



# **Colorado Measures of Academic Success (CMAS) Science 2022 Standard Setting Report**

February 6, 2024 | Prepared by Pearson



## Table of Contents

Executive Summary .....	5
1. Introduction.....	7
1.1. Assessment Background.....	7
1.2. Performance Level Descriptors .....	8
1.3. Modified ID Matching Method .....	9
1.4. Participants .....	10
1.4.1. Panelists.....	10
1.4.2. Facilitators .....	10
1.4.3. Observers.....	10
2. Standard Setting Preparation.....	11
2.1. Standard Setting Website .....	11
2.2. Standard Setting Materials .....	11
2.3. Facilitator Training.....	13
2.4. Data Preparation .....	13
2.4.1. Item Mapping and Calibration .....	13
2.4.2. Ordered Item Sets.....	14
2.4.3. Feedback Data .....	15
2.5. Meeting Security .....	16
3. Standard Setting Process.....	17
3.1. Pre-Work .....	17
3.2. General Session .....	18
3.3. Breakout Session .....	18
3.3.1. “Experience the Assessment” Activity.....	18
3.3.2. PLD Review .....	19
3.3.3. Training .....	19
3.3.4. Practice Activity.....	19
3.3.5. Judgment Rounds 1–3 .....	20
3.3.5.1. Readiness Quiz .....	20
3.3.5.2. Judgment Process .....	20
3.3.6. PLD Recommendations.....	21
3.3.7. Evaluation Surveys.....	22
4. Results.....	23
4.1. Panelist Cut Score Agreement.....	23
4.2. Round 3 Results.....	23
4.3. CDE Review .....	24
4.4. Final Approval.....	25
4.5. Reporting Scale .....	25
5. Evidence of Procedural Validity.....	27
5.1. Committee Representation .....	27
5.2. Committee Training .....	27
5.3. Perceived Validity of the Standard Setting Meeting .....	27
6. References.....	29

Appendix A: Agenda .....	30
Appendix B: Participant Information Survey Results .....	32
Appendix C: Judgment Record Form Examples .....	35
Appendix D: Readiness Quiz.....	41
Appendix E: Feedback Data Examples.....	43
Appendix F: Evaluation Survey Results .....	48
Appendix G: Panelist Cut Score Agreement .....	60
Appendix H: PowerPoint Presentations.....	63

### **List of Tables**

Table 1.1. CMAS Science Spring 2023 Test Design.....	8
Table 1.2. Process Facilitators .....	10
Table 2.1. Panelist Materials.....	12
Table 2.2. Feedback Data Provided After Each Round .....	16
Table 3.1. Meeting Schedule Overview.....	17
Table 4.1. Round 3 Cut Score Recommendations and Impact Data.....	23
Table 4.2. CDE Review Cut Score Recommendations and Impact Data .....	24
Table 4.3. Final Recommended Cut Scores on the IRT Scale.....	26
Table G.1. Panelist Cut Score Agreement—Grade 5, Round 2.....	60
Table G.2. Panelist Cut Score Agreement—Grade 5, Round 3.....	60
Table G.3. Panelist Cut Score Agreement—Grade 8, Round 2.....	61
Table G.4. Panelist Cut Score Agreement—Grade 8, Round 3.....	61
Table G.5. Panelist Cut Score Agreement—Grade 11, Round 2.....	62
Table G.6. Panelist Cut Score Agreement—Grade 11, Round 3.....	62

### **List of Figures**

Figure 2.1. Website Interface Example.....	11
Figure 2.2. Traditional OIS Example.....	14
Figure 3.1. Region of Uncertainty Example .....	21
Figure 4.1. Round 3 Impact Data.....	23
Figure 4.2. CDE Review Impact Data .....	25
Figure 5.1. Panelists’ Confidence in How Well the Cut Scores Reflect the CAS and PLDs.....	28
Figure 5.2. Panelists’ Level of Support in the Recommended Cut Scores .....	28

## Executive Summary

The Colorado Measures of Academic Success (CMAS) Science assessment in Grades 5, 8, and 11 aligned to the 2020 Colorado Academic Standards (CAS) in science was administered for the first time in Spring 2022. Cut scores must be established following the first administration of a new assessment to ensure that student performance is properly categorized into performance levels. As such, a standard setting took place from September 27–28, 2022, in Denver, Colorado, to recommend the cut scores for the new CMAS Science assessment using a modified version of the Item Descriptor (ID) Matching standard setting method (Ferrara & Lewis, 2012). The recommendations were then presented to the Colorado Department of Education (CDE) and ultimately the Colorado State Board of Education for consideration and final approval. Three grade-level panels were convened, with a total of 45 educators participating across all panels.

Student achievement on the CMAS Science assessment is classified into four performance levels that delineate the knowledge and skills expected of students to demonstrate mastery: Level 1: *Partially Met Expectations*, Level 2: *Approached Expectations*, Level 3: *Met Expectations*, and Level 4: *Exceeded Expectations*. As such, the standard setting committees recommended three cut scores to separate the score scale into the four performance levels: the Level 2 cut (between Level 1 and Level 2), the Level 3 cut (between Level 2 and Level 3), and the Level 4 cut (between Level 3 and Level 4).

The performance levels are accompanied by performance level descriptors (PLDs) that articulate what a student should know and be able to do in a performance level. The PLDs play a major role in the ID Matching standard setting process, with panelists matching each item to a descriptor and basing their recommended cut scores on these judgments.<sup>1</sup>

### Process Overview

The meeting began with an overview of the standard setting process and the expected outcomes. Panelists then reviewed items from the Spring 2022 administration, followed by a walkthrough of the meeting materials including the 2020 CAS, PLDs, and the Pearson standard setting website where the cut score recommendations were made. The panelists then participated in a practice activity before beginning the standard setting process that occurred in three rounds. In Round 1, panelists reviewed the items in administration order so items associated with the same stimuli were reviewed together. In Rounds 2 and 3, items were reviewed in order of difficulty, from easiest to most difficult.<sup>2</sup>

- In Round 1, panelists used the PLDs to assign each item a performance level that best matched an item’s response requirements (i.e., the knowledge and skills students must demonstrate to provide a correct response or receive a specific score). When matching an item to a performance level, panelists were encouraged to consider whether 50 out of 100 students at each performance level would correctly respond to the item, beginning with Level 1. For example, would 50% of Level 1 students (i.e., students with the knowledge and skills associated with this performance level) answer this item correctly? If no, move on to Level 2 (i.e., would 50% of Level 2 students answer this item correctly?). If yes, it could be considered a match with Level 1.

---

<sup>1</sup> The final CMAS Science PLDs are located online at [https://www.cde.state.co.us/assessment/cmas\\_plds](https://www.cde.state.co.us/assessment/cmas_plds).

<sup>2</sup> The Round 1 review in administration order is the only modification to the original ID Matching method.

- Round 2 began with a review of the Round 1 judgments, followed by matching each item to a performance level as was done in Round 1, adjusting their initial matches as needed. Panelists’ individual item judgments resulted in threshold regions for each performance level, where the item judgments were alternating or unclear. This is known as the region of uncertainty, and it was from this region that panelists selected the item that they felt had expectations that transitioned from the lower performance level to the next higher performance level (as described in the PLDs). This item became the cut score recommendation for each performance level.
- Round 3 began with a review of the Round 2 item judgments and cut score recommendations. Panelists then completed the same steps as Round 2, adjusting their item-descriptor matches and cut recommendations as needed. The same feedback data presented after Round 2 was provided for Round 3, with the addition of impact data (i.e., the percentage of students who would be placed into each performance level based on the recommended cut scores and Spring 2022 data). The impact data were presented only after Round 3 to ensure that panelists’ recommendations were based solely on content.

Panelists also completed three evaluation surveys throughout the process: (1) after the practice judgment activity, (2) after Round 3, and (3) at the end of the meeting following the PLD review. The purpose of the evaluations was to determine the panelists’ understanding of the process and their confidence in the results. Overall, the evaluation results indicate that the panelists understood the process and were confident about their recommendations.

## Results

The committee cut score recommendations were the median of all the panelists’ cut scores from Round 3, which were then reviewed from a policy perspective by CDE. Table E.1 presents the resulting recommended cut scores and associated impact data. The cut scores recommended by CDE were approved by the State Board of Education on December 14, 2022. The final cuts were then transformed into scale scores for the final reporting scale, as shown in Table E.2, and will be used to report student results on the CMAS Science assessments starting in Spring 2023.

**Table E.1. Final Cut Scores with Associated Impact Data**

Source	Grade	Level 2 Cut	Level 3 Cut	Level 4 Cut	%Level 1	%Level 2	%Level 3	%Level 4
Round 3	5	12	31.5*	50	38.5	42.4	17.6	1.6
	8	12	33	53	51.7	35.1	12.8	0.5
	11	7	27	49	44.2	34.8	20.7	0.4
CDE Review	5	<b>12</b>	<b>29</b>	<b>50</b>	38.5	30.9	29.1	1.6
	8	<b>9</b>	<b>26</b>	<b>53</b>	39.7	29.7	30.0	0.5
	11	<b>7</b>	<b>27</b>	<b>49</b>	44.2	34.8	20.7	0.4

*Note.* The cuts represent item numbers from the OIS. The cuts resulting from CDE’s review were approved by the State Board of Education and are considered the final cuts, as shown in bold.

\*The median for this cut fell between two items.

**Table E.2. Final Scale Score Cut Scores**

Grade	Level 1: <i>Partially Met Expectations</i>	Level 2: <i>Approached Expectations</i>	Level 3: <i>Met Expectations</i>	Level 4: <i>Exceeded Expectations</i>
5	650–724	<b>725</b> –749	<b>750</b> –788	<b>789</b> –850
8	650–724	<b>725</b> –749	<b>750</b> –796	<b>797</b> –850
11	650–724	<b>725</b> –749	<b>750</b> –786	<b>787</b> –850

*Note.* The cut score is the lowest score in the range, as shown in bold.

## 1. Introduction

Standard setting is the process whereby a group of educators is convened to recommend the cut scores (also known as performance standards) that separate an assessment's score scale into performance levels (i.e., a cut score is the minimum score students must receive to be classified into a certain performance level). This document presents the results of the standard setting meeting from September 27–28, 2022, in Denver, Colorado, to establish cut scores for the new Colorado Measures of Academic Success (CMAS) Science assessment in Grades 5, 8, and 11 using a modified version of the Item Descriptor (ID) Matching standard setting method (Ferrara & Lewis, 2012). The report includes a description of the method and the steps taken to generate the cut score recommendations, addressing the *Standards for Educational and Psychological Testing* (AERA et al., 2014, pp. 107–109) and the involvement of the expert judgment of Colorado educators following a vetted standard setting method.

### 1.1. Assessment Background

The CMAS Science assessments are online, fixed-form assessments administered in the Pearson TestNav online testing platform, with paper-accommodated forms available as needed. The assessments are aligned to the 2020 Colorado Academic Standards (CAS) in science approved by the Colorado State Board of Education in August 2018, with full instructional implementation beginning in the 2021–2022 school year.<sup>3</sup> The new CMAS Science test was administered to all tested students for the first time in Spring 2022, with the first administration with scale score and performance level reporting in Spring 2023. The cut scores were set using Spring 2022 data.

Colorado was required by state statute to revise the CAS by July 1, 2018 (CDE, 2014). As such, the 2009 CAS in science underwent a substantial update in 2018 to keep up with the shift to the Next Generation Science Standards (NGSS; NGSS Lead States, 2013) that were designed to reflect more recent research and thinking in science education. The 2020 CAS in science are considered three-dimensional in that they incorporate Disciplinary Core Ideas (DCIs), Science and Engineering Practices (SEPs), and Crosscutting Concepts (CCCs):

- DCIs form the basis for the content that students are expected to know at each grade in Physical Science, Life Science, and Earth and Space Science.
- SEPs describe how scientists investigate and build models and theories of the natural world or how engineers design and build systems. They reflect science and engineering as they are practiced and experienced.
- CCCs describe the concepts that hold true across the natural and engineered world that students can use to make connections across seemingly disparate disciplines or situations, connect new learning to prior experiences, and more deeply engage with material across the other dimensions.

The CMAS Science assessment is given in Grades 5, 8, and 11. Consistent with the standards, the Grade 5 assessment assesses the grade-level standards. Because the science standards are articulated by grade band at the middle school and high school levels rather than by grade levels, the Grade 8 CMAS Science assessment assesses all middle school science standards, and the Grade 11 assessment assesses all high school science standards.

---

<sup>3</sup> The 2020 CAS in science are located online at <https://www.cde.state.co.us/coscience/statestandards>.

The assessment is divided into item sets that present phenomenon-based scenarios as either interactive science simulations or static stimuli, followed by items related to the simulation or scenario. Separate standalone items are also included that are not associated with a stimulus to target a small number of CAS not represented in the scenarios. A phenomenon is an observable event that students can use the three dimensions (DCI, SEP, and CCC) to explain or make sense of. The items are either 1-point selected-response, 1-point technology-enhanced, or 2-point constructed-response item types.

Table 1.1 presents a summary of the CMAS Science test designs. The content for each assessment is divided across the Physical Science, Life Science, and Earth and Space Science reporting categories, as shown in the table that presents the percentage of points by reporting category.<sup>4</sup> Each assessment is administered across three sessions and is timed, with the number of minutes allotted per session provided in the table. The CR items are human scored, whereas all other item types are machine scored. Student performance is reported as an overall scale score and by reporting category, as well as an SEP subscore and a subscore for each standard. Based on the overall scale score, a student’s performance is classified into one of four performance levels: Level 1: *Partially Met Expectations*, Level 2: *Approached Expectations*, Level 3: *Met Expectations*, and Level 4: *Exceeded Expectations*.

