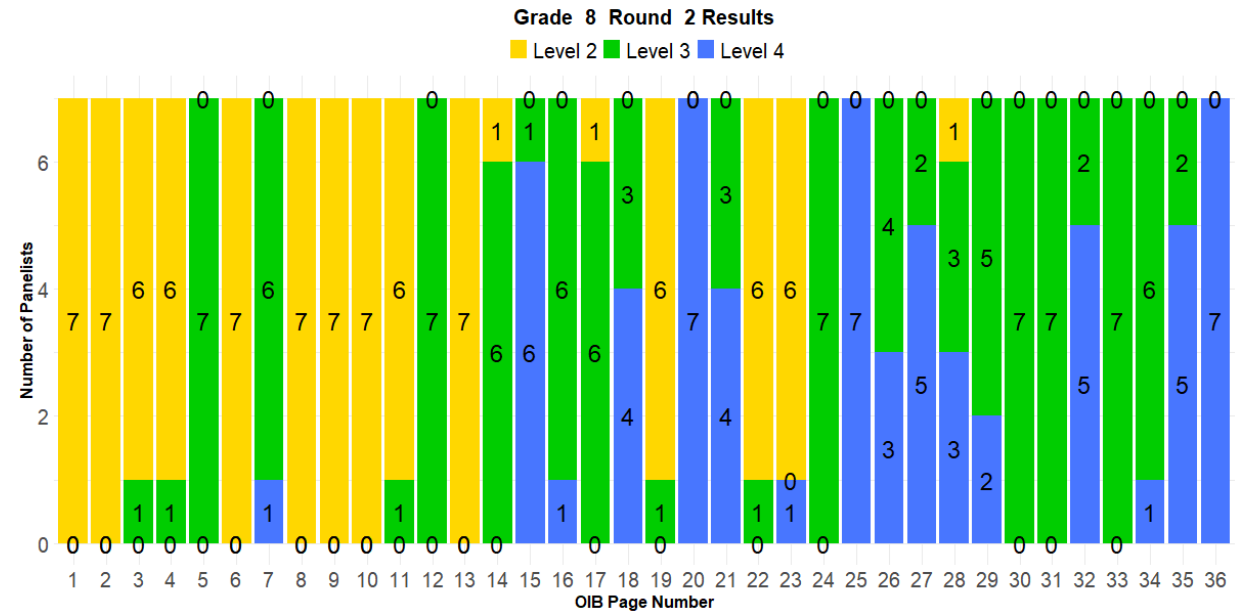
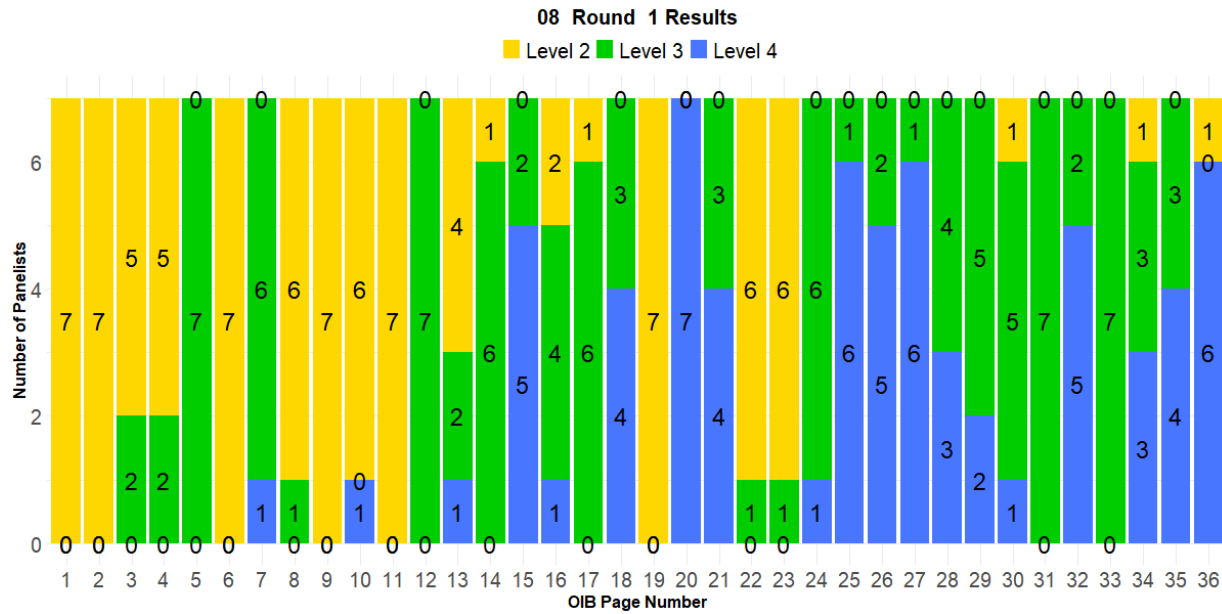# Feedback & Discussion of Round 2 Results

# Round 2 PLD judgements



Grade 8 Round 2 Results

# Round 1 vs. round 2 PLD judgements

# Introduce benchmarks

- Benchmarks are content-based information judgments from others following the same process with these items

- Benchmarks serve as additional information for your consideration.

- Will be presented as shaded rows in the OIB:

# Content-Based Benchmarks: Visual Presentation

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 10 | 555974 | 3-LS4-1 fish-teeth-rock fossils level 2 | 1 | | | -- | Detail |
| 11 | 555837 | 5-ESS2-1 atmosphere-beach-roots level 1 | 1 | | | -- | Detail |
| 12 | 555728 | 5-PS3-1 forest energy-fish-pond level 1 | 1 | | | -- | Detail |
| 13 | 557052 | 4-LS1-1 3-LS3-1 tomato cluster 4LS11 level 2 | 1 | | | -- | Detail |
| 14 | 560418 | 4-PS3-4 tea-lamp-solar light level 1 | 1 | | | -- | Detail |
| 15 | 556982 | 5-PS1-2 scale-chocolate-choc bar level 1 | 1 | | | -- | Detail |
| 16 | 555951 | 5-PS3-1 forest energy-fish-pond level 2 | 1 | | | -- | Detail |
| 17 | 555978 | 3-LS4-1 fish-teeth-rock fossils level 3 | 1 | | | -- | Detail |
| 18 | 555885 | 3-ESS2-1 thermometer-spring temp-Ohio temp level 2 | 1 | | | -- | Detail |
| 19 | 560410 | 5-PS2-1 leaves-baseball-objects level 3 | 1 | | | -- | Detail |
| 20 | 556984 | 5-PS1-2 scale-chocolate-choc bar level 2 | 1 | | | -- | Detail |
| 21 | 557000 | 5-PS2-1 faucet-raindrops-airplane level 3 | 1 | | | -- | Detail |
| 22 | 556996 | 5-PS2-1 faucet-raindrops-airplane level 1 | 1 | | | -- | Detail |
| 23 | 557429 | 5-ESS3-1 trash-deforest-habitat level 3 | 1 | | | -- | Detail |

# Content-based benchmarks

- The shaded regions are calculated based on judgments from others following the same process with these items.
- This region represents a transition area where items between two performance levels are beginning to intersect.
- The shaded region does **NOT** mean that each item in that region was judged at that performance level.
- It is vital that we have the input of educators who teach to these standards and this student population. To that end, your results may very well differ from theirs.
- The content-based benchmarks provide additional information for your consideration but are not meant to constrain or persuade your judgments.

# Preparation for Round 3

# Round 3 – Readiness survey

- In a moment, you will be redirected in the Toolkit to a short survey

- Goal: Determine if everyone understands the task at hand and is ready to proceed

- Read each question and answer yes/no

- Once everyone has completed the survey, we will review responses and proceed accordingly.

# Round 3 judgments

- You will now be redirected to Round 3
  - In the toolkit you will see the same full list of items with your work from round 2 (notes and judgments).
  - You will also see the shaded regions for the content-based benchmarks.
- Reminder: Your task
  - review items in the benchmark regions and items you were previously unsure about.
  - consider the KSAs and decide to keep or change your initial PLD match.
- Item-PLD alignment is an individual activity – please **do not discuss** your work with your colleagues at this time.

# Round 3 Judgements

- Decide to retain/adjust your judgements:
  - Review items in the benchmark regions and items you were previously unsure about
  - Consider the KSAs and decide to keep or change your initial PLD Match

- Reminder:
  - Step (a) Review the item and identify KSAs
  - Step (b) Make item-PLD alignment judgment

- Work independently
- Trust your expertise

What does a student need to know or be able to do to respond to this item?

Which PLD most closely matches the knowledge, skills, and abilities (KSAs) required by the item?

# Round 3 PLD judgements



08 Round 3 Results
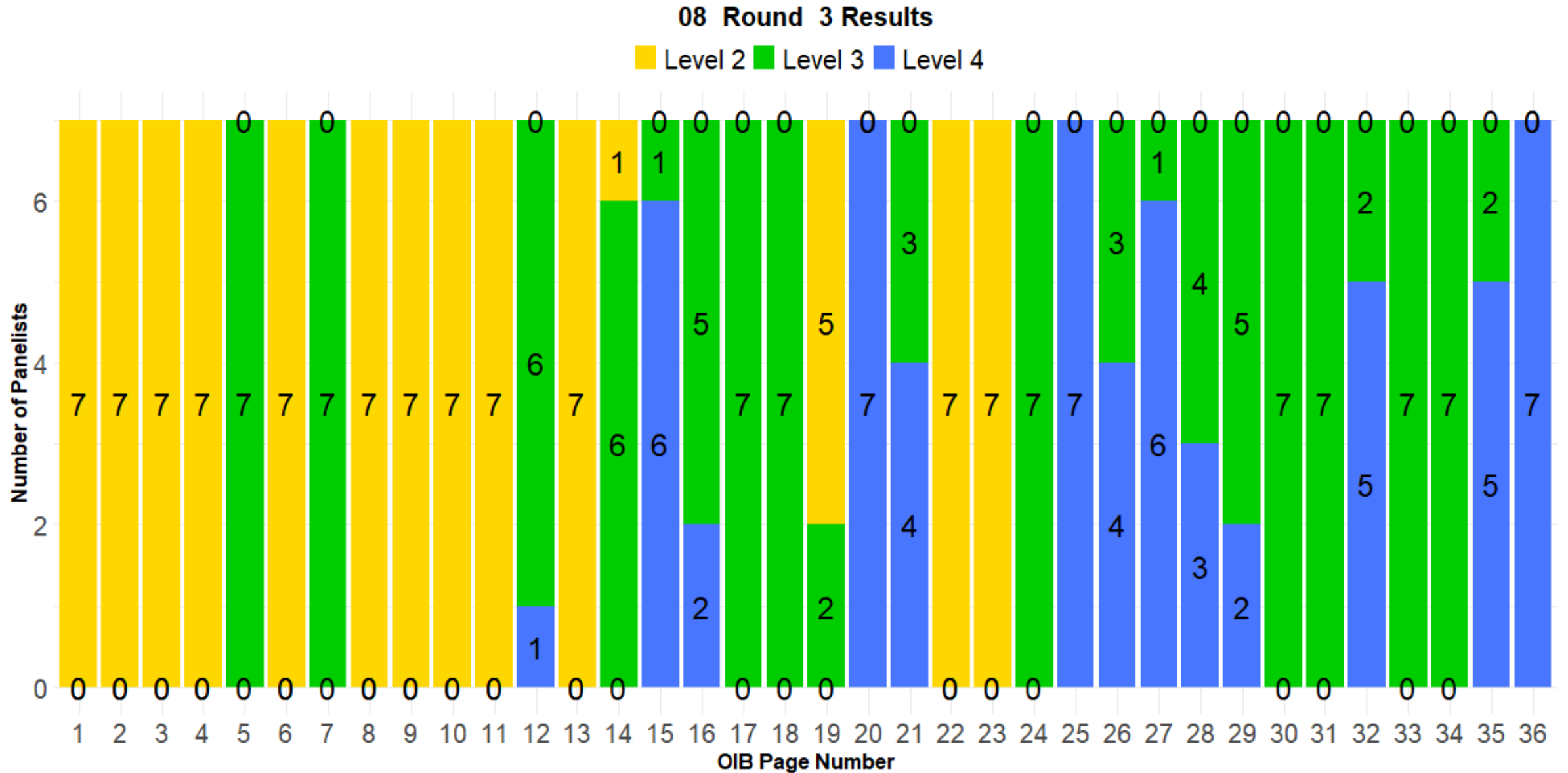
# Round 2 vs. round 3 PLD judgements
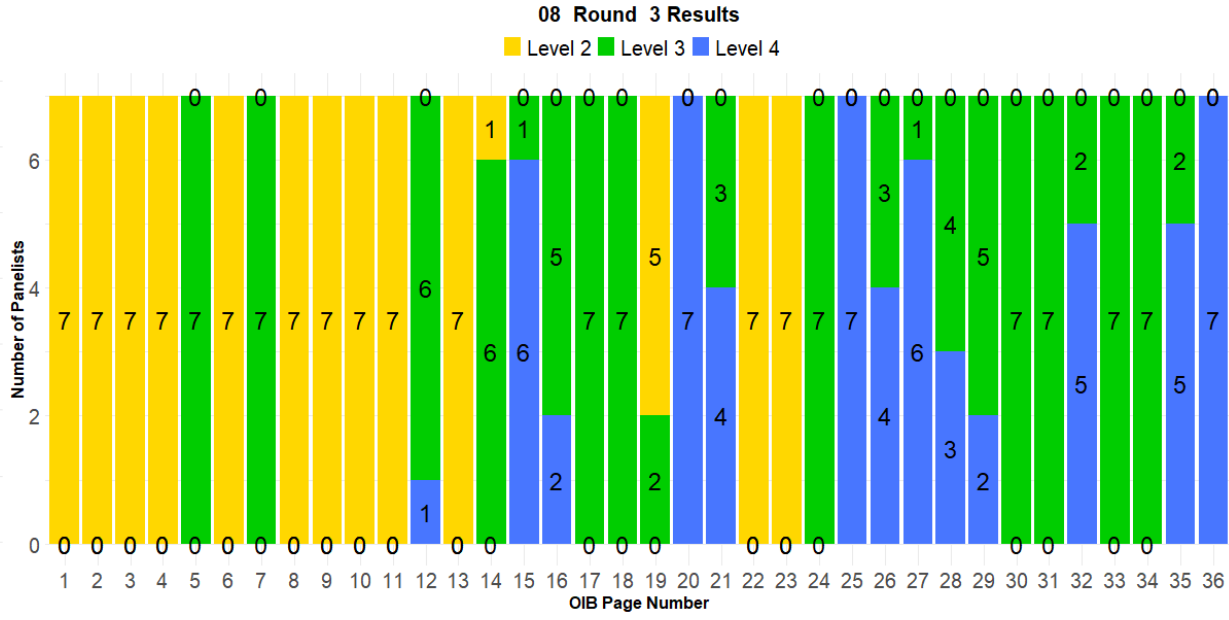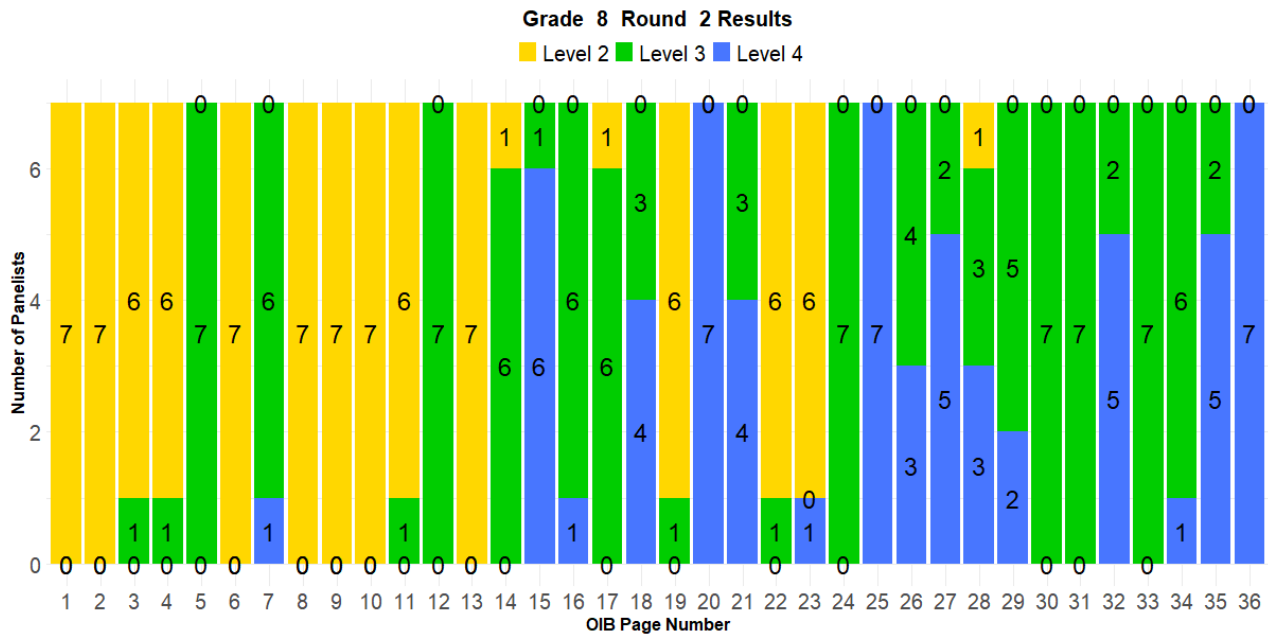
# Complete Workshop Evaluation Survey

