2022 MSAA Standard-Setting Report

Science

July 26–28, 2022—Albuquerque, New Mexico

Prepared by Cognia for the MSAA Partners

TABLE OF CONTENTS

CHAPTER 1. OVERVIEW OF STANDARD-SETTING PROCEDURES 3

CHAPTER 2. TASKS COMPLETED BEFORE STANDARD-SETTING 4

2.1 Creation of Performance Level Descriptors 4
2.2 Preparation of Materials 5
2.3 Preparation of Instructions for Facilitators 5
2.4 Creating Target Cut Scores for All Performance Levels 6
2.5 Preparation of Systems and Materials for Use During the Meeting 6
2.6 Selection of Panelists 7

CHAPTER 3. TASKS COMPLETED DURING THE STANDARD-SETTING MEETING 9

3.1 Overview of the ID Matching Method 9
3.2 General Orientation and Panelist Training 9
3.3 Review of Assessment Materials 10
3.4 Completion of the Item Map Form 10
3.5 Review of Borderline Performance Level Descriptors 11
3.6 Judgment Rounds and Feedback 11
3.6.1 Round 1 Judgments and Results 12
3.6.2 Round 2 Judgments and Results 13
3.6.3 Round 3 Judgments and Results 13

CHAPTER 4. TASKS COMPLETED AFTER THE STANDARD-SETTING MEETING 15

4.1 Analysis and Review of Panelists' Feedback 15
4.2 Target Cut Score Calculation 16
4.3 Standard-Setting Panelist Follow-Up Meetings 17
4.4 Policy Review for All Cut Scores 19
4.5 Approval of the Cut Scores 21
4.6 Preparation of Standard-Setting Report 21

REFERENCES 22

APPENDICES 23

Appendix A—Panelists
Appendix B—Performance Level Descriptors
Appendix C—PowerPoint Presentations
Appendix D—Cognia Standard-Setting Toolkit
Appendix E—Readiness surveys and results
Appendix F—Workshop Evaluation Survey and Results
Appendix G—Impact Data by Grade
Appendix H—Standard Setting Meeting Memo
Appendix I—Standard-Setting Meeting Agendas and NonDisclosure Form
Appendix J—Standard-Setting Panelist Follow-up Meeting Minutes

© 2022 Cognia, Inc.

Chapter 1. Overview of Standard-Setting Procedures

This report aims to summarize the activities involved in the standard-setting process for the Multi State Standard Assessment (MSAA) in Science for the MSAA Science Partners. The need for standard setting arises from the fact that this is a new assessment that was administered for the first time in spring 2022. The MSAA Science Partners for 2022 comprise Arizona, Bureau of Indian Education (BIE), Maine, and the U.S. Virgin Islands (USVI). For such new assessments, performance standards must be set. The primary goal of the standard setting was to determine the knowledge, skills, and abilities (KSAs) that students must demonstrate to be classified into each of the student status levels (performance levels). For the science alternate assessment, there are four performance levels named 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 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). Panelists considered externally benchmarked target cut scores during the ID Matching process.

The standard-setting meeting was held from July 26th through July 28th of 2022. In all, 25 panelists (recruited from the stakeholder partners) participated in the process and were organized into three groups of 8-9 panelists each, plus a facilitator provided by Cognia for each grade. Additionally, MSAA Science Partners from Arizona, BIE, and Maine attended to observe the standard-setting process. 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 • Preparation of Materials • Preparation of Instructions for Facilitators • Creating Target Cut Scores for All Performance Levels • Preparation of Systems and Materials for Use During the Meeting • Selection of Panelists	• Overview of the ID Matching Method • General Orientation and Panelist Training • Review of Assessment Materials • Completion of the Item Map Form • Review of Borderline Performance Level Descriptors • Judgment Rounds and Feedback	• Analysis and Review of Panelists' Feedback • Target Cut Score Calculation • Standard-Setting Panelist Follow-Up Meetings • Policy Review for All Cut Scores • Approval of the Cut Scores • Preparation of Standard-Setting Report

Chapter 2. Tasks Completed Before Standard-Setting

2.1 Creation of 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 serve as the defining descriptions for each performance level in grades 5, 8, and high school. The Policy PLDs provide the high level or overarching policy level student performance expectations for Level 1, Level 2, Level 3, and Level 4. From the Policy PLDs the Range PLDs were developed. 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. A list of the panelists is included in Appendix A. 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 were provided to the panelist via a secure File Transfer Protocol (sFTP). These 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.

Following this PLD review meeting, Cognia incorporated the input 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 B.

During standard setting, presenters provided all panelists with an overview of the Policy, Range, and Borderline PLDs. No additional edits were made to the Policy, Range, and Borderline PLDs following standard setting as these were considered final. The panelists utilized the Range and Borderline PLDs as outlined in the standard-setting process described below. Please refer to 3.1 for additional details regarding Range and Borderline PLDs.

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 (digital)
- Meeting agendas (paper)
- Nondisclosure forms (paper)
- Test booklets (digital)
- PLDs (digital & paper)
- Content standards (digital)
- Cognia Standard-Setting Toolkit (digital) which included the following:
    - Practice item booklet
    - Ordered item booklets
    - Readiness surveys
    - Judgment forms
    - Evaluation form

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, impact data, and the Standard Setting meeting memo are included in Appendices A through I.

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 C.

2.4 Creating Target Cut Scores for All Performance Levels

Cognia proposed the use of target cut scores as an option to provide initial guidance to panelists' judgment on science alternate assessment performance level cut scores. The MSAA Science Partners agreed on using the percentages of students in levels 2–4 on the MSAA ELA and Mathematics assessment to inform the setting of the target cut scores. Table 2-1 lists the percentage of students in each performance level and above (target cut scores) for Science Grades 5, 8, and 11. The target cut scores were calculated by using the average performance of MSAA ELA and Mathematics in each grade, and average performance across two content areas and the past three school years (2018–2019, 2020–2021, 2021–2022). Due to COVID-19 interruptions, no testing occurred in 2019–2020. Therefore, the 2018–2019 results were used in the target cut score calculation.

Table 2-1: Science target cut scores-Average Percentage of students in each performance level and above (2019–2022)

Grade	%Level 2 & above	%Level 3 & above	%Level 4 & above
5	66	43	12
8	70	47	16
11	75	53	14

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 bottom and the most difficult at the top. Items in the target region for the proficient performance level were shaded for easy reference.
- 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) why the item is more difficult than the previous item, (3) content-based rationales, (4) item descriptor matches, and (5) cut placements.
- 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 D.

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.

Additionally, the digital booklet was shaded to show a projected range of expected proficiency based on the target cut scores as discussed in section 2.4 of this document. The shaded region within the ordered item booklet was calculated as +/-3 pages around the target cut scores.

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 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	21	84%
	Male	4	16%
Ethnicity	White or Caucasian	8	30%
	Black or African American	6	22%
	Hispanic or Latino	2	7%
	Asian or Asian American	6	22%
	American Indian or Alaska Native	3	11%
	Native Hawaiian or Pacific Islander	1	4%
	Multiracial	1	4%
State Representation	Arizona	3	12%
	BIE	6	24%
	Maine	4	16%
	USVI	12	48%
Panelist Teaching Experience			
Grade Band*	Elementary	11	34%
	Middle	9	28%
	High	9	28%
	All Grades	3	9%
Educational Setting*	Special Education	14	41%
	General Education	15	44%
	School Administrator	3	9%
	English Language Learner Educator	2	6%
Years of Experience in Education	0-5	1	4%
	5-10	4	16%
	10-15	5	20%
	More than 15	15	60%

**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 borderline 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: (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? 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.

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. The orientation consisted of two parts. First, Cognia's Vice President of Measurement Services presented a brief overview of the ID matching procedure and the activities that would occur during the standard setting meeting. Next, Cognia's Content Development – Accessibility Director and Lead Science Content 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 Content Development–Accessibility Director and/or the Lead Science Content Specialist.

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 and to indicate why they believed each item was more difficult than the previous one. To ensure each panelist was comfortable using the provided laptop computers and understood the mechanics of data entry, Cognia Psychometricians Dr. Frank Padellaro and Dr. Sandra Sweeney reviewed the technology the panelists would use to complete their item maps. Additionally, the item map form was shaded to show a projected range of target cut score region, based on the target cut scores as discussed in section 2.4 of this document. Item map entries that would produce percentages of students at or above each performance level comparable to MSAA ELA and Mathematics assessments were identified as benchmarking locations. The shaded region within the ordered item booklet was calculated as +/-3 pages around the target cut scores.

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 Borderline 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 B.

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 median cut scores for the group based on their cut score recommendations, 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 various functions: feedback to panelists as part of the standard-setting method, reporting to Cognia and the MSAA Science Partners as intermediate evidence for the impact of panelists' judgments, and as quality control metrics.

The overall cut scores for each performance level and each round were determined by calculating the median of the individual cut scores obtained from the panelists. The Mean Absolute Difference of the panelists' cut points indicates the extent to which judgments were consistent across panelists and reflects the level of agreement among the ratings with each successive round of ratings. Conditional SEM characterizes the measurement precision for each of the scale cuts. Finally, impact data reflects the percentage of students across the MSAA Science Partners who would fall into each performance level category according to the total group median cut points. While these statistics were available, the only results revealed to panelists were those that were the median OIB page numbers.3.6.1 Modeling and Practice

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. Additionally, Cognia Psychometricians Dr. Frank Padellaro and Dr. Sandra Sweeney tested the interface/software that panelists would use to complete their judgments beforehand to make sure everything was working as expected. They were also present while panelists were completing the practice round to answer any questions from panelists about using

the Cognia Standard-Setting Tool. 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 one readiness survey (Appendix E). 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 the borderline PLDs, the standard-setting tool, and the ordered item booklet (OIB). Beginning with the first ordered item in the shaded region of 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 1, Level 2, Level 3, Level 4, or threshold) to each item. They continued in this manner until they had looked at all the items in the shaded region as well as items outside the shaded region using their content-based judgment, then placed their Level 3 cut at the item that marked the beginning of the Level 3 region based on their judgments. For the identification of this Level 3 cut point, panelists were instructed to write a content-based rationale. Panelists then repeated the process for the other two cut points (Level 2 and Level 4) and used the integrated judgment forms to record their notes and judgments.

After the completion of round one, Cognia psychometricians calculated a variety of statistics as described previously. As a reminder, the Round 1 overall cut points were determined by calculating the median of the individual cut points obtained from each panelist. Table 3-1 displays round 1 OIB page numbers, associated raw scores, median cut points on the theta scale, and impact data (percentage of students in each performance level), respectively. All statistics in Table 3-1 are listed for technical documentation purposes. Only the OIB pages were shown to panelists.

Table 3-1: Standard Setting MSAA Science Round 1 Results

Grade	Level 1	Level 2				Level 3				Level 4				Level 3 + 4
	Percent	OIB#	Raw	Theta	Percent	OIB#	Raw	Theta	Percent	OIB#	Raw	Theta	Percent	Percent
5	45.81	12	18	-0.23	8.48	20	20	-0.02	19.34	30	25	0.48	26.36	45.70
8	41.53	8	16	-0.46	25.21	16	21	0.19	11.23	28	24	0.66	22.03	33.26
11	56.50	10	17	0.11	13.77	16	19	0.34	12.48	24	22	0.68	17.25	29.73

3.6.2 Round 2 Judgments and Results

The purpose of Round 2 was for panelists to discuss their Round 1 placements and, if necessary, to revise their ratings. Before beginning their discussions, the panelists at each table were presented with the median cut points based on their Round 1 ratings for each cut point in that grade. A Cognia psychometrician presented this information to the group using a projector and laptop and explained how to use it as they completed their Round 2 discussions. The distribution of panelists' cut points was presented in terms of location in the ordered item booklet, both as numerical summaries of cut point ranges and graphically, as histograms.

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. Panelists were asked to pay particular attention to how their ratings compared to those of other panelists in their room to assess whether they were unusually stringent or lenient within the group. Once the discussions were complete, panelists completed the Round 2 Readiness Survey (Appendix E). The readiness survey was designed to ascertain whether the panelists were comfortable moving ahead to the second round of the judgment task. Once all panelists completed the Round 2 Readiness Survey, Cognia psychometricians reviewed the responses to make sure panelists were ready to undertake their second 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 2 judgments.

Once the discussions were complete, panelists were allowed to revise their Round 1 ratings on the rating form. Panelists were told to set cut score placement according to their individual best judgments; having consensus among the panelists was not necessary. They were encouraged to listen to the points made by their colleagues but not to feel compelled to change their cut score placements.

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, associated raw scores, median cut points on the theta scale, and impact data (percentage of students in each performance level), respectively.

Table 3-2: Standard Setting MSAA Science Round 2 Results

Grade	Level 1	Level 2				Level 3				Level 4				Level 3 + 4
	Percent	OIB#	Raw	Theta	Percent	OIB#	Raw	Theta	Percent	OIB#	Raw	Theta	Percent	Percent
5	45.81	12	18	-0.23	8.48	20	20	-0.02	14.93	28	24	0.37	30.77	45.70
8	41.53	9	16	-0.41	28.81	17	22	0.28	7.63	28	24	0.66	22.03	29.66
11	56.50	9	17	0.08	13.77	16	19	0.34	12.48	24	22	0.68	17.25	29.73

3.6.3 Round 3 Judgments and Results

The purpose of Round 3, the final round, was for panelists to discuss their Round 2 cut score recommendations and, if necessary, to revise their judgments. Before beginning their discussions, panelists were presented with the median cut scores based on their Round 2 judgments for each performance level cut in that grade. Facilitators presented this information to the group using a projector and laptop and explained how to use it as they completed their discussions. The distribution of panelists' cut points was presented graphically, as histograms, in terms of location in the ordered item booklet.

Within their grade-specific groups, panelists were then allowed to share their rationales for their cut placements in terms of the necessary knowledge and skills for each classification. Panelists were asked to pay particular attention to how their judgments compared to those of other panelists in their room to assess whether they were unusually stringent or lenient within the group. Once the discussions were complete, panelists completed the Round 3 Readiness Survey (Appendix E). The readiness survey was designed to ascertain whether the panelists were comfortable moving ahead to the third round of the judgment task. Once all panelists completed the Round 3 Readiness Survey, Cognia psychometricians reviewed the responses to make sure panelists were ready to undertake their third 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 3 judgments.

Once all panelists indicated that they were ready to undertake the next round, they were allowed to revise their Round 2 judgments on the judgment forms within the digital tool. Panelists were told to place cuts according to their individual best judgments; having consensus among the panelists was not necessary. They were encouraged to listen to the points made by their colleagues but not to feel compelled to change their cut placements. 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.

Table 3-3: Standard Setting MSAA Science Round 3 Results

Grade	Level 1	Level 2				Level 3				Level 4				Level 3 + 4
	Percent	OIB#	Raw	Theta	Percent	OIB#	Raw	Theta	Percent	OIB#	Raw	Theta	Percent	Percent
5	45.81	12	18	-0.23	8.48	20	20	-0.02	14.93	28	24	0.37	30.77	45.70
8	41.53	9	16	-0.41	25.21	16	21	0.19	11.23	28	24	0.66	22.03	33.26
11	56.50	9	17	0.08	13.77	15	19	0.31	12.48	24	22	0.68	17.25	29.73

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: reviewing the standard-setting process and addressing issues presented by the outcomes; presenting the results to the MSAA Science Partners; 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 were completed, the panelists' responses were reviewed. This review did not reveal any anomalies in the standard-setting process or indicate any reason that a particular panelist's data should not be included when the final cut points were calculated. In general, participants felt that the recommended cut points were appropriate and that their judgments were based on appropriate information and decision making.

As part of this survey, panelists were asked to respond to the following statement: *I am satisfied with the final group cut scores. I would not recommend changing any of the group cut scores*. Response options for the statement included *strongly disagree, disagree, undecided, agree, strongly agree,* and *not applicable*. As shown in Table 4-1, most panelists noted that they were satisfied with the final group cut score recommendations as indicated by the number of panelists who selected either *strongly agree* or *agree* in response to the statement. One member from each of the Grade 5 and Grade 11 panels selected *undecided* in response to the statement. The evaluation survey and the results of the evaluations are presented in Appendix F.

Table 4-1. Frequency of Panelist Responses to the Workshop Evaluation Survey

Grade	N	Undecided	Agree	Strongly Agree
5	9	1	5	3
8	8	0	2	6
11	8	1	3	4

Note. Since no panelists selected "strongly disagree," "disagree," or "not applicable," these categories are not included in the table.

As a follow-up question, panelists were asked if they would recommend, based on the final group of cut scores, moving a cut score up or down in the ordered item booklet (OIB), and by how many pages. Across the three grades, panelists' responses varied as follows:

- Grade 5: One panelist recommended moving the Level 2 cut score down by 2 pages (i.e., lower the cut score). In addition, two panelists suggested moving all cut scores up by 2 pages (i.e., raising the cut score).

- Grade 8: One panelist recommended moving all cut scores up by 1 page (i.e., raising the cut score).
- Grade 11: One panelist recommended moving all cut scores up by 1 page, another suggested moving all cut scores up by 2 pages, and a third recommended all cut scores be moved down by 2 pages.

4.2 Target Cut Score Calculation

Shortly following the standard-setting workshop, Cognia discovered that there was an error in the method used to calculate impact data for each line in all three grades of the ordered item booklets. The target cut score ranges in the ordered item books that were presented to panelists were mapped through this incorrect impact data. The panelists looked at these targets when they made their Round 1 cut score recommendations, and they were guided to consider those ranges as they made their cut score recommendations. This may have influenced panelists in the way they evaluated items and ultimately in their judgment process to recommend cut scores.

In addition, while at standard setting, the MSAA Science Partners were informed at the end of each day where on the MSAA reporting scale the cut scores were located and the expected impact of those cut scores. The impact data presented at these times was incorrect. Therefore, MSAA Science Partners were guided by the incorrect impact data. For example, for 5th grade science, the correct and incorrect impact data for Level 3 and above coincide almost perfectly, but the Level 2 and level 4 impact data are different between the incorrect and correct. Tables 4-2, 4-3, and 4-4 show the comparison of the incorrect (original) and corrected (updated) shaded regions by grade. See Appendix G for detailed and color-coded comparison by each grade.

Table 4-2: Science Standard Setting Benchmarking Regions-Grade 5

Level	% Benchmark At & Above	OIB Shaded Region-Original*	OIB Shaded Region- Updated**
2	66	12-18	5-11
3	43	17-23	18-24
4	12	26-32	34-39

**The incorrect shaded region; **The correct shaded region.*

Table 4-3: Science Standard Setting Benchmarking Regions-Grade 8

Level	% Benchmark At & Above	OIB Shaded Region-Original*	OIB Shaded Region- Updated**
2	70	11-17	4-9
3	47	14-20	10-14
4	16	22-28	29-35

**The incorrect shaded region; **The correct shaded region.*

Table 4-4: Science Standard Setting Benchmarking Regions-Grade 11

Level	% Benchmark At & Above	OIB Shaded Region-Original*	OIB Shaded Region- Updated**
2	75	9-15	1-6
3	53	12-18	2-8
4	14	19-25	28-34

**The incorrect shaded region; **The correct shaded region.*

Despite the incorrect target cut score ranges, there are strong suggestions that the panelists detected problems with target score ranges—because they recommended 5 out of 9 cut score OIB placements that were either at the edge of the range or outside of those ranges. The standard-setting workshop facilitators reported that panelists did not feel constrained by the target cut score ranges and they recommended final cut scores outside of the target ranges. Appendix G provides details regarding the correct and incorrect cut score ranges.

Science 5 Tab in the Workbook

The correct and incorrect target cut score ranges for Level 3 (in green) coincide almost perfectly. The correct and incorrect target cut score ranges for Level 2 (in yellow) do not overlap. However, the panelists recommended a cut score at the top of the incorrect range, one ordered item book page below the correct range. The correct and incorrect target cut score ranges for Level 4 (in blue) do not overlap. Panelists recommended a cut score that coincides with the incorrect target cut score and that does approach the correct target range. This cut score is a good candidate for policy review.

Science 8 Tab in the Workbook

The correct and incorrect target cut score ranges for all three cut scores do not overlap. However, the recommended Level 2 cut score in column L is only one ordered item book below the correct target range. (See the yellow highlighted lines in columns H and J.) The recommended Level 3 cut score (see the green highlighting) in column L is only two ordered item book pages above the correct target range. Also, the recommended cut score for Level 4 (in blue) is only one ordered item book page below the correct target range. Small adjustments in a policy review could bring these cut scores in line with the correct ranges and impact data.

Science 11 Tab in the Workbook

The correct and incorrect target cut score ranges in this grade level are widely disparate. However, the Levels 2 and 3 cut scores are in the correct target cut score range. This is not the case for the Level 4 cut score. That recommended cut score is four pages below the correct target range (in yellow). The Level 4 cut score is a good candidate for policy review adjustments.

4.3 Standard-Setting Panelist Follow-Up Meetings

Given the concerns raised following the standard-setting meeting, Cognia Chief Learning Officer Dr. Stephen Murphy, along with Arizona Department of Education Deputy Associate Superintendent of Assessment Audra Ahumada, facilitated two follow-up meetings with science standard-setting panelists on August 22, 2022. A list of the panelists is included in Appendix A. Dr. Murphy presented an overview of key areas of standard setting, walking through an outline and noting the work completed before the meeting on materials that panelists utilized and how the panelists participated. Then Dr. Murphy reviewed the purpose of target cut scores and the range miscalculation. After walking through the standard-setting process and illustrating the shaded regions from the impact data, questions were posed to panelists:

1. Can you describe how you–and your colleagues–used the target cut scores and ranges to inform your work (how did this influence your thinking)?

Feedback from panelists: Panelists used PLDs to guide their cut score recommendations. They looked at the content and went outside of the starting point cut score ranges.

2. Given what you learned about the target cut scores and ranges, how do you think that changing this information might have impacted your and/or your committee's cut score recommendations, if at all?

Feedback from panelists:

Grade 5

- One panelist felt confident about the recommendations. Her group looked at items on the lower and higher end.

Grade 8

- One panelist noted that he expected L2 would be bigger and L4 smaller. He concluded that the change would not have had a significant impact. By the end of three rounds, they were confident in their recommendations and would be unlikely to change much if the new information was introduced. There was fidelity with the process.
- One panelist noted that being presented with certain data sets served as a marker/guide in your head. The group followed the process. She is confident that the recommendations they provided are solid and would likely not have changed

Grade 11

- One panelist felt confident with the recommendations.
- One panelist noted that they looked at lower- and higher-end items
- One panelist noted that he would not change his recommendations.
- One panelist noted that they started working from L3 as a starting point then went up toward L4 and down toward L2. Facilitators encouraged panelists to look at a few items before and after the starting point. She felt that there might have been some impact if the correct info had been provided. There was a lot of variation in her initial judgments, which were unexpected. If panelists had been looking at the correct information initially, it might have brought L3 down. Ultimately, she shared that she was confident in her judgments, as related to the methodology required from them as panelists.

As noted in the bullets above, there were one or two panelists who expressed that they were unsure whether the incorrect target cut score ranges had impacted their recommendations and could not say for certain that they did not. Most, if not all, panelists were, however, satisfied with their final cut score judgments as noted in their responses to evaluation survey and as noted during the follow-up meeting. At the end of the meeting, Audra thanked the panelists and highlighted the purposes for bringing panelists back together. MSAA Science Partners wanted feedback from panelists to inform their policy decisions.

Appendix J shows the meeting minutes and PowerPoint presentation of the panelists' follow-up meeting.

4.4 Policy Review for All Cut Scores

The process of establishing performance standards for any new assessment program often continues after a standard-setting workshop. In a standard-setting workshop, content experts—in this case, educators in MSAA Science Partner states and entities—recommend cut scores to the policy makers who are responsible for an assessment program. After that, assessment program leaders have the opportunity to conduct a policy review of the recommended cut scores. In a policy review, leaders can consider the impact data,[1] vertical articulation of the impact data across grade levels,[2] and the likely effects on the policies and educational politics of the participating states and entities.

For the reasons described above, MSAA Science Partners conducted a policy review of the three recommended cut scores in all three grade levels. In this policy review, the MSAA Science Partners reviewed the appropriateness of the panelists' recommended cut scores, considering the outcomes from the standard setting as noted in the discussions above. A second goal of the policy review was to gather input from facilitators, panelists, state observers, and standard setting experts regarding the impact of the target cut scores and range miscalculations and how it may or may not have impacted panelist recommendations. A third, and important, goal of this policy review was to achieve a reasonable articulation of impact data across grades 5, 8, and 11. Here, the goal was to adjust, as necessary, cut scores recommended by the standard-setting panels—keeping true to the content-focused process that the standard-setting panelists completed in following the ID Matching process to recommend cut scores—but allowing for final and reasonable cut scores and impact data appropriate within the MSAA Science Partners policy context.

Cognia proposed policy adjustment on recommended cut scores following a generally accepted approach and practice. The rationale was to remain true to the panelists' recommendations (and their content-based judgments) as well as to consider the following, from a policy review perspective:

- Since it is not consensus, review agreement levels among panelists recommendations for each cut score by each grade
- Meet with the facilitators and discuss conversations among the panelists regarding their recommendations
- Review results from spring 2022 for Mathematics and ELA for MSAA
- Listen to reactions from MSAA Science Partners, regarding expectations of science achievement within their respective states and entities

[1] Impact is the percentages of students in Levels 1, 2, 3, and 4 after cut scores have been recommended or approved.

[2] When performance standards are vertically articulated, the percentages of examinees in performance levels are similar across grades levels or increase or decrease in understandable and meaningful explainable ways (e.g., We expect performance to be higher in grades 5 and 8 than in grade 11).

- Consider the results for other K-12 Science Frameworks (like NGSS) programs (alternate assessment or general education assessment)
- Examine the impact data across grades – how each of the grades' impact data for each performance level compares to other grades
- Review for pages in the OIBs that resulted in impact data that seemed to align with the following:
    - Expectations as expressed by MSAA Science Partners
    - Seemed appropriate given achievement on MSAA Mathematics and ELA
    - Allowed for reasonable articulation or comparisons across grades
- Examine measurement principles as related to the cut scores and ability to sustain the cut scores across time–ensuring that two adjacent cuts had an appropriate measurable "distance" – and make sure that the cut scores distinguish levels of performance

Following the standard-setting workshops, the Cognia psychometric team examined the raw scores associated with each cut recommended by panelists and made adjustments to ensure that there are appropriate distances between raw score cuts.

Using Grade 5 science as an example, the Level 2 raw score cut associated with panelists' recommendation is 18, and for Level 3 is 20. The psychometric team determined the distance between these two cuts was much closer than what we typically see. Consequently, the associated impact data shows that there are less than 10% of students in Level 2, which is also a concern to the stability of the scale in the future. In particular, we want to prevent a situation in which very few students perform at a given level in any future years. With the proposed policy adjustment cuts, we widened the raw score (and theta) range so that there is a reasonable percentage of students in each performance level, and this resulted in at least 3 raw score points (or at least 0.4 in distance in theta scale) spread among any two cut scores (see Table 4-5).

Table 4-5 contains cut scores and impact data recommended by the science alternate assessment standard-setting panelists and the policy-adjusted cut scores approved by MSAA Science Partners. Cognia recommends the use of the policy-adjusted cut scores for 2022 score reporting.

Table 4-5: Proposed Policy Adjustment and Impact Data

	Panelist Recommendations			Policy Adjusted		
Performance Level	Raw Score*	OIB Page Number	Impact Data	Raw Score	OIB Page Number	Impact Data
Grade 5						
1	--	--	46%	--	--	36%
2	18	12	9%	16	7	19%
3	20	20	15%	20	20	32%
4	24	28	31%	29	36	13%
3 + 4	--	--	46%	--	--	46%
Grade 8						
1	--	--	42%	--	--	42%
2	16	9	25%	16	7	21%

continued

Performance Level	Panelist Recommendations Raw Score*	OIB Page Number	Impact Data	Policy Adjusted Raw Score	OIB Page Number	Impact Data
3	21	16	11%	20	15	19%
4	24	28	22%	25	29	19%
3 + 4	--	--	33%	--	--	38%
Grade 11						
1	--	--	57%	--	--	49%
2	17	9	14%	16	5	21%
3	19	15	13%	19	15	16%
4	22	24	17%	23	27	15%
3 + 4	--	--	30%	--	--	30%

**Raw Score: The number of score points students received on the science alternate assessments.*

4.5 Approval of the Cut Scores

A summary of the results approved by policy review is provided in Table 4-6 below. This displays 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.

Table 4-6: Standard Setting MSAA Science Policy Review Approved Results

Grade	Level 1	Level 2				Level 3				Level 4				Level 3 + 4
	Percent	OIB#	Raw	Theta	Percent	OIB#	Raw	Theta	Percent	OIB#	Raw	Theta	Percent	Percent
5	35.63	7	16	-0.39	18.67	20	20	-0.02	32.35	36	29	1.02	13.35	45.70
8	41.53	7	16	-0.46	20.87	15	20	0.10	18.86	29	25	0.71	18.75	37.61
11	48.91	5	16	-0.05	21.36	15	19	0.31	15.19	27	23	0.85	14.54	29.73

4.6 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 2022 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

Science Alternate Assessment Performance Level Descriptor Review Meeting Participants (May 2022)

Name	State/Entity
Lacey Todd	Maine
Benjamin Altsher	Cognia
Agosto Jerusalem	US Virgin Islands
Kristen Nash	Arizona
Amanda Simcock	Cognia
Francine Galko	Cognia
Bethany Spangenberg	Arizona
Jodi Bossio-Smith	Maine
Donald Griffin	Bureau of Indian Education

Science Alternate Assessment Standard Setting Participants (July 2022)

Name	State/Entity
Renee J Atcitty	Bureau of Indian Education (school in NM)
MARLENE MURPHY	Bureau of Indian Education (school in NM)
Victoria Ostwald	Bureau of Indian Education (school in ND)
Tammy Moncel	Bureau of Indian Education (school in WI)
Lisa Forcier	Bureau of Indian Education (school in SD)
Magdalena de la Paz	Bureau of Indian Education (school in ND)
Misty Favreau	Maine
April Knippen	Arizona
Jane L Williams	Maine
Oddeth Brown	US Virgin Islands
Lawrance Nai	Arizona
Orlinda Wickham	US Virgin Islands
Joann Lewis	US Virgin Islands
Ellen Anfone	US Virgin Islands
Jessica Cana	US Virgin Islands
Carmen Bird	US Virgin Islands
Vernon Callwood	US Virgin Islands
Cynthia Locsin	US Virgin Islands
Jerwin Cruz Perez	US Virgin Islands
Angelita Cruz-Tuitt	US Virgin Islands
Liezl V. Cabanilla	US Virgin Islands
Mervin Arcillo	US Virgin Islands
Andrea Mercado	Maine
Brittney Oden	Arizona
Patricia Adams	Maine

Science Alternate Assessment Standard Setting Follow-up Meeting Participants (August 2022)

Name	State/Entity
Jane L Williams	Maine
Carmen Bird	US Virgin Islands
Oddeth Brown	US Virgin Islands
Jessica Cana	US Virgin Islands
Joann Lewis	US Virgin Islands
Vernon Callwood	US Virgin Islands
Jerwin Cruz Perez	US Virgin Islands
Renee J Atcitty	Bureau of Indian Education (school in NM)
Mervin Arcillo	US Virgin Islands
Lawrance Nai	Arizona
April Knippen	Arizona

APPENDIX—B
PERFORMANCE LEVEL DESCRIPTORS

Grade 5 PLD Table				
Policy				
	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.*	**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 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 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 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.*

Grade 5 PLD Table				
Borderline				
most situations[1] *refers to the following: Students 75% of the time can be expected to demonstrate the knowledge and skills for the level during classroom instruction and tasks.*	**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.*	**Level 2 (Approaching Expectations)** *Students performing at the borderline of Level 2 Approaching Expectations can be expected in most situations[1] 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.*	**Level 3 (Meeting Expectations)** *Students performing at the borderline of Level 3 Meeting Expectations can be expected in most situations[1] 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.*	**Level 4 (Exceeding Expectations)** *Students performing at the borderline of Level 4 Exceeding Expectations can be expected in most situations[1] 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.*
Range				
Standards **PS-1 Matter and Its Interactions** •5-PS1-2 **SEP** •Using Mathematics and Computational Thinking **CCC** •Scale, Proportion, and Quantity	Attempt to identify the appropriate tools or units of measurement (for weight, time, temperature, or volume) for a scientific task.	Identify the appropriate tools or units of measurement (for weight, time, temperature, or volume) for a scientific task.	Compare the weight of matter before and after heating, cooling, or mixing by using data.	Show that the weight of matter does not change when substances are heated, cooled, or mixed by measuring, graphing, or using mathematical relationships.