**Table 1.1. CMAS Science Spring 2023 Test Design**

Grade	Total #Points	#Test Sessions	#Minutes per Session	Physical Science	Life Science	Earth and Space Science	SEPs represented in the CAS
5	51	3	80 minutes	35%	24%	41%	65–75%
8	61	3	80 minutes	34%	36%	30%	65–74%
11	50	3	50 minutes	36%	33%	31%	65–74%

*Note.* Life Science is Physical/Life Science in Grade 5.

## 1.2. Performance Level Descriptors

A performance level descriptor (PLD) describes what a student should know and be able to do for a performance level as articulated in the content standards (e.g., the set of statements describing what it means for a Grade 5 student to have *Met Expectations* in science). They are used to support meaningful interpretations of student performance in the different performance levels.

The 2020 CAS in science provided the foundation for PLD development. Prior to the standard setting, Pearson created a draft set of PLDs representing a gradual increase in expectations across the performance levels that were then reviewed by CDE. Colorado educators then reviewed the PLDs for alignment to the 2020 CAS and consistency of expectations across performance levels in a meeting from November 6-7, 2019. The revised PLDs were reviewed and finalized by CDE and brought to the standard setting meeting for use in the modified ID Matching method.

The PLDs were a crucial part of the standard setting as they were used to match the expectations of the performance levels to the expectations of the items. Panelists also had the opportunity to provide suggestions and edits to the PLDs that were finalized after the standard setting meeting and are located online at [https://www.cde.state.co.us/assessment/cmas\\_plds](https://www.cde.state.co.us/assessment/cmas_plds).

<sup>4</sup> Based on the frameworks available online at [https://www.cde.state.co.us/assessment/cmas\\_testdesign](https://www.cde.state.co.us/assessment/cmas_testdesign)



### **1.3. Modified ID Matching Method**

Traditional standard setting approaches during the COVID pandemic and post-pandemic years presented several challenges for states transitioning to new academic standards and tests, including accounting for the impact of significant opportunity to learn issues on first-year item characteristics. These issues were acknowledged and combatted through the selection of a heavily content-based standard setting methodology after the assessments were administered to Colorado students for the first time.

The ID Matching standard setting method (Ferrara & Lewis, 2012) is a content- and item-based approach that requires panelists to match the knowledge and skill expectations of each item to those in the PLDs. The cognitive-judgmental task of identifying response demands of items and matching them to the PLDs is closely aligned with the skillsets and experience of classroom teachers (Ferrara & Lewis, 2012). The stimulus-based test design of CMAS Science also lent to the use of the modified ID Matching method that presents items in administration order during Round 1 so panelists can consider the impact of the stimuli on their judgments. This method also limited the extent to which student performance influenced the panelists' recommendations, allowing them to focus more on the content.

To recommend cut scores for each performance level based on the ID Matching method, panelists consider the association between the expectations of student proficiency as defined by the PLDs and the knowledge and skills measured by the test items. They then assign each item a performance level with the student expectations that best matches the item response requirements, followed by recommending cut scores for each performance level based on these item-descriptor matches and item difficulty order.

The original ID Matching method uses a traditional ordered item set (OIS) that orders items by difficulty to facilitate the panelists' review of the items. However, in Round 1 of the modified ID Matching method, the items are ordered by administration order to accommodate for cluster sets that include items referencing common stimuli. The traditional OIS ordered by difficulty would not allow the panelists to review the items and make their classifications based on the association of the items within the item cluster, so the administration order modification in Round 1 allows the panelists to make their classifications, including any possible item cluster effects, while still maintaining the major aspects of the ID Matching approach.

Items are ordered by difficulty, from easiest to hardest, in Rounds 2 and 3. Because the items are ordered by difficulty, the performance level classifications are expected to progress from the lowest performance level to the highest performance level. The threshold regions, or region of uncertainty, for the performance levels are the areas in the OIS where the performance level classifications are not strictly ordered as expected. Individual cut score recommendations for the performance levels are selected by the panelists from the items within this region of uncertainty.

While Round 1 does not result in cut score recommendations and instead focuses on matching items to performance levels based on the PLDs, Rounds 2 and 3 focus on determining where in the region of uncertainty best indicates progression from one performance level to the next. The committee cut score recommendations are the median of the panelists' individual recommendations from Round 3.

## 1.4. Participants

The general setup of the standard setting committee included panelists, content specialists, observers, and the facilitators. A lead data analyst was also present to perform all data analyses for the standard setting meeting, while a second analyst replicated the analyses. During the meeting, the analysts collected panelist judgment data, performed independent analysis to verify the results, and prepared panelists' feedback. Prior to the meeting, Pearson provided a staffing plan to CDE that communicated the psychometric, content, and support staff.

### 1.4.1. Panelists

Panelists were selected by CDE to participate in the standard setting to be representative of educators and stakeholders from schools and districts across the state based on various criteria such as expertise in the concepts and skills reflected in the 2020 CAS in science and in student development of the concepts and skills; working with students with disabilities; and/or working with multilingual learners. A total of 45 educators participated across the three panels, with 14–16 panelists per panel.

Appendix B presents the results from the participant information survey that panelists completed to provide background information. The panels were comprised of individuals from across the state of Colorado, including classroom teachers and administrators from K–12 districts. The panelists brought content knowledge and classroom experience to the process and played an integral role in recommending the cut scores. The panelists were divided into three panels for Grades 5, 8, and 11. Each panel was then further divided into table groups, with five or six panelists per table.

### 1.4.2. Facilitators

Each panel was led by a process facilitator with knowledge and experience facilitating standard setting meetings. A lead facilitator responsible for the overall process was also present to oversee the standard setting, assist the process facilitators as needed, and answer any questions from the panelists. The process facilitator was responsible for ensuring that appropriate processes were followed throughout all phases of the meeting and verifying that panelists had a solid understanding of the tasks they were being asked to complete. Table 1.2 presents the process facilitator for each panel. The lead facilitator was Eric Moyer, Ph.D., from Pearson.

**Table 1.2. Process Facilitators**

Panel	Facilitator
Grade 5	Ha Phan, Ph.D.
Grade 8	Lisa Ehrlich, Ph.D.
Grade 11	Edward Wolfe, Ph.D.

### 1.4.3. Observers

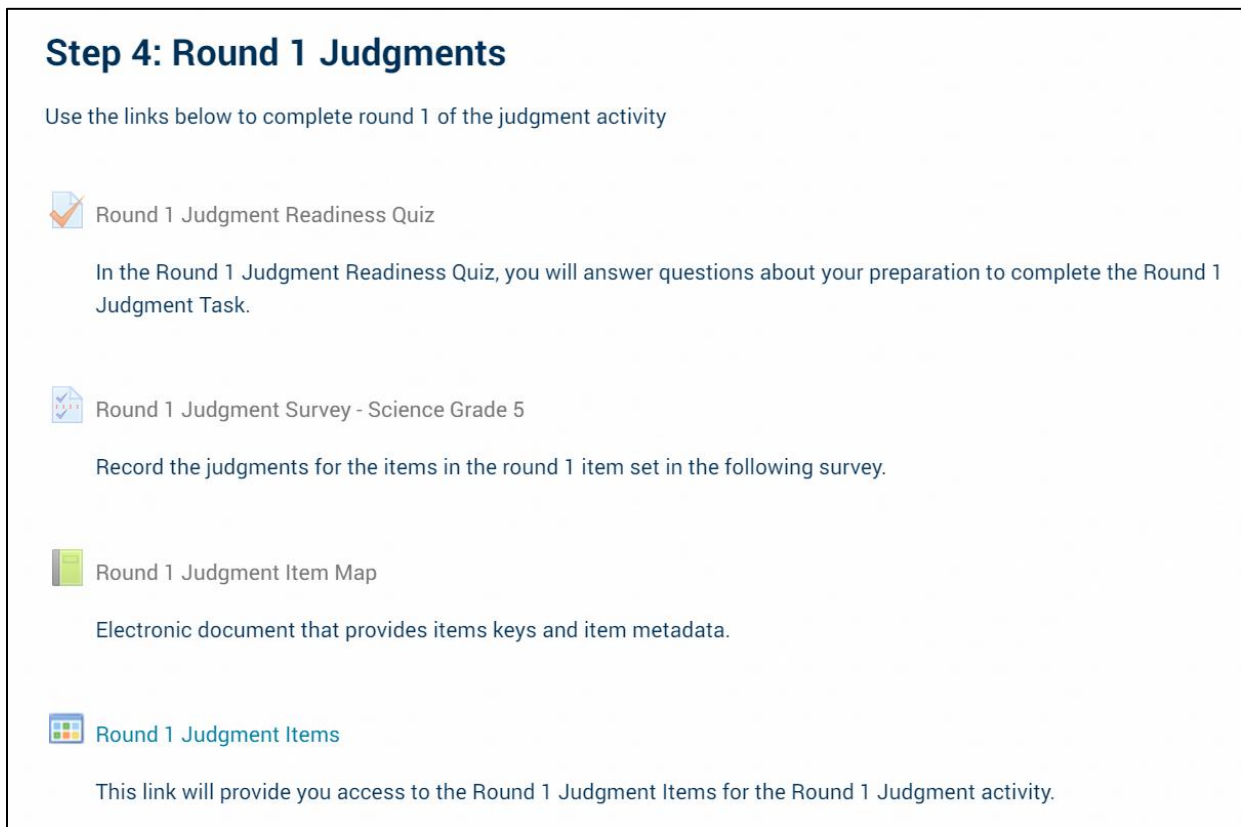
Staff from CDE were available as observers to help answer content and policy questions during the meeting and to see the standard setting process in action. Observers did not participate in the standard setting process except in isolated cases to support the facilitators when panelists had questions about the process. The purpose of observation was to allow individuals to experience the standard setting process and, in some cases, provide feedback. The number of observers was kept to a maximum of five individuals so the panelists did not feel overwhelmed.

## 2. Standard Setting Preparation

### 2.1. Standard Setting Website

The Pearson standard setting website served as the online platform during the standard setting meeting, providing a secure, central location for accessing materials and collecting panelist judgments. The standard setting website also allowed panelists to access online items within Pearson’s online testing environment, TestNav 8, during the “Experience the Assessment” activity, the practice exercise, and for each round to view the items in the OIS. The website was built using Moodle, an online, open-source collaboration and learning tool, and facilitators controlled what sections of the website were visible to panelists during the meeting to streamline the process. Figure 2.1 presents an example of the website interface to show what panelists saw during the meeting.

**Figure 2.1. Website Interface Example**



### 2.2. Standard Setting Materials

The Pearson standard setting team worked with CDE to develop the standard setting materials used during the meeting and ensure that all information provided to panelists communicated the correct information. The process for developing materials and the standard setting website began with creation of templates for each resource that were reviewed and approved by CDE. Using the approved templates, the resources were then created by Pearson and reviewed by CDE before being finalized for publication for the meetings.

Table 2.1 presents the materials provided to panelists for use during the standard setting, either online via the standard setting website and/or on paper in their panelist folder. The location of copies or examples of each material is indicated in parentheses if available. Each panelist also received a laptop computer in their meeting room to access the website and online resources.

**Table 2.1. Panelist Materials**

Material	Online	Paper
Agenda (Appendix A)	✓	✓
2020 CAS in science ( <a href="https://www.cde.state.co.us/coscience/statestandards">https://www.cde.state.co.us/coscience/statestandards</a> )	✓	
PLDs ( <a href="https://www.cde.state.co.us/assessment/cmas_plds">https://www.cde.state.co.us/assessment/cmas_plds</a> )	✓	
Item maps	✓	
PLD comment sheet		✓
Practice judgment survey	✓	
Rounds 1–3 judgment surveys	✓	
Practice judgment record form		✓
Round 1 judgment record form (Appendix C)		✓
Rounds 2 and 3 judgment record form (Appendix C)		✓
Readiness quizzes (Appendix D)	✓	
Evaluation surveys (Appendix F)	✓	
Feedback data after each round (Section 2.4.3)	✓	
PLD review form	✓	

The item map was provided for the “Experience the Assessment” activity, practice activity, and Rounds 1–3 and included a summary of the items in the OIS, including the maximum points possible, item type, item scoring keys, and content standard alignments. The judgment record form was provided in hardcopy so panelists could record their item judgments during each round (including the practice round), whereas their official record used in the analysis was captured in the judgment survey on the standard setting website. Appendix C presents an example of the judgment form for Round 1 and Round 2, with each multi-point item included once for each score point. The judgment record form for Round 1 presented the items by administration order, whereas the judgment record forms for Rounds 2 and 3 ordered the items by difficulty. The Rounds 2 and 3 judgment forms were identical.

A breakout session PowerPoint presentation for each grade was also developed to guide the facilitators through the dissemination of information and materials throughout the meeting. The Pearson standard setting team created the initial PowerPoint presentations that were then reviewed by CDE, with any suggested edits resolved by Pearson. The final presentations were reviewed and approved by CDE. The notes and scripts provided information for procedural steps, talking points, definitions to explain concepts to panelists, answers to commonly asked questions, and specific materials to distribute to panelists during the meeting. Appendix H presents the Grade 5 breakout session PowerPoint as an example.

### 2.3. Facilitator Training

The process facilitators underwent a training prior to the standard setting that included the following facets to prepare them for leading the standard setting panels:

- Overview of the assessment program, including the different item types, scoring rules, and performance levels
- Walkthrough of the Pearson standard setting website to become familiar with the platform, including how to model use of the website for panelists, how panelists gain access to the site, and how to troubleshoot issues
- Walkthrough of the standard setting agenda, with a focus on specific issues such as time management, the use of the online platform, and communicating feedback information
- A review of the breakout session PowerPoint slides and script that provided guidance throughout the presentation, including when specific language was to be used

### 2.4. Data Preparation

In preparation of the standard setting meeting, several sets of analyses were performed using data from the Spring 2022 administration to be able to order the items by difficulty and provide the impact data. The analyses included item response theory (IRT) item calibration to put the items (and subsequently any cut scores associated with those items in the OIS) and student ability on the same scale. The items could then be ordered as described in Section 2.4.2, and the items, recommended cut scores, and student ability could all be linearly transformed together to the reporting scale. Item calibration required a data file that included student scores from the Spring 2022 administration for all operational and field test items. Each student who took the Spring 2022 assessments were also assigned an ability estimate (i.e., theta). Finally, a frequency distribution of student results (i.e., student ability) on the different test forms was created based on the Spring 2022 administration to be used to generate the impact data.