# APPENDIX E
# LOGISTIC REGRESSION CALCULATION

# Logistic Regression Calculation

The proficient and advanced cut scores were computed using the logistic regression as follows:

$$log \frac{P}{1 - P} = \beta_0 + \beta_1 \theta$$

which is equivalent to:

$$P = \frac{exp\ (\beta_0 + \beta_1 \theta)}{1 + exp\ (\beta_0 + \beta_1 \theta)}$$

Where $\beta_0$ (intercept) and $\beta_1$ (slope) are two regression coefficients that need to be computed, theta $(\theta)$ is the RP67 value associated with each OIB page, and $P$ is the probability of observing a performance level (level X or above) given theta. After fitting the model with data, the theta cut score is obtained by finding which score corresponds to a probability of 0.5 for being rated above the cut as follows:

$$log \frac{0.5}{1 - 0.5} = 0 = \beta_0 + \beta_1 \theta$$

Solving the equation, the following is obtained:

$$\theta = -\frac{\beta_0}{\beta_1}$$

Additionally, the variance of the theta estimate will be computed as:

$$VAR\ (\theta) = \frac{\mu_{\beta 0}^2}{\mu_{\beta 1}^2} \left[ \frac{\sigma_{\beta 0}^2}{\mu_{\beta 0}^2} - 2\frac{Cov(\beta_0, \beta_1)}{\beta_0 \beta_1} + \frac{\sigma_{\beta 1}^2}{\mu_{\beta 1}^2} \right]$$

Therefore, the standard error of the estimate is given by:

$$SE(\theta) = \sqrt{VAR(\theta)}.$$

# APPENDIX F
# COGNIA STANDARD SETTING TOOLKIT
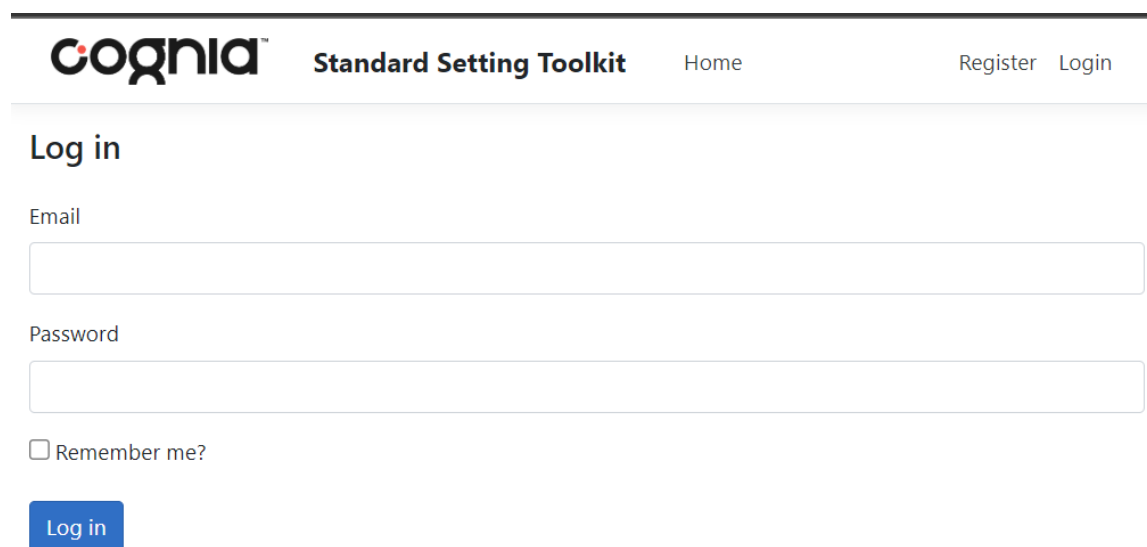
# Cognia Standard Setting Toolkit

This appendix contains sample screenshots of the Cognia Standard Setting Toolkit that panelists used for all standard setting activities during the meeting. Images provided correspond to samples (1) login screen, (2) readiness survey screen, (3) ordered item booklet view, and (4) item detail view.

**Figure 1. Sample Login Screen**

Panelists are provided with usernames and passwords to enable secure access to the toolkit.
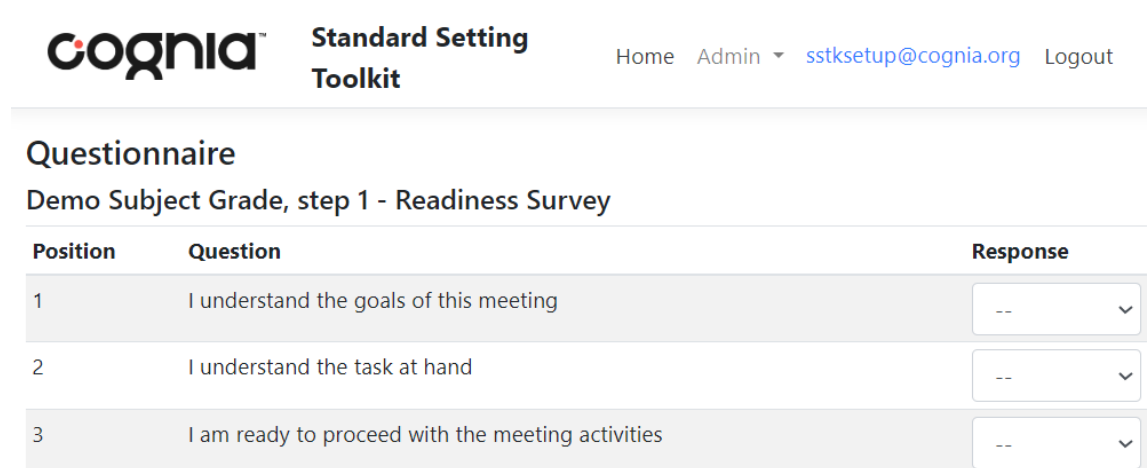


**Figure 2. Sample Readiness Survey**

**Figure 3. Sample Ordered Item Booklet View**

**Figure 4. Sample Item Detail View**

# APPENDIX G
# READINESS SURVEY RESULTS

# Grade 5 Survey Results

### Readiness—Round 1

| Question | Yes | No |
| --- | --- | --- |
| I understand the goals of the standard setting meeting. | 7 | -- |
| I understand the procedures we are using to set standards. | 7 | -- |
| I understand the differences between the performance levels. | 7 | -- |
| I understand what materials/content I should consider when making judgments. | 7 | -- |
| I understand how to make item-PLD alignment judgements. | 7 | -- |
| I understand how to use the Cognia Standard Setting Toolkit. | 7 | -- |
| I am ready to proceed with the standard setting process. | 7 | -- |

### Readiness—Round 2

| Question | Yes | No |
| --- | --- | --- |
| I understand the round 1 feedback. | 7 | -- |
| I understand that I should use the round 1 feedback as information, not persuasion, for me to consider as I make my judgements in round 2. | 7 | -- |
| I understand that I should consider the insights of my colleagues as information, but not persuasion, as I make my own independent judgments in round 2. | 7 | -- |
| I am ready to proceed with Round 2 of the standard setting process. | 7 | -- |

### Readiness—Round 3

| Question | Yes | No |
| --- | --- | --- |
| I understand the round 2 feedback. | 7 | -- |
| I understand that I should use the round 2 feedback as information, not persuasion, for me to consider as I make my judgements in round 3. | 7 | -- |
| I understand what the content-based benchmarks represent. | 7 | -- |
| I understand that I can use the content-based benchmarks as additional information, not persuasion, for me to consider as I make my judgements in round 3. | 7 | -- |
| I understand that I should consider the insights of my colleagues as information, but not persuasion, as I make my own independent judgments in round 3. | 7 | -- |
| I am ready to proceed with Round 3 of the standard setting process | 7 | -- |

# Grade 8 Survey Results

**Readiness—Round 1**

| Question | Yes | No |
|---|---|---|
| I understand the goals of the standard setting meeting. | 7 | -- |
| I understand the procedures we are using to set standards. | 7 | -- |
| I understand the differences between the performance levels. | 7 | -- |
| I understand what materials/content I should consider when making judgments. | 7 | -- |
| I understand how to make item-PLD alignment judgements. | 7 | -- |
| I understand how to use the Cognia Standard Setting Toolkit. | 7 | -- |
| I am ready to proceed with the standard setting process. | 7 | -- |

**Readiness—Round 2**

| Question | Yes | No |
|---|---|---|
| I understand the round 1 feedback. | 7 | -- |
| I understand that I should use the round 1 feedback as information, not persuasion, for me to consider as I make my judgements in round 2. | 7 | -- |
| I understand that I should consider the insights of my colleagues as information, but not persuasion, as I make my own independent judgments in round 2. | 7 | -- |
| I am ready to proceed with Round 2 of the standard setting process. | 7 | -- |

**Readiness—Round 3**

| Question | Yes | No |
|---|---|---|
| I understand the round 2 feedback. | 7 | -- |
| I understand that I should use the round 2 feedback as information, not persuasion, for me to consider as I make my judgements in round 3. | 7 | -- |
| I understand what the content-based benchmarks represent. | 7 | -- |
| I understand that I can use the content-based benchmarks as additional information, not persuasion, for me to consider as I make my judgements in round 3. | 7 | -- |
| I understand that I should consider the insights of my colleagues as information, but not persuasion, as I make my own independent judgments in round 3. | 7 | -- |
| I am ready to proceed with Round 3 of the standard setting process | 7 | -- |

# Grade HS Survey Results

### Readiness—Round 1

| Question | Yes | No |
|---|---|---|
| I understand the goals of the standard setting meeting. | 7 | -- |
| I understand the procedures we are using to set standards. | 7 | -- |
| I understand the differences between the performance levels. | 7 | -- |
| I understand what materials/content I should consider when making judgments. | 7 | -- |
| I understand how to make item-PLD alignment judgements. | 7 | -- |
| I understand how to use the Cognia Standard Setting Toolkit. | 7 | -- |
| I am ready to proceed with the standard setting process. | 7 | -- |

### Readiness—Round 2

| Question | Yes | No |
|---|---|---|
| I understand the round 1 feedback. | 7 | -- |
| I understand that I should use the round 1 feedback as information, not persuasion, for me to consider as I make my judgements in round 2. | 7 | -- |
| I understand that I should consider the insights of my colleagues as information, but not persuasion, as I make my own independent judgments in round 2. | 7 | -- |
| I am ready to proceed with Round 2 of the standard setting process. | 7 | -- |

### Readiness—Round 3

| Question | Yes | No |
|---|---|---|
| I understand the round 2 feedback. | 7 | -- |
| I understand that I should use the round 2 feedback as information, not persuasion, for me to consider as I make my judgements in round 3. | 7 | -- |
| I understand what the content-based benchmarks represent. | 7 | -- |
| I understand that I can use the content-based benchmarks as additional information, not persuasion, for me to consider as I make my judgements in round 3. | 7 | -- |
| I understand that I should consider the insights of my colleagues as information, but not persuasion, as I make my own independent judgments in round 3. | 7 | -- |
| I am ready to proceed with Round 3 of the standard setting process | 7 | -- |

# APPENDIX H
# STANDARD SETTING MEMO

# Multi-State Alternate Assessment (MSAA) Standard Setting Memo

## Science grades 5, 8, and High School
## July 18-20, 2023

## Overview

Cognia and MSAA partners collaborated to conduct a standard setting workshop from July 18-20, 2023, in Wakefield, MA. The workshop convened a diverse panel of 21 teachers, administrators, and higher-education educators. Over the course of three days, the primary objective was to provide input on the MSAA Science assessment cut scores. The workshop aimed to finalize the cut scores that were initially established in 2022, considering the newness of the assessment and incorporating an additional year of student data.