Grade 5 PLD Table				
PS-2 Motion and Stability: Forces and Interactions •3-PS2-2 •5-PS2-1 **SEP** •Planning and Carrying Out Investigations (3-PS2-2, Supporting 5-PS2-1) •Engaging in Argument from Evidence (5-PS2-1) •Analyzing and Interpreting Data (Supporting 3-PS2-2) •Developing and Using Models (Supporting 5-PS2-1) **CCC** •Patterns (3-PS2-2, Supporting 5-PS2-1) •Cause and Effect (5-PS2-1)	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.	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.	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.	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).
PS-3 Energy •4-PS3-4 •5-PS3-1 **SEP** •Constructing Explanations and Designing Solutions (4-PS3-4) •Developing and Using Models (5-PS3-1) **CCC** •Energy and Matter (4-PS3-4, 5-PS3-1) •Patterns (Supporting 5-PS3-1)	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.	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.	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.	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.
LS-1 From Molecules to Organisms: Structures and Processes •4-LS1-1 **SEP** •Engaging in Argument from Evidence •Developing and Using Models (Supporting) •Analyzing and Interpreting Data (Supporting) **CCC** •Systems and System Models •Structure and Function (Supporting)	Attempt to identify the parts of plants or animals that have specific functions by using evidence from data and/or a model.	Identify the parts of plants or animals that have specific functions by using evidence from data and/or 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 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.

Grade 5 PLD Table				
LS-3 Heredity: Inheritance and Variation of Traits •3-LS3-1 **SEP** •Analyzing and Interpreting Data •Obtaining, Evaluating, and Communicating Information (Supporting) **CCC** •Patterns	Attempt to identify patterns in trait variations between parents and their baby/babies by using data or observations.	Identify patterns in trait variations between parents and their baby/babies by using data or observations.	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 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.
LS-4 Biological Evolution: Unity and Diversity •3-LS4-1 **SEP** •Analyzing and Interpreting Data •Obtaining, Evaluating, and Communicating Information (Supporting) **CCC** •Scale, Proportion, and Quantity	Attempt to recognize that there was life on Earth long ago by using information about fossils and/or data.	Identify that plants and/or animals lived on Earth long ago by using information about fossils and/or data.	Describe how modern-day plants or animals compare to their ancestors by using observations of fossils and/or 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.
ESS-1 Earth's Place in the Universe •5-ESS1-2 **SEP** •Analyzing and Interpreting Data •Developing and Using Models (Supporting) **CCC** •Patterns •Systems and System Models (Supporting)	Attempt to identify the positions of the Sun, the Moon, and Earth in the solar system by using data or a model.	Identify the positions of the Sun, the Moon, and Earth in the solar system by using data or a model.	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.	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.

Grade 5 PLD Table				
ESS-2 Earth's Systems •3-ESS2-1 •5-ESS2-1 **SEP** •Analyzing and Interpreting Data (3-ESS2-1) •Planning and Carrying Out Investigations (Supporting 3-ESS2-1) •Developing and Using Models (5-ESS2-1) **CCC** •Patterns (3-ESS2-1) •Systems and Systems Models (5-ESS2-1)	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.	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.	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.	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.
ESS-3 Earth and Human Activity •5-ESS3-1 **SEP** •Obtaining, Evaluating, and Communicating Information **CCC** •Cause and Effect (Supporting) •Systems and System Models	Attempt to identify a natural or human impact on the environment by using data.	Identify a natural or human impact on the environment by using data.	Describe an effect (positive or negative) of human activities on the environment by using data.	Describe how humans are using science to protect Earth's resources and/or the environment by using data.

Grade 8 PLD Table				
Policy				
	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 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 (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 at 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 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 at 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 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 at Level 4.* *At Level 4, students are expected to have the knowledge and skills described in Level 3.*
Borderline				

Grade 8 PLD Table				
most situations[1] *refers to the following: Students 75% of the time can be expected to demonstrate the knowledge and skills for the level during classroom instruction and tasks.*	**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.*	**Level 2: (Approaching Expectations)** *Students performing at the borderline of Level 2 Approaching Expectations can be expected in most situations[1] 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.*	**Level 3 (Meeting Expectations)** *Students performing at the borderline of Level 3 Meeting Expectations can be expected in most situations[1] 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.*	**Level 4 (Exceeding Expectations)** *Students performing at the borderline of Level 4 Exceeding Expectations can be expected in most situations[1] 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.*
Range				
Standards **PS-1 Matter and Its Interactions** •MS-PS1-2 **SEP** •Analyzing and Interpreting Data •Planning and Carrying Out Investigations (Supporting) **CCC** •Patterns •Scale, Proportion, and Quantity (Supporting)	Attempt to identify properties of a substance by using data or observations.	Identify properties of a substance by using data or observations.	Determine the identities of substances by using data or observations on the properties of substances.	Determine whether a chemical reaction occurred by using data or observations on the properties of substances before and after an interaction.

Grade 8 PLD Table				
PS-2 Motion and Stability: Forces and Interactions •MS-PS2-2 **SEP** •Planning and Carrying Out Investigations •Analyzing and Interpreting Data (Supporting) **CCC** •Stability and Change •Cause and Effect (Supporting)	Attempt to identify the effects of pushes and pulls on objects by using data from an investigation.	Identify the effects of pushes and pulls on objects by using data from an investigation.	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.	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.
PS-3 Energy •MS-PS3-5 **SEP** •Engaging in Argument from Evidence •Asking Questions and Defining Problems (Supporting) •Analyzing and Interpreting Data (Supporting) **CCC** •Energy and Matter	Attempt to determine whether energy is being transferred in a system by asking questions or by using data.	Determine whether energy is being transferred in a system by asking questions or by using data.	Identify the forms of energy that increase or decrease when the kinetic energy of an object changes by using data as evidence.	Make or support a claim that a transfer of energy occurs when the kinetic energy of an object changes by using data as evidence.
PS-4 Waves and Their Applications in Technologies for Information Transfer •MS-PS4-2 **SEP** •Developing and Using Models •Planning and Carrying Out Investigations (Supporting) **CCC** •Structure and Function	Attempt to identify whether a wave is being reflected, absorbed, or transmitted through a material by using data or a model.	Identify whether a wave is being reflected, absorbed, or transmitted through a material by using data or a model.	Describe the path of a wave that is reflected, absorbed, or transmitted through different materials by using a model.	Represent what happens to waves when they are reflected, absorbed, or transmitted through different materials by developing a model.

Grade 8 PLD Table				
LS-1 From Molecules to Organisms: Structures and Processes •MS-LS1-3 •MS-LS1-5 **SEP** •Engaging in Argument from Evidence (MS-LS1-3) •Developing and Using Models (Supporting MS-LS1-3) •Obtaining, Evaluating, and Communicating Information (Supporting MS-LS1-3) •Constructing Explanations and Designing Solutions (MS-LS1-5) •Analyzing and Interpreting Data (Supporting MS-LS1-5) •Asking Questions and Defining Problems (Supporting MS-LS1-5) **CCC** •Systems and System Models (MS-LS1-3) •Cause and Effect (MS-LS1-5)	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.	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 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.	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.
LS-2 Ecosystems: Interactions, Energy, and Dynamics •MS-LS2-1 •MS-LS2-3 **SEP** •Analyzing and Interpreting Data (MS-LS2-1) •Developing and Using Models (MS-LS2-3) **CCC** •Cause and Effect (MS-LS2-1) •Energy and Matter (MS-LS2-3)	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.	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.	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.	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.

Grade 8 PLD Table				
ESS-1 Earth's Place in the Universe •MS-ESS1-1 **SEP** •Developing and Using Models **CCC** •Patterns •Systems and System Models (Supporting)	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.	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.	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.	Compare or show patterns in seasons, lunar phases, or eclipses by using or developing a model of the Earth-Sun-Moon system.
ESS-2 Earth's Systems •MS-ESS2-2 •MS-ESS2-4 **SEP** •Constructing Explanations (MS-ESS2-2) •Obtaining, Evaluating, and Communicating Information (Supporting MS-ESS2-2) •Developing and Using Models (MS-ESS2-4) **CCC** •Scale, Proportion, and Quantity (MS-ESS2-2) •Cause and Effect (Supporting MS-ESS2-2) •Energy and Matter (MS-ESS2-4)	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.	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 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.	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.
ESS-3 Earth and Human Activity •MS-ESS3-3 **SEP** •Constructing Explanations and Designing Solutions •Engaging in Argument from Evidence (Supporting MS-ESS3-3) •Asking Questions and Defining Problems (Supporting MS-ESS3-3) **CCC** •Cause and Effect	Attempt to identify an environmental problem caused by human activities/impact by using data.	Identify an environmental problem caused by human activities/impact by using data.	Make a claim about how a particular method would work to reduce a human impact on the environment by using data.	Select or evaluate a design for a method that is intended to minimize a human impact on the environment by using data.

Grade 11 PLD Table				
Policy				
	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 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 (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 at 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 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 at 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 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 at Level 4.* *At Level 4, students are expected to have the knowledge and skills described in Level 3.*

Grade 11 PLD Table				
Borderline				
most situations[1] *refers to the following: Students 75% of the time can be expected to demonstrate the knowledge and skills for the level during classroom instruction and tasks.*	**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.*	**Level 2 (Approaching Expectations)** *Students performing at the borderline of Level 2 Approaching Expectations can be expected in most situations[1] 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.*	**Level 3 (Meeting Expectations)** *Students performing at the borderline of Level 3 Meeting Expectations can be expected in most situations[1] 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.*	**Level 4 (Exceeding Expectations)** *Students performing at the borderline of Level 4 Exceeding Expectations can be expected in most situations[1] 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.*

Grade 11 PLD Table				
Range				
Standards **PS-1 Matter and Its Interactions** •HS-PS1-2 **SEP** •Constructing Explanations and Designing Solutions •Obtaining, Evaluating, and Communicating Information (Supporting) •Developing and Using Models (Supporting) **CCC** •Patterns •Energy and Matter (Supporting)	Attempt to show how substances react in a chemical reaction by using provided information to complete an incomplete chemical reaction model.	Show how substances react by using provided information to complete an incomplete chemical reaction model.	Identify or classify elements that will react similarly in chemical reactions by using a periodic table.	Construct an explanation for why specific chemical reactions occur by using a periodic table.
PS-2 Motion and Stability: Forces and Interactions •HS-PS2-3 •HS-PS2-5 **SEP** •Constructing Explanations and Designing Solutions (HS-PS2-3) •Engaging in Argument from Evidence (Supporting HS-PS2-3) •Developing and Using Models (Supporting HS-PS2-3) •Planning and Carrying Out Investigations (HS-PS2-5) •Analyzing and Interpreting Data (Supporting HS-PS2-5) **CCC** •Cause and Effect (HS-PS2-3, HS-PS2-5) •Systems and System Models (Supporting HS-PS2-3) •Stability and Change (Supporting HS-PS2-5)	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.	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.	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.	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.

Grade 11 PLD Table				
PS-3 Energy •HS-PS3-2 **SEP** •Developing and Using Models •Asking Questions and Defining Problems (Supporting) **CCC** •Energy and Matter	Attempt to identify questions that would determine whether an object's kinetic or potential energy is changing in a system.	Identify questions that would determine whether an object's kinetic or potential energy is changing in a system.	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 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.
LS-2 Ecosystems: Interactions, Energy, and Dynamics •HS-LS2-2 **SEP** •Using Mathematics and Computational Thinking •Obtaining, Evaluating, and Communicating Information (Supporting) **CCC** •Scale, Proportion, and Quantity •Cause and Effect (Supporting)	Attempt to identify factors that affect population size or biodiversity by using provided information.	Identify factors that affect population size or biodiversity by using provided information.	Describe how a factor affects population size or biodiversity in an ecosystem by interpreting data.	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.
LS-3 Heredity: Inheritance and Variation of Traits •HS-LS3-1 **SEP** •Asking Questions and Defining Problems •Developing and Using Models (Supporting) •Obtaining, Evaluating, and Communicating Information (Supporting) **CCC** •Cause and Effect •Structure and Function (Supporting) •Patterns (Supporting)	Attempt to identify the function of DNA or chromosomes by using provided information.	Identify the function of DNA or chromosomes by using provided information.	Describe how genes and traits are inherited from parents to offspring by using a model.	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.

Grade 11 PLD Table				
LS-4 Biological Evolution: Unity and Diversity •HS-LS4-1 •HS-LS4-3 **SEP** •Obtaining, Evaluating, and Communicating Information (HS-LS4-1, Supporting HS-LS4-3) •Analyzing and Interpreting Data (HS-LS4-3, Supporting HS-LS4-1) **CCC** •Patterns (HS-LS4-1, HS-LS4-3) •Stability and Change (Supporting HS-LS4-1)	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.	Identify how organisms have changed over time by using provided information. Identify physical traits that can vary in an organism by using provided information.	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 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.
ESS-1 Earth's Place in the Universe •HS-ESS1-6 **SEP** •Constructing Explanations and Designing Solutions •Asking Questions and Defining Problems (Supporting) •Analyzing and Interpreting Data (Supporting) **CCC** •Stability and Change •Patterns (Supporting)	Attempt to identify patterns in data about ancient Earth materials, meteorites, or other planetary surfaces by using data.	Identify patterns in data about ancient Earth materials, meteorites, or other planetary surfaces by using data.	Describe Earth's formation and early history by asking questions about ancient Earth materials, meteorites, and other planetary surfaces.	Explain Earth's formation and early history by using data about ancient Earth materials, meteorites, or other planetary surfaces.

Grade 11 PLD Table				
ESS-2 Earth's Systems •HS-ESS2-4 •HS-ESS2-5 **SEP** •Developing and Using Models (HS-ESS2-4) •Planning and Carrying Out Investigations (HS-ESS2-5) •Analyzing and Interpreting Data (Supporting HS-ESS2-5) •Asking Questions and Defining Problems (Supporting HS-ESS2-5) **CCC** •Cause and Effect (HS-ESS2-4, Supporting HS-ESS2-5) •Energy and Matter (Supporting HS-ESS2-4) •Structure and Function (HS-ESS2-5)	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.	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.	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.	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).
ESS-3 Earth and Human Activity •HS-ESS3-4 **SEP** •Constructing Explanations and Designing Solutions •Engaging in Argument from Evidence (Supporting) •Analyzing and Interpreting Data (Supporting) **CCC** •Stability and Change •Cause and Effect (Supporting)	Attempt to identify the positive or negative impacts of local human activities on natural systems by using data.	Identify the positive or negative impacts of local human activities on natural systems by using data.	Construct a claim about how a local technological solution reduces the negative impact of human activities on natural systems.	Evaluate or refine the design of a local technological solution that reduces the negative impact of human activities on natural systems.

APPENDIX—C
POWERPOINT PRESENTATIONS

Welcome and Overview

Standard Setting Workshop

MSAA Science in grades 5, 8, and 11

July 26-28, 2022

cognia™

Welcome!

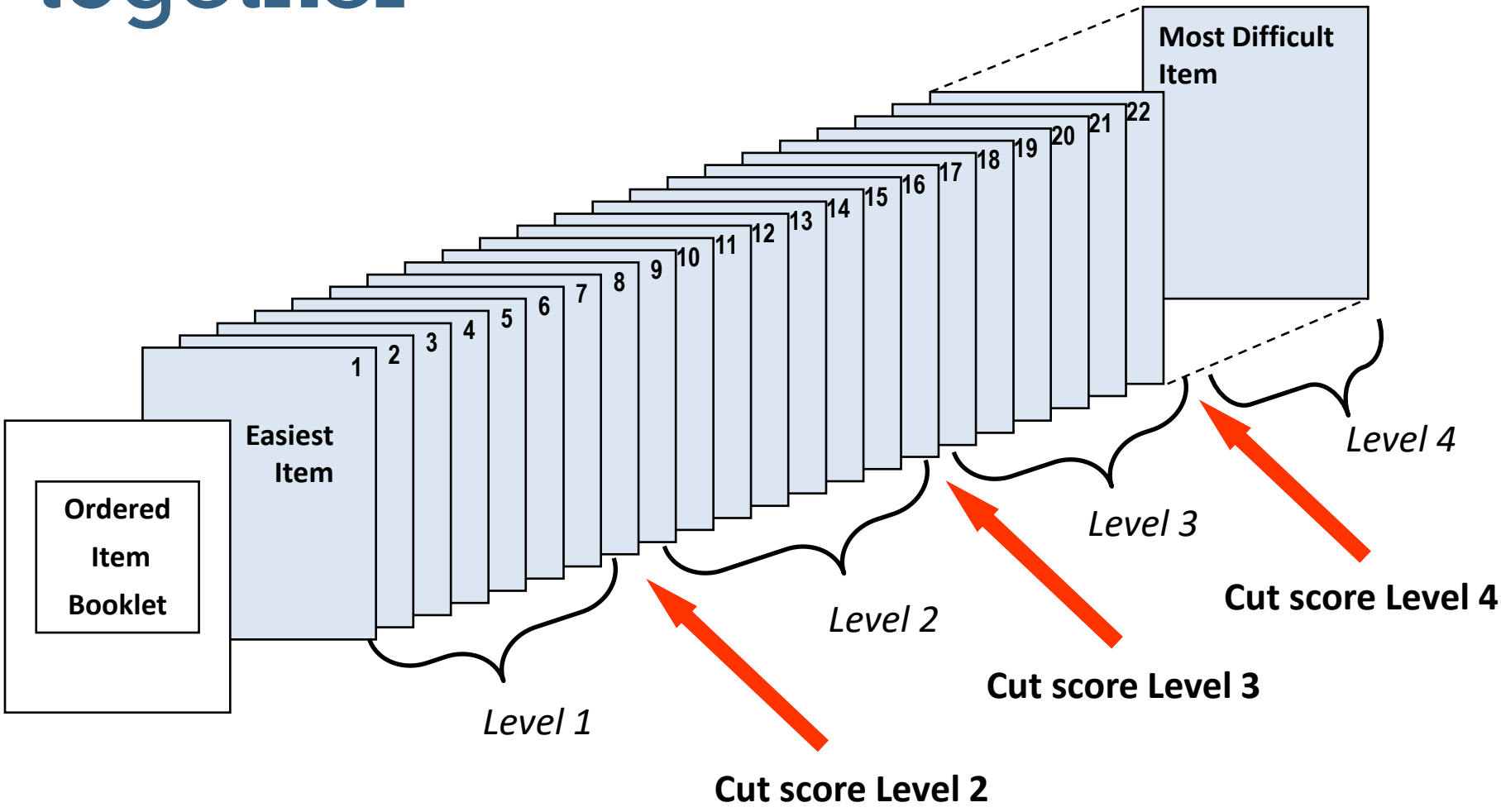
- We're here to recommend cut scores to the MSAA partners following the Item-Descriptor (ID) Matching method
- Introductions
  - MSAA partners
  - Cognia
- Show of hands
  - Who's been involved in standard setting before?
  - Which method(s)?

- Review the agenda

Overview

- Our shared goals
  - Get your recommendations for performance standards for the MSAA science assessments that provide meaningful and actionable information
- Your goals as panelists
  - Learn the concepts and procedures to recommend cut scores, following the ID Matching method
  - Follow the procedures we train you on
  - Recommend cut scores for levels 1, 2, 3, and 4
  - Rely on your expertise about the content standards, student learning, and students throughout the process

The outcome we're pursuing together



Note: Width of brackets is irrelevant

Key concepts and procedures

- Borderline PLDs
- ID Matching judgmental task
- Item map, OIB, online tool
- Target cut scores for levels 2, 3 and 4
- Become familiar with test items

- Three rounds
- Prepare, Complete, Feedback

Target cut scores

- Starting place for you to consider recommending cut scores
  - Target cut scores for levels 2, 3 and 4
- They represent MSAA partners' policy aspirations for the academic achievement of school children
- Your job
  - Follow the ID Matching judgment process
  - Apply your content expertise
  - Recommend cut scores – retain or adjust the target cuts
  - Write content-based rationales for your recommendations for levels 2, 3 and 4

Cognia Science Alternate Assessment (CSAA): Overview

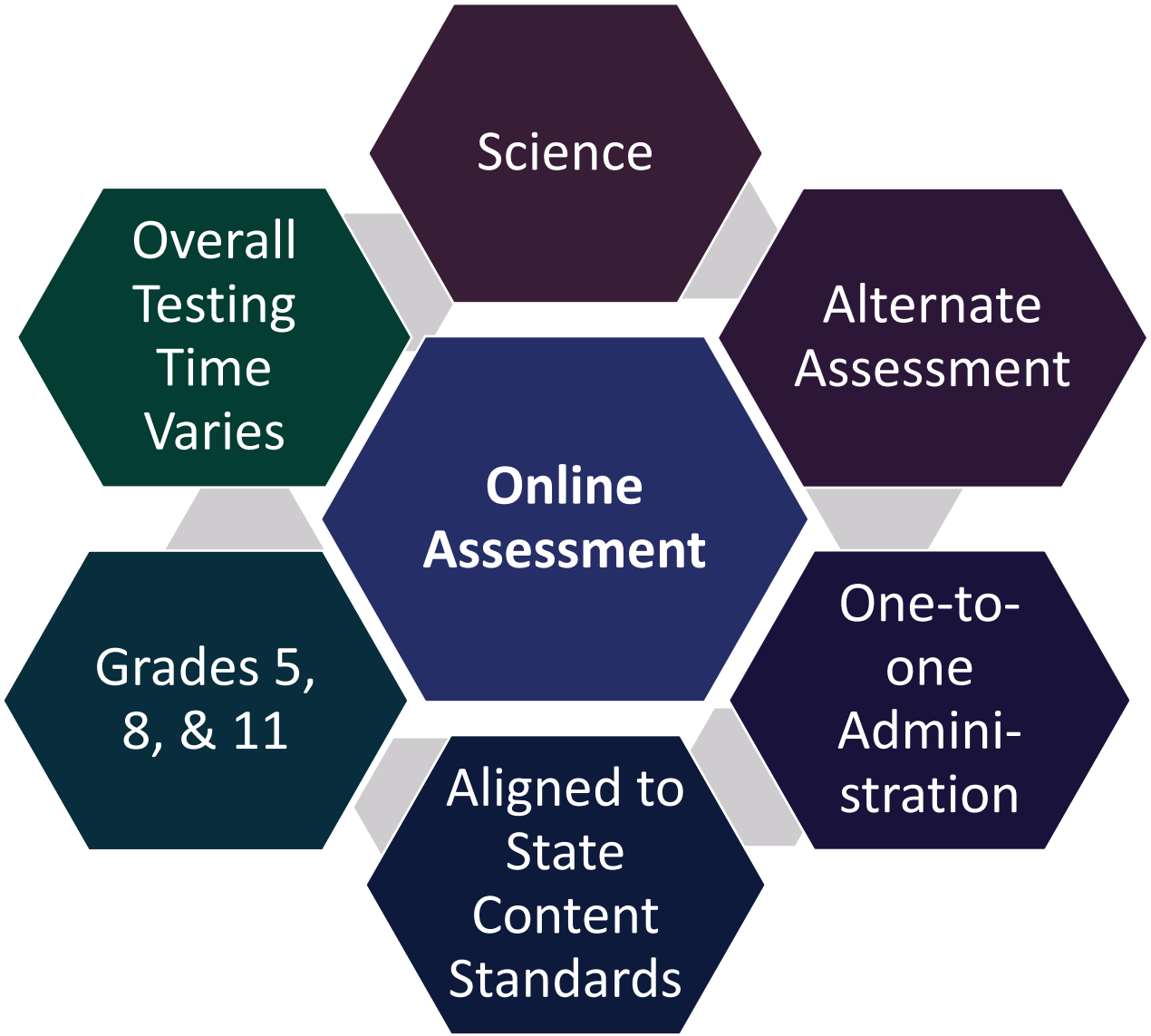
Standard Setting

July 2022

Agenda

- Assessing students on an alternate assessment
- CSAA overview
- Assessment features and accommodations
- Demo of the MSAA System
- The three dimensions of science learning
- Performance Level Descriptors (PLDs)

What is CSAA?


Science
Overall Testing Time Varies
Alternate Assessment
Online Assessment
Grades 5, 8, & 11
One-to-one Admini-stration
Aligned to State Content Standards

CSAA learner characteristics

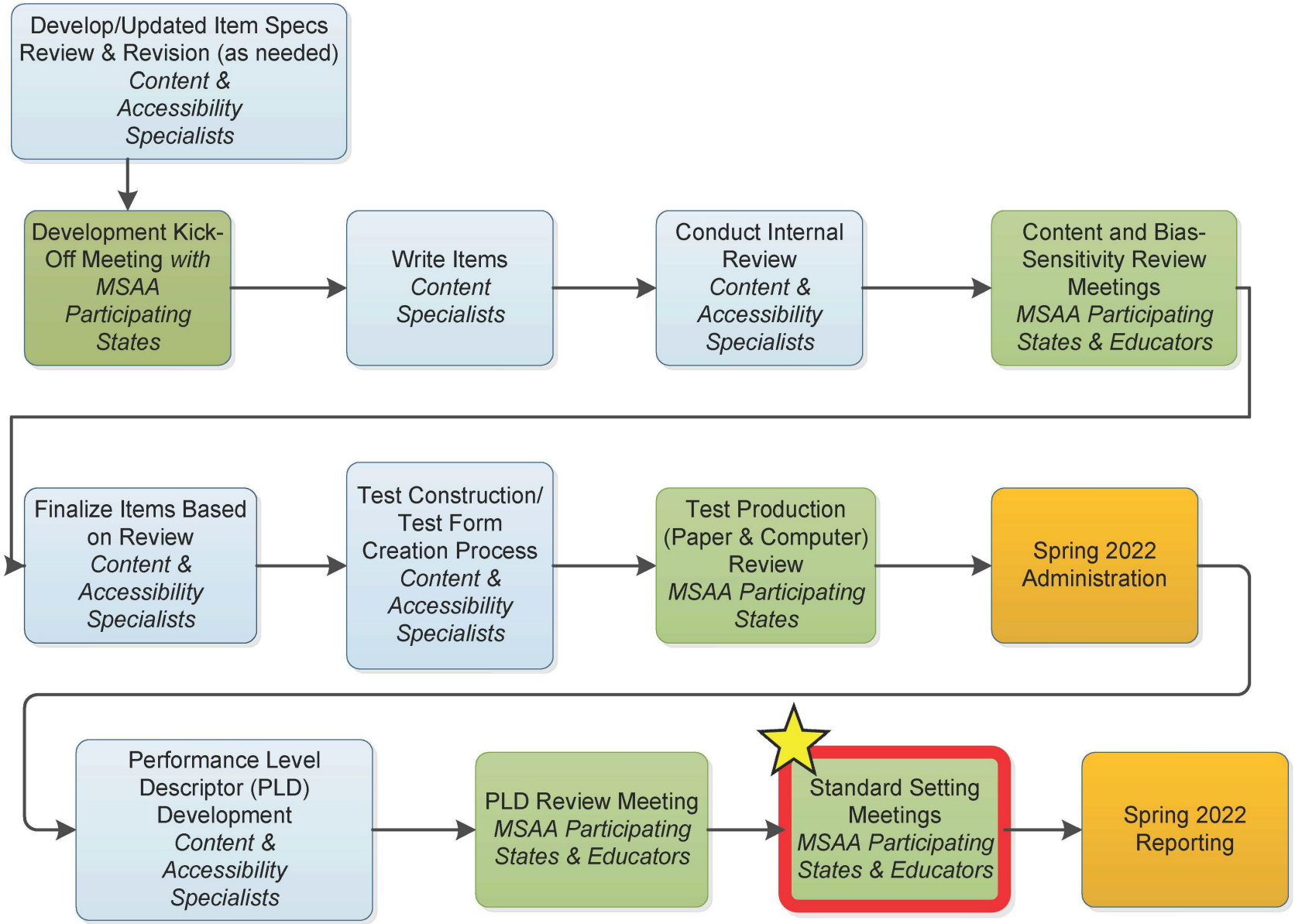
Participation criteria

Learning characteristics

CSAA purpose

Participation	Grade-level content aligned to State Content Standards
	Assesses students in grades 5, 8, and 11 (3rd yr. 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