#### 2.4.1. Item Mapping and Calibration

During the standard setting, panelists made judgments about how items were associated with the performance levels. The judgments (i.e., the item-descriptor matches) were then mapped to the underlying estimates of student ability based on the item calibrations conducted prior to the meeting using the IRT three-parameter logistic (3PL) and two-parameter logistic (2PL) models for the dichotomously scored items and the generalized partial credit (GPC) model for the polytomously scored items.<sup>5</sup> The items were calibrated by estimating the item parameters from students' item scores from the Spring 2022 operational administration using IRTPro.

Once the item judgments were mapped during the standard setting, panelists' cut score recommendations could then be placed on the IRT scale and equated to the reported scale scores. Without this step, the cut score recommendations would not be on a scale that could be maintained going forward.

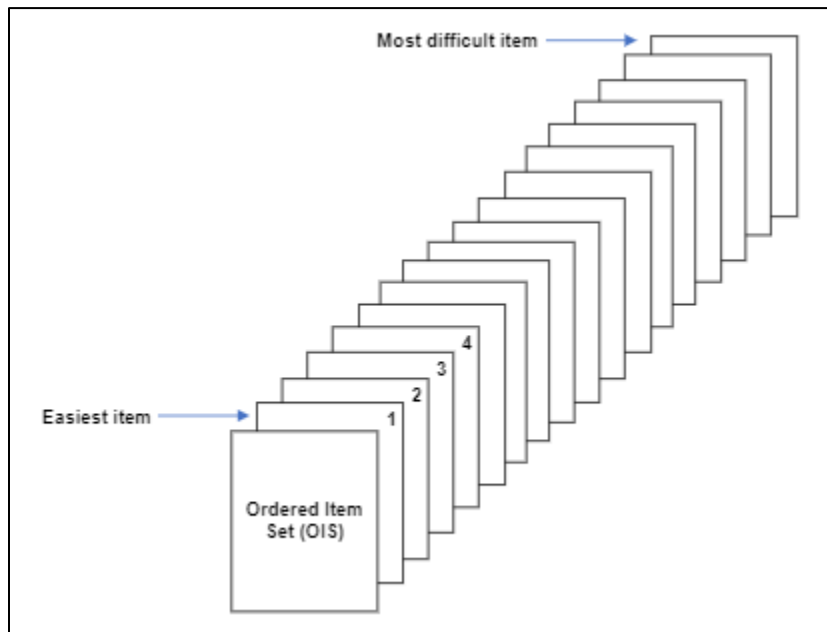
---

<sup>5</sup> For dichotomous items, the exact probability of a correct response was calculated as the number of correct responses (which is always 1 for dichotomous items) divided by the total number of possible responses. The item was calibrated under the 2PL model if this quotient was less than 0.05, otherwise the 3PL model was used.

### 2.4.2. Ordered Item Sets

In a traditional OIS in the ID Matching method, items are ordered by difficulty and appear in the OIS once for each non-0 score point. As such, dichotomously scored items appear in the OIS one time, while polytomously scored items appear in the item set multiple times. For example, a 2-point constructed-response item will appear twice for score points 1 and 2. The position of the items is based on the item difficulty for dichotomously scored items or the partial-credit threshold value for the item score point for polytomously scored items. Items are ordered by increasing item difficulty, as illustrated in Figure 2.2.

**Figure 2.2. Traditional OIS Example**



A modification to the ID Matching method occurred in Round 1 when the OIS presented the items in administration order so panelists could view the items in the same way that students were administered the items, with item clusters displayed together. The OIS for Rounds 2 and 3 consisted of the same items from Round 1 but ordered by difficulty, from easiest to hardest. The final OIS for each grade was reviewed by CDE prior to use during the standard setting meeting, and panelists accessed them through the website.

Items were ordered by difficulty in Rounds 2 and 3 based on the response probability of 0.50 (RP50), or the ability measure of the student (i.e.,  $\theta$ ) that is necessary to have a 50% chance of responding correctly or earning a specific score point under the IRT model. For dichotomous items calibrated under the 2PL model, the RP50 is expressed as follows:

$$P_{ij}(\theta) = \frac{1}{1 + \exp[-Da_j(\theta_i - b_j)]} = 0.5$$

where  $D$  is set equal to 1 when defined on the logistic scale,  $a_j$  is the discrimination parameter for item  $j$ ,  $b_j$  is the difficulty parameter for item  $j$ , and  $\theta$  is the student's ability. The above equation solved for  $\theta$  is the RP50. For dichotomous items calibrated under the 3PL model, the RP50 can be expressed as follows:

$$P_{ij}(\theta) = c_j + \frac{1 - c_j}{1 + \exp[-Da_j(\theta_i - b_j)]} = 0.5$$

where all parameters have the same interpretation as they did for the 2PL but a lower asymptote parameter,  $c_j$ , is added. Once again, the equation can be solved for  $\theta$ , resulting in the RP50. For polytomous items calibrated under the GPC, the equation is as follows:

$$P_{ij}(\theta) = \frac{\exp[\sum_{v=0}^x Da_j(\theta_i - b_j + d_{jv})]}{\sum_{k=0}^{M_j} \exp[\sum_{v=0}^k Da_j(\theta_i - b_j + d_{jv})]} = 0.5, x = 0, 1, \dots, M_j$$

where all parameters are the same as before,  $d_{jv}$  is the category parameter for category  $v$  of item  $j$ , and  $M_j$  is the maximum score on item  $j$ . As for the dichotomous items, this equation can be solved for  $\theta$  to get the RP50 for each score point. Once these were calculated for each score point for each item, the OIS could be created by ordering them from smallest to largest.

### 2.4.3. Feedback Data

Prior to the standard setting meeting, Pearson statistical analysts wrote analysis programs for summarizing the panelists' recommendations after each round using the judgment data downloaded from the Pearson standard setting website. These programs produced a report for each panelist of every judgment and cut recommendation provided after each round, in addition to summary statistics for the whole panel such as mean, range, and median that served as the panel recommendation. Table 2.2 presents the feedback data presented after each judgment round, as summarized below. Appendix E presents examples of how the feedback data were presented to the panelists during the standard setting meeting.

- *Individual Item Judgments.* Panelists' performance level designations for each item were presented in both administration and difficulty order after Round 1 and in difficulty order only for Rounds 2 and 3 to display the threshold regions. Presenting the items in difficulty order after Round 1 allowed the panelists to see the progression of the performance levels they assigned to each item as difficulty increases and the transition zones and to note if their judgments seemed at odds with student performance on the items.
- *Panelist Item Judgment Agreement.* The list of all items in the OIS and their associated performance level distributions (i.e., the percentage of panelists who assigned each performance level to an item). This feedback data also included a list of flagged items (i.e., the top 10 items with the most panelist disagreement).
- *Individual Cut Scores.* Based on their item-descriptor matches, panelists selected the item in the OIS that they believed best represented the change from the expectations of one performance level to the next.
- *Panelist Cut Score Agreement Graphs.* The cut score item selected by panelists for each performance level and the number of panelists who selected each cut were presented in bar charts. (Appendix G presents these results in table format.)

- *Committee Descriptive Information.* Panelists were shown a table that included the median cut score across all panelists for each performance level that was considered the committee-level cut score recommendation. Other descriptive information included the mean and minimum and maximum page numbers for each performance level cut. These summary statistics were provided to give the panelists an idea of how their judgments and recommended cuts compared with their peers’.
- *Impact Data.* The percentage of students who would be classified into each performance level based on the recommended cut scores and the performance of students on the Spring 2022 CMAS Science assessment, presented to panelists as stacked bar graphs. Impact data were calculated by estimating the abilities of students from the Spring 2022 administration as if they had been administered only the items in the OIS. Once the panelists made their recommendations about where the cuts between each performance level should be, the RP50 associated with the item at each cut was compared to the ability estimate of each student to categorize the students into performance levels.

**Table 2.2. Feedback Data Provided After Each Round**

Feedback Data	Round 1	Round 2	Round 3
Individual Item Judgments	✓	✓	✓
Panelist Item Judgment Agreement	✓	✓	✓
Individual Cut Scores		✓	✓
Panelist Cut Score Agreement Graphs		✓	✓
Committee Descriptive Information		✓	✓
Impact Data			✓

## 2.5. Meeting Security

Panelists reviewed test items, preliminary cut score recommendations, and associated impact data during the standard setting meeting. Due to the sensitive nature of this information, security was an essential component of the meeting. Procedures were established to ensure the security of the materials, starting with panelists signing a non-disclosure agreement (NDA) during recruitment. Throughout the meeting, the facilitators reviewed the necessity for maintaining security of the materials, discussions, and results from the meeting. The panelists were told which information they could and could not share or discuss outside of meeting rooms. All printed materials were collected at the end of each day, and the facilitators verified that all printed materials were in each panelist’s folder to ensure that all secure materials were retained.

To preserve the security of the materials and activities on the standard setting website, each panelist was provided unique login credentials for secure access to the website. The facilitators had control over the panelists’ access to each section of the website throughout the meeting. Website access was disabled at the end of each meeting day to prevent panelists from accessing secure website materials outside of the designated times. Following the meetings, Pearson archived the online materials on a secure site.



### 3. Standard Setting Process

This section describes each step that occurred during the standard setting meeting. Table 3.1 presents a high-level overview of the meeting schedule, and Appendix A presents the full meeting agenda. Once the standard setting panels provided their cut recommendations, they were presented to CDE for consideration before being brought to the State Board of Education for final approval.

**Table 3.1. Meeting Schedule Overview**

Day	Activities
Day 1	<ul style="list-style-type: none"><li>• General session<ul style="list-style-type: none"><li>○ Welcome and orientation</li><li>○ Assessment overview</li><li>○ Standard setting overview</li></ul></li><li>• Experience the assessment activity</li><li>• PLD review and discussion</li><li>• Training</li><li>• Practice activity</li><li>• Evaluation #1</li><li>• Round 1</li></ul>
Day 2	<ul style="list-style-type: none"><li>• Round 1 feedback and discussion</li><li>• Round 2</li><li>• Round 2 feedback and discussion</li><li>• Round 3</li><li>• Evaluation #2</li><li>• PLD feedback and recommendations</li><li>• Evaluation #3 and next steps</li></ul>

#### 3.1. Pre-Work

Panelists were asked to complete a set of tasks as pre-work prior to attending the standard setting meeting to maximize the efficiency of time during the meeting. As such, the panelists were registered in the Pearson standard setting website one week prior to the standard setting meeting. In an email from the website, panelists were provided with their unique user ID, a temporary password, and a link to the website. When panelists first logged in, they were required to select a unique, strong password consisting of at least eight characters, including at least one lowercase letter, one uppercase letter, one number, and one symbol.

Once panelists logged into the website, they only had limited access to certain materials designated for the pre-work that included the following activities that took approximately 15–25 minutes for the panelists to complete:

1. Standard setting overview video that provided a general overview of the standard setting meeting, including the purpose and process
2. Participant information survey to provide basic background information (see Appendix B for the results)
3. A review of the materials including the agenda, 2020 CAS, and PLDs

### **3.2. General Session**

During the opening general session, CDE welcomed the panelists, and the staff introduced themselves. Panelists were then presented an overview of the CMAS Science assessment and standard setting process, including a description of the modified ID Matching method, to ensure that all panelists began the process with a common understanding of the testing system and their role in setting cut scores. A description of the review process after the meetings was included to emphasize that the panel is making recommendations to CDE and the State Board of Education for use in determining the final cut scores. Appendix H presents the general session PowerPoint presentation.

### **3.3. Breakout Session**

After the general session, the panelists were divided by their grade-level panel to begin the breakout session that involved the work of recommending the cut scores. The facilitators, panelists, and observers first introduced themselves, followed by the distribution of materials by the facilitator and a review of the materials in the folder, the use of the website, and the use of all materials during the standard setting process.

#### *3.3.1. "Experience the Assessment" Activity*

The panelists were given an overview of the test design and item types that appeared on the Spring 2022 CMAS Science assessments. Panelists then reviewed the items through an online testing environment in TestNav that was similar to the one used during the actual test administration. The items were presented to panelists in the same administration order as presented to students. Panelists also had access to the item map that included the item keys and metadata. All field test items were removed from the test forms, and panelists experienced only the first two of the three test sessions.

While they experienced the assessment, panelists were encouraged to think from a student's perspective and take notes of the specific knowledge and skills a student would need to correctly respond to the item. Panelists could score their responses to the items to allow them to understand the scoring rules for the different item types included on the test, which also provided a good reference point for the judgment tasks that came later in the process. The system scored the selected-response and technology-enhanced items only, but rubrics were provided for the constructed-response items. The panelists were trained in any specific scoring rules used for the test. CDE staff were available to assist in the presentation and training on the scoring of items.

Panelists were given 45 minutes for this activity. This amount of time was less than the amount of time students are given to complete the test because it is expected that content experts should not need the same amount of time to complete the test as students, and the provided items only represented 66% of the test that students take. If panelists did not complete the abridged test in the time provided, they still had an opportunity to review items during the judgment tasks.