During the workshop, educators actively participated in a content-driven standard setting process. They received comprehensive training on various aspects of the assessment, including content and accessibility features, post-secondary outcomes for MSAA students, and performance level descriptors. The panel's main task was to align the knowledge, skills, and abilities required by students to answer assessment items correctly with one of the three performance-level descriptors. This alignment process involved three rounds of judgments and extensive discussions.

This memorandum provides a concise summary of the standard setting procedures and outcomes, along with a proposed set of cut scores. These cut scores were derived through collaborative inputs from the standard setting panel, science partners, and Cognia. The aim is to offer actionable information for students, parents, and educators to inform educational progress in science.

## Methods

In this workshop, the ID Matching Method was utilized as a standard setting approach. Panelists familiar with the assessment content and special education students reviewed an ordered item booklet (OIB) and performance level descriptors (PLDs) to align items with predefined PLDs. The ordering of items was based on empirical item difficulty using student data.

Afterward, the item-PLD alignments from the panelists were combined and input into statistical models, such as test characteristic curves for Level 2 cuts and logistic regression for Level 3 and 4 cuts. Cut scores for all three performance levels were derived from these models. See the accompanying documents in "*Cut Score Calculation Charts*" for more details on the statistical models.

### Standard Setting Procedure

#### Plenary Session

In a collaborative effort, Science partners, Cognia, and a member of the MSAA technical advisory committee, serving as an observer, warmly welcomed the standard setting participants. The session began with a thorough introduction, provided a comprehensive overview of the MSAA Science assessment, highlighted its key features, and explained its accessibility features. Moreover, the purpose of the standard setting was clearly articulated to set the context for the workshop.

Following the introduction, postsecondary outcomes and their direct correlation to performance levels and cut scores were introduced to participants. The link between these elements underlined the importance of establishing meaningful cut scores that align with students' readiness for inclusive college programs and integrated workforce opportunities.

## Day 1 Breakout Sessions

Following the plenary session, standard setting participants transitioned into grade-specific breakout sessions, where they gained familiarity with the MSAA content standards and the various levels of Expected Performance Expectations (EPEs).

During the breakout session, the room facilitator guided the participants through the progression between Low, Medium, and High EPE levels, emphasizing the importance of verb usage and the level of abstractness inherent in the assessed standards. The facilitator introduced the Cognia Standard Setting Toolkit to aid them in the item alignment process. Subsequently, facilitators demonstrated the alignment process by modeling three items, aligning them with PLD 2, 3, and 4, and provided clear and concise example descriptions of each item's required knowledge, skills, and abilities.

Cognia staff, science partners, and observers came together to discuss the progress made during day 1. They reviewed qualitative data provided by panelists to identify areas where there might be a lack of understanding regarding certain tasks. Items that received split views among panelists were identified for further discussion on day 2. For a few panelists who seemed to be experiencing stress, the group suggested seeking support from their state representatives to alleviate it. Additionally, a grade level with panelists needing a deeper understanding of Performance Level Descriptors (PLDs) was planned for an in-depth review to ensure clarity on day 2.

## Day 2 Breakout Sessions

All groups began by discussing the successes and areas for improvement identified on the first day, followed by a thorough examination of the specific items that were highlighted during the day 1 debrief session. Subsequently, they continued with the alignment process for round 1, ensuring that items were appropriately matched with their respective Performance Level Descriptors (PLDs).

By noon, grade 5 had completed round 1, and grades 8 and 11 were close to finishing round 1 item-PLD alignment. By the end of day, grade 5 had finished round 2, grade 8 completed the round 1 judgment discussion, and grade 11 was close to finishing round 2, with one panelist slightly behind schedule.

At the end of Day 1, partners, Cognia, and the observer collectively reviewed round 2 results from grade 5 and 11, and round 1 result for grade 8, focusing on items with split views or demonstrating all three PLD alignments (i.e., "rainbow items"). Strategies to mitigate pacing differences among the three grades were identified for grades 8 and 11.

## Day 3 Breakout Sessions

On the morning of day three, grade 5 and 11 proceeded with their round 3 review, while group 8 engaged in the round 2 judgment process. By around 10 am, grade 8 had completed the round 2 judgment. Partners, observers, and Cognia reconvened to review grade 8's round 2 results and identify items that needed further discussion. Before noon, all grades had completed their round 3 judgment. Facilitators expressed gratitude to all the panelists for their hard work and dedication during the past three days.

## Vertical Articulation

For the vertical articulation panel, two panelists were selected from each of the three grade-specific standard setting panels. The vertical articulation panel convened during the second half of day 3, and the panel was joined by relevant MSAA partners, MSAA TAC observer, and Cognia staff.

The articulation meeting process started with introductions, an overview of the articulation process, and articulation-specific training. The articulation facilitator explained differences between the standard setting and articulation goals/procedures. Next, panelists were asked to share their experiences related to the standard setting meeting. Following the feedback session, panelists delved into Performance Level Descriptors (PLDs) across all grades. They examined the expectations for students moving across the grades (5, 8, and high school) and considered the knowledge, skills, and abilities required to transition from a lower to higher PLD level (e.g., from Level 1 to Level 2). In addition, panelists considered how the standards change across grades.

2

Following the review and discussion of the PLDs, specific items based on the round 3 standard setting results from each grade were reviewed and discussed. Next, the group reviewed the percentage of students in each performance level across the three grades (based on the cut scores from the standard setting meeting) and discussed if it aligned with the expectations based on their earlier review and discussion of the PLDs. As 5 pm approached, the facilitator adjourned the meeting, thanking the panelists for their participation.

# Results

## Standard Setting Results

Table 1 and Exhibit 1 contain the standard setting panelist-recommended cut scores and associated impact data across grades 5, 8, and high school for MSAA Science. The percentage of students who participated in MSAA Science in each grade is shown for each performance level (i.e., Level 1, Level 2, Level 3, and Level 4) and the combined Level 3 and Level 4 performance levels. Exhibit 1 gives a visual representation of the same impact data.

**Table 1. Cut Scores and Impact Data from the AY23 MSAA Science Standard Setting**

| Performance Level | Grade 5 | | Grade 8 | | High School | |
|---|---|---|---|---|---|---|
| | Theta | % Students | Theta | % Students | Theta | % Students |
| Level 1 | -- | 46.02 | -- | 49.95 | -- | 46.83 |
| Level 2 | -0.310 | 10.37 | -0.370 | 10.77 | -0.060 | 16.29 |
| Level 3 | -0.060 | 20.34 | -0.074 | 23.76 | 0.108 | 24.07 |
| Level 4 | 0.508 | 23.26 | 1.059 | 15.52 | 0.926 | 12.81 |
| Level 3 + 4 | -- | 43.61 | -- | 39.28 | -- | 36.89 |

**Exhibit 1. MSAA Science Impact Data across Grades based on Standard Setting Cut Scores**
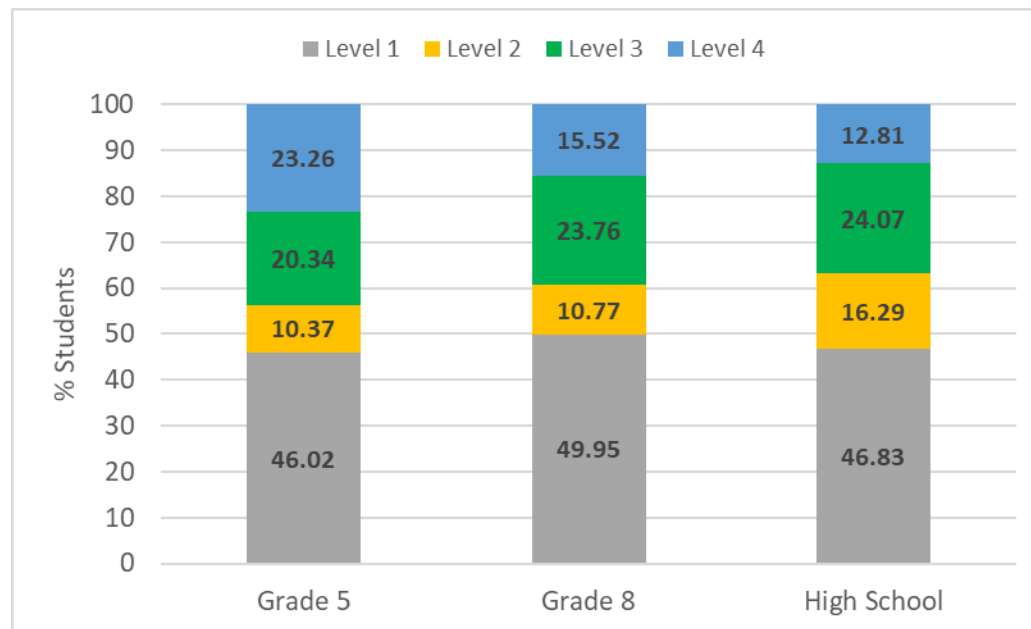


At the conclusion of the standard setting meeting, all panelists completed a workshop evaluation survey and gave their feedback regarding the workshop processes. In the final workshop evaluation, panelists expressed generally positive support for the workshop overall, workshop facilitation, training, practice,

and the workshop process. More detailed results will be presented in the Standard Setting Technical Report.

## Vertical Articulation Results

During the vertical articulation, two panelists from each grade were asked to share their experiences with standard setting. The feedback was positive, participants appreciated the thoughtfulness that went into the assessment development to provide meaningful scores for students and families.

Following the feedback session, panelists delved into Performance Level Descriptors (PLDs) and the discussion revealed general trends in each grade's PLDs. The vertical articulation panel agreed on the following:

- **Progression across performance levels:** Panelists agreed that there were "repeating patterns for verbs across grades" based on performance level. For example, one panelist noted that "level 2 largely uses the verb 'identify' across grades" (Grade 8 panelist) and the rest of the panel agreed. Building on this conversation, other panelists noted that the verbiage generally shifted from 'identify' to 'compare' to 'inference' when moving across Levels 2, 3, and 4 within each grade. Additionally, they agreed that skills required progress in complexity.
- **Progression across grades:** While similar verbs showed up in the different PLDs across grades, the context changed. Ranging from grade 5 through 8 and then high school, the context "moves further away from students' everyday experiences" (Grade 5 panelist). The panel agreed that the context was more concrete in Grade 5 and became less so (i.e., more abstract) with each higher grade. In terms of skill across grades, Grade 5 emphasized lower-level skills like recall, while grade 8 shifted to thinking of evidence and potential outcomes, and high school PLDs involved planning and investigating.
- **Progression through performance level with respect to different grades:** Panelists agreed that Level 2 required a similar amount of effort from students across the different grades, but that the required effort to progress to the next level (i.e., from Level 2 to 3, and from Level 3 to 4) became "steeper" within each higher grade. One panelist summarized the conversation by noting that "elementary [grades] was a more gentle gradient, and then each grade becomes steeper as you move from 2 to 3 to 4" (Grade 8 panelist).

Following the PLD discussion, specific items from each grade were reviewed and discussed. The discussion again highlighted similar verbs across grades, with varying content complexity. Panelists agreed to the following specific expectations regarding student achievement at each performance level across grades:

- Level 2: Similar effort required to reach Level 2 across all three grades, thus the expectation is that student achievement would be fairly consistent at this level across the three grades.
- Level 3 and Level 4: Increased effort needed to move from Level 3 to 4 within each grade, but also across grades.

After reaching consensus about the expectations for student achievement, the group reviewed the percentage of students in each performance level across the three grades based on the cut scores that resulted from the standard setting meeting. Panelists expressed surprise at the higher percentage of students classified in performance level 1 and the lower and more inconsistent percentages in performance level 2.

After the panelists left the room, Cognia, partners, and observers held a debrief session. The consensus was that the panelists' insights were valuable, and there should be a lower percentage of students classified as Level 1 in all three grades.

# Considerations for Policy Review

MSAA partners can accept the standard setting cut scores and adopt them as is. Or the partners may choose to make "policy adjustments" to the standard setting cut scores.