CSAA overview - item types

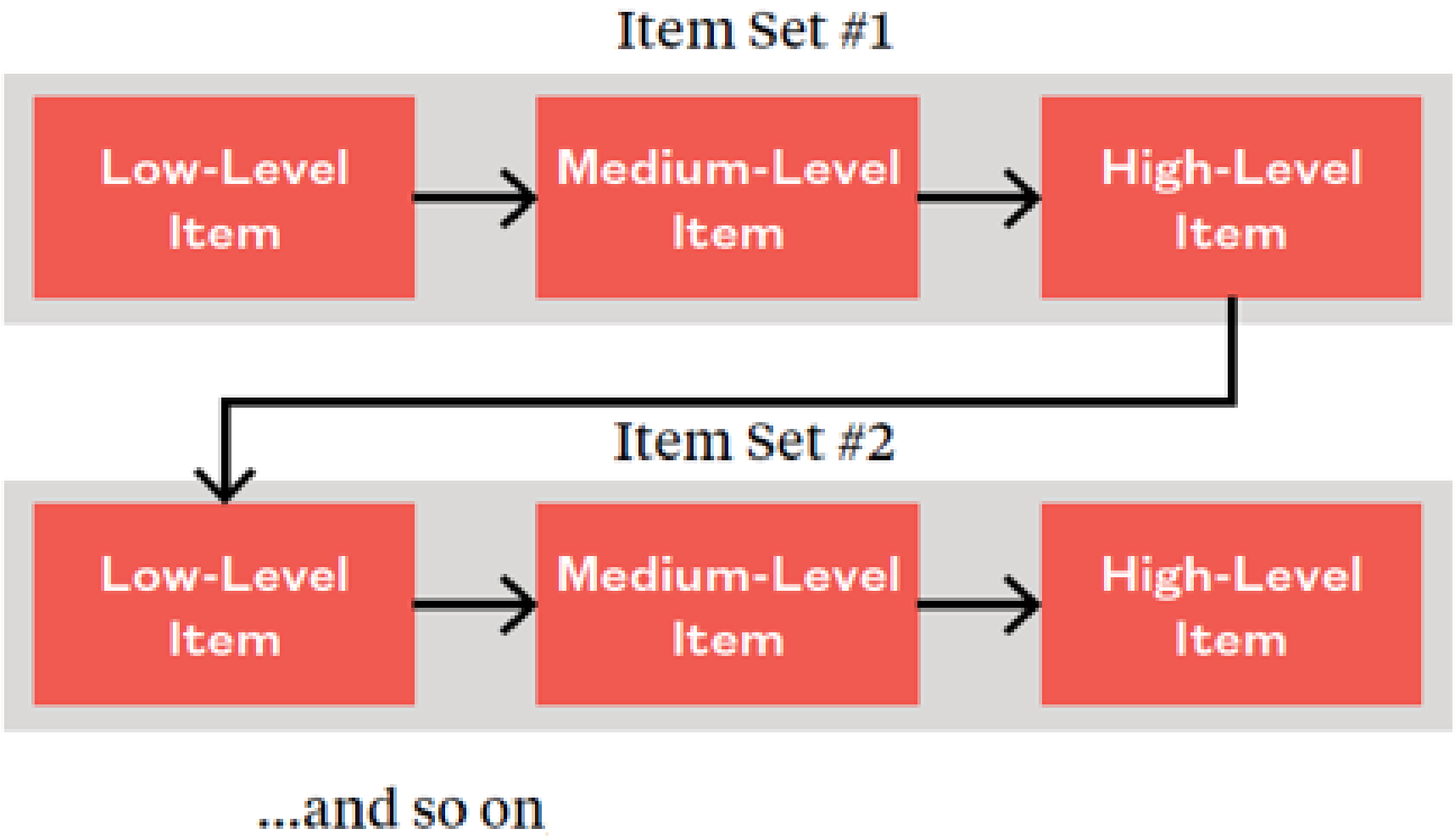
- Selected-response (multiple choice) items written at three levels with increasing cognitive demand
- Presented in a standardized format
  - Item stimulus
    - Phenomenon or context presented as the SAY section within the Test Administrator Booklet
  - Item prompt or question
    - Presented as the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

CSAA 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

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

CSAA overview - item set structure



CSAA 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

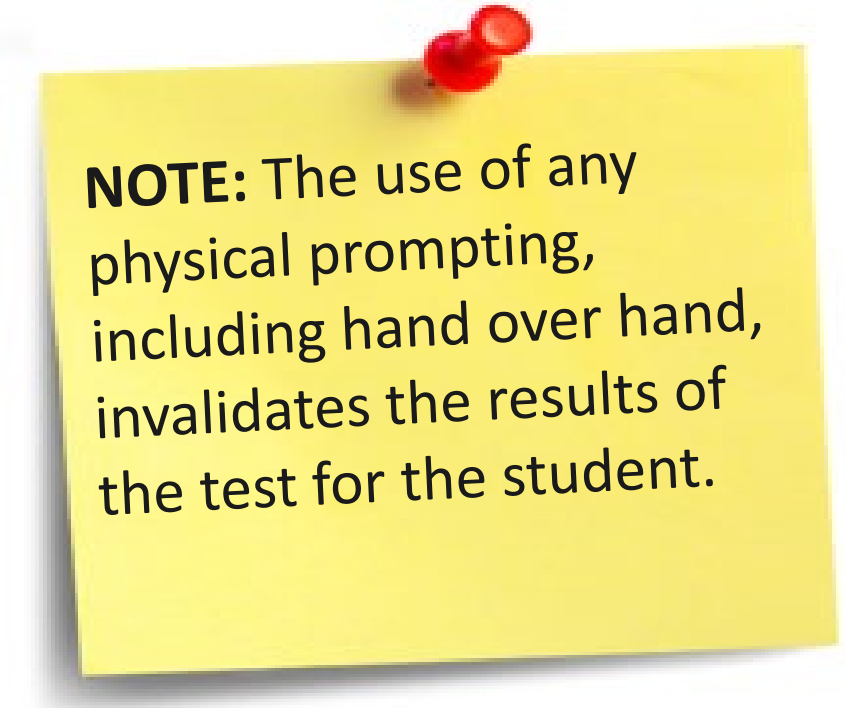
- Computer-Based Assessment Features
  - Answer Masking
  - Audio Player
  - Alternate Color Themes
  - Zoom
  - Increase Volume
  - Line Reader Tool
  - Read Aloud/Reread item directions, response options, passages

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
- Braille

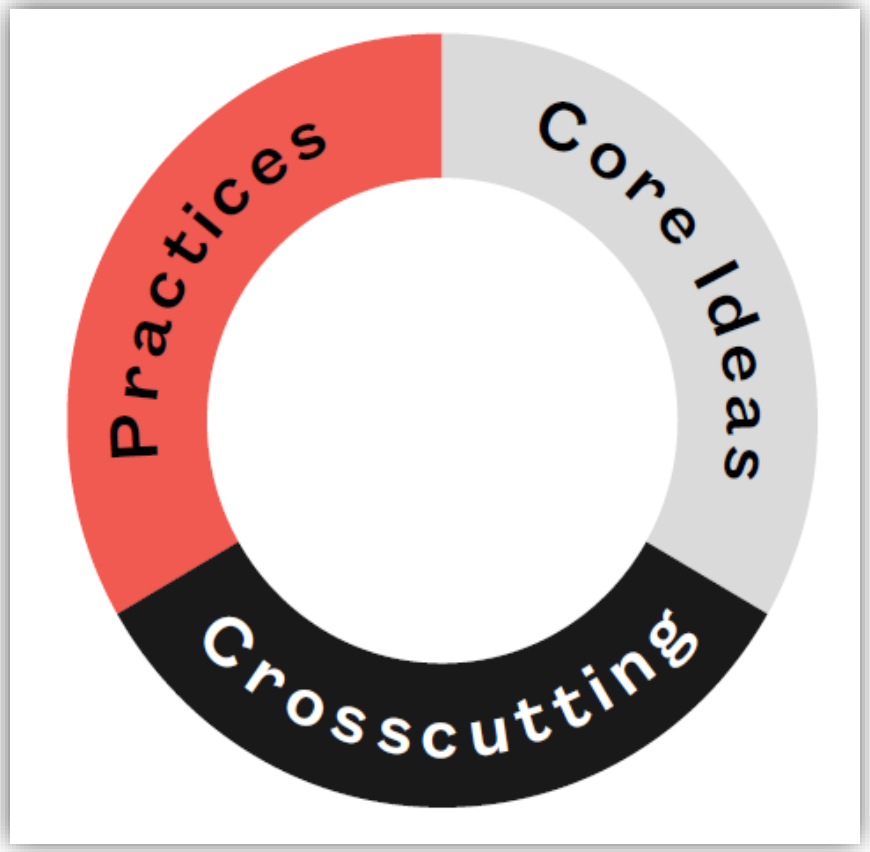


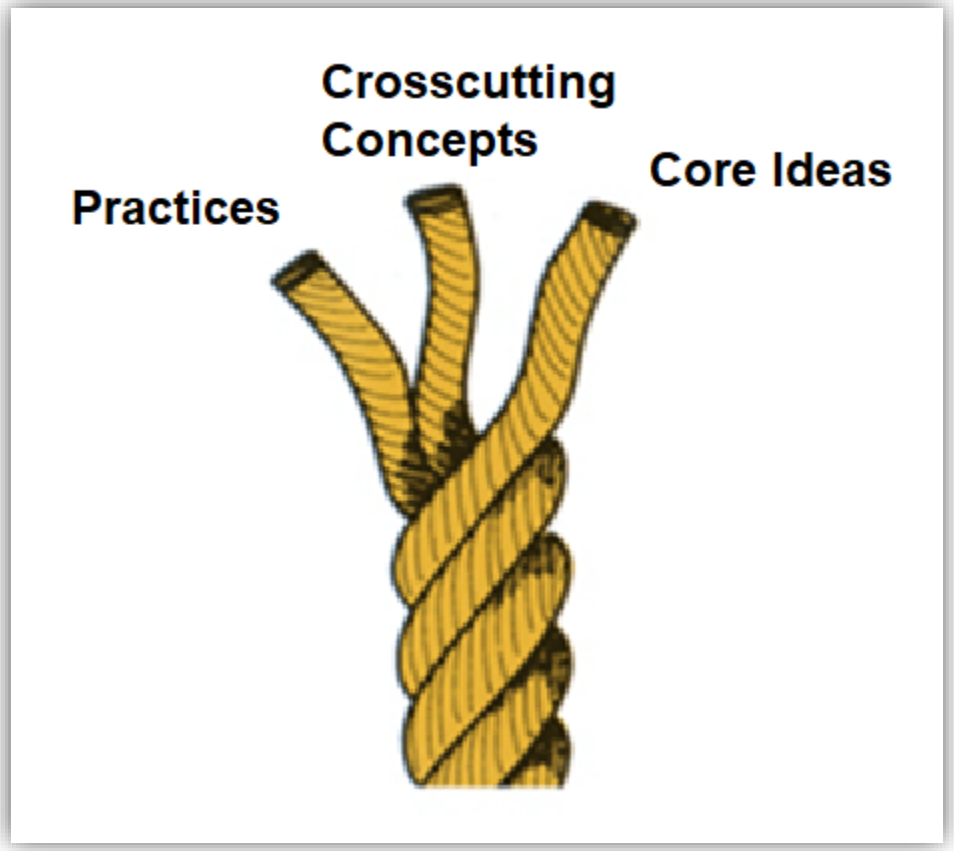
The three dimensions of science learning

- Science and Engineering Practices (SEPs)
  - What students are expected to do
- Disciplinary Core Ideas (DCIs)
  - What students are expected to know
- Crosscutting Concepts (CCCs)
  - How students think and connect ideas

Integrating dimensions

- The standards, called Performance Expectations, integrate all three dimensions in each standard.





Extended Performance Expectation (EPE)

- Aligned with *A Framework for K-12 Science Education*
- 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 band (elementary, middle, high) 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** • 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** • 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** • 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 Use observations to describe weather conditions.	**3-ESS2-1.2** Use tables or graphical displays of data to describe patterns of typical weather conditions in a particular season.	**3-ESS2-1.3** 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** • Represent data in tables and various graphical displays (bar graphs and pictographs) to reveal patterns that indicate relationships. **Supporting:** **Planning and Carrying Out Investigations**		
Disciplinary Core Ideas (DCI)	**ESS2.D: Weather and Climate** • 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** • 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: Level 1 item (3-ESS2-1 set)

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

Level 1	**3-ESS2-1.1**: Use observations to describe weather conditions.

Which word can Sofia use to describe the weather?

○ foggy ○ windy ○ sleepy

Example: Level 2 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

Level 2	**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: Level 3 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

Level 3	**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

What story do we want to tell about students?

- Students are able to use the majority of the disciplinary core ideas, practices, and crosscutting concepts to address moderately complex science phenomena and problems, some concrete and some abstract

What are PLDs?

- PLDs provide a narrative account of the knowledge, skills, and abilities that can be expected to be demonstrated by students in each level of achievement.
    - PLDs describe what students are likely to know and be able to do based on the Science Alternate Assessment EPEs.
    - PLDs inform stakeholders of how to interpret student test scores in relation to the Science Alternate Assessment EPEs.
    - PLDs are typically used for standard setting and score reporting.

How these PLDs are designed

- The PLDs you will utilize during the standard setting process are the **borderline** and **range** PLDs.
  - derived from the corresponding **policy** PLD, which describes the definition of each level.

Level 4 (Exceeding Expectations)
Students at Level 4 can be expected to demonstrate understanding and skills in 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 at Level 4.

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

How these PLDs are designed

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

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

How these PLDs are designed

- 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 •HS-PS3-2 **SEP** •Developing and Using Models •Asking Questions and Defining Problems (Supporting) **CCC** •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.

Standard Setting Panels

Grade 5	Facilitator: Matthew Gushta
	Room: Chaco
Grade 8	Facilitator: Dave Sanderson
	Room: Bandelier
Grade 11	Facilitator: Amy Leach
	Room: Tijeras

Thank you!

Instructions for Facilitators

Standard Setting Workshop

MSAA science, grade 8

July 26-28, 2022

cognia™

Welcome!

- Introductions
- One minute each panelist
  - Your name, school district, what you teach
  - Experience in other standard setting workshops
- Ask for show of hands
  - Who's been involved in SS before?
  - Which method(s)?

- Review the agenda (on next slide)

Agenda

- **Tuesday, July 26**
  - 10:45 AM – Noon: Key Concepts, Training & Practice
  - Noon – 1:00 PM: ***Lunch***
  - 1:00 PM – 2:00 PM: Key Concepts, Training & Practice
  - 2:00 PM – 3:13 PM: Familiarization with MSAA
  - 3:15 PM – 3:30 PM: ***Break***
  - 3:30 PM – 4:30 PM: Preparation for Round 1
  - 4:30 PM – 5:00 PM: Begin Round 1
  - 5:00 PM: **Adjourn**

Agenda

- **Wednesday, July 27**
  - 8:00 AM – 8:30 AM: ***Breakfast***
  - 8:30 AM – 9:00 AM: Debrief Day 1
  - 9:00 AM – 10:30 AM: Continue Round 1
  - 10:30 AM – 11:00 AM: ***Break***
  - 11:00 AM – Noon: Complete Round 1
  - Noon – 1:00 PM: ***Lunch***
  - 1:00 PM – 3:00 PM: Prepare for Round 2
  - 3:00 – 3:15 PM: ***Break***
  - 3:15 – 5:00 PM: Complete Round 2
  - 5:00 PM: **Adjourn**

Agenda

- **Thursday, July 28**
  - 8:00 AM – 8:30 AM: ***Breakfast***
  - 8:30 AM – 9:00 AM: Debrief Day 2
  - 9:00 AM – 10:00 AM: Prepare for Round 3
  - 10:00 AM – 10:15 AM: ***Break***
  - 10:15 AM – 11:30 AM: Complete Round 3
  - 11:30 AM – Noon: Debrief
  - Noon – 1:00 PM: ***Lunch***
  - 1:00 PM: **Adjourn**

Overview

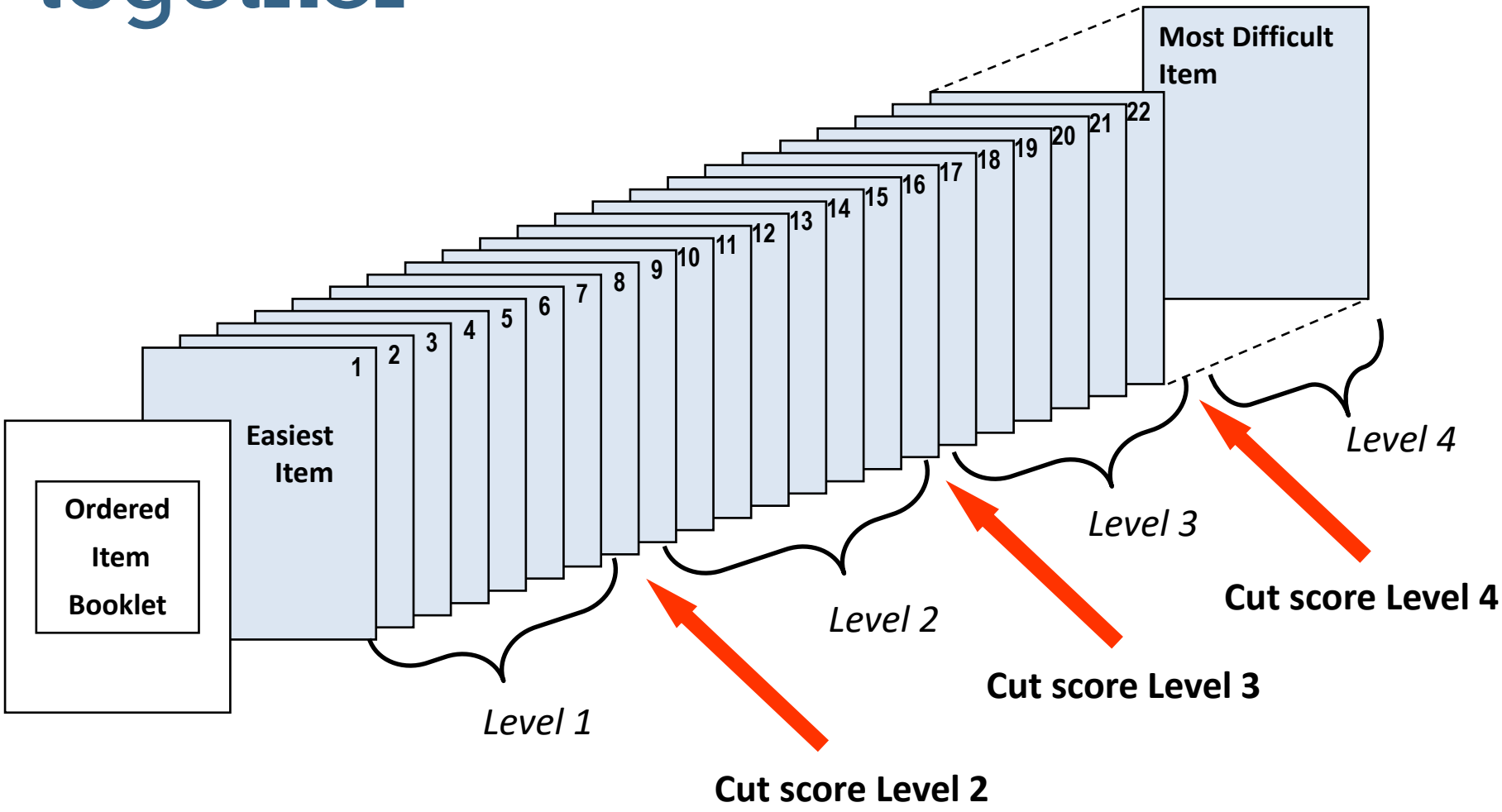
Rhythm

1. Become familiar with borderline PLDs, test items, training, and practice and using the standard setting tool
2. Prepare for round 1
3. Complete round 1
4. Review feedback from round 1, prepare for round 2
5. Complete round 2
6. Review feedback from round 2, prepare for round 3
7. Complete round 3

Overview (cont.)

- Our shared goals
  - Get your recommendations for performance standards for the MSAA science assessments that provide meaningful and actionable information
- Your goals as panelists
  - Learn the concepts and procedures to recommend cut scores, following the ID Matching method
  - Follow the procedures we train you on
  - Recommend cut scores for levels 2, 3, and 4
  - Rely on your expertise about the content standards, student learning, and students throughout the process

The outcome we're pursuing together



Note: Width of brackets is irrelevant

At each table

- Introductions
- Pick a table leader
  - Facilitate discussion at your table
  - Engage all panelists at your table
  - Ask for help from facilitator, psychometricians for tool, Cognia and MSAA content experts, as needed
  - No need to act as spokesperson for your table; individuals can speak up for themselves
  - Participate in cross-grade articulation, if needed

Key concepts and procedures

- Borderline PLDs
- ID Matching judgmental task
- Target cut scores
- Threshold regions
- Become familiar with test items
- Item map, OIB, online tool

- Rounds 1, 2, 3
- Prepare, Complete, Feedback

Performance Level Descriptors

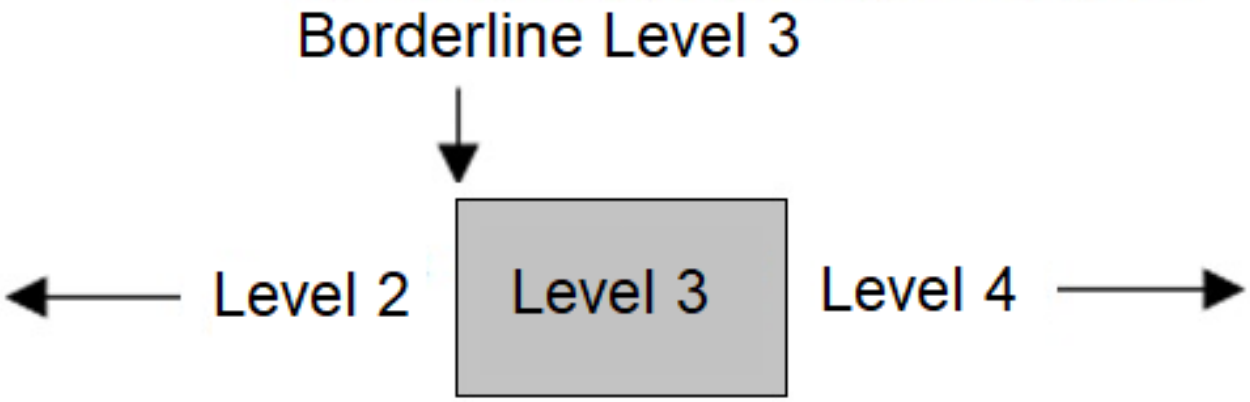
- Performance levels 1 through 4
- Performance Level Descriptors (PLDs) represent intended interpretations of solid student achievement within each level.
- Development of the PLDs began with the assumption that the grade-level content standards represent what students should know and be able to do at the end of a given grade level.

Performance Level Descriptors

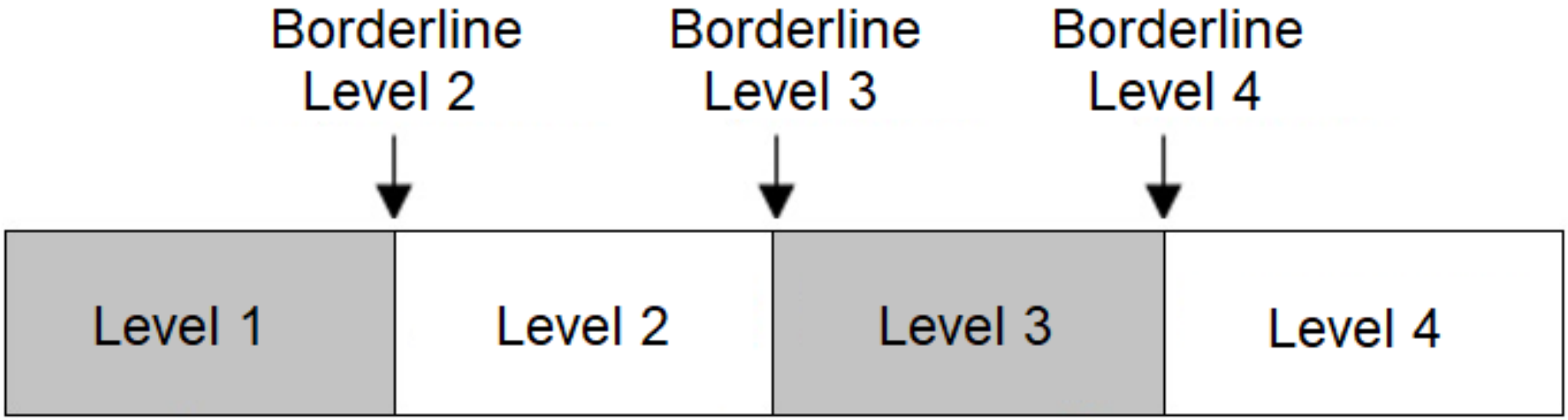
- **Range PLDs** describe the knowledge, skills, and abilities of students across the range of a specific performance level
  - For example, when the content standard says a student should be able to describe energy transfer, then the range PLD for Level 3 would explain that students should be able to describe the various ways that energy transfer can occur between everyday objects or devices.
- Our work over the next few days will be to determine the lines between students in adjacent performance levels
  - **Borderline PLDs** define knowledge and skill expectations that are *just barely* in a performance level

Borderline PLDs

- Borderline PLDs define what a student who is *just barely* in a given performance level is expected to know and be able to do
- Example: The level 3 Borderline PLD defines what a student who is *just barely* in level 3 (as opposed to level 2) can be expected to know and be able to do



Borderline PLDs



Understanding the Borderline PLDs (MS)

Level 2 (Approaching Expectations)

Students performing at the borderline of Level 2 Approaching Expectations can be expected in most situations[1] 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.

Understanding the Borderline PLDs (MS)

Level 3 (Meeting Expectations)

Students performing at the borderline of Level 3 Meeting Expectations can be expected in most situations[1] 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.

Understanding the Borderline PLDs (MS)

Level 4 (Exceeding Expectations)

Students performing at the borderline of Level 4 Exceeding Expectations can be expected in most situations1 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.

Understanding the Borderline PLDs (MS)

- MS-PS1-2 Range PLDs for Levels 2 through 4

Level 2	Level 3	Level 4
Identify properties of a substance by using data or observations.	Determine the identities of substances by using data or observations on the properties of substances.	Determine whether a chemical reaction occurred by using data or observations on the properties of substances before and after an interaction.

Understanding the Borderline PLDs (MS)

- Individual
- Table Discussion
- Room Discussion

- Highlight the progression from Borderline PLD Level 2, Borderline PLD Level 3, and Borderline PLD Level 4

Modeling: The ID matching process

1) Answer the two questions
   - What does a student need to know/be able to do to respond to this item/at this score level?
   - What makes this item more difficult than all previous items?
2) Match the items to a Borderline PLD
   - Explain how the item response demands align with Borderline PLD expectations

- I'll think out loud

Consideration

- You may judge that an item seems out of difficulty order
- There are no right or wrong answers—only your best professional judgments

Target cut scores

- What are target cut scores?
  - a score that shows a reasonable achievement level, based on a collective determination, for students taking the alternate assessment
- Why do we use them?
- How did Cognia and the Sci. Partners create them?

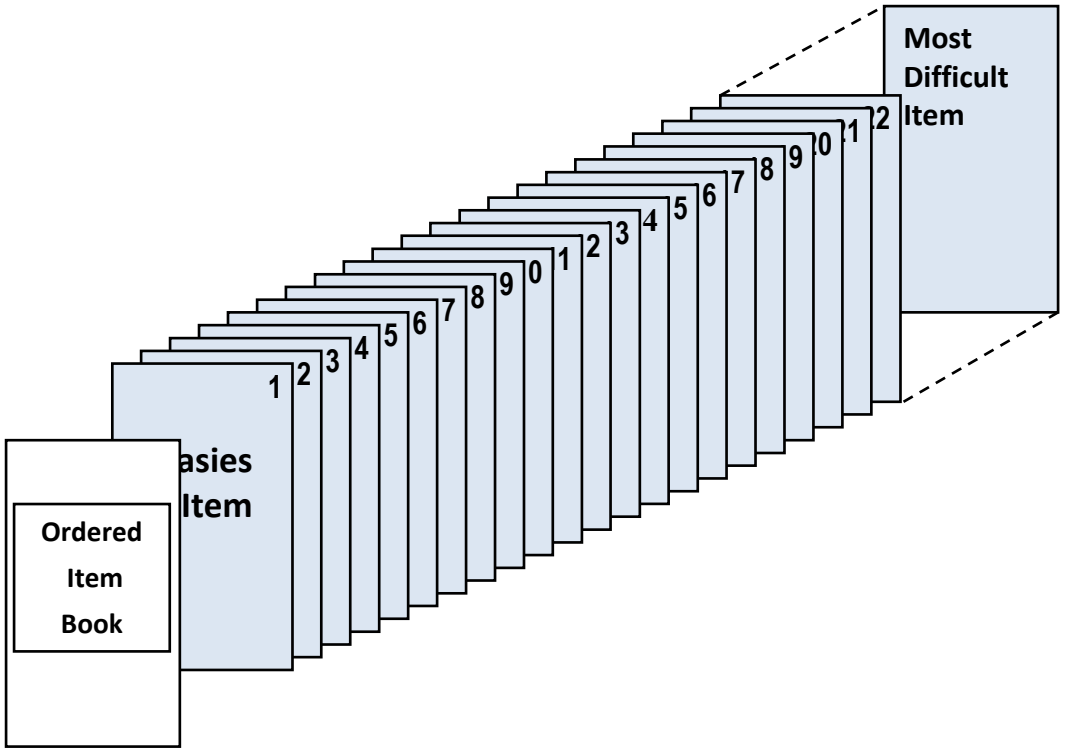
Target cut scores: Levels 2, 3 & 4

- We looked at past student performance on external tests (some gen. ed. and some alternate)
  - Specifically, percentages of students who performed at the Level 2, Level 3 and Level 4 on those tests
- We found a range of scores on the MSAA grade 8 science tests that would result in similar performance data
- The target cut score range represents aspirations for the achievement of these students.

Target cut scores: Level 2, 3 & 4

- Targets for you to consider; you can adjust them
- Your recommendation: retain or adjust the target cut scores
- Support your recommendations, whatever they may be, with your content-based rationales
- Author content-based rationales for Level 2, Level 3 and Level 4

Target cut scores, item map



28	L3	
27	L3	
26	L3	
25	L3	Target cut score range
24	L3	
23	L3	
22	**L3**	
21	L2	
20	L2	
19	L2	
18	L2	
17	L2	
16	L2	

Target cut scores: Levels 2, 3 & 4

- You're content experts
- Your job
  - Use your expertise on test content and students
  - Review response demands of items around the target cut score
  - Compare response demands to expectations in the PLDs
  - Recommend retaining or adjusting the target cut score
  - Write content-based rationales for all of your recommendations
- Throughout
  - You're free to recommend any cut score you think is appropriate
  - By following the ID Matching judgmental task
  - And writing content-based rationales

Target cut scores: Levels 2, 3 & 4

- How to think about them
  - A reasonable target to consider; a starting point
- What to do about them
  - Review at least three items above, at least three below—see the highlighted items in the tool
  - Retain or adjust the target cut scores
  - Write a content-based rationale for your cut score
- Can I move/adjust the target cut scores?
  - Yes, as long as you…
  - Support your recommendation with a content-based rationale

Content-based rationales

- For your Levels 2, 3 and 4 cut score recommendations
  - Write a content-based rationale for your recommendation to retain or adjust the cut score
- Relevant info for writing your rationale
  - Your answers to the two questions
  - The match between item response demands and expectations in the PLDs
- **Notice**: Refer to both **items** and **PLDs**

Content-based rationales (cont.)