### 3.3.2. PLD Review

Because the PLDs were a critical resource throughout the meeting, time was provided for panelists to review the PLDs. The panelists were informed that the PLDs provided a snapshot of the likely range of student's academic abilities at each performance level, including the breadth and depth of the content, skills, and abilities demonstrated by students within each level. The PLDs were written with the three-dimensional nature of the standards represented in each different performance level. Panelists discussed the PLDs with their table groups, followed by a whole-group discussion led by the facilitator to review major points and observations. This review was designed to calibrate panelist expectations for each performance level so judgments would be based on similar expectations. Based on their experience with the test items and a review of the PLDs, panelists discussed the following questions:

- In what ways do the expectations increase from lower performance levels to higher performance levels?
- Which level represents the widest range of student performance? The most narrow range? Do they represent equal ranges of performance?
- How different is student proficiency at the very bottom of a higher performance level compared to a student at the top of the adjacent lower performance level (i.e., lowest-performing Level 4 and highest-performing Level 3)?

### 3.3.3. Training

The panelists were provided thorough training on the steps to be followed to make their recommendations using the modified ID Matching method, including an orientation to each judgment process component and how they should be used such as the standard setting website, item maps, judgment record forms, and judgment surveys. Significant time was spent discussing the steps in the judgment process. To begin the process, panelists were to select the performance level with the student expectations that best matched the item response requirements, followed by choosing the items that best separated the performance levels. The training included a carefully scripted presentation to define the judgment process, with a large portion focused on the judgment question, "*Which performance level most closely matches the knowledge and skills required to likely respond successfully to this item at the score point?*"

### 3.3.4. Practice Activity

At the end of the training session, panelists made practice judgments prior to beginning the actual judgment rounds to get a feel for the range of different item types and student responses they would encounter during the judgment task, get experience reviewing and making judgments for different item types, and build their confidence in their understanding of the task they are being asked to complete. The practice items were a subset of the items from the OIS that included a range of different items, item difficulty, and scoring types. For this activity, the panelists used the practice judgment record form.

Following the practice judgments, the facilitator showed the item-level judgment results interactively through the website, including the percentage of panelists that selected each performance level for each item. The facilitator walked through the meeting materials for the first two items to make sure panelists knew where to locate key information for making their judgments. The panel also had an opportunity to discuss each practice item and to hear perspectives on why panelists selected different performance levels.

### 3.3.5. Judgment Rounds 1–3

#### 3.3.5.1. Readiness Quiz

Before making judgments for each round, panelists indicated their readiness to participate in the standard setting activity and confirm their understanding of the judgment task by responding to the questions in Appendix D. Panelists were not able to start the judgment task until they answered “yes” to each question.

#### 3.3.5.2. Judgment Process

Once a panelist answered yes to each readiness question, they could begin the judgment round. Following the modified ID Matching method, the panelists completed the following steps starting in Round 1 and continuing to Rounds 2 and 3 after group discussions. During Round 1, panelists reviewed items in administration order so the items associated with the same phenomenon-based scenario were reviewed together. For Rounds 2 and 3, the items were reviewed in order of item difficulty, from easiest to most difficult. Furthermore, while the actual judgments were made on the website, panelists were encouraged to take notes on their hardcopy judgment record sheet to support later discussion and reflection of their judgments.

- Step 1: Determine the item response requirements using your professional judgment to define the knowledge and skills that students must demonstrate to provide a correct response to the item or to receive a specific score by considering the question: *“What does a student need to know and be able to do in order to answer this item at the score point?”*
- Step 2: Review the PLDs for each performance level.
- Step 3: Consider your response to the judgment question, *“Which performance level most closely matches the knowledge and skills required to likely respond successfully to this item at the score point?”*

Likely was defined as greater than 50%. Because matching the knowledge and skills to the PLDs is rarely a perfect match, panelists were encouraged to consider whether 50 out of 100 students with the knowledge and skills associated with the performance level would answer the item correctly, beginning with the lowest performance level. If students with the knowledge and skills associated with a performance level have a 50% or greater chance of getting an item correct, it could be considered a match with that level. For example, would 50% of Level 1 students answer this item correctly? If no, consider the Level 2 students. If yes, it could be considered a match with Level 1.

- Step 4: Select a performance level for the item (Level 1, Level 2, Level 3, or Level 4) that is best associated with the knowledge and skills for the item. Record your judgment on both your paper judgment record sheet and in the judgment survey on the website.
- Step 5: Complete Steps 1–4 for all items and check the judgment pattern across the performance levels, confirming that the second point for 2-point items was not given a lower performance level than the first point for the same item.

Step 6: (Rounds 2 and 3 only) Provide your cut score recommendations by determining the items from your regions of uncertainty in the OIS that best represent each performance level (i.e., the item that comes closest to the expectations of the performance level). Enter the item number as a response to the following question beginning with the first cut: *"Based on your review of the items in the threshold region between Levels 1 and 2, type the item sequence in the box for the item that you determine 'comes closest' to the expectations for Level 2. This should be a number between 1 and 60."*<sup>6</sup> Repeat this process for the other two cuts.

The region of uncertainty includes the items not perfectly ordered by performance level (i.e., items with different performance level assignments overlapped in the OIS) and is bounded by the first time the performance level switched to the last time, as shown by the shaded region in Figure 3.1. Panelists could see this region by reviewing their responses to the item-level judgments on the same page on the website by scrolling up or referring to their judgment record sheets. The region of uncertainty was explained to them but could not be highlighted in the same way as shown in the figure, so they needed to identify it on their own. Item 14 in this example is the cut score recommendation.

**Figure 3.1. Region of Uncertainty Example**

OIS Seq.	Performance Level
10	L2
11	L3
12	L2
13	L2
14	L3
15	L2
16	L3
17	L3
18	L3

Once panelists completed their judgments for each item, they submitted their judgments for analysis. After all panelists completed the judgment activity, Pearson conducted the analyses described in Section 2.4.3, applied quality control checks, and created the feedback data to provide to panelists. Before each feedback discussion, panelists were encouraged to listen to other panelists and consider the rationales given for their judgments but not to feel pressured to change their judgment to reach consensus.

### 3.3.6. PLD Recommendations

As a final step in the standard setting meeting, panelists had the opportunity to provide feedback and recommendations on the PLDs, which would then be considered by CDE when finalizing them. Panelists were provided a response box on the standard setting website to comment in an open-response format.

<sup>6</sup> This question caused confusion among some panelists who were unsure if the number should refer to the item above the cut or the item below the cut. This confusion may have impacted the results.

### *3.3.7. Evaluation Surveys*

Panelists completed three evaluation surveys during the meeting: Evaluation #1 was administered after the practice activity and focused on the effectiveness of the training activities and PLDs, Evaluation #2 was administered after Round 3 and focused on the cut score recommendations, and Evaluation #3 was administered following the PLD review and focused on the panelists' evaluation of the overall standard setting process and their confidence in their recommended cut scores. The evaluation surveys were delivered electronically through the standard setting website. Appendix F presents the results from all three surveys. Overall, the results indicate that the panelists understood the process and were confident about their recommendations.

The Grades 5 and 11 panelists were not reminded to complete the final evaluation survey, so only a portion completed it during the meeting. Follow-up emails were sent to the panelists asking them to complete the survey afterwards, and the results in Appendix F present the combined results from all panelists (i.e., those who completed the survey during the meeting and those who completed it afterward). All the Grade 8 panelists completed the survey during the meeting.

## 4. Results

### 4.1. Panelist Cut Score Agreement

Appendix G presents the items in the OIS selected by panelists as the cut score recommendations for each performance level by round (Rounds 2 and 3 only) and the number of panelists who selected each cut score item. These results show any committee-level regions of uncertainty and panelist cut recommendation changes from Round 2 to Round 3.

### 4.2. Round 3 Results

Table 4.1 presents the final recommended committee cut score recommendations that were the Round 3 median cut scores across all panelists for each performance level, along with the associated impact data. Figure 4.1 presents the impact data in a visual presentation.

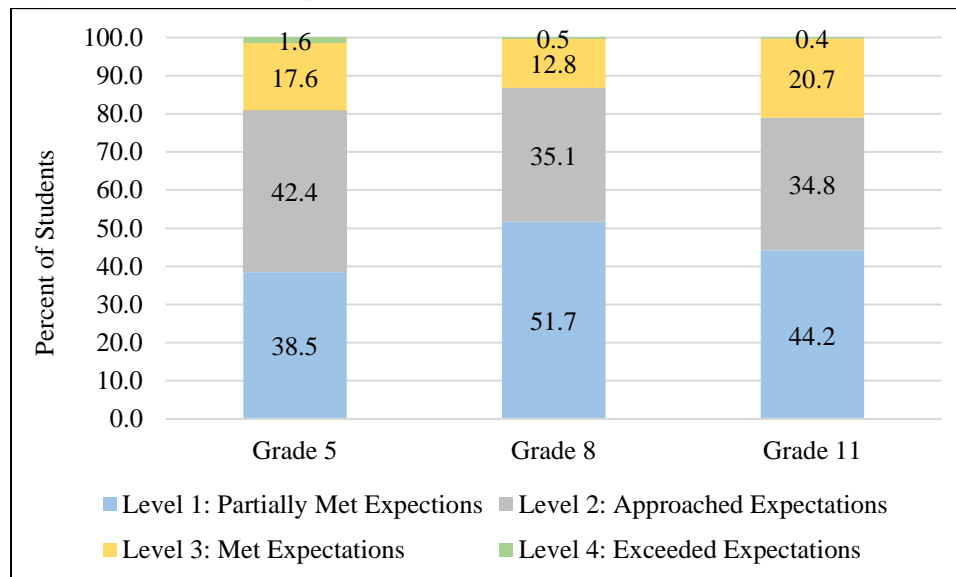
**Table 4.1. Round 3 Cut Score Recommendations and Impact Data**

Grade	Performance Level	Cut Score Item	#Students from 2022	%Students from 2022
5	Level 1: <i>Partially Met Expectations</i>	–	20,544	38.5
	Level 2: <i>Approached Expectations</i>	12	22,625	42.4
	Level 3: <i>Met Expectations</i>	31.5*	9,391	17.6
	Level 4: <i>Exceeded Expectations</i>	50	854	1.6
8	Level 1: <i>Partially Met Expectations</i>	–	25,239	51.7
	Level 2: <i>Approached Expectations</i>	12	17,135	35.1
	Level 3: <i>Met Expectations</i>	33	6,249	12.8
	Level 4: <i>Exceeded Expectations</i>	53	244	0.5
11	Level 1: <i>Partially Met Expectations</i>	–	13,047	44.2
	Level 2: <i>Approached Expectations</i>	7	10,272	34.8
	Level 3: <i>Met Expectations</i>	27	6,110	20.7
	Level 4: <i>Exceeded Expectations</i>	49	118	0.4

Note. The cut score item is the OIS page number.

\*The median for this cut fell between items.

**Figure 4.1. Round 3 Impact Data**



### 4.3. CDE Review

Before presenting cut score recommendations to the State Board of Education for approval, CDE reviewed the Round 3 cut score recommendations with an additional perspective of policy expectations. CDE considered the policy implications alongside content expectations to make sure that CDE was doing as much as possible to set standards-based expectations for Colorado students while also setting attainable benchmarks. The cut scores recommended by the committees resulted in a relatively small proportion of students meeting or exceeding expectations for each test. The outcome was not completely unexpected because Colorado schools encountered many implementation and instruction challenges during the pandemic and the 2020 CAS in science are much more complex and rigorous than Colorado’s previous science standards.

CDE reviewed the proportion of Spring 2022 students who were projected to meet or exceed expectations based on the standard setting panelists’ recommended cut scores and the proportion of students who were achieving benchmark on other science assessments, such as the previous CMAS Science assessment and the National Assessment of Educational Progress (NAEP). Based on this review and in consideration of policy implications, CDE made a few adjustments to the cut score recommendations that would increase the proportion of students meeting and exceeding expectations in Grades 5 and 8. Table 4.2 presents the resulting cut scores based on CDE’s review with their changes highlighted in the shaded cells, and Figure 4.2 presents the associated impact data.

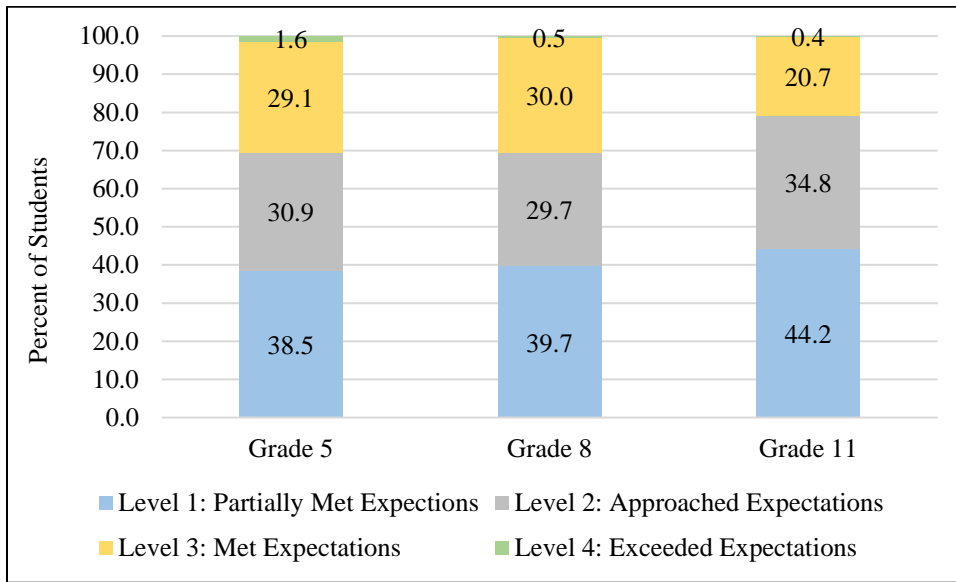
**Table 4.2. CDE Review Cut Score Recommendations and Impact Data**

Grade	Performance Level	Cut Score Item	#Students from 2022	%Students from 2022
5	Level 1: <i>Partially Met Expectations</i>	–	20,544	38.5
	Level 2: <i>Approached Expectations</i>	12	16,488	30.9
	Level 3: <i>Met Expectations</i>	29	15,528	29.1
	Level 4: <i>Exceeded Expectations</i>	50	854	1.6
8	Level 1: <i>Partially Met Expectations</i>	–	19,381	39.7
	Level 2: <i>Approached Expectations</i>	9	14,499	29.7
	Level 3: <i>Met Expectations</i>	26	14,645	30.0
	Level 4: <i>Exceeded Expectations</i>	53	244	0.5
11	Level 1: <i>Partially Met Expectations</i>	–	13,047	44.2
	Level 2: <i>Approached Expectations</i>	7	10,272	34.8
	Level 3: <i>Met Expectations</i>	27	6,110	20.7
	Level 4: <i>Exceeded Expectations</i>	49	118	0.4

*Note.* The cut score item is the OIS page number.