## Proposed Adjustment for Level 2 Cut Scores

Following the vertical articulation panel's recommendations and consensus from the MSAA Science partners, the Level 2 cut scores for all three grades were lowered, while Level 3 and Level 4 cut scores remained unchanged. The proposed adjustments for Level 2 cuts were informed by several factors:

1. The adjustments aimed to better align Level 2 cut scores with student performance expectations in the PLDs based on qualitative data from articulation panelists.
2. The adjusted Level 2 cut scores ensure improved alignment with the provisional cut from 2022, maintaining consistent percentages of students in level 1 across years.
3. Additional psychometric considerations led to lowering the Level 2 cuts by around one raw score point. For instance, in grade 5, the cut score was adjusted to align with a raw score of 15 instead of a raw score of 16. Raw score comparisons are available in Table 2 with the adjusted cut scores highlighted for easy reference.

**Table 2. Comparison of Raw Scores between Committee Recommended and Proposed Adjustment**

| Performance Level Cut | AY23 Standard Setting Recommendations | | Proposed Adjustments | |
|---|---|---|---|---|
| | Raw Score | Theta | Raw Score | Theta |
| **Grade 5** | | | | |
| Level 2 | 16 | -0.310 | 15 | -0.398 |
| Level 3 | 18 | -0.060 | 18 | -0.060 |
| Level 4 | 23 | 0.508 | 24 | 0.508 |
| **Grade 8** | | | | |
| Level 2 | 15 | -0.370 | 14 | -0.395 |
| Level 3 | 17 | -0.074 | 17 | -0.074 |
| Level 4 | 24 | 1.060 | 24 | 1.060 |
| **Grade 11** | | | | |
| Level 2 | 14 | -0.060 | 13 | -0.173 |
| Level 3 | 16 | 0.108 | 16 | 0.108 |
| Level 4 | 21 | 0.926 | 21 | 0.926 |

Table 3 displays the theta cut scores and associated impact data for each grade based on the provisional cut scores from 2022, the AY23 standard setting recommended cut scores, and the proposed adjusted cut scores. As discussed above, only the Level 2 cut scores (indicated by highlighted cells) were adjusted while the Level 2 and Level 4 cut scores remain unchanged.

Exhibit 2 gives a visual representation of the impact data across grades based on the proposed adjusted cut scores. Additionally, Exhibits 3 – 5 visually represent comparisons of impact data between the provisional, standard setting recommendation, and proposed adjusted cut scores for each grade, respectively.

**Table 3. Cut Scores and Impact Data based on Provisional, Standard Setting, and Adjusted Cut Scores**

| Performance Level | Provisional Cuts on AY23 Student Data | | AY23 Standard Setting Recommendations | | Proposed Adjustments | |
|---|---|---|---|---|---|---|
| | Cut Score (Theta) | % Students | Cut Score (Theta) | % Students | Cut Score (Theta) | % Students |
| **Grade 5** | | | | | | |
| Level 1 | -- | 39.48 | -- | 46.02 | -- | 39.48 |
| Level 2 | -0.385 | 16.92 | -0.310 | 10.37 | -0.398 | 16.92 |
| Level 3 | -0.017 | 29.00 | -0.060 | 20.34 | -0.060 | 20.34 |
| Level 4 | 1.018 | 14.60 | 0.508 | 23.26 | 0.508 | 23.26 |
| Level 3 + 4 | -- | 43.61 | -- | 43.61 | -- | 43.61 |
| **Grade 8** | | | | | | |
| Level 1 | | 42.98 | -- | 49.95 | -- | 42.98 |
| Level 2 | -0.459 | 22.81 | -0.37 | 10.77 | -0.395 | 17.74 |
| Level 3 | 0.099 | 12.14 | -0.074 | 23.76 | -0.074 | 23.76 |
| Level 4 | 0.705 | 22.07 | 1.060 | 15.52 | 1.060 | 15.52 |
| Level 3 + 4 | -- | 34.21 | -- | 39.28 | -- | 39.28 |
| **High School** | | | | | | |
| Level 1 | -- | 46.83 | -- | 46.83 | -- | 39.28 |
| Level 2 | -0.070 | 16.29 | -0.060 | 16.29 | -0.173 | 23.83 |
| Level 3 | 0.137 | 22.04 | 0.108 | 24.07 | 0.108 | 24.07 |
| Level 4 | 0.799 | 14.85 | 0.926 | 12.81 | 0.926 | 12.81 |
| Level 3 + 4 | -- | 36.89 | -- | 36.89 | -- | 36.89 |

**Exhibit 2. MSAA Science Impact Data across Grades based on Adjusted Cut Scores**
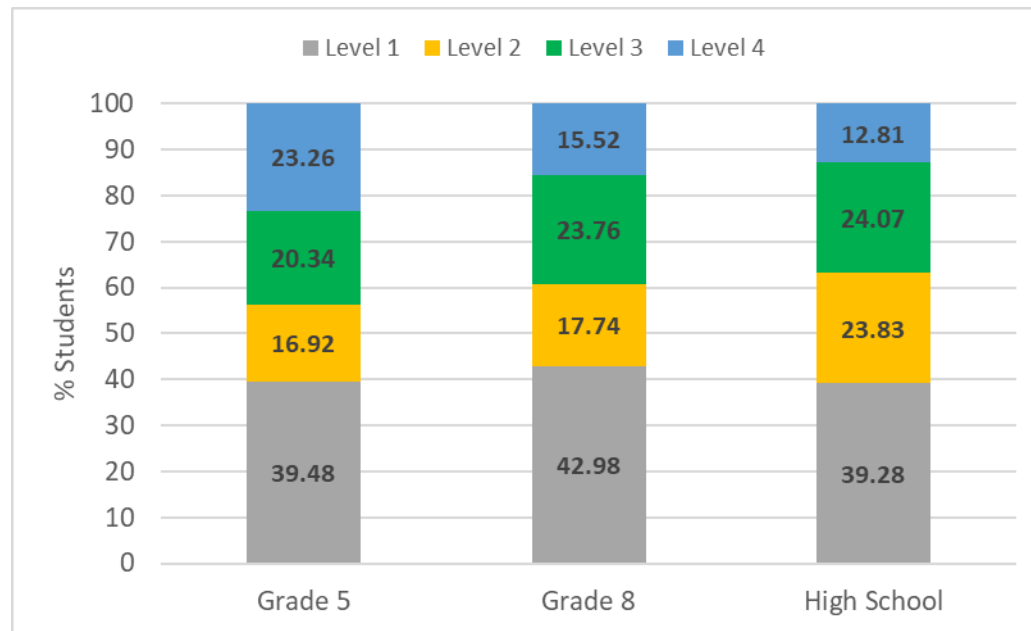
**Exhibit 3. Grade 5 Impact Data based on Provisional, Standard Setting, and Adjusted Cut Scores**



Grade 5 stacked bar chart — MSAA Science - Grade 5, % Students

| Category | Level 1 | Level 2 | Level 3 | Level 4 |
|---|---|---|---|---|
| Provisional Result | 39.48 | 16.92 | 29 | 14.6 |
| Standard Setting Recommendation | 46.02 | 10.37 | 20.34 | 23.26 |
| Proposed Adjustments | 39.48 | 16.92 | 20.34 | 23.26 |
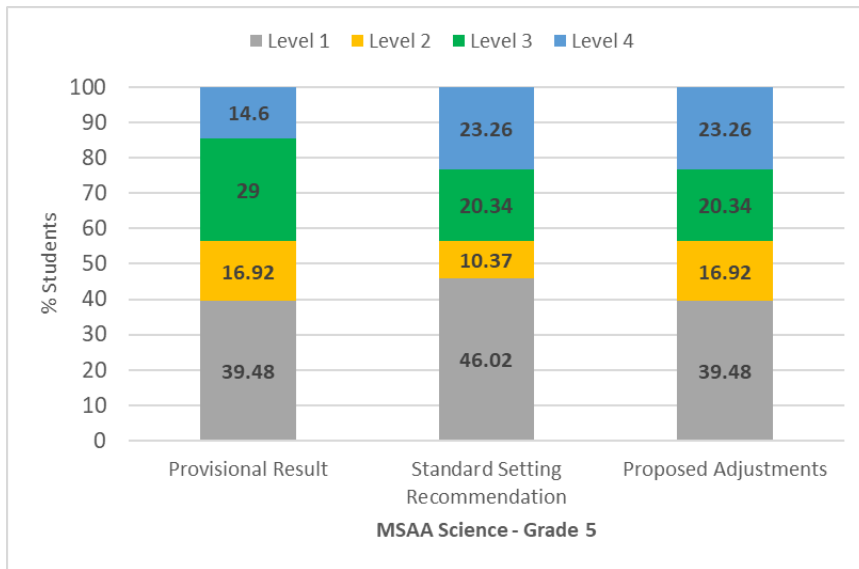
**Exhibit 4. Grade 8 Impact Data based on Provisional, Standard Setting, and Adjusted Cut Scores**



Grade 8 stacked bar chart — MSAA Science - Grade 8, % Students

| Category | Level 1 | Level 2 | Level 3 | Level 4 |
|---|---|---|---|---|
| Provisional Result | 42.98 | 22.81 | 12.14 | 22.07 |
| Standard Setting Recommendation | 49.95 | 10.77 | 23.76 | 15.52 |
| Proposed Adjustments | 42.98 | 17.74 | 23.76 | 15.52 |
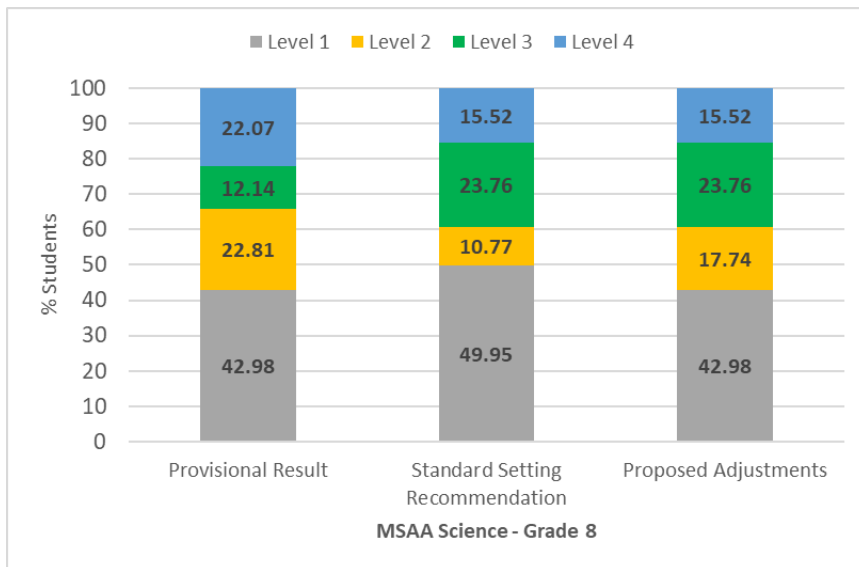
**Exhibit 5. High School Impact Data based on Provisional, Standard Setting, and Adjusted Cut Scores**
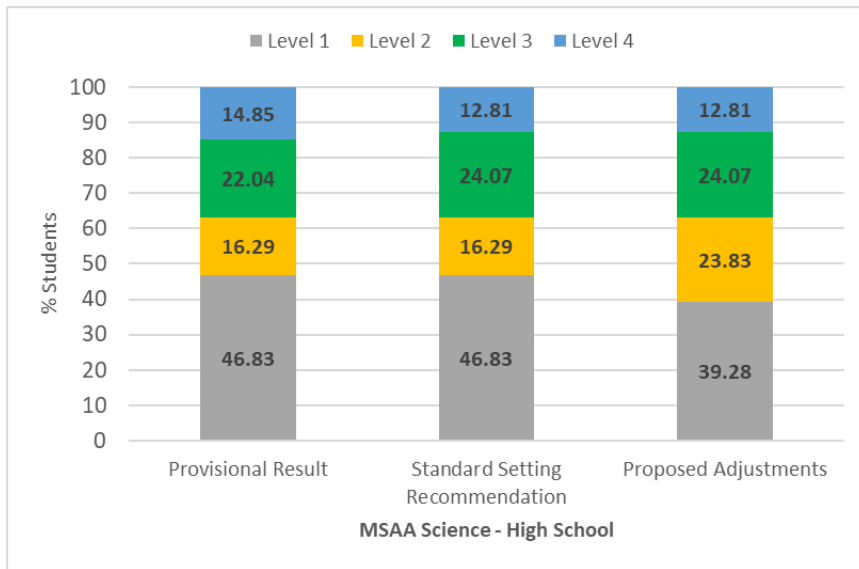


We at Cognia are grateful for the opportunity to contribute to this important conversation and will be pleased to do discuss the contents of this memo and advise MSAA Partners on psychometrically defensible ways to make policy adjustments.