- Specific and complete example
  - *The items above the Level 3 target cut score require XYZ; XYZ are described in the Level 3 PLD and not in the Level 2 PLD.*

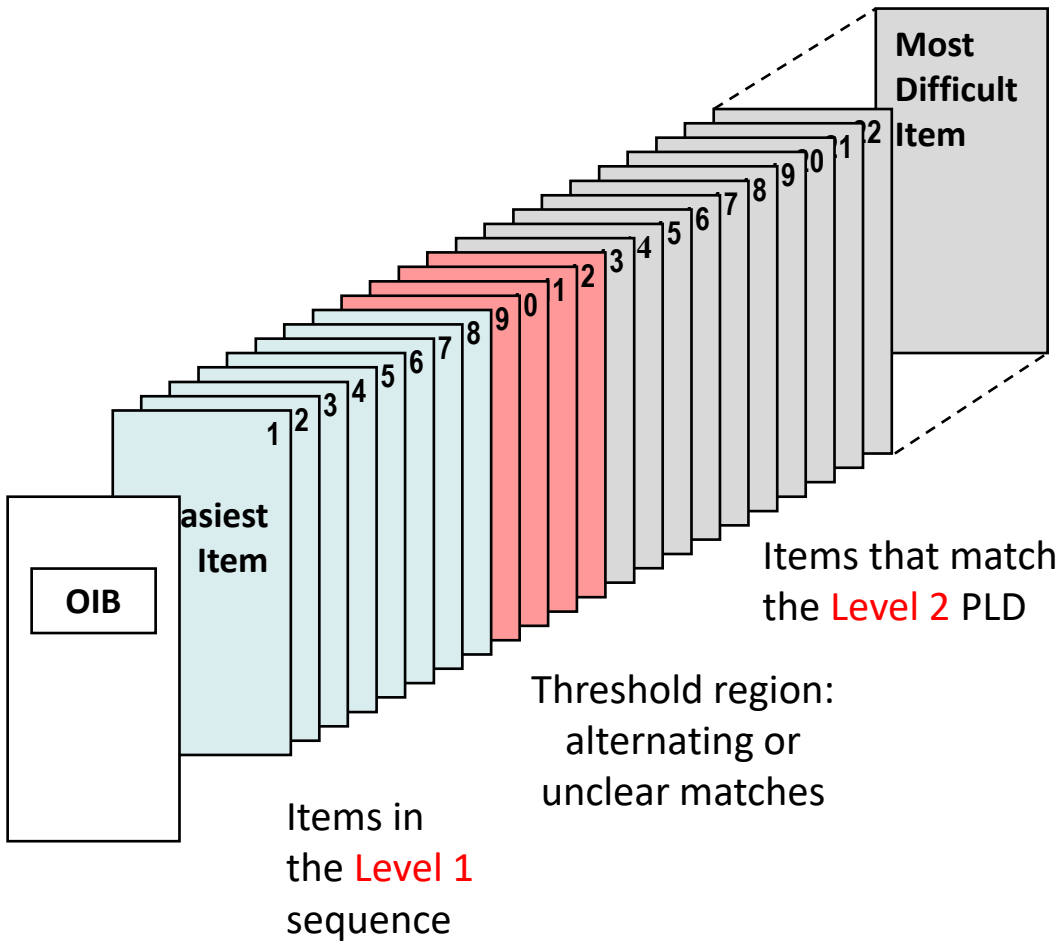
- Not adequately detailed example
  - *The items below the Level 4 cut score match the Level 3 PLD and do not match the Level 4 PLD.*

What is a threshold region?

- Threshold regions are a sequence of items that match two adjacent PLDs in an alternating and inconsistent sequence

- **Note**: If your threshold region is lengthy, go through the items at the top and bottom one more time—see if you can match some items to reduce the length
  - Don’t force it; match item response demands to PLD expectations

ID matches and threshold regions



19	L2	Item response demands clearly match the expectations in the Level 2 PLD
18	L2	
17	L2	
16	L2	
15	L2	
14	L2	
13	L2	
12	L1	Threshold region
11	L2	
10	L1	
9	L2	
8	L1	Item response demands **do not match** the expectations in the Level 2 PLD
7	L1	
6	L1	
5	L1	
4	L1	
3	L1	
2	L1	
1	L1	

Why do you end up with threshold regions?

- Reasons why panelists put some items in threshold regions
  - The response demands of these items reflect some expectations in the Level 3 PLD (for example), and some expectations in the Level 2 PLD
  - I can't make up my mind yet which PLD this item most closely matches

---

- **Note**: If your threshold region is lengthy, go through the items at the top and bottom one more time—see if you can match some items to reduce the length
  - Don’t force it; reconsider matches of item response demands to PLD expectations

End of training and practice

- We've now covered all the concepts and procedures you need to complete your tasks
  - Some are abstract and complex; we'll come back to them as needed
  - You should ask for explanations any time you want
- Do you feel ready to prepare for round 1?
- What questions, concerns, etc. remain?
- Discussion

Short Break

- Short break; we'll log your computer onto the internet

Experience the grade 8 test

- Take the grade 8 test; respond to items as if your score mattered to you
- As you process, understand, and respond to items, consider how students would do that
  - Consider their exposure to the needed content and skills
  - Try to think of students *besides* those in your classroom, school, and district
- We have 60 minutes for this activity
- Table and room, general discussion about the experience
- **Once you have completed your test, you can take a break. Please be back by 3:30 pm.**

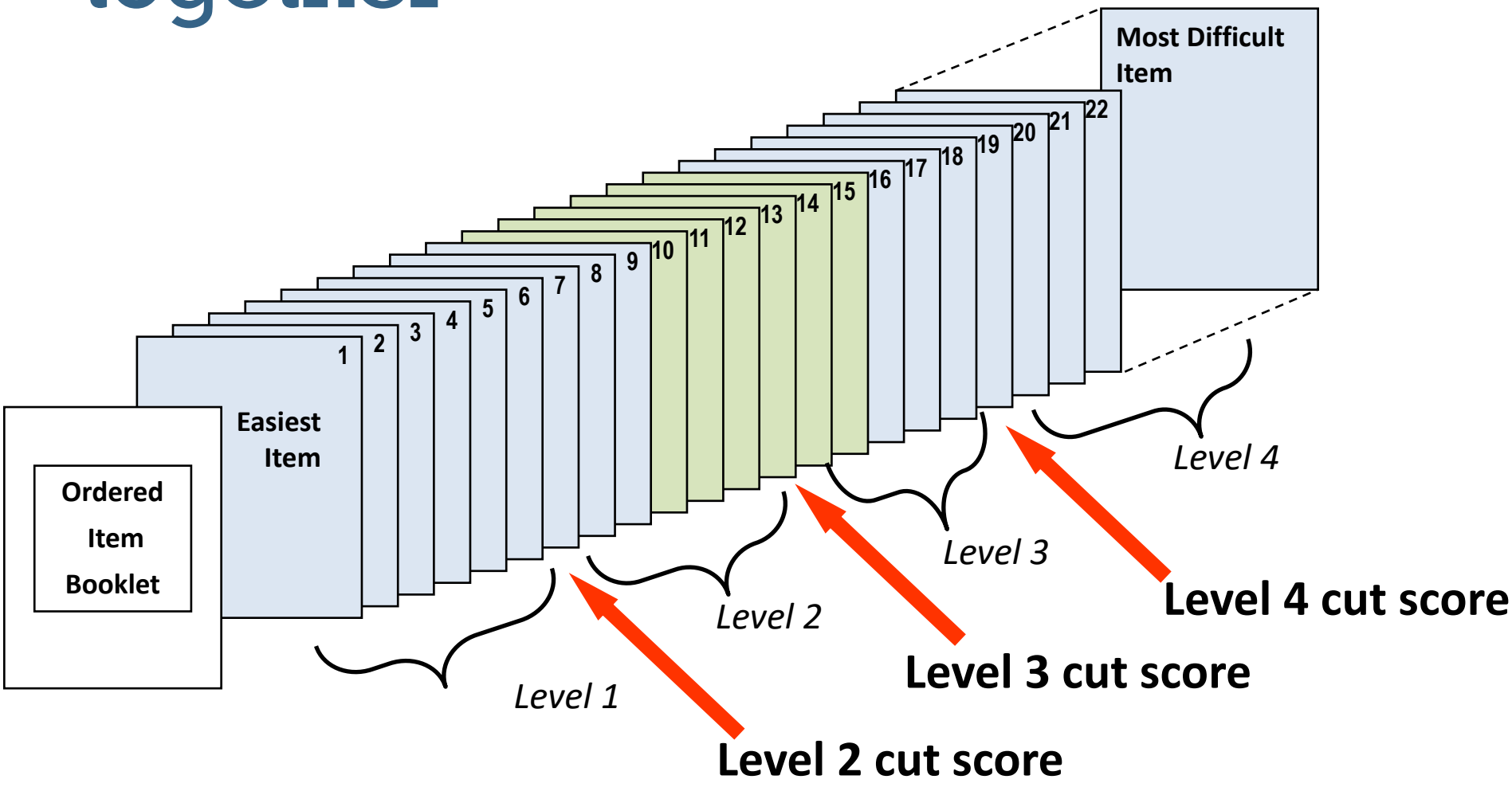
Break

- Once you have completed your test, you can take a break. Please be back by 3:30 pm.

Prepare for Round 1: Review

- The ID matching judgmental task
- Recommend retaining or adjusting the Level 3 target cut score
  - Write a content-based rationale for your recommendation
- Then move on to Level 4 and Level 2 cut scores

The outcome we're pursuing together



Note: Width of brackets is irrelevant

Standard setting tool

- Chrome Book navigation
  - Maximize browser screen
  - Log into the tool
  - View borderline PLDs, content standards
  - Tabs at top of browser screen
- Item map—all items, ordered by difficulty
  - Enter notes and judgments
  - Enter PLD matches and cut scores (later)
- Item view
  - Items, stimuli, rubrics
  - Multiple stimuli (left to right)

cognia™

Panelist practice

- Review of Process
  - Answer the two questions,
  - Match items to PLDs
  - Explain your thinking as a content-based rationale
- Panelists practice **independently**; enter answers to two Qs in the tool
- Discussion: Share insights, look for shared understandings, no persuasion

(1) What does a student need to know and be able to do in order to respond to this item?
(2) What makes this item more difficult than the preceding items?

Which PLD most closely matches the knowledge and skill demands for each item?

Hint: Items are ordered by difficulty.

cognia™

Round 1 steps

1. Answer the two questions
   - Start at the bottom item in the highlighted range, work above
2. Match items to PLDs
3. Recommend cut scores, write content-based rationales

(1) What does a student need to know and be able to do in order to respond to this item?
(2) What makes this item more difficult than the preceding items?

- Work independently
- Trust your expertise
- Take notes in the tool

Are you ready to undertake round 3?

- Ask final questions
- Ask for more explanation, demonstration of steps

- Complete the Readiness Survey for round 2
  - Open the survey in the tool

- Work independently
- Pacing guidance/time allotment – provided by facilitator

Grade 8, round 1

Level 3 (Orange/Green)

1. Answer the two questions
   - Start at the bottom item in the highlighted range, work up
2. Match items to PLDs
3. Recommend cut score, write content-based rationales

The items above the Level 3 target cut score require XYZ; XYZ are described in the Level 3 PLD and not in the Level 2 PLD.

Level 4 (Blue) and Level 2 (Yellow/Orange)

1. Answer the two questions,
   - Start at the bottom item in the highlighted range, work up
2. Match items to PLDs
3. Recommend cut scores , write content-based rationales

All Three Cut Scores
Work independently
Trust your judgment
Take notes in the tool as helpful

cognia™

Round 1 feedback

- Median for the room
  - Each anonymous panelist: highest and lowest OIB page
- Using the feedback
  - Demonstrate reasoning for OIB pages around the median cut score and outlier pages
  - Share insights
  - **No right or wrong**, **no persuasion to change**
  - What matters is *sharing the reasoning* for each page
- Table and room discussion

Table and room discussions

- In all discussions with your colleagues
- Your goals:
  - Share your insights
  - Listen to your colleagues' insights
  - Develop shared understandings
  - Support independent decision making
  - Courtesy and respect are paramount
- **Not** your goals:
  - Agree with your colleagues
  - Persuade your colleagues to agree with you
  - Reach consensus

Are you ready to undertake round 2?

- Ask final questions
- Ask for more explanation, demonstration of steps

- Complete the Readiness Survey
- Work independently
- 20-30 minutes? (more if needed)

Grade 8, round 2

Level 3

1. Answer the two questions
   - For items affected if you are changing your cut score
2. Match items to PLDs
3. Update cut scores, if needed, write content-based rationales

Level 4 and 2

1. Answer the two questions, starting at…
   - For items affected if you are changing your cut score
2. Match items to PLDs
3. Update cut scores, if needed

All Three Cut Scores

Work independently

Trust your judgment

Take notes in the tool as helpful

cognia™

Round 2 feedback

- Median for the room
  - Each anonymous panelist: highest and lowest OIB page
- Using the feedback
  - Demonstrate reasoning for OIB pages around the median cut score and outlier pages
  - Share insights
  - **No right or wrong**, **no persuasion to change**
  - What matters is *sharing the reasoning* for each page
- Table and room discussion

Insert feedback displays

- Insert three cut score bar charts from Word doc

Table and room discussions

- In all discussions with your colleagues
- Your goals:
  - Share your insights
  - Listen to your colleagues' insights
  - Develop shared understandings
  - Support independent decision making
  - Courtesy and respect are paramount
- **Not** your goals:
  - Agree with your colleagues
  - Persuade your colleagues to agree with you
  - Reach consensus

Are you ready to undertake round 3?

- Ask final questions
- Ask for more explanation, demonstration of steps

- Complete the Readiness Survey
- Work independently
- 20-30 minutes? (more if needed)

Grade 8, round 3

Level 3

1. Answer the two questions
   - For items affected if you are changing your cut score
2. Match items to PLDs
3. Update cut scores, if needed, write content-based rationales

Level 4 and 2

1. Answer the two questions, starting at…
   - For items affected if you are changing your cut score
2. Match items to PLDs
3. Update cut scores, if needed

All Three Cut Scores

Work independently

Trust your judgment

Take notes in the tool as helpful

cognia™

Round 3 feedback

- Median for the room
  - Each anonymous panelist: highest and lowest OIB page

Insert feedback displays

- Insert three cut score bar charts from Word doc

Closing session

- Review final results; discussion
  - Will there be cross-grade articulation?
- Complete the workshop evaluation in the tool
- Dismissal

Thank you for your insightful discussions and excellent work!

APPENDIX—D
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 sample (1) login screen, (2) practice item booklet, (3) readiness survey screen, (4) ordered item booklet view, (5) item view, and (6) completion survey. A brief description accompanies each image.

Figure 1. Sample Login Screen

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

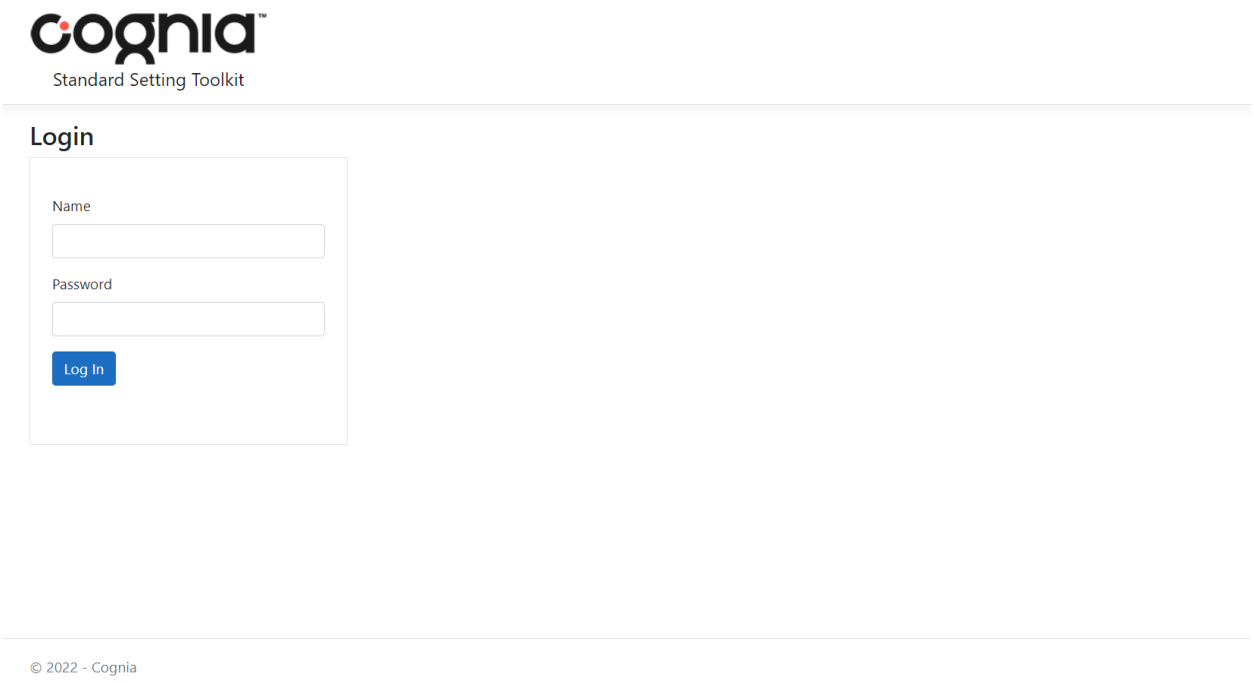


Figure 2. Sample Practice Item Booklet

This image shows a list of sample practice items as a truncated ordered item booklet. Panelists use the practice item booklet during the practice round to become familiar with use of the tool and to practice the ID matching process.

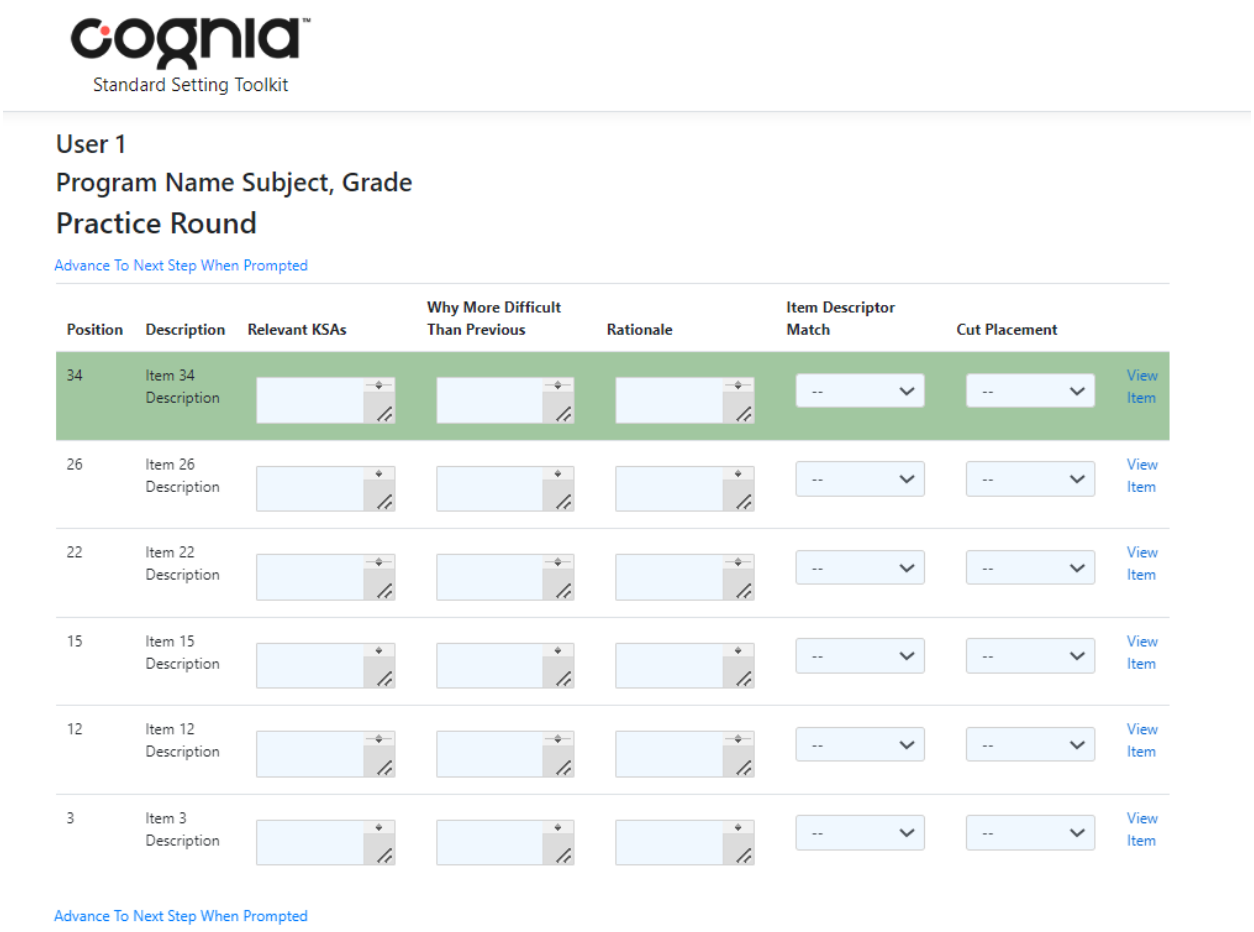


Figure 3. Sample Readiness Survey

Before each round of judgements, panelists complete a readiness survey to indicate whether they are ready to undertake the associated judgement round.

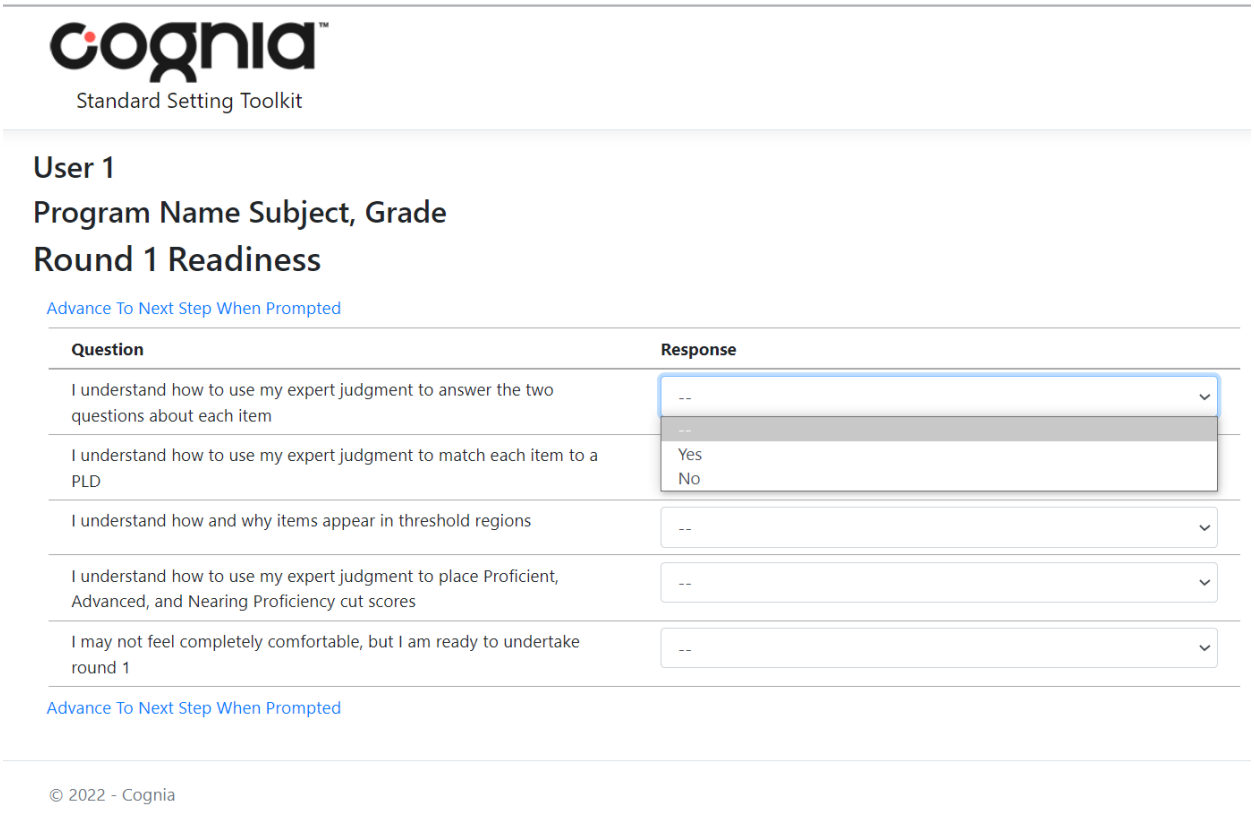


Figure 4. Sample Ordered Item Booklet View (truncated)

This image shows a sample truncated view of the complete ordered item booklet as displayed on panelists' screens. As a reminder, the item list is ordered from easiest (at the bottom) to most difficult (at the top).

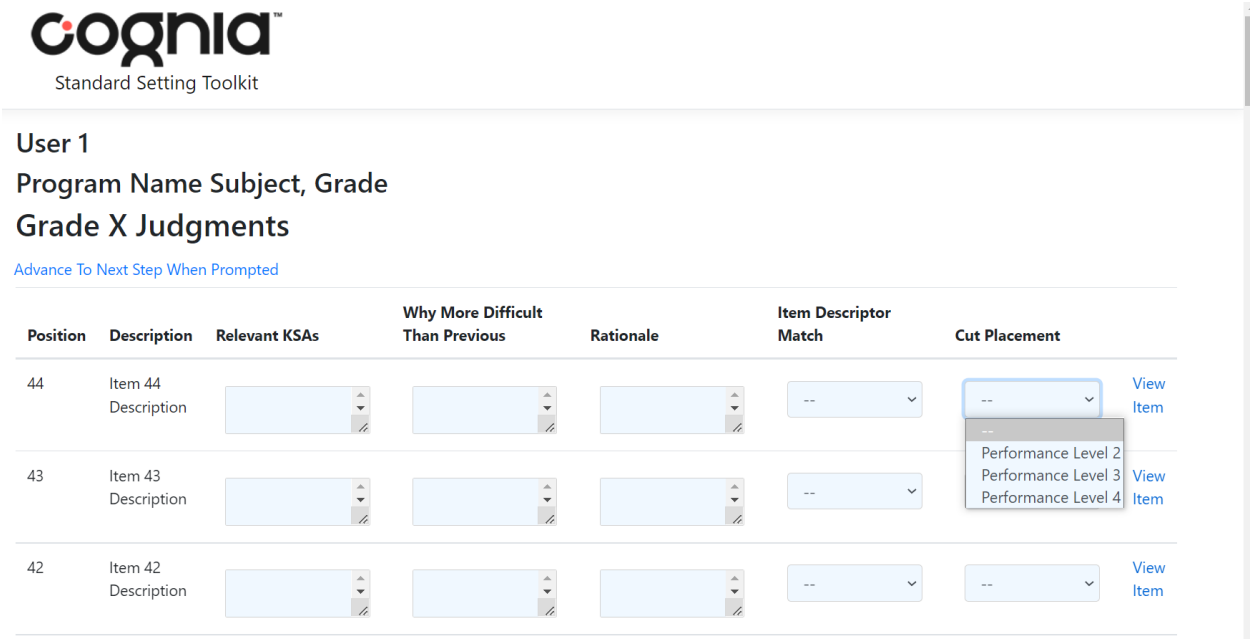


Figure 5. Sample Item View

The item view displays a PDF of the item along with links to any associated stimuli and/or rubrics. In addition, notes below the item provide the item description, the associated standard, and (when relevant) notes about possible score points for the item.

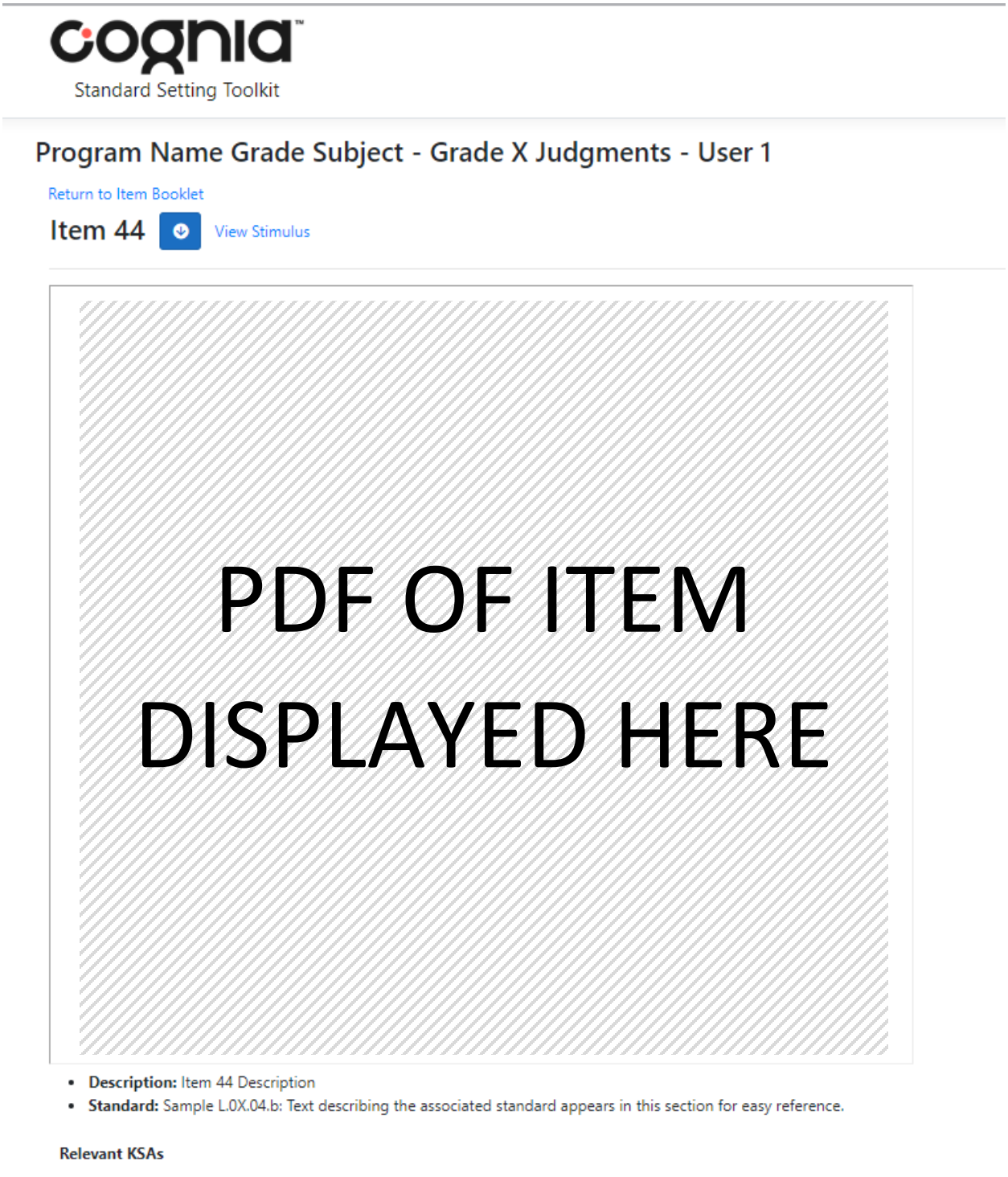


Figure 6. Sample Judgement Form

The judgement form provides space for panelists to write notes about (1) the relevant knowledge, skills, and abilities (KSAs) needed to respond to the item, (2) why the item is more difficult than the previous item, and (3) content-based rationales. In addition, dropdown menus are provided for the item descriptor matches and the cut placements. Note that the judgement form can be accessed through both the booklet view and the item view.