**Figure 4.2. CDE Review Impact Data**



#### **4.4. Final Approval**

The State Board of Education reviewed the CMAS Science assessment cut scores and impact data resulting from the panels' recommendations and from CDE's recommendations. The Board approved the cuts recommended by CDE on December 14, 2022.

#### **4.5. Reporting Scale**

Following the State Board of Education approval, the final cut scores were transformed to scale scores on the reporting scale. The process of determining the transformation rules from the IRT scale to the final reporting scale was guided by several principles:

1. The final cut scores determined while selecting the final scaling solution should respect the cut score recommendations from standard setting as closely as possible.
2. The scaling solution should involve a single linear transformation, from the IRT scale to the reporting scale.
3. The reporting scale score range should be the same across grades and tests.
4. The cut scores on the reporting scale for Level 2 should be the same across grades.
5. The cut scores on the reporting scale for Level 3 should be the same across grades.

The reporting scale also needed to have the following properties across all grades:

- The lowest obtainable scale score (LOSS) would be set at 650.
- The highest obtainable scale score (HOSS) would be set at 850.
- The Level 2 cut score would be set at 725.
- The Level 3 cut score would be set at 750.

The Level 4 scale score cut was found empirically by transforming the approved cut on the theta scale to the reporting scale using the slope and intercept that relate the Level 2 and Level 3 theta cuts to their fixed values on the reporting scale (725 and 750). Table 4.3 presents the ability (theta)-to-scale score conversions after transforming the performance level cut scores to the reporting scale, along with the theta cuts and Slope  $a$  and Intercept  $b$  scaling constants after the conversions of student ability (theta) to scale scores. Please note that while the cut scores were defined with the same scale score cuts for Level 2 and Level 3 across grades, they are not identical, and direct comparisons through averaging and aggregation across grades should not be made without study and/or statistical adjustments. The scale scores and distributions of students resulting from the cuts were not designed for direct comparison.

**Table 4.3. Final Recommended Cut Scores on the IRT Scale**

Grade	Level 2	Level 3	Level 4	Level 2	Level 3	Level 4	Slope $a$	Intercept $b$
5	-0.3244	0.5078	1.8184	725	750	789	30.0406	734.7465
8	-0.2275	0.6003	2.1619	725	750	797	30.2033	731.8699
11	-0.2354	0.7802	2.2962	725	750	787	24.6161	730.7900

*Note.* The first set of cuts are the ability (theta) scale cut scores, whereas the second set of cuts are the scale score cut scores.

## **5. Evidence of Procedural Validity**

This section details various evidence for the validity of the process used during the standard setting meeting, including committee representation, committee training, and panelists' perceived validity of the meeting and cut score recommendations.

### **5.1. Committee Representation**

Panelists completed a participant information survey that collected information about their background relevant to educational experience, including their current position and their number of years teaching a course related to their standard setting committee. As shown in Appendix B that presents the results of the self-reported demographic characteristics, most panelists were teachers in Grades K–12, had more than 10 years of experience in education, and had educational experience with students receiving mainstream special education services, English language learners (ELLs), and students receiving general education instruction. Most panelists were also currently working in school districts and were representative of the various types of districts across the state, including size, type, and socioeconomic status.

### **5.2. Committee Training**

It was essential that panelists understood how to make judgments as part of the modified ID Matching method. Prior to the standard setting meeting, panelists were invited to participate in pre-work where they received an overview of standard setting and an understanding of their role as a panelist. During the meeting, panelists received more training on the standard setting process and their role as a panelist during both the general and breakout sessions. Appendix H presents the Grade 5 breakout PowerPoint presentation used during this training as an example. They then were able to participate in a practice round to implement the standard setting methodology to a reduced set of items without consequence before beginning the actual judgments, including making judgments within the standard setting website.

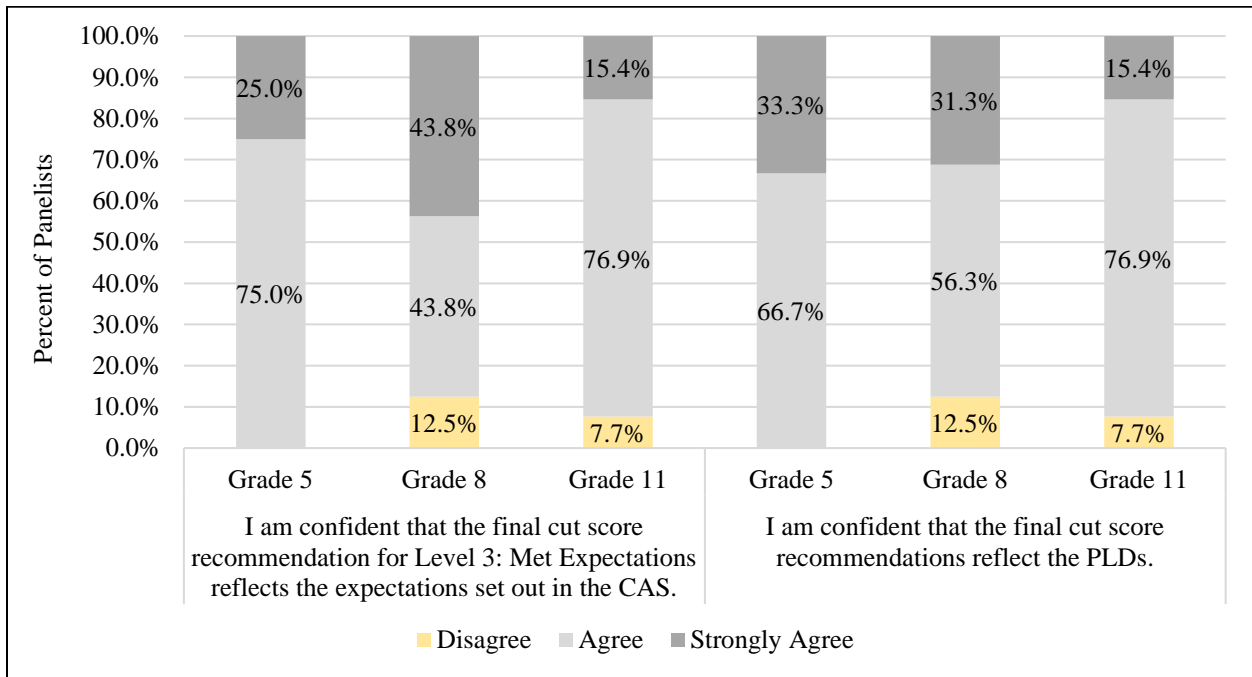
Before each judgement round, panelists responded to a readiness survey to ensure that they were prepared to make their judgments. Panelists were not able to continue to the judgment survey unless they answered yes to all questions on the readiness survey.

### **5.3. Perceived Validity of the Standard Setting Meeting**

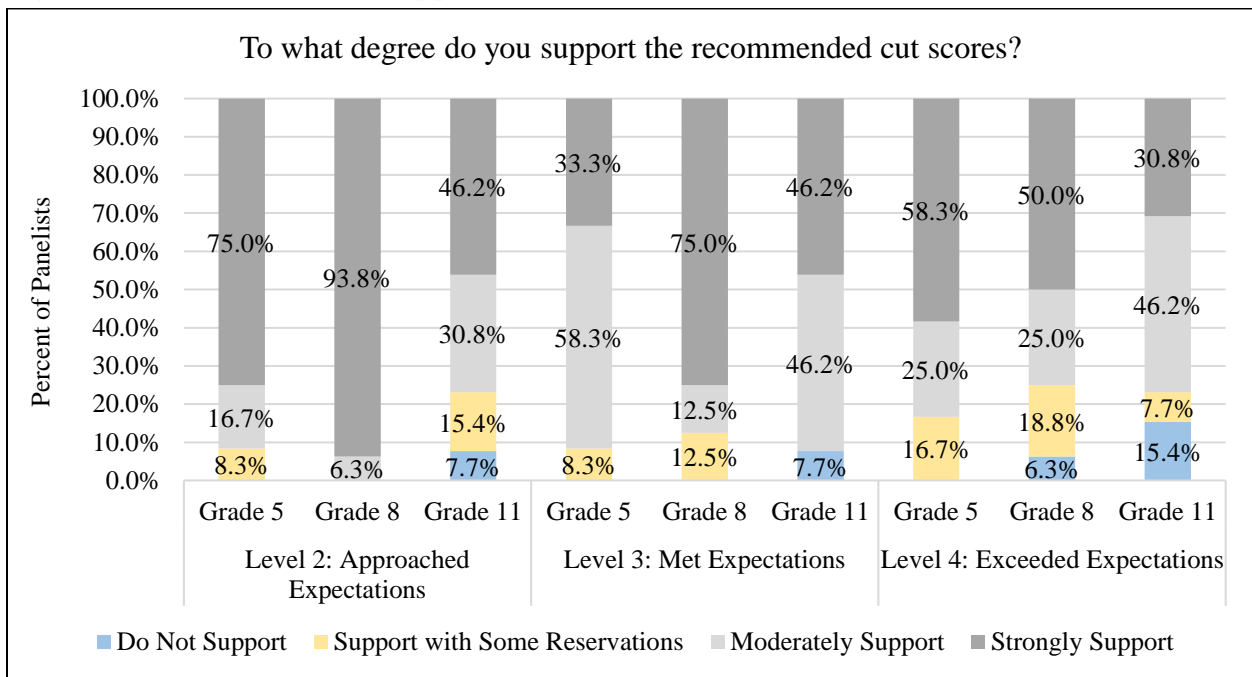
At the end of the practice round, panelists completed a process evaluation to record their opinion on the training provided. Appendix F presents the results of this survey. Most panelists indicated that the introduction to the standard setting process and the practice exercise were successful or very successful. Panelists also communicated their perceived validity of the recommended cut scores and the overall standard setting process in the second and third evaluation surveys administered after Round 3 and at the end of the meeting following the PLD recommendations, respectively. Appendix F presents the results of these surveys as well.

Most panelists agreed or strongly that the recommended cut scores reflected the content standards and PLDs and support the cut scores for each performance level, as shown in Figure 5.1 and Figure 5.2. Overall based on the process evaluation results, the panelists were satisfied with their recommendations and with the standard setting meeting, providing for the validity of the cut score recommendations.

**Figure 5.1. Panelists' Confidence in How Well the Cut Scores Reflect the CAS and PLDs**



**Figure 5.2. Panelists' Level of Support in the Recommended Cut Scores**



## 6. References

- American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (2014). *Standards for educational and psychological testing*. AERA.
- Colorado Department of Education (CDE). (2014, February). *SB08-212 goals and accomplishments. CAP4K: Colorado Achievement Plan for Kids*.  
<https://spl.cde.state.co.us/artemis/edserials/ed278internet/ed2782014internet.pdf>
- Ferrara, S., & Lewis, D. M. (2012). The item-descriptor (ID) matching method. In G. J. Cizek (Ed.), *Setting performance standards: Foundations, methods, and innovations* (2nd ed., pp. 255–282). Routledge.
- NGSS Lead States. (2013). *Next generation science standards: For states, by states*. The National Academic Press. <https://www.nextgenscience.org/search-standards>

**Appendix A: Agenda**

# Colorado Measures of Academic Success (CMAS) Science Test

## Standard Setting Meeting – Agenda

### Day 1 (Tuesday – September 27, 2022)

- |          |   |
|----------|---|
| 8:00 AM  | General Session<br>CMAS Science Overview<br>Standard Setting Process Overview                         |
| 9:30 AM  | Break   |
| 9:45 AM  | Breakout Sessions<br>Experience the Assessment<br>Performance Level Descriptors Review and Discussion |
| 12:00 PM | Lunch   |
| 1:00 PM  | Breakout Sessions<br>Standard Setting Training<br>Standard Setting Practice Activity                  |
| 2:15 PM  | Break   |
| 2:30 PM  | Breakout Sessions<br>Round 1 Judgments  |
| 5:00 PM  | End of Day 1  |

Day 2 (Wednesday – September 28, 2022)

- 8:00 AM Breakout Sessions  
Round 1 Judgment Feedback and Discussion  
Round 2 Judgments
- 11:00 AM Break
- 11:30 AM Breakout Sessions  
Round 2 Judgment Feedback and Discussion
- 12:00 PM Lunch
- 1:00 PM Breakout Sessions  
Round 2 Judgment Feedback and Discussion (continued)  
Round 3 Judgments
- 2:30 PM Break
- 3:00 PM Breakout Sessions  
Round 3 Judgment Feedback and Discussion  
Performance Level Descriptors Recommendations
- 4:00 PM End of Day 2

## Appendix B: Participant Information Survey Results

### 1. What is your current position?

Answer Option	Grade 5	Grade 8	Grade 11
Teacher (K–12 Education)	12	12	11
Teacher (Higher Education)	–	–	–
Administrator (School)	3	–	–
Administrator (District)	–	–	1
Other Position	1	3	2
<b>Total</b>	<b>16</b>	<b>15</b>	<b>14</b>