# APPENDIX I
# WORKSHOP EVALUATION AND SURVEY RESULTS

# Grade 5

| | Question | Strongly Agree | Agree | Undecided | Disagree | Strongly Disagree | Not Applicable |
|---|---|---|---|---|---|---|---|
| 1 | I understood the goals of the standard setting workshop. | 5 | 2 | | | | |
| 2 | I understood the procedures we followed to recommend standards. | 6 | 1 | | | | |
| 3 | I understood that my role was to make content-based judgements about the alignment between the items and the performance level descriptors. | 6 | 1 | | | | |
| 4 | The workshop procedures made sense to me, and I learned how to apply them efficiently. | 6 | 1 | | | | |
| 5 | I am confident about my understanding of this standard setting process. | 5 | 1 | 1 | | | |
| 6 | The workshop facilitator explained things clearly to us. | 5 | 2 | | | | |
| 7 | The workshop facilitator encouraged us to raise questions and put our understandings into our own words. | 6 | 1 | | | | |
| 8 | The workshop facilitator provided clear and helpful responses to my questions and other requests for clarification. | 6 | 1 | | | | |
| 9 | The workshop facilitator took steps to help the standard setting process run smoothly. | 6 | 1 | | | | |
| 10 | Sufficient time was allotted for training and practice on the standard setting concepts, tasks, and procedures. | 6 | 1 | | | | |
| 11 | I understood the progressions in expectations across the Level 2, Level 3, and Level 4 performance levels as defined by the Performance Level Descriptors (PLDs). | 5 | 2 | | | | |
| 12 | I became sufficiently familiar with the assessment to make item-PLD judgements, based on responding to items on the test and considering the knowledge, skills, and abilities required by the items. | 6 | 1 | | | | |
| 13 | I understood the ID Matching task, including considering the knowledge, skills, and abilities required by each item, and matching those item response demands to PLDs. | 6 | 1 | | | | |
| 14 | I understood how to use the Cognia Standard Setting Toolkit to record my responses regarding skills and notes as instructed. | 6 | 1 | | | | |

| | Question | Strongly Agree | Agree | Undecided | Disagree | Strongly Disagree | Not Applicable |
|---|---|---|---|---|---|---|---|
| 15 | I understood how to use the Standard Setting Toolkit to record my item-PLD alignments. | 6 | 1 | | | | |
| 16 | I understood how to write content-based rationales for my item-PLD alignment judgements. | 4 | 2 | | | | 1 |
| 17 | I understood that the cut scores were calculated based on all item-PLD alignment judgements from all panelists. | 4 | 2 | 1 | | | |
| 18 | I understood how to use the feedback after each round, in preparation for the next round. | 6 | 1 | | | | |
| 19 | I understood what the content-based benchmarks represented. | 3 | 4 | | | | |
| 20 | I understood how to consider the content-based benchmarks in round 3, as I made my item-PLD alignment judgements. | 4 | 3 | | | | |

| Item ID | Question | Responses |
|---|---|---|
| 21 | Please indicate any parts of the standard setting training and process that you felt worked well. | I like that we were able to use a chrome book. It made things easier. I also appreciated the expertise of the facilitators and experts. Their inputs made everything so much easier |
| | | NA |
| | | Our group was amazing. Great facilitator and wonderful participants. There were awesome conversations! |
| | | Our group discussions were rich and extremely educational for understanding about the process and purpose |
| | | We had a great grade level group with thoughtful discussions and respectful and knowledgeable panelists |
| | | I think that the dialogue between the panelist made me more confident that I was learning from those that have more experience |

| Item ID | Question | Responses |
|---|---|---|
| 21 | | I was truly impressed by how at ease our group was at talking about why we chose certain levels. The discussions were so very helpful and have helped me learn how to look at the standards better. |
| 22 | Please indicate any parts of the standard setting training and process that we should improve. | This was an excellent process. As of now, I wouldn't change anything. |
| | | NA |
| | | The first day had a lot of whole group information. A lot of that information would have been better received in the small group setting. I also think that having the group together for lunch and breakfast like at IRC would be really nice. Having the span for 6:30-10 made it hard to run into other people from the workshop at breakfast. |
| | | Remember to inform panelists about stimulus/how to use all tabs available |
| | | There was frequent downtime when our group was ahead of schedule. |
| | | none that I could think of |
| | | I can't think of a thing. |
| 23 | Please note any other feedback you would like us to consider. | Thank you for allowing me to participate and be a part of this group. |
| | | This was all new to me as I had never done this before. Interesting work and I am glad I was able to be a part of it. |
| | | I really enjoy this work. Thank you for the opportunity to participate. |
| | | Great experience. Thanks so much! |
| | | I really enjoyed working with the group and am interested in participating again. |
| | | Thank you for all the wonderful feedback. It was important to take into consideration what other thoughts went into the dialogue. |
| | | Thank you for selecting me. This has been a great experience! |

# Grade 8

| | Question | Strongly Agree | Agree | Undecided | Disagree | Strongly Disagree | Not Applicable |
|---|---|---|---|---|---|---|---|
| 1 | I understood the goals of the standard setting workshop. | 4 | 3 | | | | |
| 2 | I understood the procedures we followed to recommend standards. | 4 | 3 | | | | |
| 3 | I understood that my role was to make content-based judgements about the alignment between the items and the performance level descriptors. | 6 | 1 | | | | |
| 4 | The workshop procedures made sense to me, and I learned how to apply them efficiently. | 3 | 3 | 1 | | | |
| 5 | I am confident about my understanding of this standard setting process. | 4 | 3 | | | | |
| 6 | The workshop facilitator explained things clearly to us. | 6 | 1 | | | | |
| 7 | The workshop facilitator encouraged us to raise questions and put our understandings into our own words. | 6 | 1 | | | | |
| 8 | The workshop facilitator provided clear and helpful responses to my questions and other requests for clarification. | 5 | 2 | | | | |
| 9 | The workshop facilitator took steps to help the standard setting process run smoothly. | 6 | 1 | | | | |
| 10 | Sufficient time was allotted for training and practice on the standard setting concepts, tasks, and procedures. | 1 | 4 | | 2 | | |
| 11 | I understood the progressions in expectations across the Level 2, Level 3, and Level 4 performance levels as defined by the Performance Level Descriptors (PLDs). | 5 | 2 | | | | |
| 12 | I became sufficiently familiar with the assessment to make item-PLD judgements, based on responding to items on the test and considering the knowledge, skills, and abilities required by the items. | 4 | 3 | | | | |
| 13 | I understood the ID Matching task, including considering the knowledge, skills, and abilities required by each item, and matching those item response demands to PLDs. | 3 | 4 | | | | |
| 14 | I understood how to use the Cognia Standard Setting Toolkit to record my responses regarding skills and notes as instructed. | 5 | 2 | | | | |

| | Question | Strongly Agree | Agree | Undecided | Disagree | Strongly Disagree | Not Applicable |
|---|---|---|---|---|---|---|---|
| 15 | I understood how to use the Standard Setting Toolkit to record my item-PLD alignments. | 5 | 2 | | | | |
| 16 | I understood how to write content-based rationales for my item-PLD alignment judgements. | 3 | 4 | | | | |
| 17 | I understood that the cut scores were calculated based on all item-PLD alignment judgements from all panelists. | 4 | 3 | | | | |
| 18 | I understood how to use the feedback after each round, in preparation for the next round. | 6 | 1 | | | | |
| 19 | I understood what the content-based benchmarks represented. | 5 | 1 | | | 1 | |
| 20 | I understood how to consider the content-based benchmarks in round 3, as I made my item-PLD alignment judgements. | 4 | 3 | | | | |

| Item ID | Question | Responses |
|---|---|---|
| 21 | Please indicate any parts of the standard setting training and process that you felt worked well. | The discussions with my peers helped me understand the task and the process. I feel the rounds were great as I was able to review and make corrections as needed. |
| | | I felt like the discussions following each round of review were very helpful. Hearing the ideas and thoughts of my fellow panelists was helpful in pushing my thinking forward. |
| | | -- |
| | | This is my first time attending this training and I have learned a lot from this training. I had difficulties during the first round because it was a lot of information to take in but after the discussion, I understand the process of the standard setting. |
| | | -- |
| | | Jill was a great facilitator, she ensured we all felt comfortable and a part of the group as well as that our feedback and knowledge was important to the process. The small group instruction was broken down into manageable pieces. The process was clearly outlines and practiced with the "experts" |

| Item ID | Question | Responses |
|---|---|---|
| 21 | | The practice, discussion, and review activities. |
| 22 | Please indicate any parts of the standard setting training and process that we should improve. | I feel like we should have 1/2 day to go over the introductions, directions and the expectations and practice so that on day 1 we can start the process. |
| | | I would suggest having a different format for the training. It would have been helpful to have a half day for the plenary information. A thought would be to have the welcome and overview then have a break for lunch then return for the small group introduction. It would be better to start the Rounds the next day. It would have been very helpful to have a chance to let the information filter through and start fresh on a new day. |
| | | I think we could have spent more time on our round two discussion. That felt rushed but ultimately many of the earlier items in the OIB that were not aligned became more aligned by the end of round three, so it did work itself out. |
| | | N/A |
| | | Because our Round 1 took much longer than expected, and there were technical difficulties, our Round 1 discussion felt very rushed. |
| | | The information provided at the beginning as a whole group was overwhelming, and I feel that I didn't retain the information because so much was given at one time. I think it would have been helpful to receive some of that information on Monday then be able to jump right in on Tuesday. There were many breaks due to some technological issues which didn't feel like the best use of our time and made the discussion parts have to be very short. I would have liked to have more time to discuss. |
| | | I think we should consider having a half-day Orientation and Practice Session, separate from the actual work session required of us to participate in. Day 1 was information overload for many of us. |
| 23 | Please note any other feedback you would like us to consider. | Jill's knowledge and expertise really made the process so easy and made the work we had to do go smoothly. Having a facilitator like this is great as it makes us feel confident as we were doing the work. |

| Item ID | Question | Responses |
|---|---|---|
| 23 | | The selection of snacks was great. The hotel accommodations were pleasant, and the staff was helpful. |
| | | In the training it might be helpful to point out the SEP DCI and CCC language earlier on so that all participants are on the same page regarding 3D alignment. I took notes in the plenary that helped me to understand that, and perhaps putting a column on the printed Range PLD's could assist with that. A 2 is just the DCI, a 3 is the DCI + SEP or CCC and the PLD 4 must be 3D aligned. |
| | | Me coming back to attend another training.... lol just kidding not applicable! |
| | | It would have been helpful and less overwhelming to have the plenary session in the afternoon on Monday so that we had more time for the specific training/practice and, therefore, more time to complete every portion of each round. |
| | | Thank you for a great training! |
| | | Thank you very much for allowing us to participate in this work session! This is one of the most useful work sessions I've ever participated in. The level of expertise that the members brought to the groups is just phenomenal! Learning the standard setting of the cut scores process was an amazing journey! |

# Grade HS

| | Question | Strongly Agree | Agree | Undecided | Disagree | Strongly Disagree | Not Applicable |
|---|---|---|---|---|---|---|---|
| 1 | I understood the goals of the standard setting workshop. | 6 | | | | 1 | |
| 2 | I understood the procedures we followed to recommend standards. | 6 | | | | 1 | |
| 3 | I understood that my role was to make content-based judgements about the alignment between the items and the performance level descriptors. | 4 | 2 | | | 1 | |
| 4 | The workshop procedures made sense to me, and I learned how to apply them efficiently. | 3 | 4 | | | 1 | |
| 5 | I am confident about my understanding of this standard setting process. | 3 | 4 | | | 1 | |
| 6 | The workshop facilitator explained things clearly to us. | 4 | 3 | | | 1 | |
| 7 | The workshop facilitator encouraged us to raise questions and put our understandings into our own words. | 4 | 3 | | | 1 | |
| 8 | The workshop facilitator provided clear and helpful responses to my questions and other requests for clarification. | 5 | | 1 | | 1 | |
| 9 | The workshop facilitator took steps to help the standard setting process run smoothly. | 4 | 1 | 1 | | 1 | |
| 10 | Sufficient time was allotted for training and practice on the standard setting concepts, tasks, and procedures. | 6 | 1 | | | | |
| 11 | I understood the progressions in expectations across the Level 2, Level 3, and Level 4 performance levels as defined by the Performance Level Descriptors (PLDs). | 4 | 2 | | 1 | | |
| 12 | I became sufficiently familiar with the assessment to make item-PLD judgements, based on responding to items on the test and considering the knowledge, skills, and abilities required by the items. | 4 | 2 | | | 1 | |
| 13 | I understood the ID Matching task, including considering the knowledge, skills, and abilities required by each item, and matching those item response demands to PLDs. | 4 | 2 | | | 1 | |
| 14 | I understood how to use the Cognia Standard Setting Toolkit to record my responses regarding skills and notes as instructed. | 5 | 2 | | | | |

| | Question | Strongly Agree | Agree | Undecided | Disagree | Strongly Disagree | Not Applicable |
|---|---|---|---|---|---|---|---|
| 15 | I understood how to use the Standard Setting Toolkit to record my item-PLD alignments. | 6 | 1 | | | | |
| 16 | I understood how to write content-based rationales for my item-PLD alignment judgements. | 3 | 3 | | | 1 | |
| 17 | I understood that the cut scores were calculated based on all item-PLD alignment judgements from all panelists. | 6 | | | | 1 | |
| 18 | I understood how to use the feedback after each round, in preparation for the next round. | 6 | | | | 1 | |
| 19 | I understood what the content-based benchmarks represented. | 5 | 1 | | | 1 | |
| 20 | I understood how to consider the content-based benchmarks in round 3, as I made my item-PLD alignment judgements. | 5 | 1 | | | 1 | |

| Item ID | Question | Responses |
|---|---|---|
| 21 | Please indicate any parts of the standard setting training and process that you felt worked well. | Information given and resources were helpful and easy to understand. |
| | | is understanding the PDL and how I should match the KSA. And be able to identify what are the skills students should know for each content. |
| | | I felt starting with a wide range overview, and then narrowing it down to go into more detail really helped me to understand what was expected of me. |
| | | - continuous review of previous activities completed and reinforcement of the expectations; taking the test from the beginning helped with familiarizing myself with the content |
| | | Having to go through taking the test helped a lot in evaluating the items throughout round 1, 2, and 3; including the facilitator to consistently review what we needed to do and having the discussion to ensure everyone is on the same page. |

| Item ID | Question | Responses |
|---|---|---|
| 21 | | The time allotted for us to complete our judgment was sufficient and we were able to utilize the time to complete the work. Sufficient mental breaks were allotted for us to reset and then reconvene with our work. |
| | | David really made this process enjoyable. He ensured that we understood the process and made it a fun environment. |
| 22 | Please indicate any parts of the standard setting training and process that we should improve. | none. |
| | | N/A |
| | | None. The expectations were clear. |
| | | Maybe have more people involved with this as there were only 7 of us. I am thinking that having more people providing their input will allow for better reliability |
| | | Making sure technology works. :) |
| | | I think that the KSA framework needs to be clearer. I understand that these are knowledge, skills, and abilities that the students need to have to perform well or be able to answer the question correctly. |
| | | I can't think of any. I've been to a lot of standardize testing content review and standard setting, and David was by far the best facilitator. |
| 23 | Please note any other feedback you would like us to consider. | none. |
| | | Thank you for having us. I appreciate your patience trying to help and let me understand everything about this workshop. |
| | | Ice cream every day |
| | | * Enjoyed meeting everyone from different areas * great opportunities in this area to see other places on this side of the world, especially historical sites * the sessions can be a little more interactive with more encouragement and enthusiasm |

| Item ID | Question | Responses |
|---------|----------|-----------|
| 23 | | Not a big fan of ice cream but it helps others, then more ice cream at break time. |
| | | As a science educator, I sometimes felt like my input was somewhat dismissed by the facilitator, especially during the KSA discussion for the items. I feel that in some instances when I shared the knowledge that students needed relating to content, it was not valued at some point or was not taken into consideration. I strongly believe that we must consider the background knowledge students pertaining to the content for the students to do well in assessments in addition to the skills needed and their cognitive abilities. |
| | | ICE CREAM, all day every day! Can the next location be Guam? But for real, this venue was great in terms of snacks. |