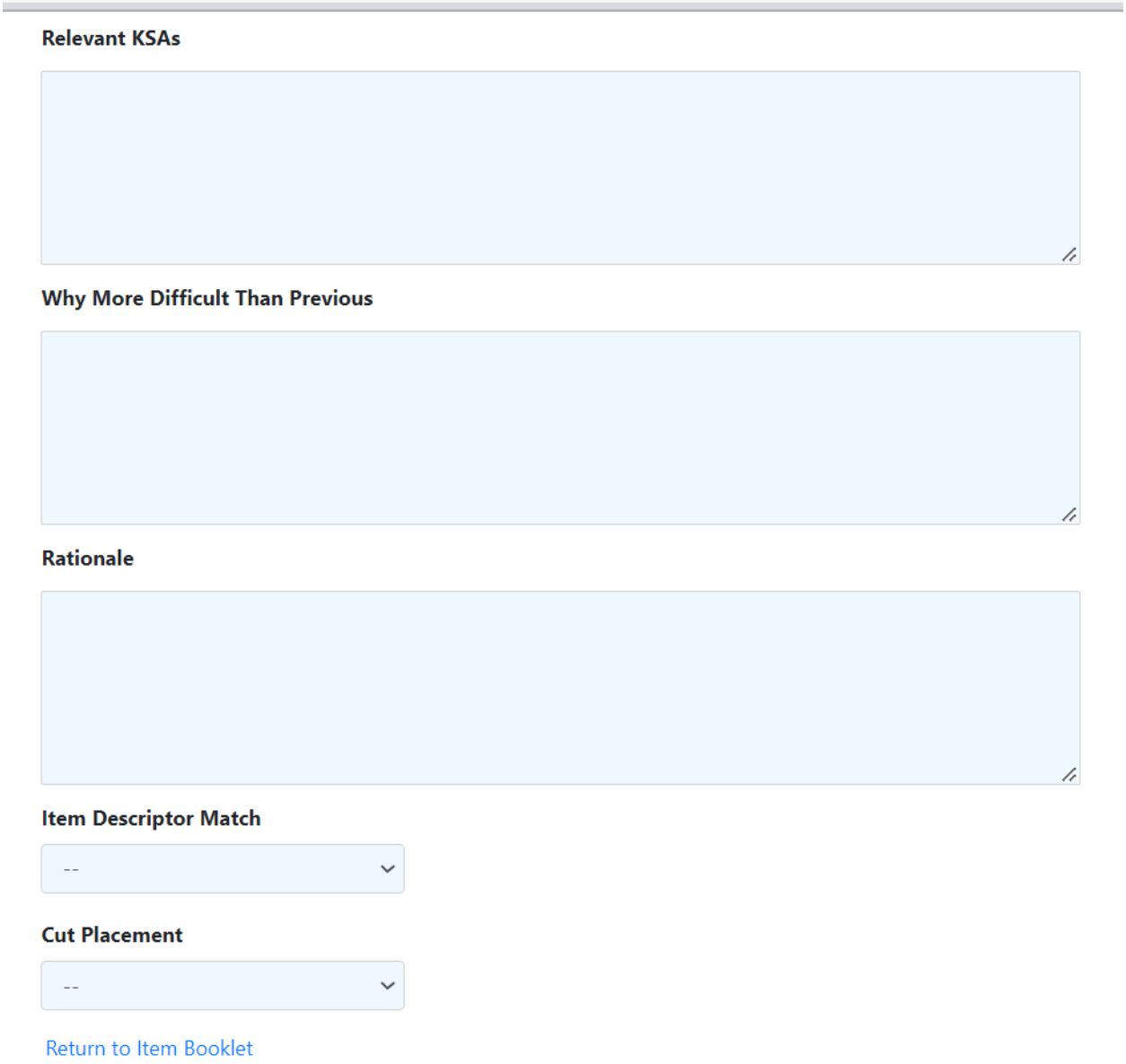
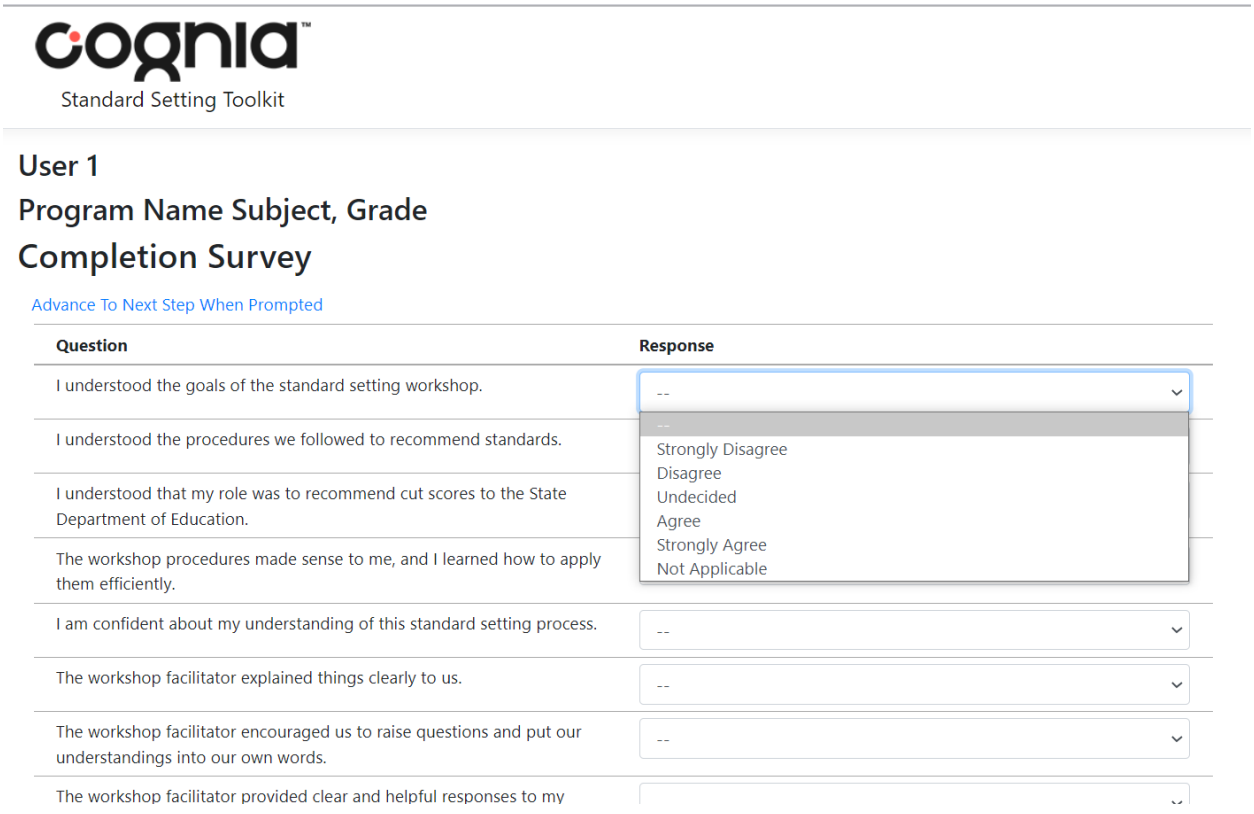


Figure 7. Sample Completion Survey (truncated)

This image provides a truncated view of the completion survey provided to panelists at the end of the standard setting meeting to collect their final evaluations and feedback on various aspects of the meeting.



APPENDIX—E

READINESS SURVEYS AND RESULTS

Readiness Surveys and Results

Round 1 Readiness Survey

Survey Questions	Response Options	
	Yes	No
I understand how to use my expert judgment to answer the two questions about each item	25	0
I understand how to use my expert judgment to match each item to a PLD	25	0
I understand how and why items appear in threshold regions	25	0
I understand how to use my expert judgment to place Level 2, Level 3, and Level 4 cut scores	25	0
I may not feel completely comfortable, but I am ready to undertake round 1	25	0

Round 2 Readiness Survey

Survey Questions	Response Options	
	Yes	No
I know that feedback and discussion in preparation for round 2 will help me feel even more comfortable	25	0
I understand the round 1 feedback about (a) our group cut scores for Level 2, Level 3, and Level 4, and (b) the highest and lowest panelist cut scores for each level	25	0
I understand the ground rules for discussing feedback in preparation for round 2: sharing information, avoiding persuasion	25	0
I understand that I should use the round 1 feedback as information, not persuasion, for me to consider as I place my cut scores in round 2	25	0
I'm ready to undertake round 2	25	0

Round 3 Readiness Survey

Survey Questions	Response Options	
	Yes	No
I know that feedback and discussion in preparation for round 3 will help me feel even more comfortable	25	0
I understand the round 2 feedback about (a) our group cut scores for Level 2, Level 3, and Level 4, and (b) the highest and lowest panelist cut scores for each level	25	0
I understand the ground rules for discussing feedback in preparation for round 3: sharing information, avoiding persuasion	25	0
I understand that I should use the round 2 feedback as information, not persuasion, for me to consider as I place my cut scores in round 3	25	0
I'm ready to undertake round 3	25	0

APPENDIX—F

WORKSHOP EVALUATION SURVEY AND RESULTS

Workshop Evaluation Survey

This appendix provides the list of survey questions and response options for the workshop evaluation survey. As a reminder, panelists complete the workshop evaluation survey at the end of the standard setting meeting. The survey is delivered and completed within the Cognia Standard Setting Toolkit.

Survey Questions and Response Options:

Note: Questions 1 – 20 were on a Likert scale and included the following response options: Strongly Disagree, Disagree, Undecided, Agree, Strongly Agree, and Not Applicable.

1. I understood the goals of the standard setting workshop.
2. I understood the procedures we followed to recommend standards.
3. I understood that my role was to recommend cut scores to the State Department of Education.
4. The workshop procedures made sense to me, and I learned how to apply them efficiently.
5. I am confident about my understanding of this standard setting process.
6. The workshop facilitator explained things clearly to us.
7. The workshop facilitator encouraged us to raise questions and put our understandings into our own words.
8. The workshop facilitator provided clear and helpful responses to my questions and other requests for clarification.
9. The workshop facilitator took steps to help the standard setting process run smoothly.
10. Sufficient time was allotted for training and practice on the standard setting concepts, tasks, and procedures.
11. I understood the progressions in expectations across Level 2, Level 3, and Level 4 as defined by the borderline Performance Level Descriptors.
12. I became sufficiently familiar with the assessment to recommend cut scores, based on responding to items on the test and answering the two questions about items.
13. I understood the ID Matching task, including answering the two questions about each item, matching those item response demands to PLDs, and how to place cut scores.
14. I understood how to use the standard setting tool to record my responses regarding skills and item difficulties as instructed.
15. I understood how to use the standard setting tool to record my recommended cut scores.
16. I understood that I could recommend retaining or adjusting the target cut scores.
17. I understood how to write content-based rationales for my cut score recommendations.
18. I understood that the group recommended cut score was the average (i.e., the median) of all 13 recommended cut scores.
19. I understood how to use the feedback after round 1 on the group recommended cut score and the individual panelist highest and lowest cut scores, in preparation for round 2.
20. I am satisfied with the final group cut scores. I would not recommend changing any of the group cut scores.

Note: Question 21 was on a Likert scale and included the following response options: Up 2 Pages, Up 1 Page, Do Not Move, Down 1 Page, Down 2 Pages, Not Applicable

21. If no, would you recommend moving a cut score up or down in the OIB, and by how many pages?

Note: Questions 22 and 23 were open-ended.

22. Please indicate any parts of the standard setting training and process that we should improve.
23. Please indicate any concerns you may have about the workshop process and the final recommended cut scores.

MSAA Science Standard Setting Workshop Evaluation Results

Table F-1. Frequency of Evaluation Responses for Grade 5

	Not Applicable	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree
I understood the goals of the standard setting workshop.					5	4
I understood the procedures we followed to recommend standards.		1			2	6
I understood that my role was to recommend cut scores to the State Department of Education.					1	8
The workshop procedures made sense to me, and I learned how to apply them efficiently.				1	4	4
I am confident about my understanding of this standard setting process.					4	5
The workshop facilitator explained things clearly to us.		1			3	5
The workshop facilitator encouraged us to raise questions and put our understandings into our own words.		1			2	6
The workshop facilitator provided clear and helpful responses to my questions and other requests for clarification.		1		1		7
The workshop facilitator took steps to help the standard setting process run smoothly.		1		1	1	6
Sufficient time was allotted for training and practice on the standard setting concepts, tasks, and procedures.				1	1	7
I understood the progressions in expectations across the Nearing Proficiency, Proficient, and Advanced levels as defined by the borderline Performance Level Descriptors.					5	4
I became sufficiently familiar with the assessment to recommend cut scores, based on responding to items on the test and answering the two questions about items.					3	6
I understood the ID Matching task, including answering the two questions about each item, matching those item response demands to PLDs, and how to place cut scores.					4	5
I understood how to use the standard setting tool to record my responses regarding skills and item difficulties as instructed.					5	4
I understood how to use the standard setting tool to record my recommended cut scores.					5	4
I understood that I could recommend retaining or adjusting the target cut scores.					5	4
I understood how to write content-based rationales for my cut score recommendations.				1	4	4
I understood that the group recommended cut score was the average (i.e., the median) of all 13 recommended cut scores.		1			2	6
I understood how to use the feedback after round 1 on the group recommended cut score and the individual panelist highest and lowest cut scores, in preparation for round 2.	1				4	4
I am satisfied with the final group cut scores. I would not recommend changing any of the group cut scores.				1	5	3
	Up 2 Pages	Up 1 Page	Do Not Move	Down 1 Page	Down 2 Pages	Not Applicable
If no, would you recommend moving a cut score up or down in the OIB, and by how many pages?	2		4	1		2

Table F-2. Frequency of Evaluation Responses for Grade 8

	Not Applicable	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree
I understood the goals of the standard setting workshop.					2	6
I understood the procedures we followed to recommend standards.					5	3
I understood that my role was to recommend cut scores to the State Department of Education.					4	4
The workshop procedures made sense to me, and I learned how to apply them efficiently.					2	6
I am confident about my understanding of this standard setting process.					2	6
The workshop facilitator explained things clearly to us.					1	7
The workshop facilitator encouraged us to raise questions and put our understandings into our own words.						8
The workshop facilitator provided clear and helpful responses to my questions and other requests for clarification.						8
The workshop facilitator took steps to help the standard setting process run smoothly.						8
Sufficient time was allotted for training and practice on the standard setting concepts, tasks, and procedures.					4	4
I understood the progressions in expectations across the Nearing Proficiency, Proficient, and Advanced levels as defined by the borderline Performance Level Descriptors.					4	4
I became sufficiently familiar with the assessment to recommend cut scores, based on responding to items on the test and answering the two questions about items.					3	5
I understood the ID Matching task, including answering the two questions about each item, matching those item response demands to PLDs, and how to place cut scores.					3	5
I understood how to use the standard setting tool to record my responses regarding skills and item difficulties as instructed.					4	4
I understood how to use the standard setting tool to record my recommended cut scores.					3	5
I understood that I could recommend retaining or adjusting the target cut scores.					3	5
I understood how to write content-based rationales for my cut score recommendations.					5	3
I understood that the group recommended cut score was the average (i.e., the median) of all 13 recommended cut scores.					4	4
I understood how to use the feedback after round 1 on the group recommended cut score and the individual panelist highest and lowest cut scores, in preparation for round 2.					4	4
I am satisfied with the final group cut scores. I would not recommend changing any of the group cut scores.					2	6
	Up 2 Pages	**Up 1 Page**	**Do Not Move**	**Down 1 Page**	**Down 2 Pages**	**Not Applicable**
If no, would you recommend moving a cut score up or down in the OIB, and by how many pages?		2	3			3

Table F-3. Frequency of Evaluation Responses for Grade 11

	Not Applicable	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree
I understood the goals of the standard setting workshop.		1				7
I understood the procedures we followed to recommend standards.					2	6
I understood that my role was to recommend cut scores to the State Department of Education.						8
The workshop procedures made sense to me, and I learned how to apply them efficiently.					4	4
I am confident about my understanding of this standard setting process.					2	6
The workshop facilitator explained things clearly to us.					1	7
The workshop facilitator encouraged us to raise questions and put our understandings into our own words.					2	6
The workshop facilitator provided clear and helpful responses to my questions and other requests for clarification.					2	6
The workshop facilitator took steps to help the standard setting process run smoothly.						8
Sufficient time was allotted for training and practice on the standard setting concepts, tasks, and procedures.					3	5
I understood the progressions in expectations across the Nearing Proficiency, Proficient, and Advanced levels as defined by the borderline Performance Level Descriptors.					1	7
I became sufficiently familiar with the assessment to recommend cut scores, based on responding to items on the test and answering the two questions about items.					2	6
I understood the ID Matching task, including answering the two questions about each item, matching those item response demands to PLDs, and how to place cut scores.					2	6
I understood how to use the standard setting tool to record my responses regarding skills and item difficulties as instructed.					2	6
I understood how to use the standard setting tool to record my recommended cut scores.					1	7
I understood that I could recommend retaining or adjusting the target cut scores.					1	7
I understood how to write content-based rationales for my cut score recommendations.					4	4
I understood that the group recommended cut score was the average (i.e., the median) of all 13 recommended cut scores.					2	6
I understood how to use the feedback after round 1 on the group recommended cut score and the individual panelist highest and lowest cut scores, in preparation for round 2.					2	6
I am satisfied with the final group cut scores. I would not recommend changing any of the group cut scores.				1	3	4
	Up 2 Pages	**Up 1 Page**	**Do Not Move**	**Down 1 Page**	**Down 2 Pages**	**Not Applicable**
If no, would you recommend moving a cut score up or down in the OIB, and by how many pages?	1	1	5		1	

APPENDIX—G
IMPACT DATA BY GRADE

Grade 5

Item Number	Op Position	OIB Position	Step	rp50	Benchmark Incorrect	Benchmark Cut Incorrect	Benchmark Correct	Benchmark Cut Correct	SS Final Cut	As Reported	Corrected
557040	23	1	1	-0.58149						12%	42%
556998	19	2	1	-0.52275							
555746	2	3	1	-0.45421							
557035	21	4	1	-0.43084							
780756	37	5	1	-0.42491			3				
780723	34	6	1	-0.39953			3				
780734	35	7	1	-0.38509			3				
555799	4	8	1	-0.35025			3	Level 2 Cut			
555792	3	9	1	-0.32398			3				
555837	5	10	1	-0.29946			3				
557052	25	11	1	-0.2566			3				
560418	28	12	1	-0.22532	3				**Level 2 Cut**	28%	5%
555737	1	13	1	-0.21927	3						
555974	13	14	1	-0.20515	3						
555891	11	15	1	-0.15936	3	Level 2 cut					
555849	6	16	1	-0.15699	3						
556982	16	17	1	-0.11477	6						
557037	22	18	1	-0.04904	6		2				
555874	8	19	1	-0.03496	2		2				
556204	15	20	1	-0.01657	2	Level 3 cut	2		**Level 3 Cut**	41%	13%
555978	14	21	1	0.14967	2		2	Level 3 Cut			
556996	18	22	1	0.20887	2		2				
555872	7	23	1	0.2194	2		2				
557000	20	24	1	0.22934			2				
557054	26	25	1	0.2316							
556984	17	26	1	0.24949	1						
557429	27	27	1	0.3491	1						
780745	36	28	1	0.3718	1				**Level 4 Cut**	19%	40%
555894	12	29	1	0.44476	1	Level 4 cut					
780778	39	30	1	0.47728	1						
560422	29	31	1	0.48654	1						
555880	9	32	1	0.51675	1						
555889	10	33	1	0.59775							
780767	38	34	1	0.7142			1				
557043	24	35	1	0.74549			1				
556986	32	36	1	1.01755			1				
556202	31	37	1	1.13517			1	Level 4 Cut			
555882	30	38	1	1.32959			1				
670555	33	39	1	1.38633			1				

Grade 8

Item Number	Op Position	OIB Position	Step	rp50	Benchmark Incorrect	Benchmark Cut Incorrect	Benchmark Correct	Benchmark Cut Correct	SS Final Cut	As Reported	Corrected
572799	25	1	1	-0.74067						6%	49%
558487	16	2	1	-0.72323							
558366	1	3	1	-0.68301							
558517	19	4	1	-0.57171			3				
572767	23	5	1	-0.5345			3				
652184	27	6	1	-0.52893			3				
558480	13	7	1	-0.45912			3	Level 2			
558453	9	8	1	-0.45647			3				
558368	2	9	1	-0.41005			6		**Level 2 Cut**	45%	7%
558373	4	10	1	-0.16261			6				
558536	21	11	1	-0.07825	3		2	Level 3			
558492	17	12	1	0.01972	3		2				
558469	10	13	1	0.02633	3		2				
572771	24	14	1	0.08819	3	Level 2	2				
558377	6	15	1	0.099	6						
652251	36	16	1	0.18536	6				**Level 3 Cut**	44%	12%
558401	7	17	1	0.28194	2	Level 3					
780906	37	18	1	0.29025	2						
781029	39	19	1	0.29056	2						
652195	34	20	1	0.3346	2						
558370	3	21	1	0.34885							
652191	28	22	1	0.35733	1						
558482	14	23	1	0.38351	1						
652199	29	24	1	0.42258	1						
558484	15	25	1	0.5544	1	Level 4					
558451	8	26	1	0.58338	1						
558475	12	27	1	0.59687	1						
558524	32	28	1	0.65797	1				**Level 4 Cut**	4%	32%
558457	31	29	1	0.70519			1				
558375	5	30	1	0.97762			1				
652241	35	31	1	1.02172			1				
558540	22	32	1	1.15383			1	Level 4			
558404	30	33	1	1.22477			1				
558526	20	34	1	1.26904			1				
572774	33	35	1	1.29801			1				
558472	11	36	1	1.33207							
780996	38	37	1	1.59458							
558496	18	38	1	1.88793							
572809	26	39	1	2.00774							

Grade 11

Item Number	Op Position	OIB Position	Step	rp50	Benchmark Incorrect	Benchmark Cut Incorrect	Benchmark Correct	Benchmark Cut Correct	SS Final Cut	As Reported	Corrected
560460	12	1	1	-1.06858			3			7%	55%
555703	8	2	1	-0.50577			3				
781231	33	3	1	-0.4105			3	Level 2			
560444	11	4	1	-0.07521			2				
560535	13	5	1	-0.04877			2	Level 3			
556348	9	6	1	-0.02085			2				
555699	6	7	1	0.04204			2				
555678	1	8	1	0.04512			2				
555685	3	9	1	0.08479	3				**Level 2 Cut**	43%	9%
561844	16	10	1	0.11239	3						
561880	19	11	1	0.1331	3						
561846	17	12	1	0.15569	3	Level 2					
781193	30	13	1	0.16295	6						
560537	14	14	1	0.23222	6						
560539	15	15	1	0.31312	2	Level 3			**Level 3 Cut**	35%	18%
556352	10	16	1	0.34147	2						
575290	20	17	1	0.3578	2						
555687	4	18	1	0.35984	2						
561848	18	19	1	0.48479	1						
781270	35	20	1	0.50282	1						
555680	2	21	1	0.53042	1						
781246	34	22	1	0.58649	1	Level 4					
575296	29	23	1	0.62128	1						
555689	5	24	1	0.6805	1				**Level 4 Cut**	14%	19%
560468	26	25	1	0.74062	1						
781204	31	26	1	0.75841							
556350	22	27	1	0.84906							
555701	7	28	1	0.93242			1				
561899	27	29	1	0.94588			1				
781215	32	30	1	0.98215			1				
781285	36	31	1	1.11502			1	Level 4			
555682	21	32	1	1.14158			1				
560452	24	33	1	1.35921			1				
781298	37	34	1	1.38521			1				
560464	25	35	1	1.39328							
575293	28	36	1	1.50109							
560448	23	37	1	1.62943							

APPENDIX H
STANDARD SETTING
MEETING MEMO

Multi-State Alternate Assessment (MSAA)

Standard Setting Memo

Science grades 5, 8, and 11

July 26-29, 2022

BACKGROUND

Cognia and the MSAA partners convened a panel of elementary, middle, and high school teachers during July 26-29, 2022, to participate in a standard setting workshop for the MSAA Science assessments.

- Educators were trained to recommend Level 2, Level 3, and Level 4 cut scores to enable reporting of student performance on the MSSA Science assessments.
- Eight to nine educators per grade participated in 2 ½ days of training and decision-making with Cognia standard setting specialists.

The standard setting panelists reviewed test content and performance level descriptors and followed the Item-Descriptor (ID) Matching standard setting method to recommend these cut scores.

In this memo, we summarize the standard setting procedures and results. We also identify and describe in some detail a data error and propose a solution that can be implemented right away.

METHOD

In following the ID Matching method, educators review the knowledge and skill response demands of MSAA Science items in ordered item books (i.e., ordered from the easiest to the most difficult item). in addition, panelists were provided target cut scores and ranges in the ordered item map, which accompanied the ordered item book.

- Panelists matched those item response demands to knowledge and skill expectations in borderline performance level descriptors for the Level 2, Level 3, and Level 4 performance levels.[1]
- Panelists were guided on how to consider the target cut score ranges; that is, to use them as a reasonable starting point for their review and, as needed, to recommend cut scores above or below the target cut score ranges.
- Working independently, the standard setting panelists conducted the ID matching process in two rounds and recommended cut scores for each of the three levels in each round. After round 1, the Cognia workshop facilitator led panelists through a discussion of agreements and disagreements among the panelists on cut score locations and rationales for the various cut scores they recommended.
- The ensuing discussion enabled panelists to consider their colleagues' insights about item response demands and matches to performance level descriptors and to consider adjusting their cut score recommendations in the second, final round.

At the end of the workshop, panelists completed a workshop evaluation survey in which they provided feedback about various aspects of the workshop. As part of this survey, panelists were asked to respond to the following statement: *I am satisfied with the final group cut scores. I would not recommend changing any of the group cut scores*. Response options for the statement included *strongly disagree, disagree, undecided, strongly agree, agree,* and *not applicable*. As shown in Table 1, most panelists noted that they

[1] *Borderline performance level descriptors define knowledge and skills that students who are just barely in a performance level are expected to know and be able to demonstrate.*

were satisfied with the final group cut score recommendations as indicated by the number of panelists who selected either *strongly agree* or *agree* in response to the statement. One member from each of Grade 5 and Grade 11 panels selected *undecided* in response to the statement.

Table 1. Frequency of Panelist Responses to the Workshop Evaluation Survey

Grade	N	Undecided	Strongly Agree	Agree
5	9	1	3	5
8	8	0	6	2
11	8	1	4	3

Note. Since no panelists selected "strongly disagree," "disagree," or "not applicable," these categories are not included in the table.

As a follow-up question, panelists were asked if they would recommend moving a cut score up or down in the ordered item booklet (OIB), and by how many pages. Across the three grades, panelists' responses varied as follows:

- Grade 5: One panelist recommended moving the Level 2 cut score down by 2 pages (i.e., lower the cut score). In addition, 2 panelists suggested moving all cut scores up by 2 pages (i.e., raise the cut score).
- Grade 8: One panelist recommended moving all cut scores up by 1 page (i.e., raise the cut score).
- Grade 11: One person recommended moving all cut scores up by 1 page, another suggested moving all cut scores up by 2 pages, and a third recommended all cut scores be moved down by 2 pages.

RESULTS: Panelist Recommendations and Proposed Policy Review Adjustments

Table 2 contains the panelist recommended cut scores and associated impact data across grades 5, 8, and 11 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), as well as for the combined Level 3 and Level 4 performance levels. Table 2 also includes proposed cut scores for a policy review adjusted and the associated impact data with these cut scores. Exhibits 1 and 2 provide stacked bar charts displaying impact data for the panelist recommended and proposed policy adjusted cut scores.

Table 2. OIB Pages and Impact Data from the MSAA Science Standard Setting: Preliminary

	Panelist Recommendations		Proposed Policy Adjustments	
Performance Level	OIB Page Number	Impact	OIB Page Number	Impact
Grade 5				
1	--	42%/47%	--	36%
2	12	5%/8%	7	19%
3	20	13%/16%	20	29%
4	28	40%/29%	36	16%
3 + 4	--	53%/45%	--	45%
Grade 8				
1	--	49%/43%	--	43%
2	9	7%/24%	7	21%
3	16	12%/12%	15	19%

continued

	Panelist Recommendations		Proposed Policy Adjustments	
Performance Level	OIB Page Number	Impact	OIB Page Number	Impact
4	28	32%/21%	29	17%
3 + 4	--	44%/33%	--	36%
Grade 11				
1	--	55%/55%	--	49%
2	9	9%/17%	5	22%
3	15	18%/14%	15	16%
4	24	19%/15%	27	13%
3 + 4	--	37%/29%	--	29%

Table Notes

(1) Impact data in column 3, to the left of the forward slash, are calculated based on the 2022 standard setting ordered item books.

(2) Impact data in column 3, to the right of the forward slash, are calculated based on projections to the preliminary proposed operational test forms.

(3) The impact data includes all students who will receive score reports in 2022.

(4) Once policy adjustments are approved by the MSAA partners, we will check the cut scores and impact data against the final, approved operational test forms.