### 2. How many years of professional experience in education do you have?

Answer Option	Grade 5	Grade 8	Grade 11
None	–	–	–
1 to 5 years	1	–	–
6 to 10 years	2	2	3
11 to 15 years	2	3	5
16 to 20 years	3	6	1
More than 20 years	8	4	5
<b>Total</b>	<b>16</b>	<b>15</b>	<b>14</b>

### 3. How many years of professional experience do you have teaching Science Grade [5, 8, 11]?

Answer Option	Grade 5	Grade 8	Grade 11
None	2	–	3
1 to 5 years	5	6	2
6 to 10 years	5	4	4
11 to 15 years	2	3	3
16 to 20 years	1	1	–
More than 20 years	1	1	2
<b>Total</b>	<b>16</b>	<b>15</b>	<b>14</b>

### 4. For which of the following populations do you have educational experience with? (Check all that apply.)

Answer Option	Grade 5	Grade 8	Grade 11
Students receiving mainstream special education services	15	14	11
Students receiving self-contained special education services	5	8	8
Students who are English language learners	16	13	12
Students who are receiving general education instruction	16	13	13
Students who are receiving vocational technical instruction	1	3	3



**5. What is the highest degree you have completed?**

Answer Option	Grade 5	Grade 8	Grade 11
High School Diploma	–	–	–
Associates degree (A.A., A.S.)	–	–	–
Bachelors degree (B.A., B.S.)	3	5	1
Masters degree (M.A., M.S.)	11	10	11
Doctoral degree (Ph.D., Eh.D.)	2	–	2
Total	16	15	14

**6. What is your gender?**

Answer Option	Grade 5	Grade 8	Grade 11
Male	–	1	3
Female	15	13	10
No answer	–	–	–
Total	15	14	13

**7. What is your ethnicity?**

Answer Option	Grade 5	Grade 8	Grade 11
Hispanic or Latino	2	1	–
Not Hispanic or Latino	14	14	12
No answer	–	–	–
Total	16	15	12

**8. What is your race?**

Answer Option	Grade 5	Grade 8	Grade 11
American Indian or Alaskan Native	–	–	–
Asian	–	1	–
Black or African American	1	–	1
Native Hawaiian or Pacific Islander	–	–	–
White	14	13	11
No answer	–	–	–
Total	15	14	12

**9. Do you currently work in a school district?**

Answer Option	Grade 5	Grade 8	Grade 11
Yes	15	15	14
No	1	–	–
Total	16	15	14

**10. Which word best describes the size of the school district where you work?**

Answer Option	Grade 5	Grade 8	Grade 11
Small	4	2	5
Medium	5	2	3
Large	6	11	6
Total	15	15	14

**11. Which word best describes the type of school district where you work?**

Answer Option	Grade 5	Grade 8	Grade 11
Rural	3	2	7
Metropolitan/Urban	5	4	3
Suburban	7	9	4
Total	15	15	14

**12. Which word best describes the socioeconomic status of the school district where you work?**

Answer Option	Grade 5	Grade 8	Grade 11
Low	3	6	6
Moderate	9	6	6
High	3	3	2
Total	15	15	14

## Appendix C: Judgment Record Form Examples

## CMAS Science Grade 5 Standard Setting

## Round 1 Judgment Record Form

Which performance level most closely matches the knowledge and skills required to likely respond successfully to the item at this score point?

SEQUENCE	UIN	JUDGMENT			
		Level 1	Level 2	Level 3	Level 4
1	XXXXXXXX	Level 1	Level 2	Level 3	Level 4
2	XXXXXXXX - 1 point	Level 1	Level 2	Level 3	Level 4
2	XXXXXXXX - 2 points	Level 1	Level 2	Level 3	Level 4
3	XXXXXXXX	Level 1	Level 2	Level 3	Level 4
4	XXXXXXXX - 1 point	Level 1	Level 2	Level 3	Level 4
4	XXXXXXXX - 2 points	Level 1	Level 2	Level 3	Level 4
5	XXXXXXXX-XXXXXXXX	Level 1	Level 2	Level 3	Level 4
6	XXXXXXXX-XXXXXXXX	Level 1	Level 2	Level 3	Level 4
7	XXXXXXXX-XXXXXXXX	Level 1	Level 2	Level 3	Level 4
8	XXXXXXXX-XXXXXXXX	Level 1	Level 2	Level 3	Level 4
9	XXXXXXXX-XXXXXXXX - 1 point	Level 1	Level 2	Level 3	Level 4
9	XXXXXXXX-XXXXXXXX - 2 points	Level 1	Level 2	Level 3	Level 4
10	XXXXXXXX-XXXXXXXX	Level 1	Level 2	Level 3	Level 4
11	XXXXXXXX-XXXXXXXX	Level 1	Level 2	Level 3	Level 4
12	XXXXXXXX-XXXXXXXX - 1 point	Level 1	Level 2	Level 3	Level 4

<b>12</b>	XXXXXXXX-XXXXXXXX - 2 points	Level 1	Level 2	Level 3	Level 4
<b>13</b>	XXXXXXXX-XXXXXXXX - 1 point	Level 1	Level 2	Level 3	Level 4
<b>13</b>	XXXXXXXX-XXXXXXXX - 2 points	Level 1	Level 2	Level 3	Level 4
<b>14</b>	XXXXXXXX-XXXXXXXX - 1 point	Level 1	Level 2	Level 3	Level 4
<b>14</b>	XXXXXXXX-XXXXXXXX - 2 points	Level 1	Level 2	Level 3	Level 4
<b>15</b>	XXXXXXXX	Level 1	Level 2	Level 3	Level 4
<b>16</b>	XXXXXXXX	Level 1	Level 2	Level 3	Level 4
<b>17</b>	XXXXXXXX	Level 1	Level 2	Level 3	Level 4
<b>18</b>	XXXXXXXX	Level 1	Level 2	Level 3	Level 4
<b>19</b>	XXXXXXXX-XXXXXXXX	Level 1	Level 2	Level 3	Level 4
<b>20</b>	XXXXXXXX-XXXXXXXX - 1 point	Level 1	Level 2	Level 3	Level 4
<b>20</b>	XXXXXXXX-XXXXXXXX - 2 points	Level 1	Level 2	Level 3	Level 4
<b>21</b>	XXXXXXXX-XXXXXXXX - 1 point	Level 1	Level 2	Level 3	Level 4
<b>21</b>	XXXXXXXX-XXXXXXXX - 2 points	Level 1	Level 2	Level 3	Level 4
<b>22</b>	XXXXXXXX-XXXXXXXX	Level 1	Level 2	Level 3	Level 4
<b>23</b>	XXXXXXXX-XXXXXXXX - 1 point	Level 1	Level 2	Level 3	Level 4
<b>23</b>	XXXXXXXX-XXXXXXXX - 2 points	Level 1	Level 2	Level 3	Level 4
<b>24</b>	XXXXXXXX-XXXXXXXX	Level 1	Level 2	Level 3	Level 4
<b>25</b>	XXXXXXXX-XXXXXXXX	Level 1	Level 2	Level 3	Level 4
<b>26</b>	XXXXXXXX-XXXXXXXX	Level 1	Level 2	Level 3	Level 4
<b>27</b>	XXXXXXXX-XXXXXXXX	Level 1	Level 2	Level 3	Level 4
<b>28</b>	XXXXXXXX-XXXXXXXX	Level 1	Level 2	Level 3	Level 4
<b>29</b>	XXXXXXXX-XXXXXXXX	Level 1	Level 2	Level 3	Level 4
<b>30</b>	XXXXXXXX-XXXXXXXX	Level 1	Level 2	Level 3	Level 4

<b>31</b>	XXXXXXXX-XXXXXXXX	Level 1	Level 2	Level 3	Level 4
<b>32</b>	XXXXXXXX-XXXXXXXX	Level 1	Level 2	Level 3	Level 4
<b>33</b>	XXXXXXXX-XXXXXXXX	Level 1	Level 2	Level 3	Level 4
<b>34</b>	XXXXXXXX-XXXXXXXX	Level 1	Level 2	Level 3	Level 4
<b>35</b>	XXXXXXXX-XXXXXXXX	Level 1	Level 2	Level 3	Level 4
<b>36</b>	XXXXXXXX-XXXXXXXX	Level 1	Level 2	Level 3	Level 4
<b>37</b>	XXXXXXXX-XXXXXXXX	Level 1	Level 2	Level 3	Level 4
<b>38</b>	XXXXXXXX-XXXXXXXX	Level 1	Level 2	Level 3	Level 4
<b>39</b>	XXXXXXXX-XXXXXXXX - 1 point	Level 1	Level 2	Level 3	Level 4
<b>39</b>	XXXXXXXX-XXXXXXXX - 2 points	Level 1	Level 2	Level 3	Level 4
<b>40</b>	XXXXXXXX-XXXXXXXX	Level 1	Level 2	Level 3	Level 4
<b>41</b>	XXXXXXXX-XXXXXXXX	Level 1	Level 2	Level 3	Level 4
<b>42</b>	XXXXXXXX-XXXXXXXX - 1 point	Level 1	Level 2	Level 3	Level 4
<b>42</b>	XXXXXXXX-XXXXXXXX - 2 points	Level 1	Level 2	Level 3	Level 4
<b>43</b>	XXXXXXXX-XXXXXXXX - 1 point	Level 1	Level 2	Level 3	Level 4
<b>43</b>	XXXXXXXX-XXXXXXXX - 2 points	Level 1	Level 2	Level 3	Level 4
<b>44</b>	XXXXXXXX-XXXXXXXX	Level 1	Level 2	Level 3	Level 4
<b>45</b>	XXXXXXXX-XXXXXXXX - 1 point	Level 1	Level 2	Level 3	Level 4
<b>45</b>	XXXXXXXX-XXXXXXXX - 2 points	Level 1	Level 2	Level 3	Level 4
<b>46</b>	XXXXXXXX-XXXXXXXX - 1 point	Level 1	Level 2	Level 3	Level 4
<b>46</b>	XXXXXXXX-XXXXXXXX - 2 points	Level 1	Level 2	Level 3	Level 4

## CMAS Science Grade 5 Standard Setting

### Round 2 Judgment Record Form

Which performance level most closely matches the knowledge and skills required to likely respond successfully to the item at this score point?

SEQUENCE	UIN	JUDGMENT			
1	XXXXXXXX	Level 1	Level 2	Level 3	Level 4
2	XXXXXXXX-XXXXXXXX	Level 1	Level 2	Level 3	Level 4
3	XXXXXXXX-XXXXXXXX - 1 point	Level 1	Level 2	Level 3	Level 4
4	XXXXXXXX-XXXXXXXX	Level 1	Level 2	Level 3	Level 4
5	XXXXXXXX	Level 1	Level 2	Level 3	Level 4
6	XXXXXXXX	Level 1	Level 2	Level 3	Level 4
7	XXXXXXXX	Level 1	Level 2	Level 3	Level 4
8	XXXXXXXX-XXXXXXXX	Level 1	Level 2	Level 3	Level 4
9	XXXXXXXX-XXXXXXXX	Level 1	Level 2	Level 3	Level 4
10	XXXXXXXX-XXXXXXXX	Level 1	Level 2	Level 3	Level 4
11	XXXXXXXX-XXXXXXXX	Level 1	Level 2	Level 3	Level 4
12	XXXXXXXX-XXXXXXXX	Level 1	Level 2	Level 3	Level 4
13	XXXXXXXX-XXXXXXXX - 1 point	Level 1	Level 2	Level 3	Level 4
14	XXXXXXXX-XXXXXXXX	Level 1	Level 2	Level 3	Level 4
15	XXXXXXXX-XXXXXXXX - 1 point	Level 1	Level 2	Level 3	Level 4
16	XXXXXXXX-XXXXXXXX	Level 1	Level 2	Level 3	Level 4

<b>17</b>	XXXXXXXX-XXXXXXXX	Level 1	Level 2	Level 3	Level 4
<b>18</b>	XXXXXXXX-XXXXXXXX	Level 1	Level 2	Level 3	Level 4
<b>19</b>	XXXXXXXX-XXXXXXXX - 1 point	Level 1	Level 2	Level 3	Level 4
<b>20</b>	XXXXXXXX-XXXXXXXX	Level 1	Level 2	Level 3	Level 4
<b>21</b>	XXXXXXXX-XXXXXXXX	Level 1	Level 2	Level 3	Level 4
<b>22</b>	XXXXXXXX-XXXXXXXX	Level 1	Level 2	Level 3	Level 4
<b>23</b>	XXXXXXXX	Level 1	Level 2	Level 3	Level 4
<b>24</b>	XXXXXXXX-XXXXXXXX - 2 points	Level 1	Level 2	Level 3	Level 4
<b>25</b>	XXXXXXXX-XXXXXXXX - 1 point	Level 1	Level 2	Level 3	Level 4
<b>26</b>	XXXXXXXX-XXXXXXXX	Level 1	Level 2	Level 3	Level 4
<b>27</b>	XXXXXXXX - 1 point	Level 1	Level 2	Level 3	Level 4
<b>28</b>	XXXXXXXX-XXXXXXXX - 1 point	Level 1	Level 2	Level 3	Level 4
<b>29</b>	XXXXXXXX - 1 point	Level 1	Level 2	Level 3	Level 4
<b>30</b>	XXXXXXXX-XXXXXXXX - 1 point	Level 1	Level 2	Level 3	Level 4
<b>31</b>	XXXXXXXX-XXXXXXXX - 1 point	Level 1	Level 2	Level 3	Level 4
<b>32</b>	XXXXXXXX-XXXXXXXX - 2 points	Level 1	Level 2	Level 3	Level 4
<b>33</b>	XXXXXXXX - 2 points	Level 1	Level 2	Level 3	Level 4
<b>34</b>	XXXXXXXX-XXXXXXXX - 1 point	Level 1	Level 2	Level 3	Level 4
<b>35</b>	XXXXXXXX-XXXXXXXX	Level 1	Level 2	Level 3	Level 4
<b>36</b>	XXXXXXXX-XXXXXXXX	Level 1	Level 2	Level 3	Level 4
<b>37</b>	XXXXXXXX-XXXXXXXX	Level 1	Level 2	Level 3	Level 4
<b>38</b>	XXXXXXXX-XXXXXXXX - 1 point	Level 1	Level 2	Level 3	Level 4
<b>39</b>	XXXXXXXX-XXXXXXXX	Level 1	Level 2	Level 3	Level 4
<b>40</b>	XXXXXXXX-XXXXXXXX - 1 point	Level 1	Level 2	Level 3	Level 4