# APPENDIX J
# STANDARD SETTING MEETING AGENDA
# AND NON-DISCLOSURE FORM

# cognia®

# MSAA Science Standard Setting

## Panelist Meeting Agenda

### Day 1: Tuesday, July 18

| Time | Agenda Item | Activities |
|------|-------------|------------|
| 07:30 – 08:30 | Breakfast | Registration & Check In |
| 08:30 – 10:15 | Plenary Session: Welcome & Overview | Introductions, overview of workshop goals, MSAA science assessment; test design, standard setting, and the ID Matching method. |
| 10:15 – 10:30 | Break & transition to grade-specific breakout rooms | |
| 10:30 – 12:00 | Breakout Session: Welcome & Overview | Introductions, meeting norms, and experience the test. |
| 12:00 – 01:00 | Lunch | |
| 01:00 – 02:00 | Familiarization with the MSAA Assessment | Review performance level descriptors (PLDs) and Extended Performance Expectations (EPEs). |
| 02:00 – 03:00 | Key Concepts/Processes, Training & Practice | Training on the ID-Matching method and the ordered item booklet (OIB) |
| 03:00 – 03:15 | Break | |
| 03:15 – 04:15 | Key Concepts/Processes, Training & Practice | Practice: Facilitator models ID-Matching judgmental task; Panelists practice and discussion; Prepare for round 1: Complete readiness survey |
| 04:15 – 05:00 | Begin Round 1 | |
| 05:00 | Adjourn for the day | |

### Day 2: Wednesday, July 19

| Time | Agenda Item | Activities |
|------|-------------|------------|
| 07:30 – 08:30 | Breakfast | After breakfast, convene in grade-specific breakout rooms |
| 08:30 – 09:00 | Debrief Day 1 | Check-in on the process, challenges, etc. |
| 09:00 – 12:00 | Complete Round 1 | |
| 12:00 – 01:00 | Lunch | |
| 01:00 – 03:00 | Discussion & Preparation for Round 2 | Review and discussion of round 1 feedback/results, introduce benchmarks, and complete round 2 readiness survey. |
| 03:00 – 03:15 | Break | |
| 03:15 – 05:00 | Complete Round 2 | |
| 05:00 | Adjourn for the day | |

| Time | Agenda Item | Activities |
|---|---|---|
| **07:30 – 08:30** | Breakfast | After breakfast, convene in grade-specific breakout rooms |
| **08:30 – 09:00** | Debrief Day 2 | Check-in on the process, challenges, etc. |
| **09:30 – 10:30** | Discussion & Preparation for Round 3 | Review and discussion of round 2 feedback/results, and complete round 3 readiness survey. |
| **10:30 – 11:30** | Complete Round 3 | |
| **11:30 – 12:00** | Complete Final Evaluation Survey | |
| **12:00 – 01:00** | Adjourn* and lunch | |

*Lunch will be provided for all panelists. Panelists selected to stay for the Articulation meeting will resume at 1:30 PM

# Articulation Meeting Agenda

**Day 3: Thursday, July 20**

| Time | Agenda Item | Activities |
|---|---|---|
| **01:30 – 03:30** | Vertical articulation | Key Concepts/processes and training; Complete readiness survey; proceed with vertical articulation process |
| **03:30 – 03:45** | Break | |
| **03:45 – 04:30** | Vertical articulation | Continuation |
| **04:30 – 05:00** | Complete Articulation Evaluation Survey | |
| **05:00** | Adjourn | |

### Personal Information

* 1. Contact Information

**First Name**

**Last Name**

**Email Address**

**Phone Number**

2. Please select the State or Entity you are representing.

* 3. Are you able to attend the MSAA Science Standard Setting Meeting in person?



## 2023 MSAA Science Standard Setting Registration for Panelists

### MSAA NonDisclosure Agreement

4. Please check the box below to indicate that you have read and agree to the Nondisclosure agreement that is required prior to being able to participate in this review or access any materials:

I understand these test materials are restricted. I understand that all test questions and other materials which are considered part of the Multi-State Alternate Assessment (MSAA), including but not limited to all reading passages, graphics/images, charts, graphs, and tables, are considered secure.

Note: Additional test security measures may be required for your state.

I agree not to disclose or discuss the contents of the tests in any way that would compromise the validity of the test questions.

I understand that I may not save, print, make a copy, take a screenshot, photograph, or use any other means of duplicating any of the materials without the express permission of the facilitator or an MSAA representative. I also agree to permanently delete all files upon completion of my review.

○ Yes, I have read and agree to adhere to the Nondisclosure agreement.

○ No, I do not agree.



| 2023 MSAA Science Standard Setting Registration for Panelists |
| --- |

Overnight Accommodations

**If you are flying, your check-in and check-out date will be based on your flight arrival and departure. If you are driving, your check-in date will be 7/17 unless noted otherwise. This information will be communicated to you in the final logistics email.**

5. Special Requirements for Hotel (i.e., accessible hotel room)

[                    ]



| 2023 MSAA Science Standard Setting Registration for Panelists |
| --- |

Traveling to Meeting

6. Please select the option that applies to you:

- ◯ I am flying to this meeting.

- ◯ I am driving to this meeting and will plan to check-in Monday, 7/17.

- ◯ I am driving and need a different check-in date.

[                                    ]



| 2023 MSAA Science Standard Setting Registration for Panelists |
|---|
| Travel Accommodations |

* 7. Legal Name (as it appears on your identification)

[                                    ]

* 8. Date of Birth

D.O.B.

Date

[ MM/DD/YYYY ]  [   ]

* 9. Preferred Departure Airport

[                                    ]

10. Please enter your Known Traveler Number (KTN#).

[                  ]

11. If traveling internationally, please enter your passport number.

[                  ]

12. If traveling internationally, please enter the country where your passport was issued.

[                  ]

13. If traveling internationally, please enter your passport expiration date.

Expiration Date

Date

| MM/DD/YYYY | |

14. Special Travel Considerations

Please contact Cognia Events at events@cognia.org with any questions.

Dietary Requirements and Special Accommodations

15. Please indicate any food allergies or dietary restrictions. Please include if your allergy is airborne.

16. Please indicate any special accommodations you may need to participate in this meeting.

Please contact Cognia Events at events@cognia.org with any questions.



2023 MSAA Science Standard Setting Registration for Panelists

**Please review your responses below to ensure everything is correct.**

- **If so, please save a copy for your records and select the "Submit" button below.**
- **If not, please select the "Prev" button below to go back and select the correct responses.**

**Contact Information:** {{ Q1 }}

**State/Entity:** {{ Q2 }}

**Availability:** {{ Q3 }}

**NonDisclosure Agreement:** {{ Q4 }}

**Overnight Accommodations:** {{ Q5 }}

**Traveling to meeting:** {{ Q6 }}

**Legal Name:** {{ Q7 }}

**Date of Birth:** {{ Q8}}

**Preferred Airport:** {{ Q9 }}

**Known Traveler Number:** {{ Q10 }}

**Passport Number:** {{ Q11 }}

**Country Passport Issued:** {{ Q12 }}

**Passport Expiration:** {{ Q13 }}

**Special Travel Considerations:** {{ Q14 }}

**Food Allergies and/or Dietary Requirements:** {{ Q15 }}

**Special Accommodations:** {{ Q16 }}

Please contact Cognia Events at events@cognia.org with any questions.

# APPENDIX K
## SYNOPSIS OF PROCEDURAL VALIDITY EVIDENCE

# Synopsis of Procedural Validity Evidence for the Cutscores Resulting from Grades 5, 8, and 11 Standard Setting for MSAA Science Alternate Assessment July 18-20, 2023

Prepared by Rachel Quenemoen, Member of the MSAA Technical Advisory Committee (TAC)

Final Report August 11, 2023

# Synopsis of Procedural Validity Evidence for the Cutscores Resulting from Grades 5, 8, and 11 Standard Setting for MSAA Science Alternate Assessment July 18-20, 2023

## Table of Contents

# Synopsis of Procedural Validity Evidence for the Cutscores Resulting from Grades 5, 8, and 11 Standard Setting for MSAA Science Alternate Assessment July 18-20, 2023

## Part I: MSAA Lead, TAC, and Vendor Plan and Preparatory Procedures

The AY23 Cognia Science Alternate Assessment (CSAA)/Multi-State Alternate Assessment (MSAA) Science Standard Setting Final Plan was the culmination of several months and iterative reviews by the MSAA Partners and the MSAA Technical Advisory Committee (TAC). Preliminary standards had been set in AY22, but MSAA members and the TAC had flagged procedural issues that were compounded by the discovery of data errors in preparation for the event. To ensure a quality outcome, intensive efforts on the part of Cognia, MSAA Partners, and the MSAA TAC resulted in a quality plan and preparatory procedures in AY23. The agreed upon plan used the Item-Descriptor (ID) Matching Method with content-based benchmarks to recommend cut scores. Recommended cut scores for four achievement levels were to provide clear indicators of student progress toward MSAA grade-level science expectations, including clear indicators of student preparedness for the next grade level as well as inclusive college programs and competitive integrated employment.

**Summary validity evidence from Plan and Preparatory Procedures.**
Extensive and iterative review by MSAA Partners and MSAA TAC resulted in consensus that the plan was compliant with current standard setting practices.
**Any remaining threats to be resolved.**
None.

## Part II: Implementation of Standard Setting Panelist Procedures

MSAA Partners asked that an external observer be engaged for the 2023 standard-setting meeting. A member of the MSAA TAC was invited to do so to help the state leads identify and correct issues during implementation to ensure high quality outcomes. Fidelity of implementation was the primary driver for the observational protocol, but the focus was on:
- capturing barriers or anomalies to flag for state leads and vendor as the plan was implemented,
- documenting the nature of issues and the identified interventions, and
- observing the effectiveness of agreed upon solutions.

See Appendix A for the scope of work and general participant observation protocol. After the meeting, summary field notes were provided to MSAA partners and the vendor and reviewed by them in draft form for appropriateness and accuracy. The observer made final decisions on all edits. The Summary Field Notes in Appendix B provide additional evidence to support the conclusions reached in this Synopsis. This Synopsis summarizes the identified issues, interventions, and outcomes of the three-day standard setting meeting as part of the validity evidence for the resulting cutscores.

## Large Group Orientation

- **Fidelity of implementation:** As planned.
- **Issues that emerged**: Panels of ten members had been recruited, but during the last days prior several panelists dropped out, in part due to severe weather in some of the partner states, flights being canceled with no alternative available, and personal concerns.
- **Interventions that occurred:** Partners in nearby states were able to reach out to additional panelists who had applied, and the meeting began with seven panelists at each grade level. All but one of the partner entities was represented.
- **Outcomes:** A general overview was provided. With the orientation and training, all seven panelists in each grade level completed the work showing qualitative evidence of understanding of the task and compliance with the training. Some issues emerged among panelists, but interventions were able to resolve them, and resulted in recommendations for future panel work/training. See below.

## Panel Training

- **Fidelity of implementation:** The facilitators were trained on and given a script to carefully work through the content resources and supporting information, prior to working through the sample items. That being said, one panel was trained differently from the other two, deviating from the agreed upon script.
- **Issues that emerged:** Both MSAA leaders and the test vendor were alerted to a potential issue with shortened training in one room. Because not all panelists were volunteering evidence of their understanding, a review of panelist work was planned for day's end, conducted by specifically sampling from panelists who had been quiet and not checked for understanding directly.
- **Interventions that occurred:** During the end of day debrief with MSAA leaders and vendor staff, preliminary qualitative data were reviewed, specifically looking at quiet vs. verbal panelists in all three rooms. Each room had at least one example of minimal understanding, all related to not sufficiently using evidence from both item knowledge, skills, and abilities (KSAs) and Performance Level Descriptor (PLD) resources in their notes. Some quiet panelists clearly had good understanding, some did not. It was not dramatically different across the three rooms. Two of the facilitators noted that at least one panelist had said that they learned the most about how to do this through discussing the sample items, and hearing their peers talk about their decisions. In addition, two facilitators noted a panelist who shared that they were not confident in their decisions when asked verbally. State partners wondered whether panelists felt comfortable with saying how they felt, given all had indicated readiness to begin Round 1 on the survey provided.
- **Outcomes:** Two resolutions occurred to address the issues.
  - The first was to design a retraining plan for all three groups before continuing with the process, because evidence from the panelists across the three rooms was similar, and each room had at least 1-2 panelists for whom understanding was not clear. The facilitators agreed to emphasize the resources again in the opening session with panels, with the shortened training room planning for an additional review of materials that may have been gone over too quickly.
  - The second resolution was to delay the use of benchmarks until after Round 2, instead of after Round 1. MSAA leaders raised the issue of introducing benchmarks prior to Round 2 in light of expressed concern by some panelists that having time to learn from their peers was important and could build their

confidence in providing independent judgments. Although vendor staff suggested the plan was fixed and couldn't be changed, after further discussion they noted it would be easy to change the process (a click of a button to reorder) and the only technical argument would be that there is some evidence that panelists harden their opinions after Round 2.