Exhibit 1. Stacked Bar Charts of Impact Data by Grade (Standard Setting Committee Recommendations)

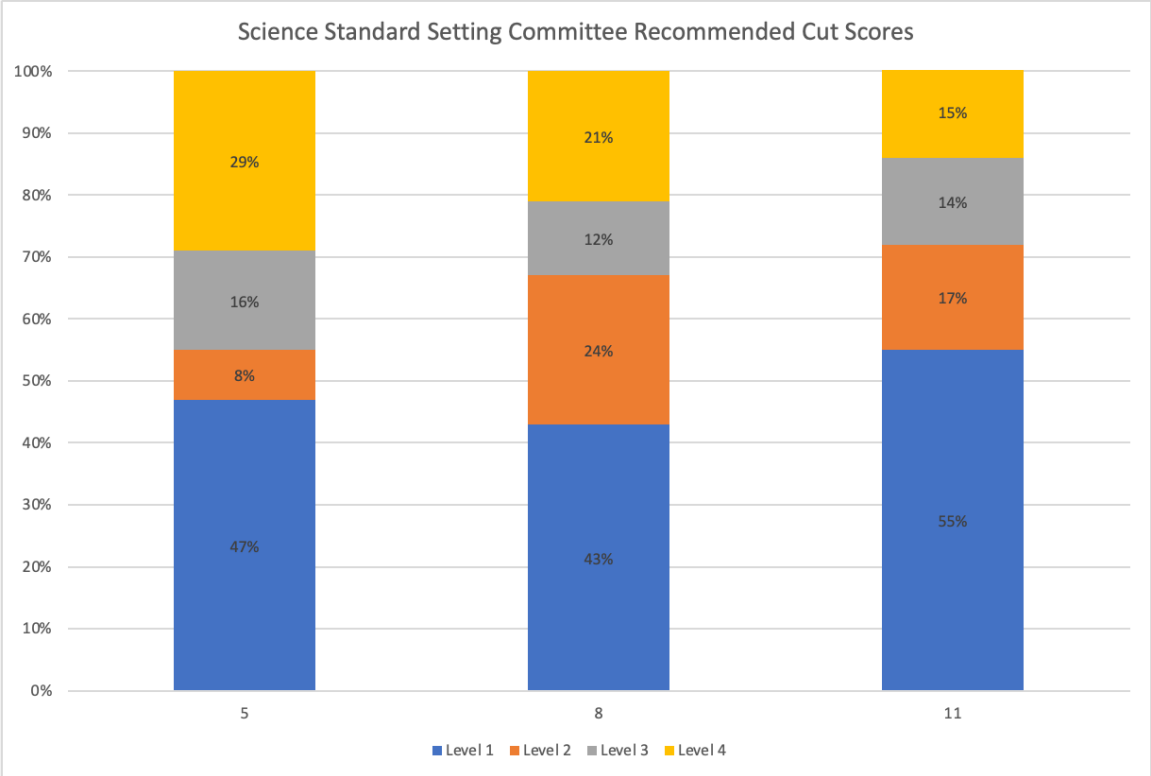
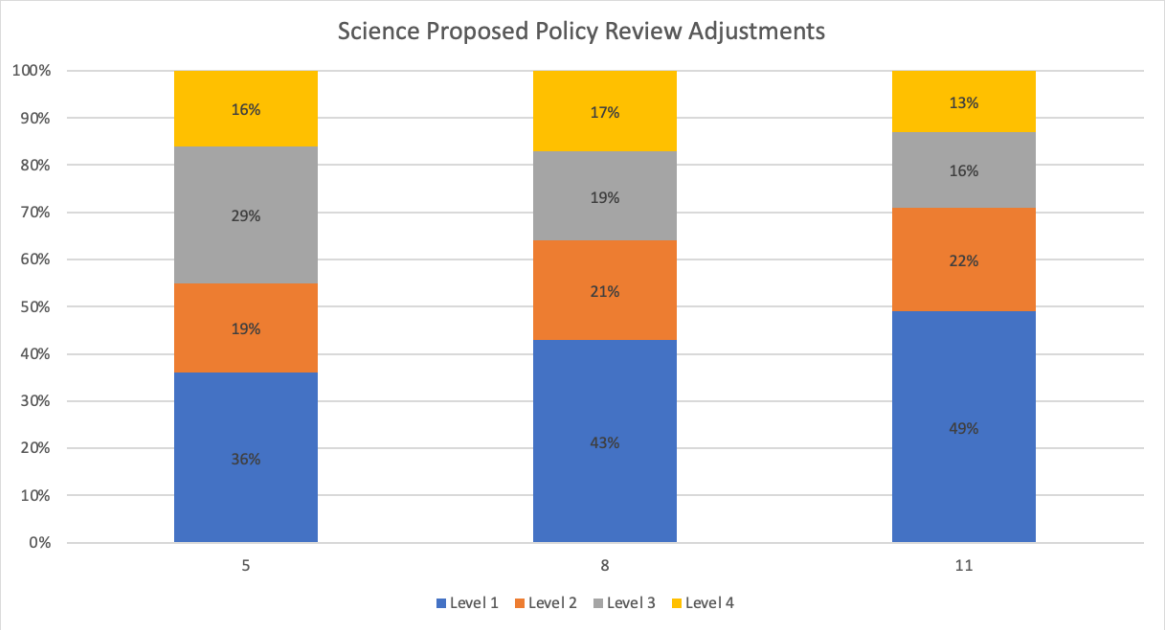


Exhibit 2. Stacked Bar Charts of Impact Data by Grade (Proposed Policy Review Adjustments)



Policy Review for All Cut Scores

Often, the full process of establishing performance standards for any new assessment program continues after a standard setting workshop. In a standard setting workshop, content experts—in this case, educators from MSAA partner states and entities—recommend cut scores to the policy makers who are responsible for an assessment program. After that, assessment program leaders can conduct a policy review of the recommended cut scores. In a policy review, leaders can consider the impact data,[2] vertical articulation of the impact data across grade levels,[3] and the likely effects on the policies and educational politics of the participating states and entities.

For the reasons described above and as noted in the errata noted below, Cognia has proposed to the MSAA partners states starting places to conduct a policy review of the three recommended cut scores in all three grade levels. In this policy review, the MSAA partners would review the appropriateness of the panelists' recommended cut scores, the proposed adjusted cut scores, and consider in context of other programs in their states. A second, and important goal of this policy review would be to achieve reasonable articulation of impact data across grades 5, 8, and 11. Here, the goal is to adjust, as necessary, some cut scores recommended by the standard setting panels to achieve articular—while preserving as much as possible the content-focused process that the standard setting panelists completed in following the ID-Matching process to recommend cut scores.

As previously noted, Table 2 and Exhibit 2 contain proposed policy review adjusted cut scores and impact data for the MSAA state partners to consider. After the policy review, adjusted cut scores that are approved by the MSAA partners are referred to as "policy-adjusted" cut scores. Once the partners make policy adjustments, MSAA can approve these as final cut scores, ready for use for 2022 score reporting.

---

[2] *Impact is the percentages of students in Levels 1, 2, 3, and 4 after cut scores have been recommended or approved.*

[3] *When performance standards are vertically articulated, the percentages of examinees in performance levels are similar across grades levels or increase or decrease in understandable and meaningfully explainable ways (e.g., We expect performance to be higher in grades 5 and 8 than in grade 11).*

ERRATA: Target Cut Score Calculation

In preparing this memo, Cognia discovered that there was an error in the method used to calculate impact data for each line in the ordered item map in the MSAA science standard setting. The target cut score ranges in the ordered item maps that we presented to panelists were mapped through this incorrect impact data. The panelists looked at these targets when they made their round 1 cut score recommendations, and they were guided to consider those ranges as they made their cut score recommendation. This may have influenced panelists in the way they evaluated items and ultimately in their judgment process to recommend cut scores.

In addition, while at standard setting, the MSAA partners were informed at the end of each day where on the MSAA reporting scale the cut scores were located and the expected impact from those cut scores. The impact data presented at these times was incorrect.

Despite the incorrect target cut score ranges, there are strong suggestions that the panelists detected problems with target score ranges—because they recommended cut scores outside of those ranges. The standard setting workshop facilitators reported that panelists did not feel constrained by the target cut score ranges and they recommended final cut scores outside of the target ranges. The Excel workbook that you received from Cognia provides details regarding the correct and incorrect cut score ranges.

Here is how to interpret the color-coded columns in the Workbook tabs *Science 05*, *Science 08*, and *Science 11*.

Column H. Incorrect target cut score ranges which the workshop panelists received.

Column I. Incorrect target cut scores.

Column J. Correct cut score ranges.

Column K. Correct target cut scores.

Column L. Final cut scores recommended by the standard setting panel as a group (i.e., the median of their recommended cut scores).

Column M. Impact data that corresponds to the panelists' recommended cut scores. This is the incorrect impact data that the MSAA partners received at the end of the workshop.

Column N. Corrected impact data, based on the panelists' recommended cut scores.

Science 05 Tab in the Workbook

The correct and incorrect target cut score ranges for Level 3 (in green) coincide almost perfectly. The correct and incorrect target cut score ranges for Level 2 (in yellow) do not overlap. However, the panelists recommended a cut score at the top of the incorrect range, one ordered item book page below the correct range. The correct and incorrect target cut score ranges for Level 4 (in blue) do not overlap. Panelists recommended a cut score that coincides with the incorrect target cut score and that does approach the correct target range. This cut score is a good candidate for policy review.

Science 08 Tab in the Workbook

The incorrect and correct target cut score ranges for all three cut scores do not overlap. However, the recommended level 2 cut score in column L is only one ordered item book below the correct target range. (See the yellow highlighted lines in columns H and J.) The recommended Level 3 cut score (see the green highlighting) in column L is only two ordered item book pages above the correct target range. Also, the recommended cut score for Level 4 (in blue) is only one ordered item book page below the correct target range. Small adjustments in a policy review could bring these cut scores in line with the correct ranges and impact data.

Science 11 Tab in the Workbook

The incorrect and correct target cut score ranges in this grade level are widely disparate. However, the Levels 2 and 3 cut scores are in the correct target cut score range. This is not the case for the Level 4 cut score. That recommended cut score is four pages below the correct target range (in yellow). The Level 4 cut score is a good candidate for policy review adjustments.

APPENDIX—I
STANDARD-SETTING MEETING AGENDAS AND NONDISCLOSURE FORM

MSAA Science Standard Setting 2022

Agenda

Tuesday, 7/26

Schedule	Activity	Notes
7:30 AM – 8:30 AM	Registration, Check-In & Breakfast	
8:30 AM – 10:30 AM	Welcome & Overview	
10:30 AM –10:45 AM	Break & Transition to Breakout Rooms	
10:45 AM –12:00 PM	Key Concepts/Processes, Training & Practice	
12:00 PM –1:00 PM	Lunch	
1:00 PM – 2:00 PM	Key Concepts/Processes, Training & Practice	
2:00 PM – 3:15 PM	Taking the Science Test	
3:15 PM – 3:30 PM	Break	
3:30 PM – 4:30 PM	Preparation for Round 1	
4:30 PM – 5:00 PM	Begin Round 1	
5:00 PM	Adjourn	

Wednesday, 7/27

Schedule	Activity	Notes
7:30 AM – 8:30 AM	Continental Breakfast	
8:30 AM – 9:00 AM	Debrief Day 1	
9:30 AM –10:30 AM	Complete Round 1	
10:30 AM – 11:00 AM	Break	
11:00 AM –12:00 PM	Complete Round 1	
12:00 PM – 1:00 PM	Lunch	
1:00 PM – 3:00 PM	Prepare for Round 2	
3:00 PM – 3:15 PM	Break	
3:15 PM – 5:00 PM	Complete Round 2	
5:00 PM	Adjourn	

Thursday, 7/28

Schedule	Activity	Notes
7:30 AM – 8:30 AM	Continental Breakfast	
8:30 AM – 9:00 AM	Debrief Day 2	
9:00 AM –10:00 AM	Prepare for Round 3	
10:00 AM – 10:15 AM	Break	
10:15 AM –11:30 PM	Complete Round 3	
11:30 AM – 12:00 PM	Debrief	
12:00 PM –1:00 PM	Lunch	
1:00 PM	Adjourn	

MSAA Nondisclosure Form

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, writing prompts, graphics/images, charts, graphs, and tables, etc. 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 or print materials downloaded or notes taken during this meeting without the express permission of the facilitator or an MSAA representative.

_____ _____

Full Name **State/Entity**

_____ _____

Signature **Date**

APPENDIX—J
STANDARD-SETTING PANELIST FOLLOW-UP MEETING MINUTES

MSAA Meeting Minutes

Science Standard Setting

Panelist Follow-Up

Date: Monday, 8/22/22 at 12:00 PM – 2:00 PM EST

Meeting Recording:
https://cognia.zoom.us/rec/share/gWO3bSuI1SYSU6UM3Zmp40kzJfUim3182flyU84bNhctq3M402NJyCf9uyvLO4A.fIIMZYpLec59gVwe

Passcode: LfX4^@jN

PARTICIPANTS *(X signifies attendance; * indicates facilitator/scribe)*

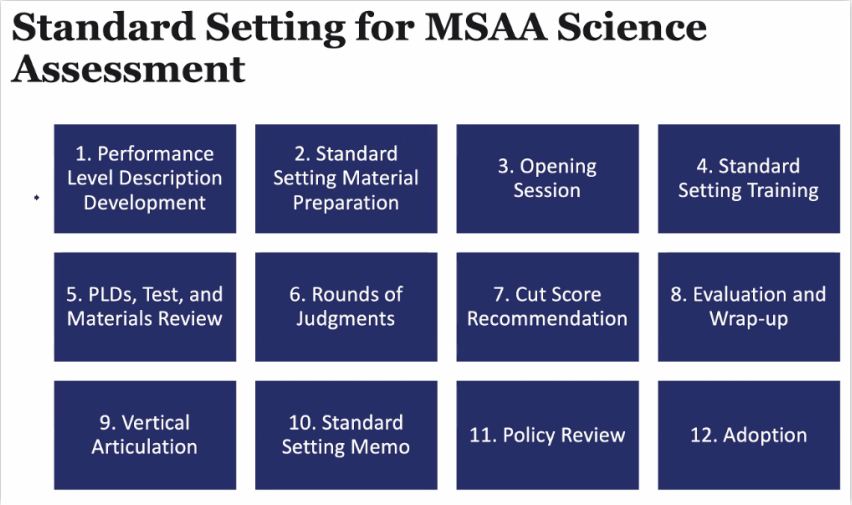
	SCIENCE PARTNERS		PANELISTS GRADE 5		PANELISTS GRADE 8		PANELISTS GRADE 11
x	Anju Kuriakose (AZ)	x	Carmen Bird (USVI)		Andrea Mercado (ME)		April Knippen (AZ)
x	Audra Ahumada (AZ)	x	Jane L Williams (ME)		Angelita Cruz-Tuitt (USVI)		Brittney Oden (AZ)
x	Bethany Spangenberg (AZ)		Jerwin Cruz Perez (USVI)		Liezl V. Cabanilla (USVI)		Cynthia Locsin (USVI)
x	Don Griffin (BIE)		Magdalena de la Paz (BIE)		Marlene Murphy (BIE)		Ellen Anfone (USVI)
x	Hansley Mussotte (AZ)		Misty Favreau (ME)		Mervin Arcillo (USVI)	x	Jessica Cana (USVI)
x	Janette Kirk (ME)		Orlinda Wickham (USVI)		Renee J Atcitty (BIE)	x	Joann Lewis (USVI)
x	Jodi Bossio-Smith (ME)		Patricia Adams (ME)		Tammy Moncel (BIE)		Lisa Forcier (BIE)
x	Sarah Han (AZ)	x	Oddeth Brown (USVI)		Lawrance Nai (AZ)	x	Vernon Callwood (USVI)
x	Sabiha Klepk (AZ)		Victoria Ostwald (BIE)				
	Tamarah Pfeiffer (BIE)						
	COGNIA						
x	Amy Leach						
x	Debbian James						
x	Erin Maskwa						
x	Jill Stepanek						
	Julie DiBona						
x	Karen Whisler						
x	Kelly Ickes						
x	Lauren Copp						
x	Louis Roussos						
x	Matthew Gushta						
x	Qi Qin						
x	Sandra Sweeney						
x	*Stephen Murphy						

Agenda

- Welcome and Introductions
- Overview of Purpose
- Review of Range and Target Cut Scores
- Policy Review
- Conclusion and Wrap-Up

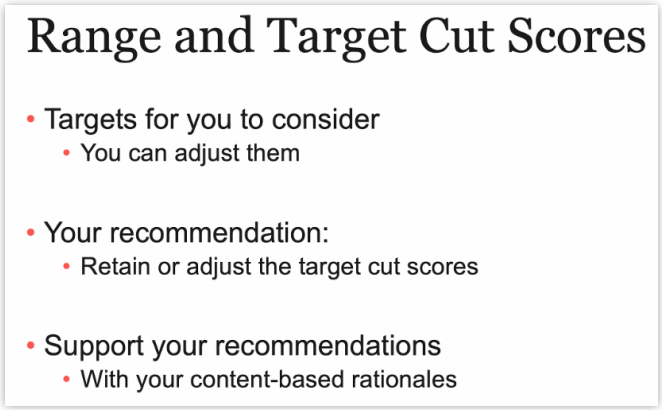
Meeting Notes

- Welcome and Introductions
  - Stephen Murphy (Cognia Chief Learning Officer) facilitated the meeting. He began with a welcome and expressed gratitude to the panelists for the work they did.
  - Everyone introduced themselves: MSAA partners, panelists and Cognia staff
  - Stephen presented an overview of key areas of standard setting, walking through an outline and noting the work completed before the meeting on materials that panelists utilized (#1 & #2 in graphic below) and how the panelists participated (#3-8 in graphic below)

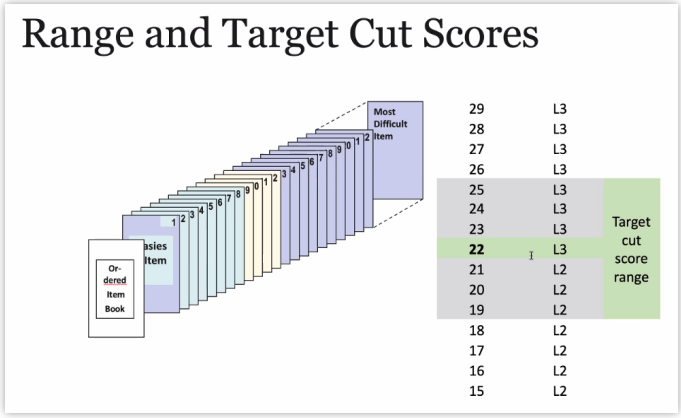


    - Panelists made judgments on Levels 2, 3 & 4 through lens of OIB and provided content-based cut-score recommendations.
  - Tasks that followed the standard setting (#9-12 in graphic above)
    - Vertical Articulation – evaluation of the impact data to determine to what extent it does or does not make sense when we view it across grades.
    - Standard Setting Memo
    - Policy Review: Partners
    - Adoption: decision yet to be made in this case
- Overview of Purpose
  - Review of target cut score and range miscalculation
    - Panelists provided their recommendations for the cut scores based on targets given for consideration
    - Highlighted OIB and target cut score range
      - Miscalculation was in the starting place for the range
  - The three bullets below are from the slides that were presented to panelists within the committees which outline the process. Panelists were provided with targets to consider which could be adjusted as

determined and guided through not only their own decision making, but also the discussions they had with other panelists. Recommendations could also be retained or adjusted with content-based rationales to support the recommendations.



- Specific to #10 in the graphic above, the Memo is a brief summary that provides a summary of what occurred during the meeting and outcomes.
- In preparing the memo, Cognia discovered a miscalculation with the computation that went into determining these target range and the actual cut scores themselves (graphic below). Panelists would have started at a different point had we not made that miscalculation.
- Panelists were given a range with which they would then go into the ordered item booklet, review the item, and review the proficiency level description and begin that alignment-based process to assign a level to each of the items.



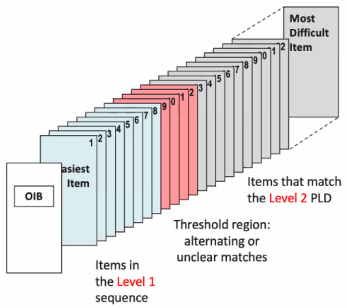
- In terms of the cut score recommendations (slide snip below), panelists had discussion and then proceeded with round two in which they were asked to go back into the content, review the content, review the proficiency level descriptions, and at the conclusion, make content based rationales. Cognia provided guidance regarding how to structure the content-based rationale and what information to draw from.
- Panelists were asked to address two questions and, most importantly, referenced both the items and the PLDs.

Cut Score Recommendations

- For your Level 2, 3, and 4 cut score recommendations
  - Write a content-based rationale for your recommendation to retain or adjust the cut scores
- Relevant for writing your rationales
  - Your answers to the two questions
  - The match between item response demands and expectations in the PLDs
- Notice: Refer to both **items** and **PLDs**

  - Highlighted graphic (below) about recommendations
    - Panelists accessed the OIB through the item map and through reviewing the items using the standard setting tool.
    - OIB ordered from easiest to most difficult with threshold being the transition point where cut score recommendations would fall

Cut Score Recommendations



19	L2	Item response demands clearly match the expectations in the Level 2 PLD
18	L2	
17	L2	
16	L2	
15	L2	
14	L2	
13	L2	
12	L1	Threshold region
11	L2	
10	L1	
9	L2	
8	L1	Item response demands **do not match** the expectations in the Level 2 PLD
7	L1	
6	L1	
5	L1	
4	L1	
3	L1	
2	L1	
1	L1	

- Review of Range and Target Cut Scores
  - Below are the recommendations that came out of the standard setting. These are important data points that are applied to student data from Science test.
  - Panelists engaged in the process appropriately and followed all procedures as outlined.

Standard Setting Committee Recommendations

Subject	Grade	Round Information	Level 2 Cut OIB Page #	Level 3 Cut OIB Page #	Level 4 Cut OIB Page #
SCI05	5	SS Final Round	12	20	28
SCI08	8	SS Final Round	9	16	28
SCI11	11	SS Final Round	9	15	24

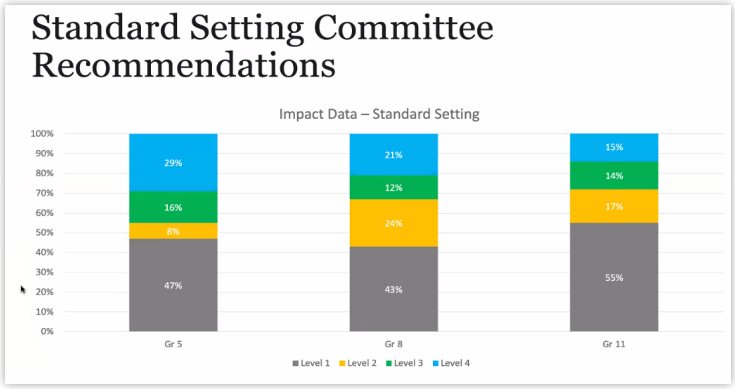
  - Cognia then shared charts of item maps for each grade and each level with the following:
    - Shaded region (solid color) shown at the standard setting where panelists entered the OIB
    - Correct shaded region (gray) that should have been provided to panelists
    - Standard setting cut score from the committee (diagonal color pattern)
    - Proposed policy adjusted cut (checkered pattern)

- From Cognia understanding of what occurred during the committee meetings, we feel the panelists completed the work appropriately. Page numbers that panelists recommended was based on that work.
- Cognia shared graphics (below) of the resulting impact data that panelists have not seen before.
  - Data is across the states
  - Audra emphasized that this data has not yet been approved and cannot be shared at this time.
  - These graphs show the percentage of students in each performance category based on the cut scores that came out of the standard setting meeting.
    - First graphic below has final page number recommendations from working through the OIB
    - Classification using the page numbers applied to student data from the most recent science test results to determine percentages of students by grade across the four performance levels.

Standard Setting Committee Recommendations

Subject	Grade	Round Information	Level 2 Cut OIB Page #	Level 3 Cut OIB Page #	Level 4 Cut OIB Page #	Percent Level 1	Percent Level 2	Percent Level 3	Percent Level 4
SCI05	5	SS Final Round	12	20	28	47%	8%	16%	29%
SCI08	8	SS Final Round	9	16	28	43%	24%	12%	21%
SCI11	11	SS Final Round	9	15	24	55%	17%	14%	15%

  - Second graphic below illustrates percentages by grade for impact data



- Judgment Task and Discussion: Questions for Panelists
  - Can you describe how you – and your colleagues – used the target cut scores and ranges to inform your work (how did this influence your thinking)?
    - No specific responses provided to this question.
  - Given what you learned about the target cut scores and ranges, how do you think that changing this information might have impacted your and/or your committee's cut score recommendations, if at all?
    - If move, would it be earlier in the OIB or later relative to its current location?
  - Response and discussion by grade
    - Panelists used PLDs to guide their cut score recommendations. They looked at the content and also outside of the starting point cut score ranges.
    - Grade 5
      - One panelist felt confident with the recommendations. Her group looked at items on the lower and higher end.
    - Grade 8
      - There were no panelists from Grade 8 in attendance.

        - Grade 11
            - One panelist felt confident with the recommendations.
            - One panelist noted that they looked at lower and higher end items.
- Policy Review (not included in this meeting)
- Conclusion and Wrap-Up
    - Audra thanked the panelists and highlighted the purposes for bringing panelists back together
        - Partners need to make policy decisions.
        - Partners wanted feedback from panelists to inform those decisions.

MSAA Meeting Minutes

Science Standard Setting

Panelist Follow-Up

Date: Monday, 8/22/22 at 3:00 PM – 5:00 PM EST

Meeting Recording:
https://cognia.zoom.us/rec/share/I6k62vM02BGaniKK3XG6I3EEmsZXuw3Nz3yQcf6PczYaCpBXP9G6Sn_tJ9uIdGFW.hIPjyCz3ly4q3xbb

Passcode: 9K#8CGW0

PARTICIPANTS *(X signifies attendance; * indicates facilitator/scribe)*

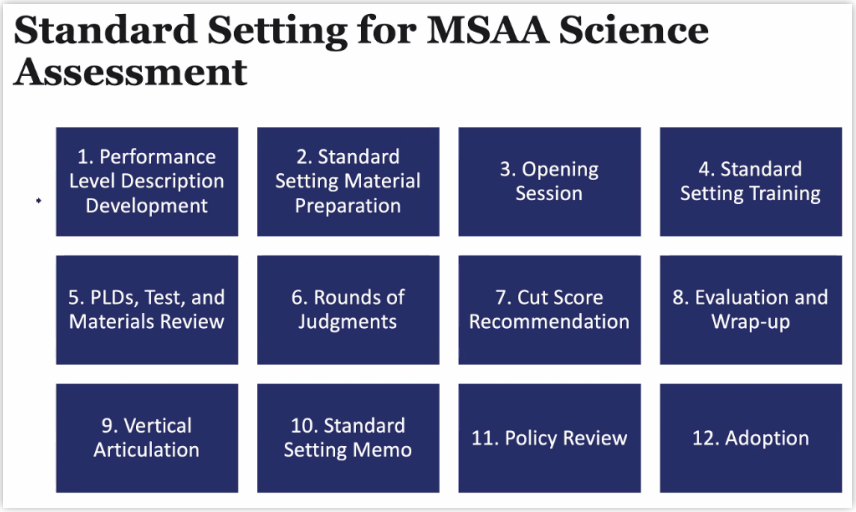
	SCIENCE PARTNERS		PANELISTS GRADE 5		PANELISTS GRADE 8		PANELISTS GRADE 11
	Anju Kuriakose (AZ)		Carmen Bird (USVI)		Andrea Mercado (ME)	x	April Knippen (AZ)
x	Audra Ahumada (AZ)		Jane L Williams (ME)		Angelita Cruz-Tuitt (USVI)		Brittney Oden (AZ)
x	Bethany Spangenberg (AZ)	x	Jerwin Cruz Perez (USVI)		Liezl V. Cabanilla (USVI)		Cynthia Locsin (USVI)
x	Don Griffin (BIE)		Magdalena de la Paz (BIE)		Marlene Murphy (BIE)		Ellen Anfone (USVI)
x	Hansley Mussotte (AZ)		Misty Favreau (ME)	x	Mervin Arcillo (USVI)		Jessica Cana (USVI)
	Janette Kirk (ME)		Orlinda Wickham (USVI)	x	Renee J Atcitty (BIE)		Joann Lewis (USVI)
x	Jodi Bossio-Smith (ME)		Patricia Adams (ME)		Tammy Moncel (BIE)		Lisa Forcier (BIE)
x	Sarah Han (AZ)		Oddeth Brown (USVI)	x	Lawrance Nai (AZ)	x	Vernon Callwood (USVI)
	Sabiha Klepk (AZ)		Victoria Ostwald (BIE)				
	Tamarah Pfeiffer (BIE)						
	COGNIA						
x	Amy Leach						
x	Debbian James						
	Erin Maskwa						
	Jill Stepanek						
	Julie DiBona						
	Karen Whisler						
x	Kelly Ickes						
x	Lauren Copp						
x	Louis Roussos						
	Matthew Gushta						
x	Qi Qin						
x	Sandra Sweeney						
x	*Stephen Murphy						

Agenda

- Welcome and Introductions
- Overview of Purpose
- Review of Range and Target Cut Scores
- Policy Review
- Conclusion and Wrap-Up

Meeting Notes

- Welcome and Introductions
  - Stephen Murphy (Cognia Chief Learning Officer) facilitated the meeting. He began with a welcome and expressed gratitude to the panelists for the work they did.
  - Everyone introduced themselves: MSAA partners, panelists and Cognia staff
  - Stephen presented an overview of key areas of standard setting, walking through an outline and noting the work completed before the meeting on materials that panelists utilized (#1 & #2 in graphic below) and how the panelists participated (#3-8 in graphic below)



    - Panelists made judgments on Levels 2, 3 & 4 through lens of OIB and provided content-based cut-score recommendations.
  - Tasks that followed the standard setting (#9-12 in graphic above)
    - Vertical Articulation – evaluation of the impact data to determine to what extent it does or does not make sense when we view it across grades.
    - Standard Setting Memo
    - Policy Review: Partners
    - Adoption: decision yet to be made in this case
- Overview of Purpose
  - Review of target cut score and range miscalculation
    - Panelists provided their recommendations for the cut scores based on targets given for consideration
    - Highlighted OIB and target cut score range
      - Miscalculation was in the starting place for the range

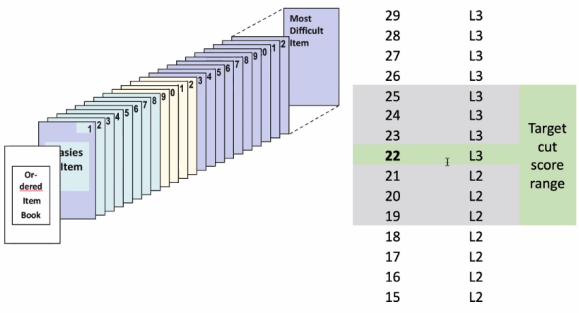
- The three bullets below are from the slides that were presented to panelists within the committees which outline the process. Panelists were provided with targets to consider which could be adjusted as determined and guided through not only their own decision making, but also the discussions they had with other panelists. Recommendations could also be retained or adjusted with content-based rationales to support the recommendations.