<b>41</b>	XXXXXXXX-XXXXXXXX - 2 points	Level 1	Level 2	Level 3	Level 4
<b>42</b>	XXXXXXXX-XXXXXXXX	Level 1	Level 2	Level 3	Level 4
<b>43</b>	XXXXXXXX-XXXXXXXX - 2 points	Level 1	Level 2	Level 3	Level 4
<b>44</b>	XXXXXXXX-XXXXXXXX	Level 1	Level 2	Level 3	Level 4
<b>45</b>	XXXXXXXX-XXXXXXXX - 2 points	Level 1	Level 2	Level 3	Level 4
<b>46</b>	XXXXXXXX-XXXXXXXX - 2 points	Level 1	Level 2	Level 3	Level 4
<b>47</b>	XXXXXXXX - 2 points	Level 1	Level 2	Level 3	Level 4
<b>48</b>	XXXXXXXX	Level 1	Level 2	Level 3	Level 4
<b>49</b>	XXXXXXXX-XXXXXXXX - 2 points	Level 1	Level 2	Level 3	Level 4
<b>50</b>	XXXXXXXX-XXXXXXXX	Level 1	Level 2	Level 3	Level 4
<b>51</b>	XXXXXXXX-XXXXXXXX	Level 1	Level 2	Level 3	Level 4
<b>52</b>	XXXXXXXX-XXXXXXXX	Level 1	Level 2	Level 3	Level 4
<b>53</b>	XXXXXXXX-XXXXXXXX - 1 point	Level 1	Level 2	Level 3	Level 4
<b>54</b>	XXXXXXXX-XXXXXXXX	Level 1	Level 2	Level 3	Level 4
<b>55</b>	XXXXXXXX-XXXXXXXX - 2 points	Level 1	Level 2	Level 3	Level 4
<b>56</b>	XXXXXXXX-XXXXXXXX - 2 points	Level 1	Level 2	Level 3	Level 4
<b>57</b>	XXXXXXXX-XXXXXXXX - 2 points	Level 1	Level 2	Level 3	Level 4
<b>58</b>	XXXXXXXX-XXXXXXXX	Level 1	Level 2	Level 3	Level 4
<b>59</b>	XXXXXXXX-XXXXXXXX - 2 points	Level 1	Level 2	Level 3	Level 4
<b>60</b>	XXXXXXXX-XXXXXXXX - 2 points	Level 1	Level 2	Level 3	Level 4



## Appendix D: Readiness Quiz

### Round 1:

1. I understand the steps I am to follow to make my judgments for the Round 1 Judgment activity.
  - a. Yes
  - b. No
2. I understand how to use the Performance Level Descriptors (PLDs) to guide my judgments.
  - a. Yes
  - b. No
3. I understand how to use the system to see each item's content and its correct answer/scoring rubric in the item map.
  - a. Yes
  - b. No
4. I understand that for multiple point items, the performance level for 2 points must be the same as or higher than the performance level for 1 point.
  - a. Yes
  - b. No
5. I understand how to use the system to record my judgments.
  - a. Yes
  - b. No
6. I am ready to complete my Round 1 judgments.
  - a. Yes
  - b. No

### Round 2:

1. I understand the steps I am to follow to make my judgments for the Round 2 Judgment activity.
  - a. Yes
  - b. No
2. I understand how to use the Performance Level Descriptors (PLDs) to guide my judgments.
  - a. Yes
  - b. No
3. I understand items are in the order of difficulty and how to use the system to see each item's content and its correct answer/scoring rubric in the item map.
  - a. Yes
  - b. No

4. I understand that for multiple point items, the performance level for 2 points must be the same as or higher than the performance level for 1 point.
  - a. Yes
  - b. No
5. I understand how to use the system to record my judgments.
  - a. Yes
  - b. No
6. I understand the committee-level feedback provided and how my judgments compare.
  - a. Yes
  - b. No
7. I am ready to complete my Round 2 judgments.
  - a. Yes
  - b. No

**Round 3:**

1. I understand the steps I am to follow to make my judgments for the Round 3 Judgment activity.
  - a. Yes
  - b. No
2. I understand how to use the Performance Level Descriptors (PLDs) to guide my judgments.
  - a. Yes
  - b. No
3. I understand items are in the order of difficulty and how to use the system to see each item's content and its correct answer/scoring rubric in the item map.
  - a. Yes
  - b. No
4. I understand that for multiple point items, the performance level for 2 points must be the same as or higher than the performance level for 1 point.
  - a. Yes
  - b. No
5. I understand how to use the system to record my judgments.
  - a. Yes
  - b. No
6. I understand the committee-level feedback provided and how my judgments compare.
  - a. Yes
  - b. No
7. I am ready to complete my Round 3 judgments.
  - a. Yes
  - b. No

## Appendix E: Feedback Data Examples

The following examples show how the feedback data were presented to panelists during the standard setting meeting.

### Individual Item Judgments (Rounds 1–3)

---

This provided each panelist with the item-level judgments that were recorded in the Pearson standard setting website. This was provided so that the panelist could check that the system recorded their judgments correctly.

#### Science Grade 5 – Individual Rating – Round 1

Table=1 Full\_name=

AdminSeq	FeedbackUIN	Judgment	Max_Score_Points
01		Level 2	1
02		Level 3	1
03		Level 1	1
04		Level 2	2
05		Level 4	2
06		Level 3	1
07		Level 2	1
08		Level 1	2
09		Level 1	2
10		Level 3	1

### Panelist Item Agreement (Rounds 1–3)

---

This provided the panelists with item-level judgment distributions for the committee for each item. The top 10 items with the greatest level of judgment disagreement were also identified for each performance level.

#### Science Grade 5 Round 1 Flagged Items

UIN	Max Points	Level 1	Level 2	Level 3	Level 4
XXXXXXXX-XXXXXXXX	1	31%	38%	19%	13%
XXXXXXXX-XXXXXXXX	1	38%	31%	25%	6%
XXXXXXXX-XXXXXXXX_2pts	2	6%	31%	38%	25%
XXXXXXXX-XXXXXXXX	1	31%	38%	31%	.
XXXXXXXX-XXXXXXXX_2pts	2	.	38%	31%	31%
XXXXXXXX-XXXXXXXX_1pt	2	31%	38%	31%	.
XXXXXXXX-XXXXXXXX	1	19%	38%	38%	6%
XXXXXXXX-XXXXXXXX	1	19%	38%	38%	6%
XXXXXXXXXXXXXXXXXX_1pt	2	19%	38%	38%	6%
XXXXXXXXXXXXXXXXXX_2pts	2	.	38%	38%	25%

**Science Grade 5 Round 1 (Administration Order)**

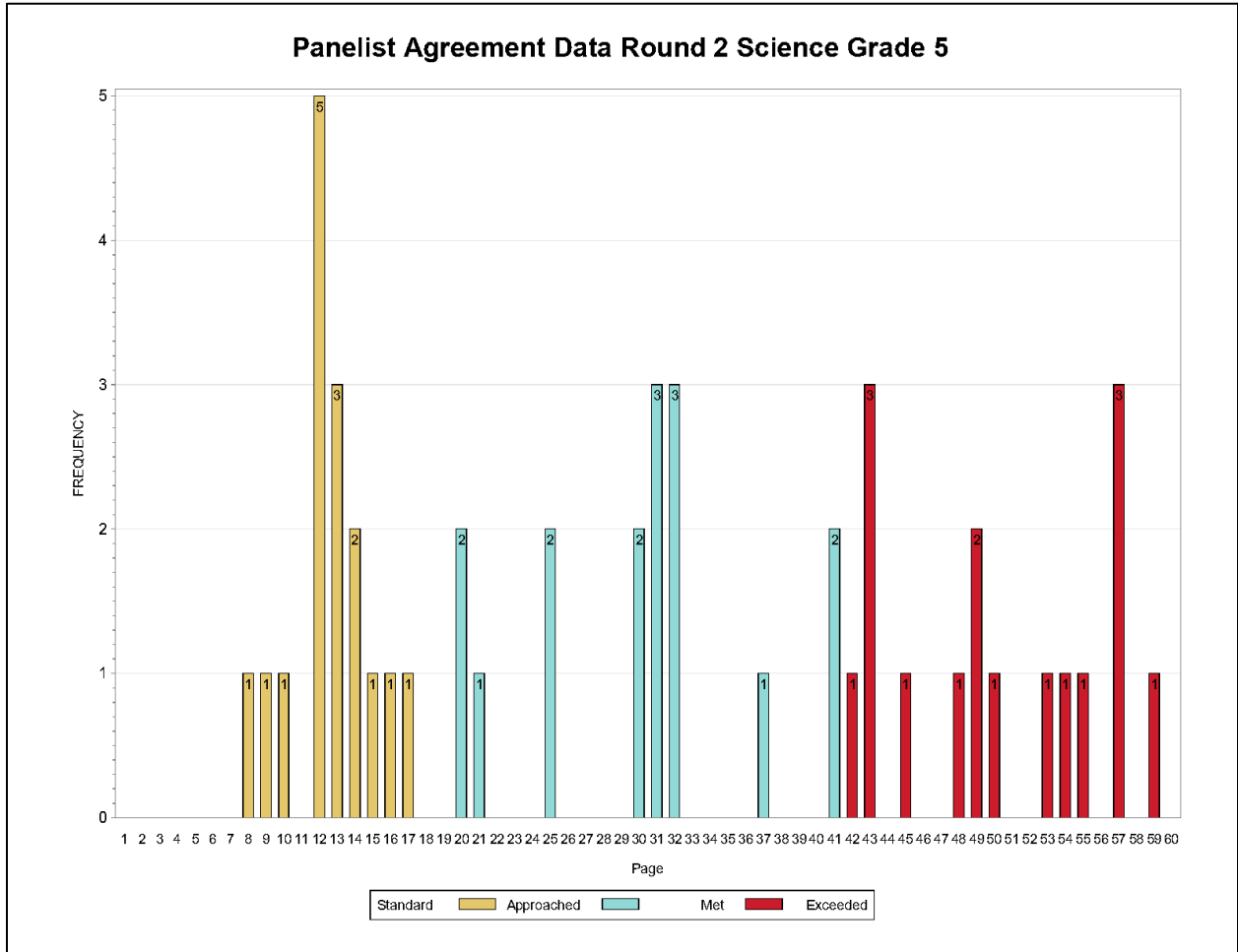
UIN	Max Points	Level 1	Level 2	Level 3	Level 4
XXXXXXXX	1	50%	50%	.	.
XXXXXXXX_1pt	2	100%	.	.	.
XXXXXXXX_2pts	2	6%	88%	6%	.
XXXXXXXX	1	75%	25%	.	.
XXXXXXXX_1pt	2	94%	.	6%	.
XXXXXXXX_2pts	2	.	88%	6%	6%
XXXXXXXX-XXXXXXXX	1	56%	44%	.	.
XXXXXXXX-XXXXXXXX	1	81%	19%	.	.
XXXXXXXX-XXXXXXXX	1	25%	63%	13%	.
XXXXXXXX-XXXXXXXX	1	.	13%	44%	44%
XXXXXXXX-XXXXXXXX_1pt	2	6%	44%	50%	.
XXXXXXXX-XXXXXXXX_2pts	2	.	13%	31%	56%
XXXXXXXX-XXXXXXXX	1	50%	31%	19%	.
XXXXXXXX-XXXXXXXX	1	6%	50%	38%	6%
XXXXXXXX-XXXXXXXX_1pt	2	25%	44%	31%	.
XXXXXXXX-XXXXXXXX_2pts	2	.	13%	56%	31%
XXXXXXXX-XXXXXXXX_1pt	2	19%	63%	19%	.
XXXXXXXX-XXXXXXXX_2pts	2	.	6%	69%	25%
XXXXXXXX-XXXXXXXX_1pt	2	6%	38%	50%	6%
XXXXXXXX-XXXXXXXX_2pts	2	.	6%	44%	50%
XXXXXXXX	1	81%	13%	6%	.
XXXXXXXX	1	63%	31%	6%	.
XXXXXXXX	1	19%	75%	6%	.
XXXXXXXX	1	31%	44%	19%	6%
XXXXXXXX	1	56%	25%	13%	6%
XXXXXXXX-XXXXXXXX_1pt	2	25%	31%	44%	.
XXXXXXXX-XXXXXXXX_2pts	2	6%	19%	25%	50%
XXXXXXXX-XXXXXXXX_1pt	2	25%	50%	25%	.
XXXXXXXX-XXXXXXXX_2pts	2	6%	31%	38%	25%
XXXXXXXX-XXXXXXXX	1	13%	50%	19%	19%
XXXXXXXX-XXXXXXXX_1pt	2	19%	38%	38%	6%
XXXXXXXX-XXXXXXXX_2pts	2	.	13%	38%	50%
XXXXXXXX-XXXXXXXX	1	38%	31%	25%	6%
XXXXXXXX-XXXXXXXX	1	13%	75%	6%	6%
XXXXXXXX-XXXXXXXX	1	25%	19%	44%	13%
XXXXXXXX-XXXXXXXX	1	31%	56%	13%	.