On the flip side, MSAA leaders noted that there were observed cultural tendencies for quiet participation, expressed panelist concerns about not having confidence, and expressed panelists preference to learn from discussions with peers. They proposed that Round 1 discussion may result in all panelists, not just the verbally expressive ones, having confidence in their independent judgments, thus yielding better reflection of all panelist expertise on Round 2, especially if that is prior to sharing external benchmarks. MSAA leaders also asked the vendor to avoid any language when presenting the benchmarks prior to Round 3 that these judgments were "expert" or "actual" judgements but were instead simply a second set of panelists with content expertise who had also reviewed the items, and it "may" help panelists review those items that remained uncertain to them.

The vendor agreed to move the benchmarks from prior to Round 2 until after Round 2 and to use careful language to respect the work done by panelists, representing the benchmarks as one more resource to consult if panelists are struggling with any final determinations on individual items.

**Rounds 1, 2, 3**
- **Fidelity of implementation:** Adjustments due to the training issues above resulted in retraining during Round 1 or prior to Round 2, depending on the group. The introduction of benchmarks was also delayed from prior to Round 2 to prior to Round 3. With those agreed to adjustments at the end of the first day, implementation of the three rounds occurred within standard procedures.
- **Issues that emerged:** Two panelists in two different groups continued to work very slowly on Rounds 1 and 2, threatening timelines.
- **Interventions that occurred:** Additional time was provided for the two panelists, since review of their work in progress indicated increased understanding after retraining, but a continued slow pace. They were also provided encouragement and support from their state lead. Their state lead verified that both speak English as a second language, and that cultural expectations would encourage them to be very quiet and reserved, possibly not asking for the support they need in this novel setting. One stayed two hours later than other panelists on day 2. Both came an hour early on day 3, demonstrating commitment to the tasks in the face of the challenges. Both finished their tasks successfully within time available on day 3.
- **Outcomes**: In the future, it is important to work among state leads and with the vendors to ensure that cultural, language, and disability preferences and support are planned into the protocols for training and monitoring panelists. Cross state/entity work inevitably brings differences among the panelists, which are strengths that must be part of systematic planning and training. For example, group discussion and concrete examples made a major difference in understanding for the two panelists who have English as a second language and who are culturally reserved. Round 3 included introductions of the benchmark data, at a point where all panelists were comfortably engaged. Based on the preliminary data review, the panels' work was very consistent with the benchmark data,

and thus the benchmarks were treated as simply another data point to consider as they did their final ratings.

Each panel proceeded quickly with Round 3 and completed in a timely way. Because of the congruence with benchmarks, evidence from qualitative entries indicating appropriate use of the resources to support judgements, and general appropriate appearance of the preliminary cuts, no validation step was needed, which was agreed to by the vendor and MSAA partners.

**Summary validity evidence from Implementation of Standard Setting Panelist Procedures.**
Although the procedures needed adjusting as issues arose, the MSAA partners and the vendor worked effectively to define issues, identify resolutions, and implement them successfully. Implementation was compliant with current standard-setting practices.
**Any remaining threats to be resolved.** None.

# Part III: Cross Grade Articulation
## Fidelity of implementation

Two panelists from each grade level remained after Round 3 for articulation. The facilitator script and agendas that were provided to the external observer in advance did not reflect the extensive content-based orientation that was actually used. Instead, the advance agenda reflected a brief orientation and a more traditional data-based approach, providing preliminary cuts and asking whether adjustments should be made.

**Issues that emerged:** Given this standard-setting method was driven by content, it seems likely that the articulation panel could better conceptualize building on a content-focused model. In the opinion of the observer, given that the revised approach seemed a better entry point for the panelists, no comments or questions were made by the observer as it played out, but on alert for any signs the panelists struggled with the tasks.

The vendors had the panelists go through all three grade levels of content resources comparing and contrasting what they saw, in a facilitated conversation that drew out all six panelists. The panelists used all the content language that they had been trained on and eloquently built common understanding of within and across grade content. Their discussion and interactions were good evidence that the panelists were well trained and focused on the tasks they were given.

Then, together the six were facilitated by the vendor to build a conceptual model of what they would expect to see in terms of performance within and across grades, at a high level, including discussion of whether as the grades progressed if it became harder or easier to move up the levels. The vendor report will show this conceptual model, along with the rationale the panelists gave for why the model looked as it did. That documentation should be studied to possibly point to an effective content-focused methodology for articulation in content-driven standard-setting methodologies.

At that point, the vendor switched to the more traditional model, showing cuts on the content bar graphs, then switching to a visual that had lines for the preliminary cuts and impact data. The panelists showed confusion and concern and began to seemingly arbitrarily suggest that cuts be moved, including expressing concern that too many kids were in the lowest level. MSAA partners were observing and showing concern as the process lost focus. One of the panelists

raised the concern that the group had lost their grounding in the goals of the work, and the model they had worked on, and at that point, the lead vendor facilitator stopped the process and asked panelists to step out briefly so the MSAA partners and vendor could discuss next steps.

**Interventions that occurred:** Once panelists left, the vendor noted to the MSAA Partners that in order to bring the process full circle as he had intended - starting with content and building a conceptual model, moving to adjusting cuts, and then returning to content - would take more time than they had, since panelists were scheduled to leave. In turn, the MSAA partners noted the high quality of content work that had been done in the first segment, and the thoughtful insights the six panelists had shared.  There was a collective sense that the conceptual model the panelists had built as part of the initial segment gave MSAA leaders the direction they needed to continue the work in a policy meeting. In addition, the quality of the preliminary cuts coming out of the initial data indicated a pattern very close to the model the panelists built already, so MSAA partners anticipated very few if any policy adjustments would be needed.

With that, panelists were recalled to the room, told that the MSAA Partners were very impressed with their work on the model, and that they would be able to leave as scheduled since the model provided the guidance needed to finalize the cuts at the policy level.

**Outcomes:** The content-focused process used in this standard-setting was judged as successful in the discussion that followed the panelists' departures, as a consensus opinion of the MSAA partners and the vendor. Rethinking the articulation process and redesigning it to maintain the focus on the content in articulation processes can contribute to the field as they move forward in this methodology.

## Summary validity evidence from the Cross Grade Articulation.

MSAA Partners, vendor, and observer agreed that the model and qualitative feedback provided by the panelists gave policymakers a clear pathway to finalize cuts that strongly reflects the voice of the panelists. No more time was available for state policy review and final articulation of the results, which is in process at the writing of this report.

**Any remaining threats to be resolved.** None, pending the policy adjustments made by the MSAA Partners.

**Additional reports and data to contribute to the overall validity argument.**
Post meeting procedures and reports, including review of preliminary data sets and final cuts procedures will be reviewed through established MSAA Partner and MSAA TAC procedures.

The vendor standard setting report will be reviewed through established MSAA Partner and MSAA TAC procedures.

# APPENDIX A

**MSAA Science Assessment Standard Setting: Observational Protocol and Scope of Work
Rachel Quenemoen, July 2023**

Pre-meeting summary of activity, including very brief reference to initial standard setting, TAC review and recommendations resulting in second meeting, extensive review and discussion from 2-3 TAC meetings, brief overview of proposed plans and consistency with established practices, documenting that the entire TAC and MSAA Partners have reviewed and approved the plan.

Panel observations will be focused on:
Fidelity of implementation of agreed to standard-setting plan;
Identification of any substantive issues raised by panelists, vendor, or partners, to share with partners who can address them in end of day meetings or as necessary;
Document adjusted plans and rationale for each based-on end of day partner/vendor discussions;
Cross-grade articulation process and summary.

Similar to panelist process, note consistency of cross-grade articulation process with current practices and planning, flagging any discrepancy to partners as needed. If necessary, document any implications for policy review post-meeting.

Final report with overall summary of meeting, high level, including a judgement on whether there are any procedural reasons to question the validity of the resultant cutscores produced by these panels. Includes a bullet summary of evidence that may be used in validity argument, as basis for discussion at future TAC meeting and eventual inclusion in Peer Review submission/technical report.

POST MEETING: High level review of Cognia standard setting technical report (in consultation with other TAC members), which is outside of the observational protocol in the role of one of five TAC members.

## APPENDIX B

**Observational Evidence from the Grades 5, 8, and 11 Standard Setting for MSAA Science Alternate Assessment July 18-20, 2023, for Partner Documentation and to Support Validity Evidence Synopsis**

**Prepared by Rachel Quenemoen, member of the MSAA Technical Advisory Committee (TAC)**


## DAY ONE Field Notes

DAY ONE OPENING SESSION: Opening session was an informational fire hose, but got people assembled and possibly is typical. There are alternatives that can be explored for future meetings to make the session more engaging or even somewhat participatory, but nothing was observed that would affect validity of the end results. For Peer Review submission, partners may find it useful to note that it included an overview of WIOA requirements for instruction and assessment, with emphasis on post-secondary and competitive employment opportunities as context for their work.

PANEL MEMBERSHIP: Introductions included membership is from every entity except for USVI. Self-introductions in the panels reflected a mix of science and special education backgrounds, some citing both. See vendor standard setting report for more information. Panels of ten members had been recruited, but during the last days prior, several panelists dropped out, in part due to severe weather in some of the partner states, flights being canceled with no alternative available, and personal concerns. Partners in nearby states were able to reach out to additional panelists who had applied, and the meeting began with seven panelists at each grade level. All seven panelists in each grade level completed the work showing qualitative evidence of understanding of the task and compliance with the training. Some issues emerged among panelists, but interventions were able to resolve them, and resulted in recommendations for future panel work/training.

CONVENING OF PANELS: As panels convened and began work, there were glitches, mostly technology problems, that were annoying but resolved. These glitches affected taking student test, but in the end, didn't prevent all panelists from completing the task within allotted time. There were appropriate numbers and skills of vendor staff to intervene on the problems, there were experienced, trained facilitators for the three panels. There was a detailed facilitator script, which two of three appeared to follow consistently, with minor variations. The one facilitator who did not was within professional boundaries, but it did result in a different training experience for one panel. See Issue discussion and resolution below.

All three panel rooms felt respectful and safe overall. There was a marked tendency for similar panelists to sit together, and some cultural differences observed in the way they interacted (see issue discussion below) with the primary effect that not all panelist comprehension or understanding was extroverted and observable. One facilitator seemed skilled at drawing all out, others relied on volunteers to show understanding to queries, which meant that some panelists remained quiet throughout. A check with MSAA lead observers found consensus that facilitators knew how to facilitate but may not have utilized expected instructional practices to identify learning needs and encourage an inclusive environment.

TRAINING ISSUE: One substantive issue emerged in the inconsistent training of panels. That is, the orientation to the resources for science content standards (EPEs) and performance level descriptors, training on use of the materials, sample items done as a model, as a group, and independently, was schedule to run from 1-4 pm, with one break, or about 2.75 hours of actual work prior to quiz beginning Round 1. One facilitator/room completed the training by 2:15, with no break, or 1.25 hours of actual work prior to quiz beginning Round 1. Most of the panelists in that shortened training panel appeared to be tracking well and in fact one panelist essentially trained other panelists on use of the technology interface, so some of the pace may have been set by the group, but the facilitator script was not implemented. Two panel members sitting together were quiet throughout and gave no clear indication of either understanding or confusion, just were not reacting visibly.

On observation in another panel where training continued on track with the facilitator script, there was also one table with two members who were quiet throughout and gave no clear visible indications of status. In that panel, the facilitator systematically called on and drew out those panelists, engaging them in the process and checking on understanding. In the third room, which was observed during the opening orientation and student test experience but not during training on the resources, there was also one table with two very quiet panelists, who were not drawn out in other observed sessions in that room but may have been during the resource training.

The facilitators were trained on and given a script to carefully work through the content resources, including the EPEs and the PLDs and supporting information, prior to working through the sample items. One can argue that lengthy abstract training on the resources prior to actually working through sample items and application of the judgment procedure may not be the best use of time for all panelists. That is, it may be that the "shortened" version of the training was adequate to prepare the panelists for the work, but that could not be assured based on observations in the room.

That being said, one panel was trained differently from the other two in this case, deviating from the agreed upon script. Both MSAA leaders and the test vendor were alerted to a potential issue with shortened training in one room. Because not all panelists were volunteering evidence of their understanding, a review of panelist work was planned for day's end, conducted by specifically sampling from panelists who had been quiet and not checked for understanding. Panelists were also required to complete a readiness survey. All indicated that they were ready to conduct the matching. However, one panelist later indicated they were not confident (as described below).

QUALITATIVE DATA REVIEW END OF DAY: During the end of day debrief with MSAA leaders and vendor staff, preliminary qualitative data were reviewed, specifically looking at quiet vs. verbal panelists in all three rooms. Each room had at least one example of minimal understanding, all related to not sufficiently using evidence from both item KSAs and PLD resources in their notes. Some quiet panelists clearly had good understanding, some did not. It was not dramatically different across the three rooms.