Range and Target Cut Scores

- Targets for you to consider
  - You can adjust them
- Your recommendation:
  - Retain or adjust the target cut scores
- Support your recommendations
  - With your content-based rationales

- Specific to #10 in the graphic above, the Memo is a brief summary that provides a summary of what occurred during the meeting and outcomes.
- In preparing the memo, Cognia discovered a miscalculation with the computation that went into determining these target range and the actual cut scores themselves (graphic below). Panelists would have started at a different point had we not made that miscalculation.
- Panelists were given a range with which they would then go into the ordered item booklet, review the item, and review the proficiency level description and begin that alignment-based process to assign a level to each of the items.

Range and Target Cut Scores



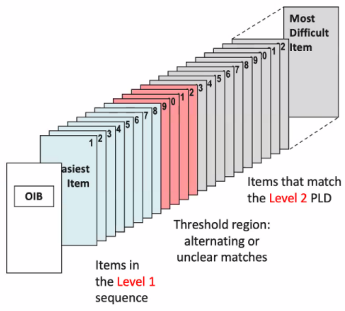
- In terms of the cut score recommendations, panelists had discussion and then proceeded with round two in which they were asked to go back into the content, review the content, review the proficiency level descriptions, and at the conclusion, make content based rationales. Cognia provided guidance regarding how to structure the content-based rationale and what information to draw from.
- Panelists were asked to address two questions and, most importantly, referenced both the items and the PLDs.

Cut Score Recommendations

- For your Level 2, 3, and 4 cut score recommendations
  - Write a content-based rationale for your recommendation to retain or adjust the cut scores
- Relevant for writing your rationales
  - Your answers to the two questions
  - The match between item response demands and expectations in the PLDs
- Notice: Refer to both **items** and **PLDs**

- Highlighted graphic (below) about recommendations
  - Panelists accessed the OIB through the item map and through reviewing the items using the standard setting tool.
  - OIB ordered from easiest to most difficult with threshold being the transition point where cut score recommendations would fall

Cut Score Recommendations



19	L2	Item response demands clearly match the expectations in the Level 2 PLD
18	L2	
17	L2	
16	L2	
15	L2	
14	L2	
13	L2	
12	L1	Threshold region
11	L2	
10	L1	
9	L2	
8	L1	Item response demands **do not match** the expectations in the Level 2 PLD
7	L1	
6	L1	
5	L1	
4	L1	
3	L1	
2	L1	
1	L1	

  - Below are the recommendations that came out of the standard setting. These are important data points that are applied to student data from Science test.
  - Panelists engaged in the process appropriately and followed all procedures as outlined.

Standard Setting Committee Recommendations

Subject	Grade	Round Information	Level 2 Cut OIB Page #	Level 3 Cut OIB Page #	Level 4 Cut OIB Page #
SCI05	5	SS Final Round	12	20	28
SCI08	8	SS Final Round	9	16	28
SCI11	11	SS Final Round	9	15	24

- Cognia then shared charts of item maps for each grade and each level with the following:
  - Shaded region (solid color) shown at the standard setting where panelists entered the OIB
  - Correct shaded region (gray) that should have been provided to panelists
  - Standard setting cut score from the committee (diagonal color pattern)
  - Proposed policy adjusted cut (checkered pattern)

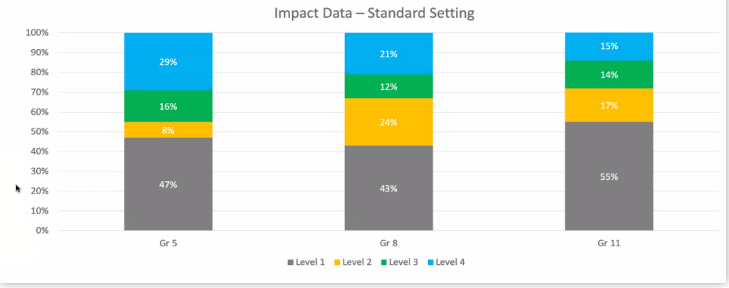
- From Cognia understanding of what occurred during the committee meetings, we feel the panelists completed the work appropriately. Page numbers that panelists recommended was based on that work.
- Cognia shared graphics (below) of the resulting impact data that panelists have not seen before.
  - Data is across the states
  - Stephen emphasized that this data has not yet been approved and cannot be shared at this time.
  - These graphs show the percentage of students in each performance category based on the cut scores that came out of the standard setting meeting.
    - First graphic below has final page number recommendations from working through the OIB
    - Classification using the page numbers applied to student data from the most recent science test results to determine percentages of students by grade across the four performance levels.

Standard Setting Committee Recommendations

Subject	Grade	Round Information	Level 2 Cut OIB Page #	Level 3 Cut OIB Page #	Level 4 Cut OIB Page #	Percent Level 1	Percent Level 2	Percent Level 3	Percent Level 4
SCI05	5	SS Final Round	12	20	28	47%	8%	16%	29%
SCI08	8	SS Final Round	9	16	28	43%	24%	12%	21%
SCI11	11	SS Final Round	9	15	24	55%	17%	14%	15%

  - Second graphic below illustrates percentages by grade for impact data

Standard Setting Committee Recommendations



- Judgment Task and Discussion: Questions for Panelists
  - Can you describe how you – and your colleagues – used the target cut scores and ranges to inform your work (e.g., used primarily in round one and then not in subsequent rounds)?
  - Given what you learned about the target cut scores and ranges, how do you think that changing this information might have impacted your and/or your committee's cut score recommendations, if at all?
    - If move, would it be earlier in the OIB or later relative to its current location?
  - Response and discussion by grade
    - Panelists used PLDs to guide their cut score recommendations. They looked at the content and also outside of the starting point cut score ranges.
    - Grade 5
      - No feedback provided.
    - Grade 8
      - One panelist noted that he expected L2 would be bigger and L4 smaller. He concluded that the change would not have had a significant impact. By end of three rounds, they were

confident in their recommendations and would be unlikely to change much if the new information was introduced. There was fidelity with the process.
        - One panelist noted that being presented with certain data sets served as a marker/guide in your head. The group followed the process. She is confident that the recommendations they provided are solid and would likely not have changed.
    - Grade 11
        - One panelist noted that he would not change his recommendations.
        - One panelist noted that they started working from L3 as a starting point then went up toward L4 and down toward L2. Facilitators encouraged panelists to look a few items before and after the starting point. She felt that there might have been some impact if the correct info had been provided. There was a lot of variation in her initial judgments which was unexpected. If panelists were looking at the correct info initially, it might have brought L3 down. Ultimately, she shared that she was confident in her judgments as related to the metholodogy required from them as panelists.
- Policy Review (not included in this meeting)
- Conclusion and Wrap-Up
    - Audra thanked the panelists and highlighted the purposes for bringing panelists back together
        - Partners need to make policy decisions.
        - Partners wanted feedback from panelists to inform those decisions.

Agenda

- Welcome and Introductions
- Overview of Purpose
- Review of Range and Target Cut Scores
- Policy Review
- Conclusion and Wrap-Up

Standard Setting for MSAA Science Assessment

1. Performance Level Description Development
2. Standard Setting Material Preparation
3. Opening Session
4. Standard Setting Training
5. PLDs, Test, and Materials Review
6. Rounds of Judgments
7. Cut Score Recommendation
8. Evaluation and Wrap-up

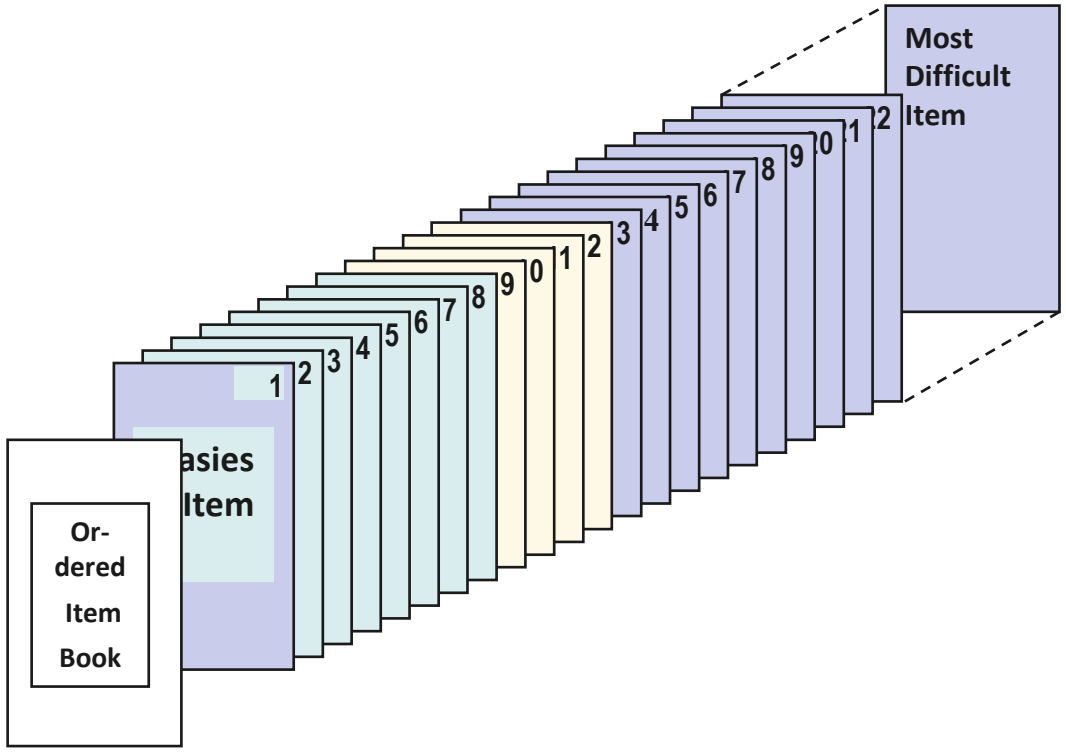
Standard Setting for MSAA Science Assessment

1. Performance Level Description Development
2. Standard Setting Material Preparation
3. Opening Session
4. Standard Setting Training
5. PLDs, Test, and Materials Review
6. Rounds of Judgments
7. Cut Score Recommendation
8. Evaluation and Wrap-up
9. Vertical Articulation
10. Standard Setting Memo
11. Policy Review
12. Adoption

Range and Target Cut Scores

- Targets for you to consider
  - You can adjust them
- Your recommendation:
  - Retain or adjust the target cut scores
- Support your recommendations
  - With your content-based rationales

Range and Target Cut Scores



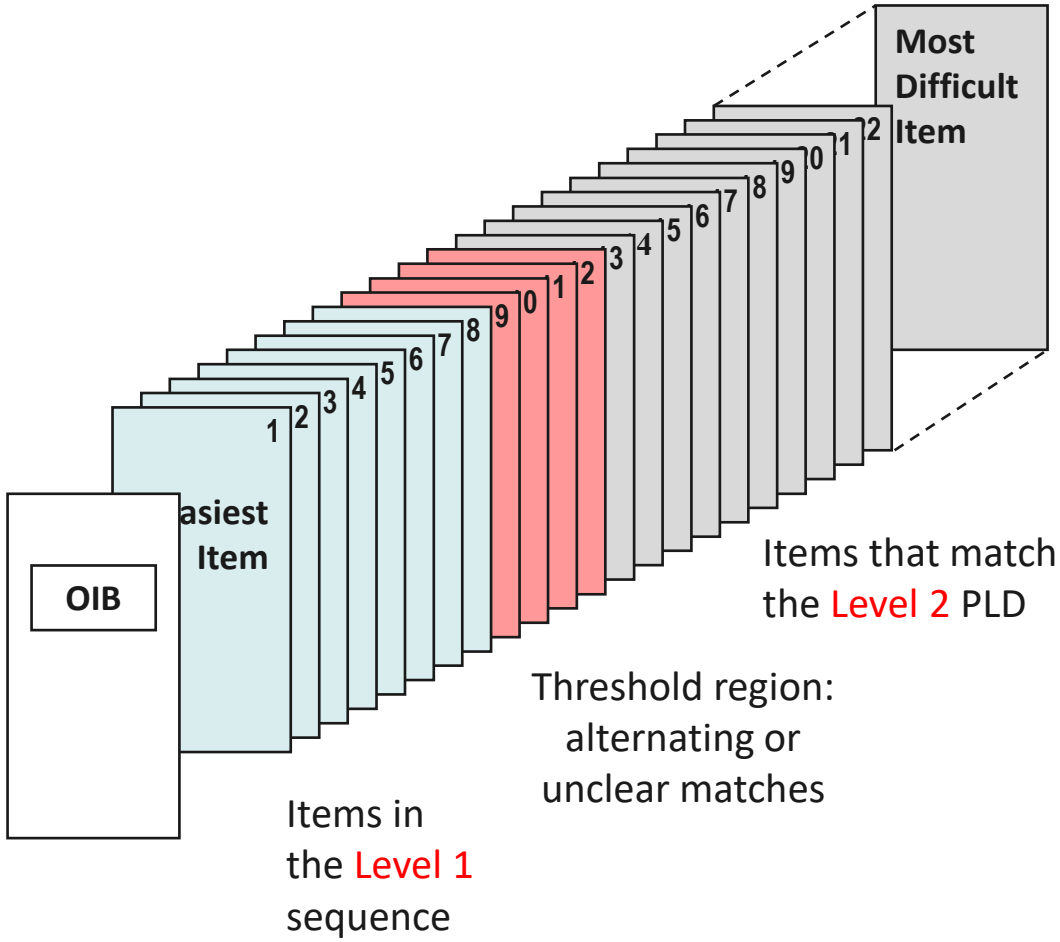
29	L3	
28	L3	
27	L3	
26	L3	
25	L3	Target cut score range
24	L3	
23	L3	
22	L3	
21	L2	
20	L2	
19	L2	
18	L2	
17	L2	
16	L2	
15	L2	

Cut Score Recommendations

- For your Level 2, 3, and 4 cut score recommendations
  - Write a content-based rationale for your recommendation to retain or adjust the cut scores

- Relevant for writing your rationales
  - Your answers to the two questions
  - The match between item response demands and expectations in the PLDs
- Notice: Refer to both **items** and **PLDs**

Cut Score Recommendations

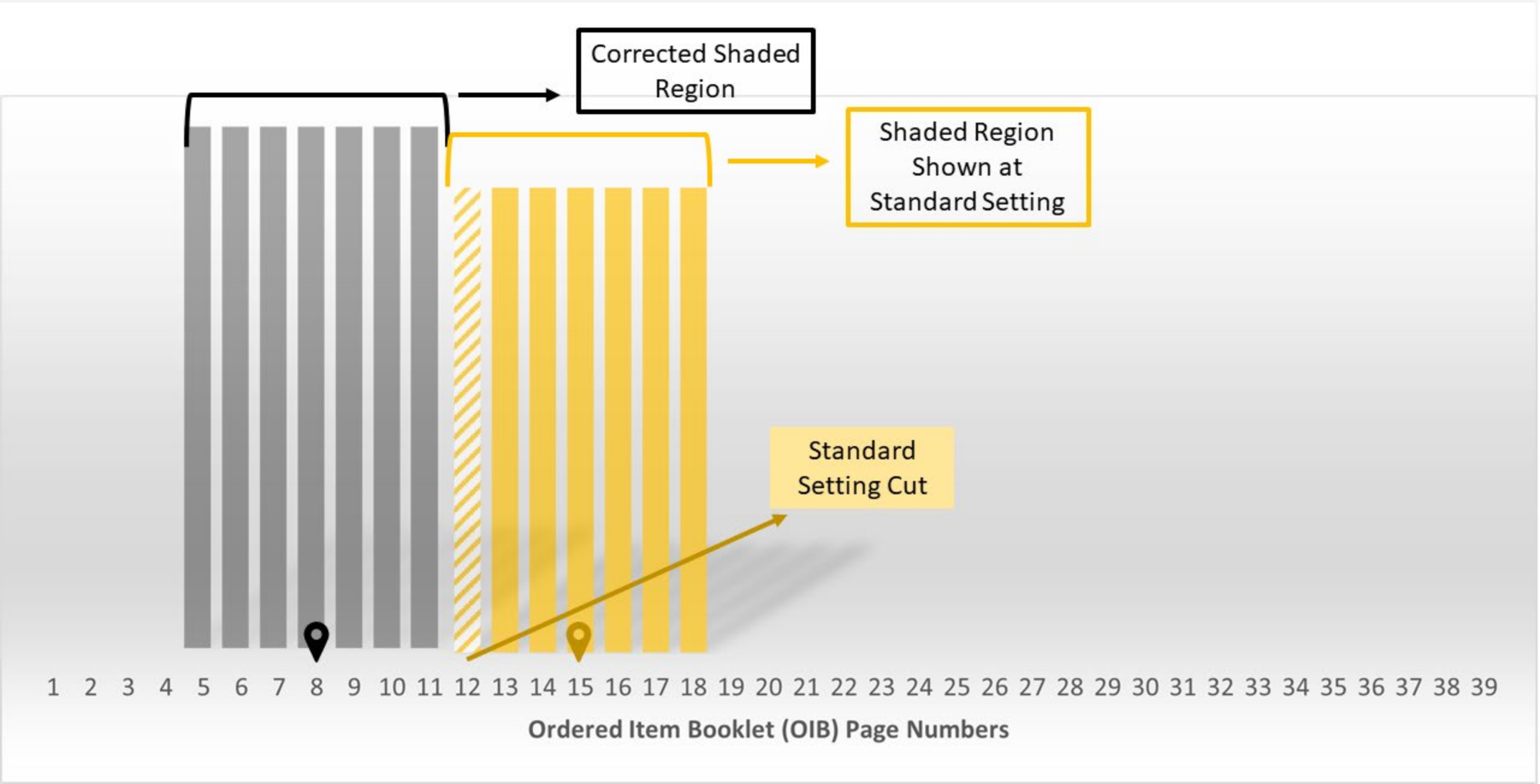


19	L2	Item response demands clearly match the expectations in the Level 2 PLD
18	L2	
17	L2	
16	L2	
15	L2	
14	L2	
13	L2	
12	L1	Threshold region
11	L2	
10	L1	
9	L2	
8	L1	Item response demands **do not match** the expectations in the Level 2 PLD
7	L1	
6	L1	
5	L1	
4	L1	
3	L1	
2	L1	
1	L1	

Standard Setting Committee Recommendations

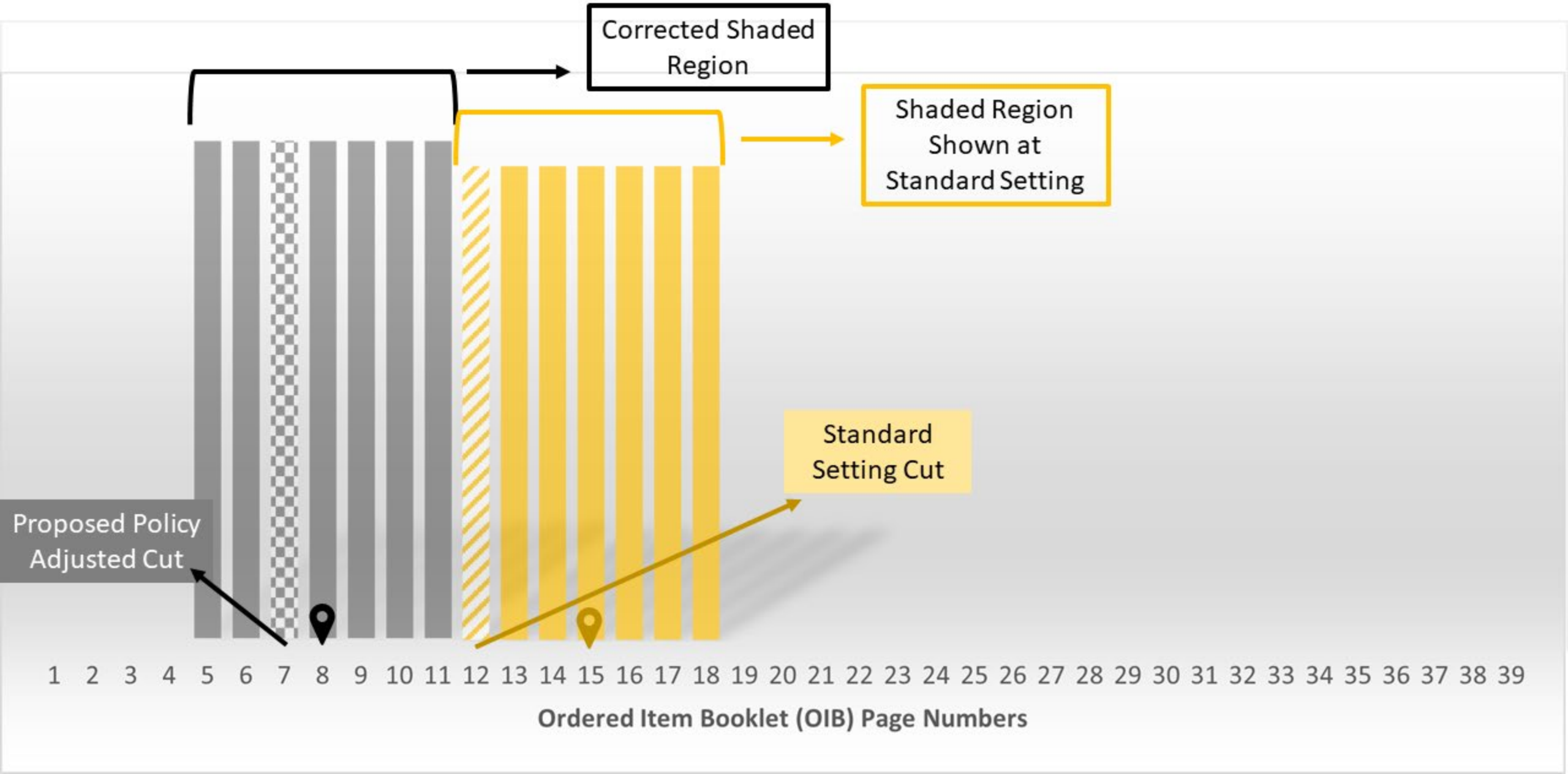
Subject	Grade	Round Information	Level 2 Cut OIB Page #	Level 3 Cut OIB Page #	Level 4 Cut OIB Page #
SCI05	5	SS Final Round	12	20	28
SCI08	8	SS Final Round	9	16	28
SCI11	11	SS Final Round	9	15	24