ADDITIONAL END OF DAY QUALITATIVE DATA FROM FACILITATORS: Two of the facilitators noted that at least one panelist had said that they learned the most about how to do this through discussing the sample items, and hearing their peers talk about their decisions. In addition, two facilitators noted a panelist who shared that they were not confident in their

decisions when asked verbally. State partners wondered whether panelists felt comfortable with saying how they felt – day 2 retraining suggests perhaps not.

RESOLUTION - RETRAINING PLAN: The evidence from the panelists across the three rooms was similar, and each room had at least 1-2 panelists where their understanding wasn't clear. The facilitators agreed to emphasize the resources again in the opening session with panels, with the shortened training room planning for a review of materials that may have been gone over too quickly.

RESOLUTION - DELAY OF BENCHMARK USE UNTIL AFTER ROUND 2

MSAA leaders raised the issue of introducing benchmarks prior to Round 2 in light of expressed concern by some panelists that having time to learn from their peers was important and could build their confidence in providing independent judgments. Although vendor staff suggested the plan was fixed and couldn't be changed, after discussion they noted it would be easy to change the process (a click of a button to reorder) and the only technical argument would be that there is some evidence that panelists harden their opinions after Round 2.

On the flip side, MSAA leaders noted that there were observed cultural tendencies for quiet participation, expressed panelist concerns about not having confidence, and expressed panelists preference to learn from discussions with peers. They proposed that Round 1 discussion may result in all panelists, not just the verbally expressive ones, having confidence in their independent judgments, thus yielding better reflection of all panelist expertise on Round 2, especially if that is prior to sharing external benchmarks. MSAA leaders also asked the vendor to avoid any language when presenting the benchmarks prior to Round 3 that these judgments were "expert" or "actual" judgements but were instead simply a second set of panelists with content expertise who had also reviewed the items, and it "may" help panelists review those items that remained uncertain to them.

The vendor agreed to move the benchmarks from prior to Round 2 until after Round 2 and to use careful language to respect the work done by panelists and representing the benchmarks as one more resource to consult if panelists are struggling with any final determinations on individual items.

IMPLEMENTATION OF RETRAINING PLAN PRIOR TO START OF DAY 2: The plan for the retraining in the shortened training panel was shared with MSAA leaders in the room and one of state leaders agreed to observe and provide additional feedback from grade 8 to the external observer.
Because the facilitator was receptive to the feedback, the external observer continued sampling observations in each room during similar activities the next day, rather than focusing solely on the room with the shortened training period.

## DAY TWO Field Notes

RETRAINING OBSERVATIONS: Given the concerns about training differences across panels, and the observations that very quiet panelists in each room may or may not have full understanding, each of the facilitators began their sessions with a review of the resources and queries of the members. In grade 8, the facilitator spent additional time to better ensure all panelists were prepared. The external observer focused on grades 5 and 11 per observation

protocol and one of the state leaders was in grade 8 and represented the observer. Her notes are below, after grade 5 summary sentences.

GRADE 5 RETRAINING:  The quiet table in grade 5 participated actively, with all panelists engaged and a full group discussion. Panelists appeared comfortable and ready to proceed.

STATE LEAD OBSERVATION OF GRADE 8 RETRAINING:
Facilitator opened with an exercise where panelists could anonymously provide feedback on their current comfort level/understanding of the concepts. Of the 7 panelists, two reported somewhat comfortable and one reported discomfort.  (Deep end of the pool). The majority of panelists reported middle of the range.

The facilitator then asked panelists to jot down on post-its "positives and negatives" of the experience so far. One Partner Observer suggested reframing "negatives" as "opportunities to learn." The panelists posted their thoughts and were then asked to select and share another person's Post-it with the larger group. The comments recorded by panelists are included below in bullet points.

The facilitator then moved panelists through a revisiting of the sample items.  She shared a KSA written in more of a "student objective" format and compared this to a KSA for the same sample item which included greater detail and language lifted from the corresponding PLD. Facilitator emphasized the incorporation of PLD language into the KSA/notes as opposed to writing it as what a student should know. The facilitator shared similar examples for the additional two sample items.  The facilitator emphasized importance of taking time and also encouraged panelists who are moving more quickly to revisit their judgments on earlier items.

The facilitator then repeated opening activity. All panelists now reported being in the comfortable-neutral range with the concepts/work. One panelist shared that knowing to include PLD language in KSA section was helpful and represented "a missing piece." Another panelist reported it was a helpful suggestion to go back and review earlier judgments from the overview page of the toolkit as time allowed. Panelists moved into Round 1.

Additional observations by state lead:
Language is a cultural force and it could be helpful to establish language norms with facilitators/partners for future, similar work involving panelists. One panelist expressed slight disbelief at the word "expert" (inferring this meant self-doubt in expertise. Does this word sometimes serve to intimidate rather than empower?) Additionally, statements such as "We want you to do good work" could be interpreted as, we were not doing 'good work' before.  In fact, all panelists are putting forth a best-faith effort and therefore producing 'good work.' We should avoid the implication that speed or other factors could in fact result in panelists not producing 'good work.' Just a thought!

**Post-its, Positives and Opportunities shared by Grade 8 Panelists during retraining:**

Positives
Appreciated Rachel's topic – who we are doing this work for
Work is important, I enjoy it
Team is supportive, comfortable safe
Interaction with peers
Aligning standards to items

Learning importance of alignment of items to PLDs
Learning process of SS

Opportunities
Doubt myself
Not sure I'm doing this correctly
Overwhelming amount of training
A lot of info, short amount of time
Writing KSAs, determining PLDs
Hope state sticks with this test for a length of time
Making sure there is rigor without being too challenging

GRADE 11 RETRAINING: In grade 11, all panelists seem engaged, not all are speaking, one panelist who was working slowly is carefully considering discussion, looking at resources, nodding. She and one other panelist remain quiet.

ROUND 1 CONTINUES:
Grade 5 finishes quickly and moves on to round 1 discussion, which involves all panelists, engaged and positive.

Grade 8 has one panelist unable to complete task as of 2:15 and the delay is threatening the ability to complete the tasks by midday on the third day. After conferring with Cognia, with recommendations to move things along, state lead reps who have been observing grade 8 decide:
1. Panelist will stop her work, which evidences random answers as stress/pressure mounts. Given it is in the middle of the day, having other panelists wait has added to her stress. She should benefit from the discussion on round 1, get additional modeling and insights. She will be encouraged to go back to any and all items as part of round 2.
2. Although her data will be reflected in the bar graphs showing responses, they will not be included in preliminary cuts based on round 1.
3. The goal is to help her catch up during round 2, benefitting from the debriefing to make the work quality go up, and be sure she knows that she can stay late or begin early tomorrow if needed.
4. State leads/observer reached out to that panelist's state lead to encourage her to reach out to support her, ensure that she makes use of any available extra time. Assuming she completes round 2 successfully, her data will be added in again at that point.
5. If she cannot complete the task or has outlying data that is of concern, worst case scenario is that there may be a need to run data without her, with some type of expert review for cuts post meeting.
6. The panelist's state lead sat in on the round 1 debrief at that point.

Grade 11: In grade 11, another panelist has been slow - if she needs more time to finish, we could propose their state lead working with both slower panelists to get them to take extra time to finish.

TECHNOLOGY DELAY: Another major technology glitch affected that grade 8 room, delaying round 1 review for an hour. The slow panelist had closed her notebook so the tech team had migrated the data, but then she began re-entering information. That resulted in two versions that were different, and in order to merge them, it required three verifications of the data, which took time. The grade 8 panelists ended up playing an online team building game in the interim and

seemed to enjoy it. When they commenced round 1 debrief, the room was relaxed, seemingly happy, and the slow-paced panelist was integrated into the group. With three states and observer in the room observing round 1 debrief, all agreed that there was active, relaxed participation, with what appeared like an a-ha moment for the struggling panelist. At the end, they were able to finish review at end of day and start round 2 in morning, in hopes of finishing both round 2 and 3 by noon.

OBSERVED TECHNIQUES FOR DRAWING OUT PANELISTS: In grade 11, the facilitator used one confident, verbal panelist sitting next to the quiet one struggling with pace of work to model each step of the round 1 report out. The panelist appeared to benefit from the close proximity of her seat mate working through the process and the group discussion. In grade 8, the facilitator made use of a vendor content expert standing adjacent to the quiet panelist to ask him to work through one of the judgments. He asked the quiet panelist to share her screen and resources, which meant the panelist was able to work with him one-on-one without being singled out. Rather, the panelist was in the role of helping the content expert when he didn't have the resources he needed. It appeared comfortable and non-threatening, and ultimately successful.

END OF DAY 2 DEBRIEF: At the vendor MSAA partner debrief, all agreed to move forward with backup plan for use of data should the two panelists continue to struggle. In grade 11, the quiet panelist sat another 2 hours at the end of day 2 working over her answers after she had initially entered them. Both are from the same partner - that leader will work with them to reassure them and help them prepare for day 3. The data look reasonable, with half hour early start for grade 8, plan is to plan for normal finish.

## DAY THREE Field Notes

Day 3 began in each room with panelists working on round 2. In those observations, it became clear that both slow-paced panelists do not use English as their first language. In addition, their state lead verified that cultural expectations would encourage them to be very quiet and reserved, possibly not asking for the support they need in this novel setting. One stayed two hours later than other panelists yesterday. Both came at 7:30 this morning. They are both showing an admirable commitment to doing their part.

LESSONS LEARNED: In the future, it is important to work among state leads and with the vendors to ensure that cultural, language, and disability preferences and support are planned into the protocols for training and monitoring panelists. Cross state/entity work inevitably brings differences among the panelists that must be part of systematic planning and training. Group discussion and concrete examples made a major difference in understanding for the two panelists who have English as a second language and who are culturally reserved.

Round 2 completion went quickly for the grade 5 panel, and within time parameters for grade 8 (which started half an hour early, with one panelist coming an hour early), and on time for Grade 11 (also one panelist came an hour early). The retraining, interventions, opportunities for additional group discussion and concrete examples, along with extra time appeared to bring all panelists to the point they could effectively do their work and participate in round 2 debriefs.

Round 3 included introductions of the benchmark data. Based on the preliminary data review, the panels' work was very consistent with the benchmark data, and thus the benchmarks were treated as simply another data point to consider as they did their final ratings.

Each panel proceeded quickly with round 3 and completed in a timely way. Because of the congruence with benchmarks, evidence from qualitative entries indicating appropriate use of the resources to support judgements, and general appropriate appearance of the preliminary cuts, no validation step was needed, which was agreed to by vendor and MSAA partners.

## DAY THREE ARTICULATION PANEL Field Notes

Two panelists from each grade level remained for articulation. The facilitator script and agendas that were provided to the external observer in advance did not reflect the extensive content-based orientation that was actually used. Instead, the advance agenda reflected a traditional data-based approach, providing preliminary cuts and asking whether adjustments should be made.

That change was fortunate, even though the observer was not aware in advance. Perhaps MSAA partners were aware of the change. Given this standard setting method was driven by content, it seems likely that the articulation panel should better conceptualize building on that model. In the opinion of the observer, given that the revised approach seemed a better entry point for the panelists, no comments or questions were made by the observer as it played out.

The vendors had the panelists go through all three grade levels of content resources, comparing and contrasting what they saw, in a facilitated conversation that drew out all six panelists. The panelists used all of the content language that they had been trained on and eloquently build common understanding of within and across grade content. It was impressive, and another strong piece of evidence that the panelists were well trained and focused on the tasks they were given.

Then, together the six were facilitated by the vendor to build a model of what they would expect to see in terms of performance within and across grades, at a high level, including discussion of whether as the grades progressed if it became harder or easier to move up the levels. The vendor report will show this conceptual model, along with the rationale the panelists gave for why the model looked as it did. That documentation should be studied to possibly point to an effective methodology for articulation in content driven standard setting.

At that point, the vendor switched to the more traditional model, showing cuts on the content bar graphs, then switching to a visual that had lines for the preliminary cuts and impact data. The panelists showed confusion and concern and began arbitrarily suggesting cuts be moved, including expressing concern that too many kids were in the lowest level. MSAA partners were observing and showing concern as the process lost focus. One of the panelists expressed concern that the group had lost their grounding in the goals of the work, and the model they had worked on, and at that point, the lead vendor facilitator stopped the process and asked panelists to step out briefly so the MSAA partners and vendor could discuss next steps.

Once panelists left, the vendor noted to the MSAA Partners that to bring the process full circle as he had intended - starting with content, moving to adjusting cuts, and then returning to content - would take more time than they had, since panelists were scheduled to leave. In turn, the MSAA partners noted the high quality of content work that had been done in the first segment, and the thoughtful collective insights the six panelists had shared.  There was a collective sense that the conceptual model the panelists had built as part of the initial segment gave MSAA leaders the direction they needed to continue the work in a policy meeting. In addition, the quality of the preliminary cuts coming out of the initial data indicated a pattern very

close to the model the panelists built already, so MSAA partners anticipated very few if any policy adjustments would be needed.

With that, panelists were recalled to the room, told that the MSAA Partners were very impressed with their work on the model, and that they would be able to leave as scheduled since the model provided the guidance needed to finalize the cuts at the policy level.

LESSON LEARNED CONSENSUS: The content driven process used in this standard-setting was highly successful in the opinion of the MSAA partners and the vendor, even with the technology glitches and the struggles of a few panelists. Rethinking the articulation process and redesigning it to maintain the focus on the content will be a big step forward in this methodology. That said, MSAA Partners, vendor, and observer were all pleased with what the results were, given the consensus that the model and qualitative feedback provided by the panelists gave policymakers a clear pathway to finalize cuts that strongly reflect the voice of the panelists.

OVERALL END POINT OF SSS MEETING: Issues resolutions worked through with vendor and MSAA partners were successful. No remaining procedural threats to validity remain based on observation of the process. MSAA Partners will reconvene to make any policy adjustments. The full TAC and MSAA partners will review the vendor standard-setting report to make final validity judgements.