Grade 5 – Level 2

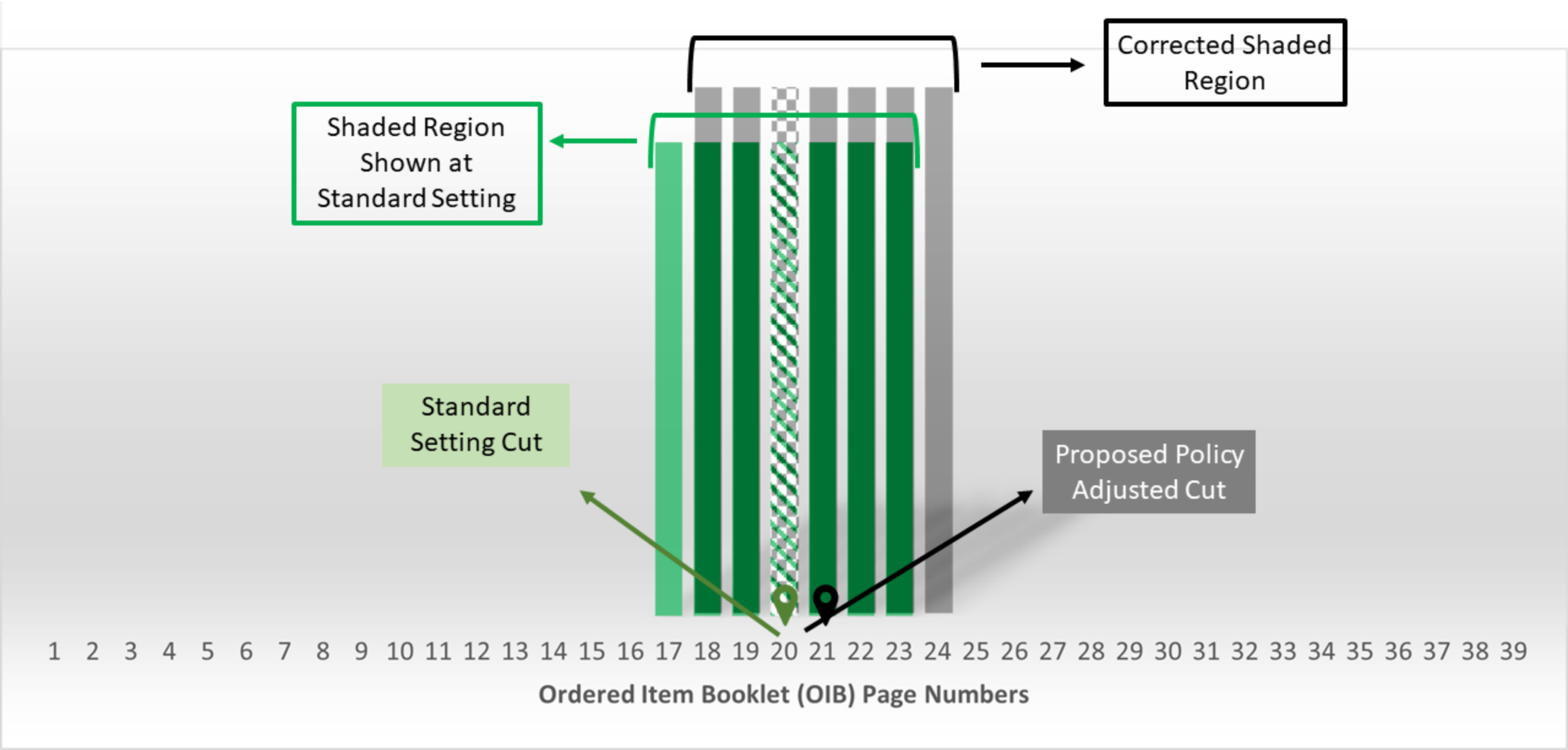


Grade 5

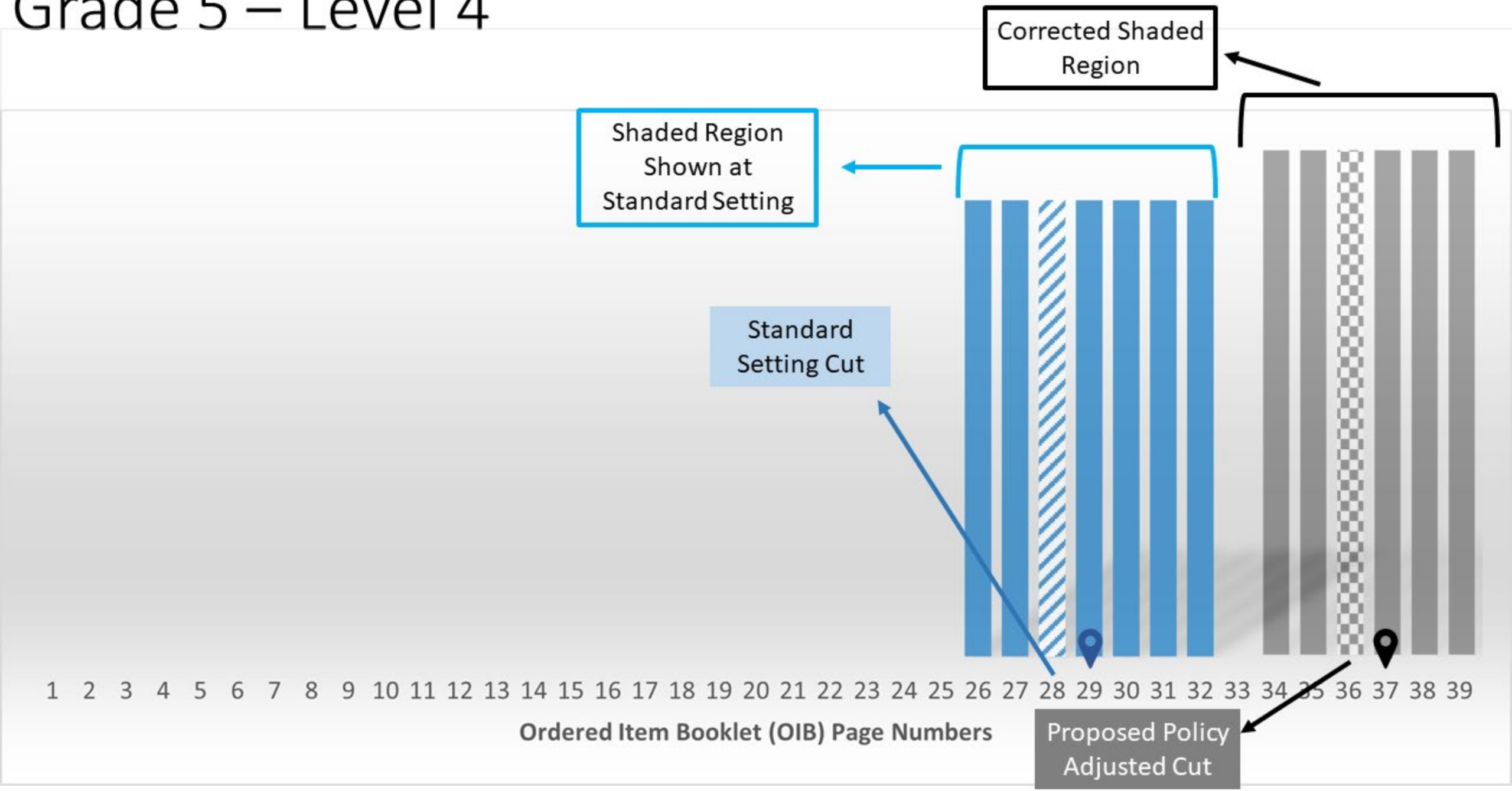
Grade 5 – Level 2



Grade 5 – Level 3

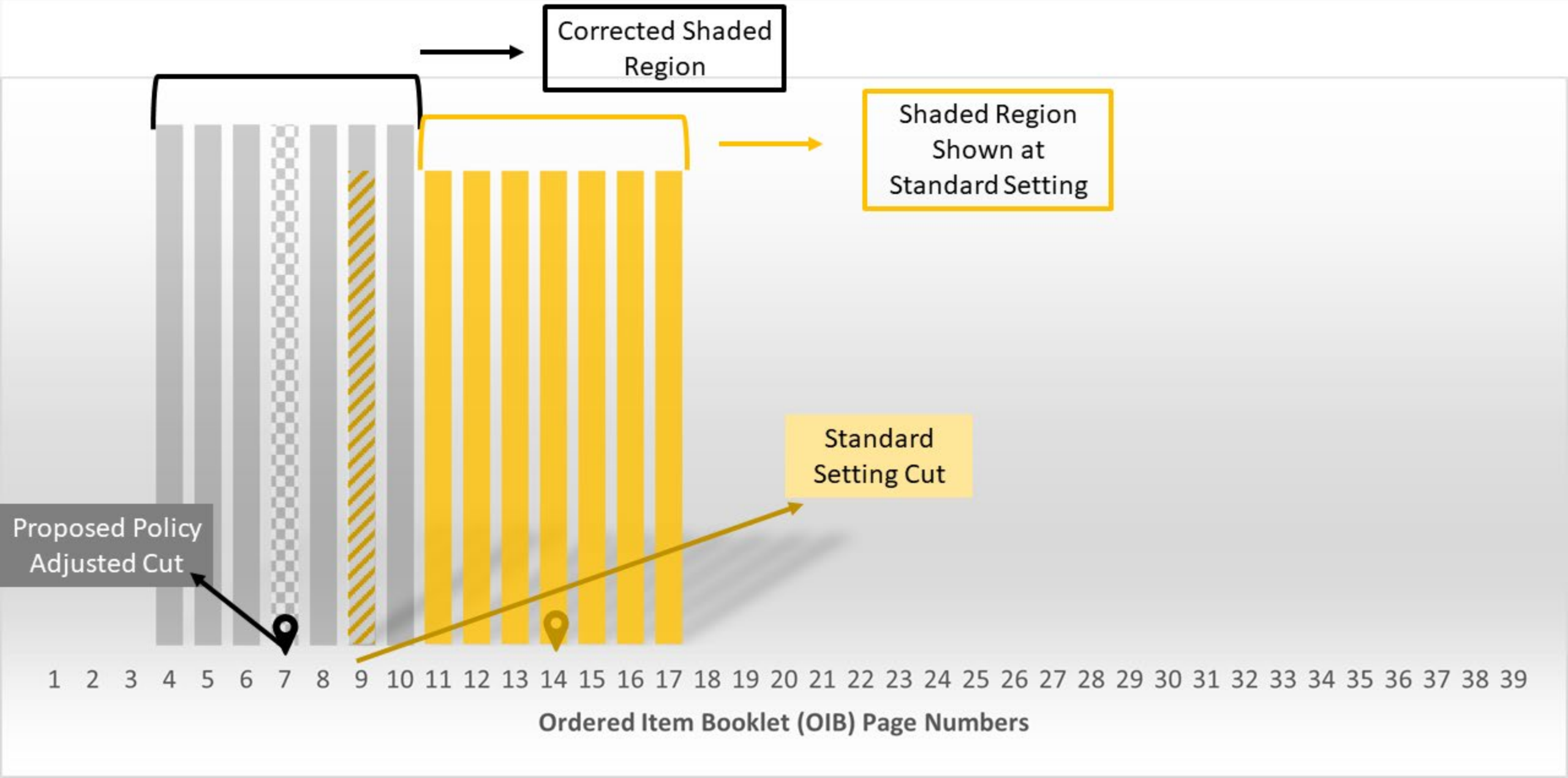


Grade 5 – Level 4

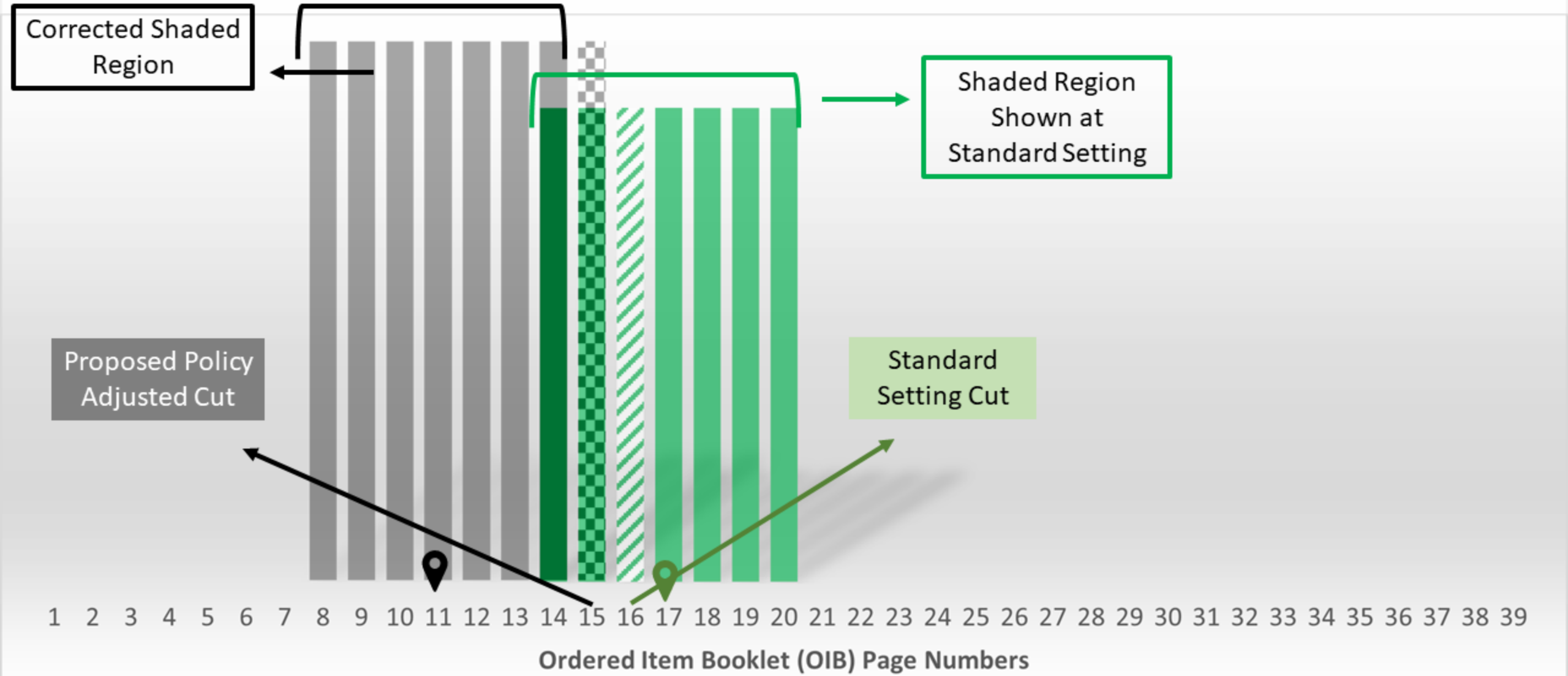


Grade 8

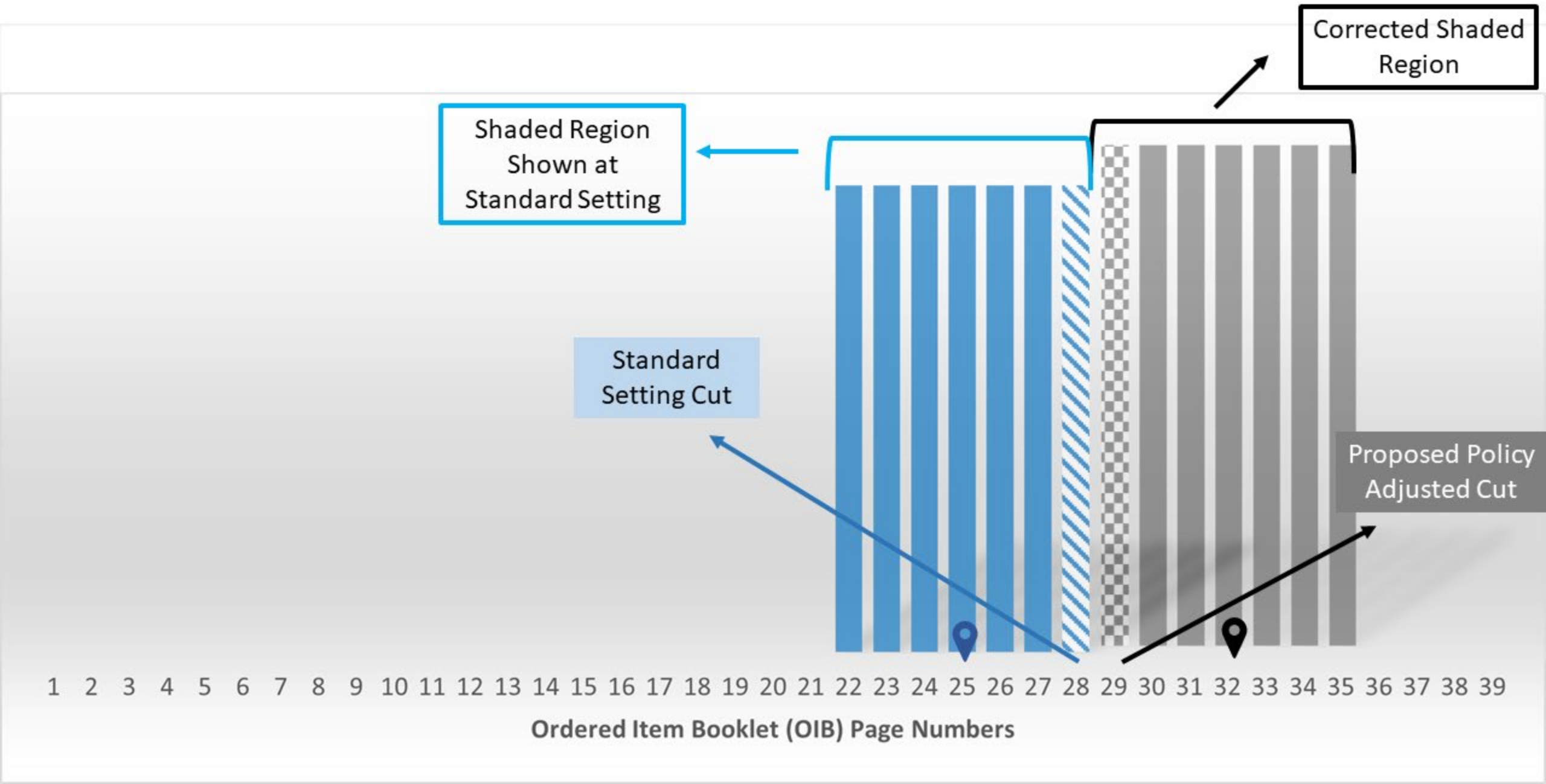
Grade 8 – Level 2



Grade 8 – Level 3

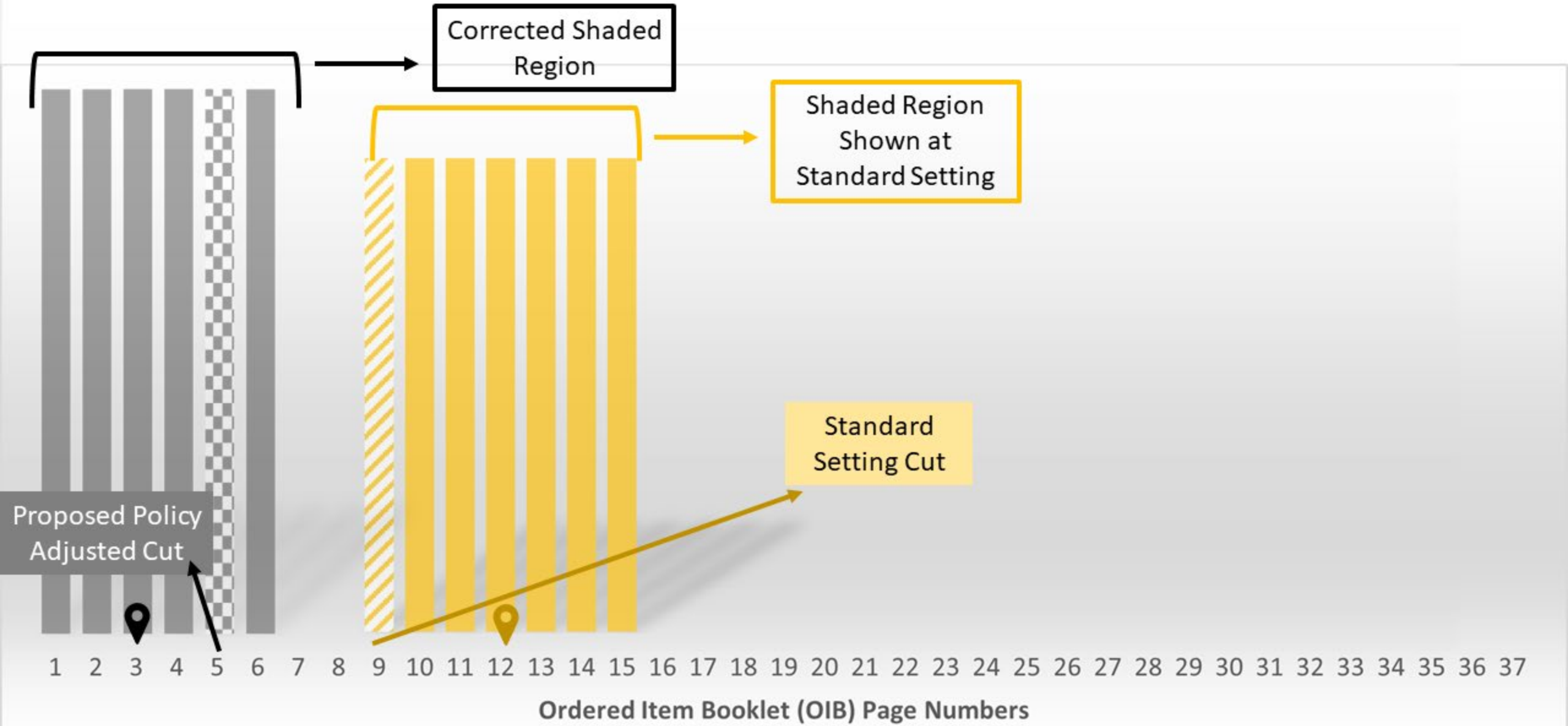


Grade 8 – Level 4

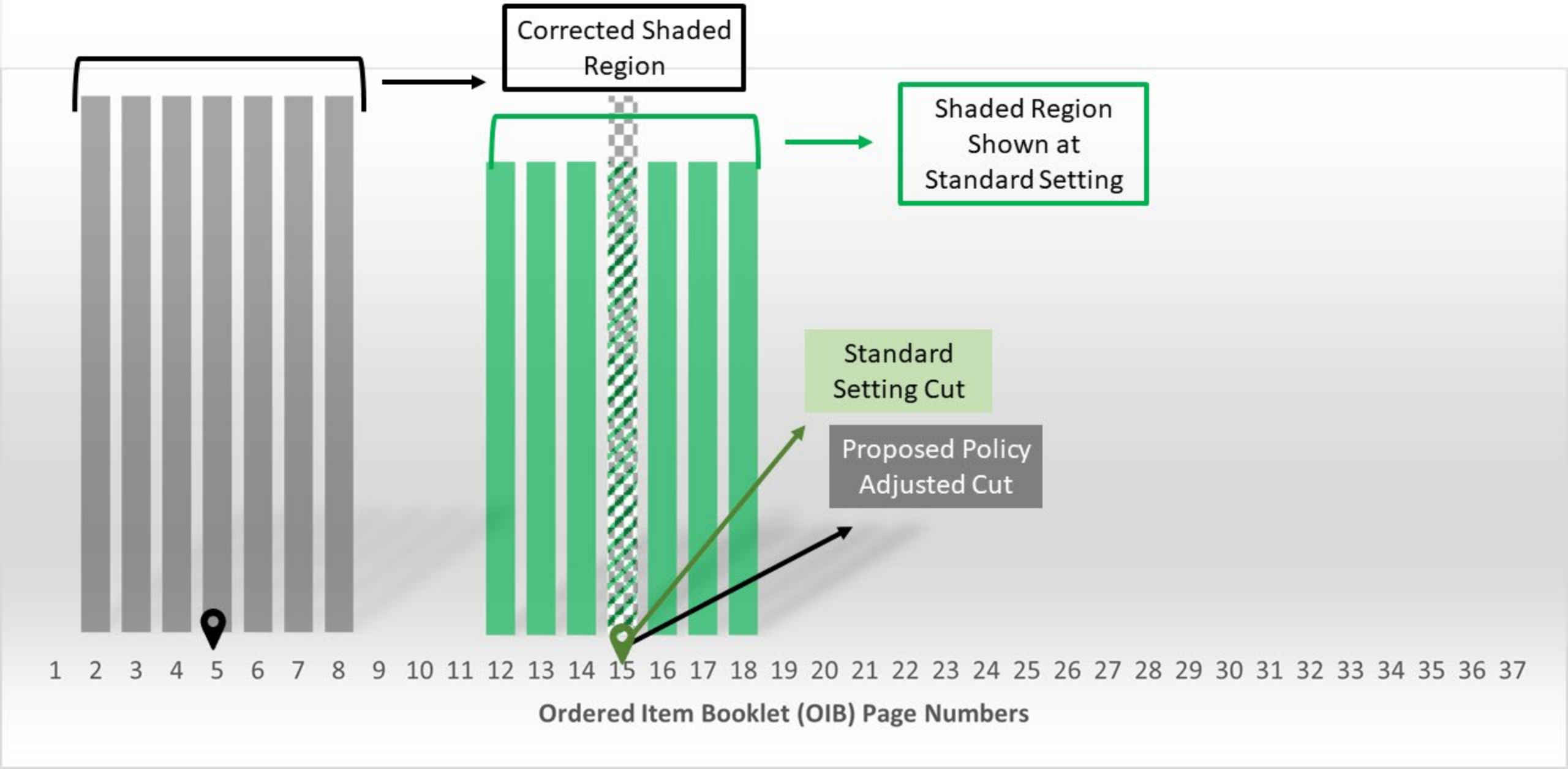


Grade 11

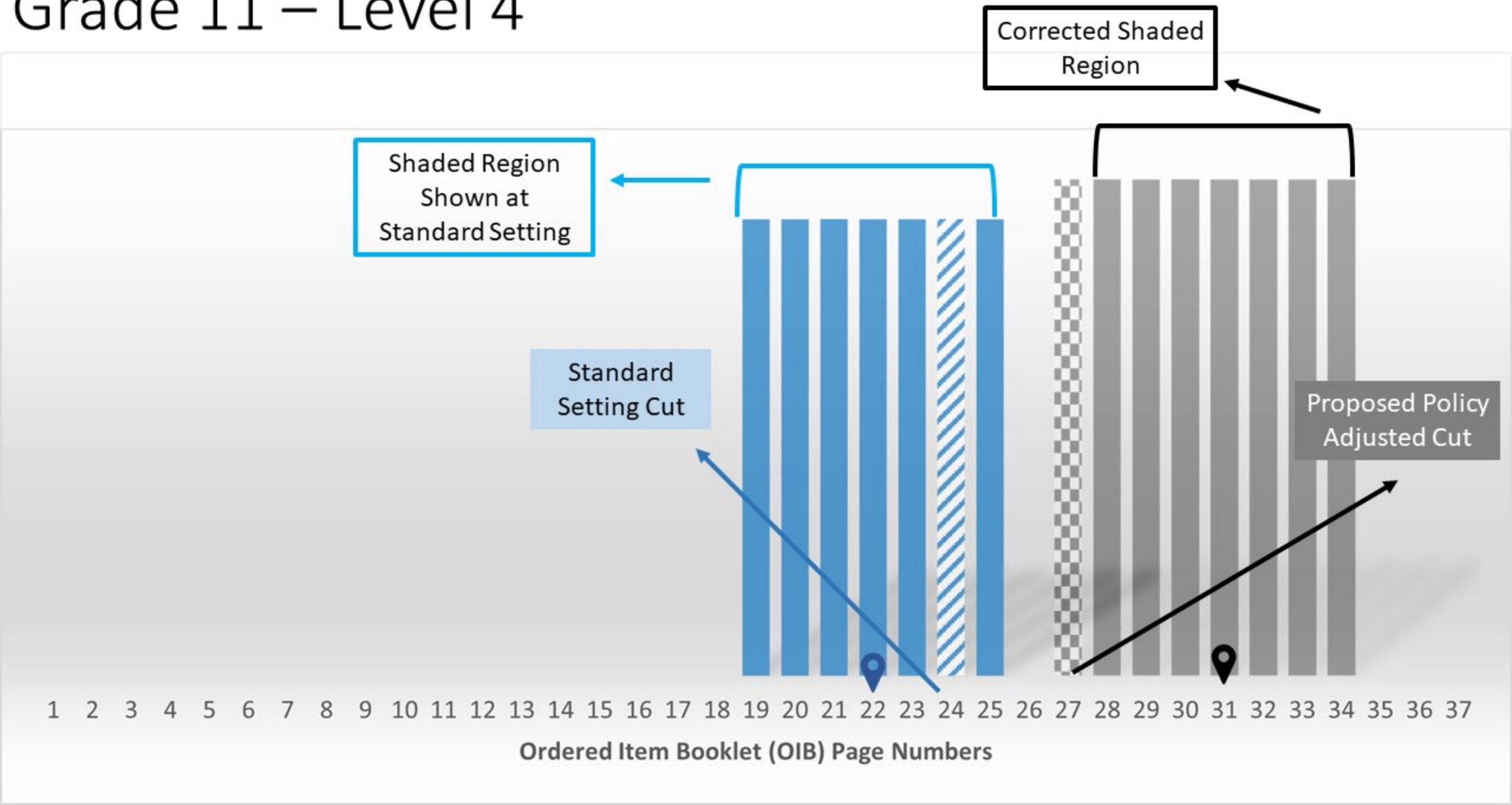
Grade 11 – Level 2



Grade 11 – Level 3



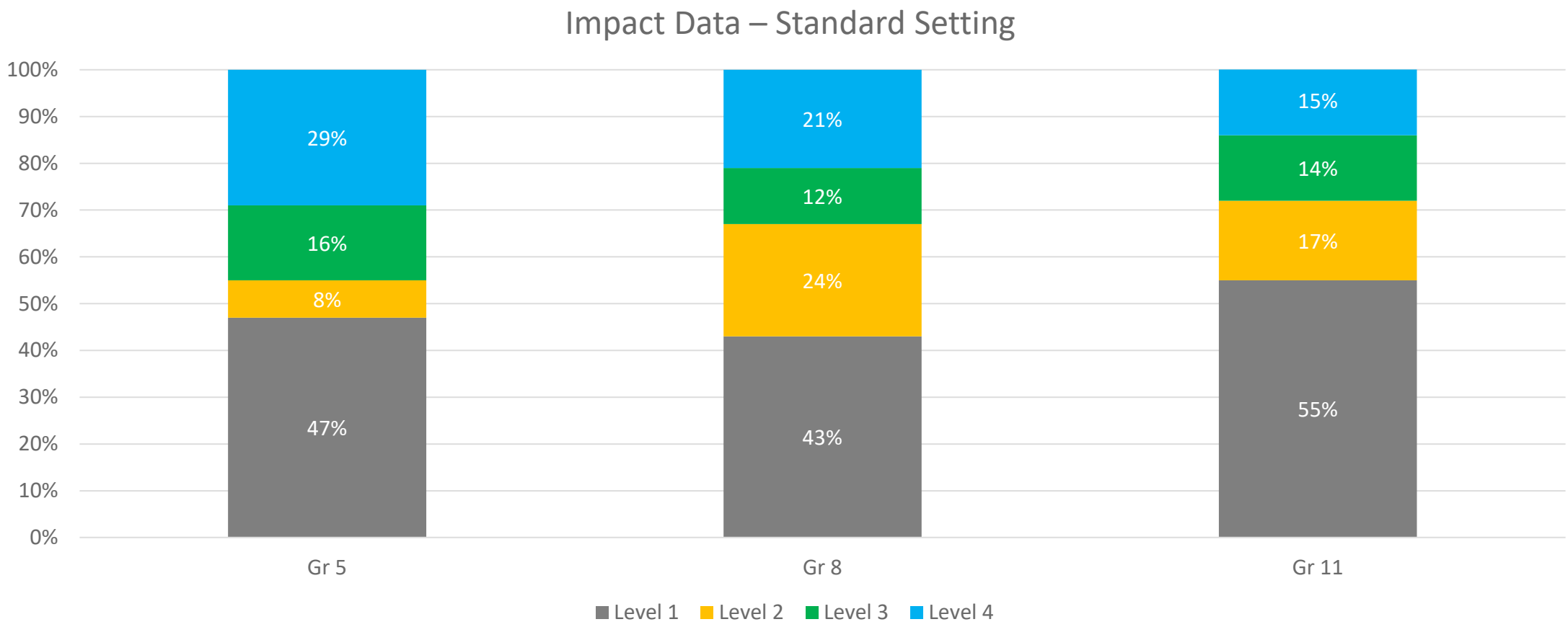
Grade 11 – Level 4



Standard Setting Committee Recommendations

Subject	Grade	Round Information	Level 2 Cut OIB Page #	Level 3 Cut OIB Page #	Level 4 Cut OIB Page #	Percent Level 1	Percent Level 2	Percent Level 3	Percent Level 4
SCI05	5	SS Final Round	12	20	28	47%	8%	16%	29%
SCI08	8	SS Final Round	9	16	28	43%	24%	12%	21%
SCI11	11	SS Final Round	9	15	24	55%	17%	14%	15%

Standard Setting Committee Recommendations



Judgment Task and Discussion

- Can you describe how you – and your colleagues – used the target cut scores and ranges to inform your work?
- Given what you learned about the target cut scores and ranges, how do you think that changing this information might have impacted your and/or your committee's cut score recommendations, if at all?
  - If move, would it be earlier in the OIB or later relative to its current location?
- Response and discussion by grade
- Confirm or adjust recommendations

Any Additional Questions?

Vertical Articulation and Policy Review

- In a policy review, leaders can consider the impact data,2 vertical articulation of the impact data across grade levels,3 and the likely effects on the policies and educational politics of the participating states and entities.
- Here, the goal is to adjust, as necessary, some cut scores recommended by the standard setting panels to achieve articulation—while preserving as much as possible the content-focused process that the standard setting panelists completed in following the ID-Matching process to recommend cut scores.

Vertical Articulation and Policy Review

- Rationale for Policy Review Recommendations
  - Panelist Recommendations
  - Vertical Articulation
  - Policy Expectations
  - Measurement Properties

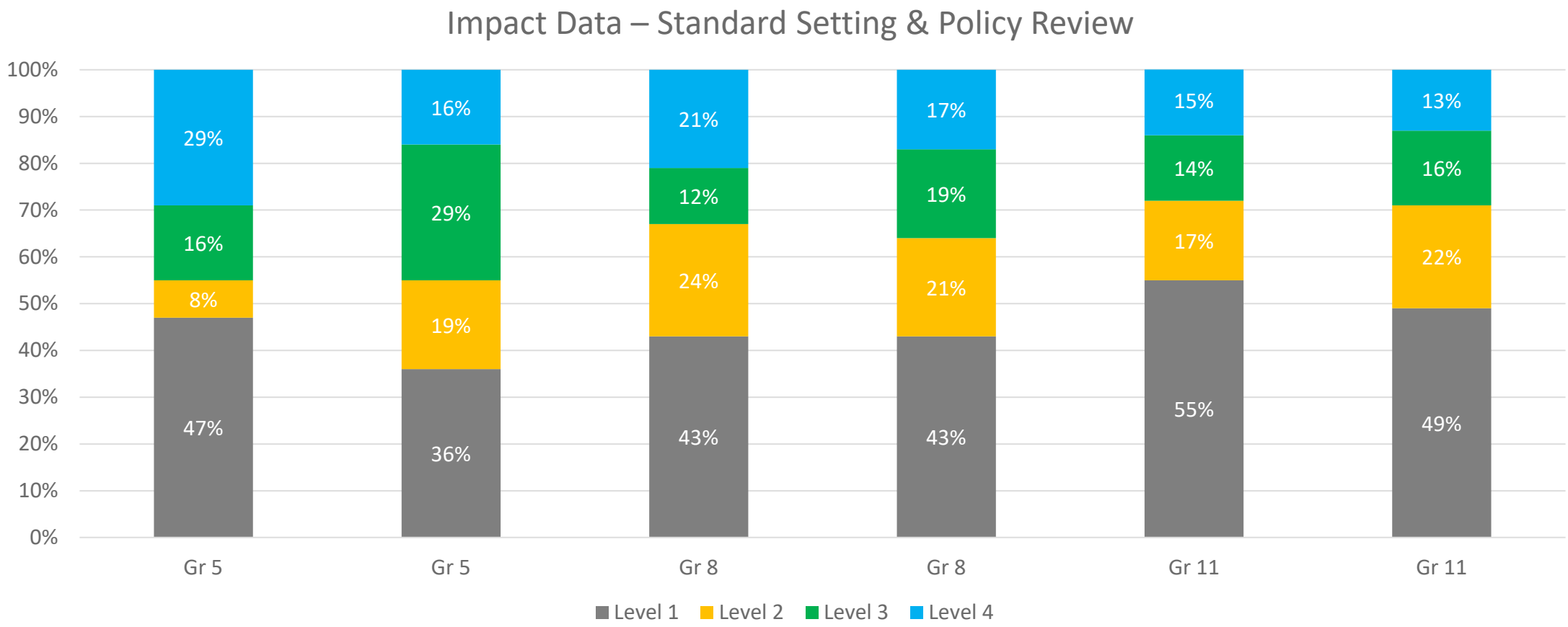
Vertical Articulation and Policy Review

- Review of results from standard setting committee recommendations
  - OIB page number cuts
  - Impact data
- Review of proposed policy review adjustments
  - OIB page number cuts
  - Impact data

Policy Review

Subject	Grade	Round Information	Level 2 Cut OIB Page #	Level 3 Cut OIB Page #	Level 4 Cut OIB Page #	Percent Level 1	Percent Level 2	Percent Level 3	Percent Level 4
SCI05	5	SS Final Round	12	20	28	47%	8%	16%	29%
SCI05	5	Policy Adjustment	7	20	36	36%	19%	29%	16%
SCI08	8	SS Final Round	9	16	28	43%	24%	12%	21%
SCI08	8	Policy Adjustment	7	15	29	43%	21%	19%	17%
SCI11	11	SS Final Round	9	15	24	55%	17%	14%	15%
SCI11	11	Policy Adjustment	5	15	27	49%	22%	16%	13%

Policy Review



Policy Review

- Question:
  - Do you agree with adopting the proposed policy review adjustments?
    - Why or why not? Any further adjustments?
  - Response and discussion by grade
    - Grade 5
    - Grade 8
    - Grade 11

Conclusion and Wrap Up