Appendix E: Feedback Data Examples

UIN	Max Points	Level 1	Level 2	Level 3	Level 4
XXXXXXXX-XXXXXXXX	1	6%	6%	75%	13%
XXXXXXXX-XXXXXXXX	1	31%	38%	31%	.
XXXXXXXX-XXXXXXXX	1	88%	13%	.	.
XXXXXXXX-XXXXXXXX	1	38%	63%	.	.
XXXXXXXX-XXXXXXXX	1	6%	19%	75%	.
XXXXXXXX-XXXXXXXX	1	25%	63%	13%	.
XXXXXXXX-XXXXXXXX	1	19%	38%	38%	6%
XXXXXXXX-XXXXXXXX	1	19%	63%	13%	6%
XXXXXXXX-XXXXXXXX	1	31%	38%	19%	13%
XXXXXXXX-XXXXXXXX	1	19%	38%	38%	6%
XXXXXXXX-XXXXXXXX	1	63%	19%	19%	.
XXXXXXXX-XXXXXXXX_1pt	2	25%	56%	19%	.
XXXXXXXX-XXXXXXXX_2pts	2	.	25%	69%	6%
XXXXXXXX-XXXXXXXX	1	69%	19%	13%	.
XXXXXXXX-XXXXXXXX	1	56%	31%	13%	.
XXXXXXXX-XXXXXXXX_1pt	2	31%	38%	31%	.
XXXXXXXX-XXXXXXXX_2pts	2	.	38%	31%	31%
XXXXXXXX-XXXXXXXX_1pt	2	25%	31%	44%	.
XXXXXXXX-XXXXXXXX_2pts	2	.	19%	44%	38%
XXXXXXXX-XXXXXXXX	1	63%	19%	13%	6%
XXXXXXXX-XXXXXXXX_1pt	2	13%	69%	19%	.
XXXXXXXX-XXXXXXXX_2pts	2	.	6%	75%	19%
XXXXXXXX-XXXXXXXX_1pt	2	31%	50%	19%	.
XXXXXXXX-XXXXXXXX_2pts	2	.	38%	38%	25%

**Panelist Cut Score Agreement Graphs (Rounds 2 and 3)**

This feedback was presented to panelists as bar graphs displaying the distribution of panelist recommendations for the cut score for each performance level.



**Committee Descriptive Information (Rounds 2 and 3)**

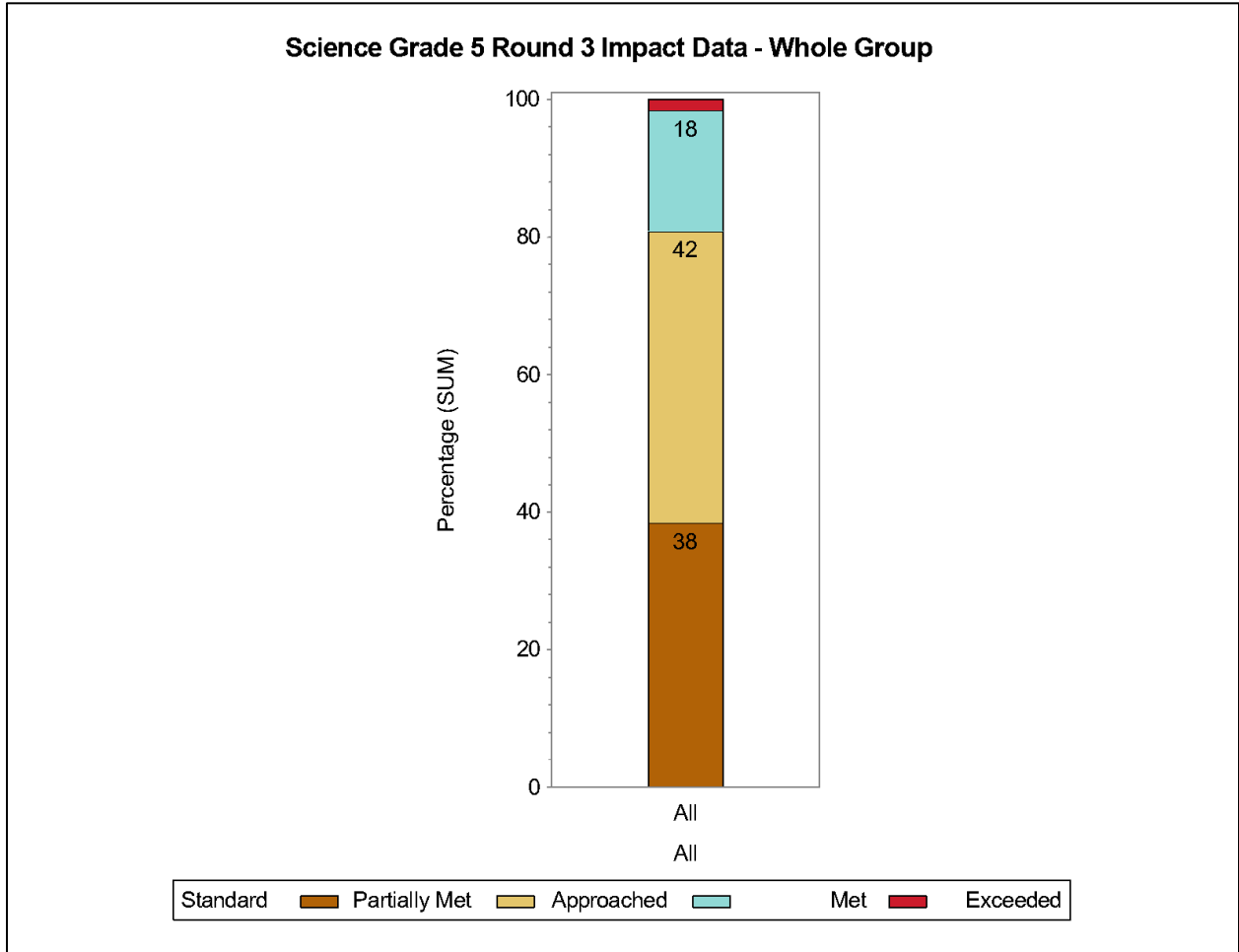
These summary statistics gave panelists an idea of how their judgments and recommended cuts compared with their peers’.

**Distributions of Cut Pages for Round 2 Science Grade 5 - Whole Group**

	Performance Level		
	Approached	Met	Exceeded
<b>Number of Individuals</b>	16	16	16
<b>Median Page Number</b>	12.5	31.0	49.5
<b>Mean Page Number</b>	12.6	29.9	50.3
<b>Minimum Page Number</b>	8	20	42
<b>Maximum Page Number</b>	17	41	59

**Impact Data (Round 3)**

Impact data presented the percentage of students who would be classified into each performance level based on the recommended cut scores and the performance of students on the Spring 2022 CMAS Science assessment. This showed panelists how their cut recommendations might impact the operational performance level distributions.



## Appendix F: Evaluation Survey Results

### Process Evaluation #1

The purpose of this evaluation is to collect information about your experience in recommending cut scores associated with the achievement levels for the CMAS assessments. Your opinions provide an important part of our evaluation of this meeting.

**1. Select the option that best reflects your opinion about the level of success of the various components of the meeting in which you participated. The activities were designed to help you both understand the process and be supportive of the recommendations made by the committee.**

#### Grade 5

Answer Option	Not Successful	Partially Successful	Successful	Very Successful
Overview of the CMAS assessments	–	1	8	7
Introduction to the standard setting process	–	1	9	6
Experiencing the actual assessment	–	–	6	10
Discussion of the scoring of items on the assessment	1	1	8	6
Discussion of performance level descriptors (PLDs)	–	1	10	5
Overview of the standard setting procedure	–	2	9	5
Practice exercise for the standard setting procedure	–	4	4	8

#### Grade 8

Answer Option	Not Successful	Partially Successful	Successful	Very Successful
Overview of the CMAS assessments	–	–	8	7
Introduction to the standard setting process	1	–	10	4
Experiencing the actual assessment	–	2	4	9
Discussion of the scoring of items on the assessment	1	1	7	6
Discussion of performance level descriptors (PLDs)	1	1	9	4
Overview of the standard setting procedure	1	2	7	5
Practice exercise for the standard setting procedure	–	2	8	5

#### Grade 11

Answer Option	Not Successful	Partially Successful	Successful	Very Successful
Overview of the CMAS assessments	–	–	9	6
Introduction to the standard setting process	–	–	10	5
Experiencing the actual assessment	–	2	8	5
Discussion of the scoring of items on the assessment	–	–	11	4
Discussion of performance level descriptors (PLDs)	–	1	10	4
Overview of the standard setting procedure	–	2	9	4
Practice exercise for the standard setting procedure	–	–	11	4



**2. How useful do you feel the following activities or information were in assisting you to make your recommendations?**

**Grade 5**

Answer Option	Not Useful	Somewhat Useful	Useful	Very Useful
Performance Level Descriptors (PLDs)	1	–	4	11
Standard Setting Training	1	1	8	6

**Grade 8**

Answer Option	Not Useful	Somewhat Useful	Useful	Very Useful
Performance Level Descriptors (PLDs)	–	2	5	8
Standard Setting Training	–	5	3	7

**Grade 11**

Answer Option	Not Useful	Somewhat Useful	Useful	Very Useful
Performance Level Descriptors (PLDs)	–	–	10	5
Standard Setting Training	–	1	10	4

**3. How adequate were the following elements of the session?**

**Grade 5**

Answer Option	Not Adequate	Somewhat Adequate	Adequate	More Than Adequate
Training provided on the standard setting process	–	1	11	4
Amount of time spent training	–	1	9	6
Total amount of time to discuss the PLDs	–	4	9	3
Total amount of time to discuss the practice judgments	1	2	11	2

**Grade 8**

Answer Option	Not Adequate	Somewhat Adequate	Adequate	More Than Adequate
Training provided on the standard setting process	1	1	10	3
Amount of time spent training	1	2	7	5
Total amount of time to discuss the PLDs	1	3	8	3
Total amount of time to discuss the practice judgments	1	3	8	3

**Grade 11**

Answer Option	Not Adequate	Somewhat Adequate	Adequate	More Than Adequate
Training provided on the standard setting process	–	–	13	2
Amount of time spent training	–	1	12	2
Total amount of time to discuss the PLDs	1	1	11	2
Total amount of time to discuss the practice judgments	–	1	11	3

**Process Evaluation #2 (Cut Score Recommendations Evaluation)****1. Rate your confidence in how well the cut scores reflect the Colorado Academic Standards (CAS) and the Performance Level Descriptors (PLDs).**

## Grade 5

Answer Option	Strongly Disagree	Disagree	Agree	Strongly Agree
I am confident that the final cut score recommendation for Level 3: <i>Met Expectations</i> reflects the expectations set out in the Colorado Academic Standards	–	–	9	3
I am confident that the final cut score recommendations reflect the performance level descriptors	–	–	8	4

## Grade 8

Answer Option	Strongly Disagree	Disagree	Agree	Strongly Agree
I am confident that the final cut score recommendation for Level 3: <i>Met Expectations</i> reflects the expectations set out in the Colorado Academic Standards	–	2	7	7
I am confident that the final cut score recommendations reflect the performance level descriptors	–	2	9	5

## Grade 11

Answer Option	Strongly Disagree	Disagree	Agree	Strongly Agree
I am confident that the final cut score recommendation for Level 3: <i>Met Expectations</i> reflects the expectations set out in the Colorado Academic Standards	–	1	10	2
I am confident that the final cut score recommendations reflect the performance level descriptors	–	1	10	2

**2. Taking into consideration the Colorado Academic Standards (CAS) and the Performance Level Descriptors (PLDs), indicate to what degree you support the recommended cut score for each performance level.**

## Grade 5

Answer Option	Do Not Support	Support with Some Reservations	Moderately Support	Strongly Support
To what degree do you support the recommended cut score for Level 2: <i>Approached Expectations</i>	–	1	2	9
To what degree do you support the recommended cut score for Level 3: <i>Met Expectations</i>	–	1	7	4
To what degree do you support the recommended cut score for Level 4: <i>Exceeded Expectations</i>	–	2	3	7

## Grade 8

Answer Option	Do Not Support	Support with Some Reservations	Moderately Support	Strongly Support
To what degree do you support the recommended cut score for Level 2: <i>Approached Expectations</i>	–	–	1	15
To what degree do you support the recommended cut score for Level 3: <i>Met Expectations</i>	–	2	2	12
To what degree do you support the recommended cut score for Level 4: <i>Exceeded Expectations</i>	1	3	4	8

## Grade 11

Answer Option	Do Not Support	Support with Some Reservations	Moderately Support	Strongly Support
To what degree do you support the recommended cut score for Level 2: <i>Approached Expectations</i>	1	2	4	6
To what degree do you support the recommended cut score for Level 3: <i>Met Expectations</i>	1	–	6	6
To what degree do you support the recommended cut score for Level 4: <i>Exceeded Expectations</i>	2	1	6	4

**3. Taking into consideration the Colorado Academic Standards (CAS) and the Performance Level Descriptors (PLDs), indicate your evaluation of the appropriateness of the recommended cut score for each performance level.**

## Grade 5

Answer Option	Way Too Low	A Bit Low	Appropriate	A Bit High	Way Too High
The recommended cut score for Level 2: <i>Approached Expectations</i> is	–	–	12	–	–
The recommended cut score for Level 3: <i>Met Expectations</i> is	–	3	8	1	–
The recommended cut score for Level 4: <i>Met Expectations</i> is	–	2	9	1	–

## Grade 8

Answer Option	Way Too Low	A Bit Low	Appropriate	A Bit High	Way Too High
The recommended cut score for Level 2: <i>Approached Expectations</i> is	–	1	14	1	–
The recommended cut score for Level 3: <i>Met Expectations</i> is	–	–	12	4	–
The recommended cut score for Level 4: <i>Met Expectations</i> is	1	3	10	2	–

## Grade 11

Answer Option	Way Too				Way Too High
	Low	A Bit Low	Appropriate	A Bit High	
The recommended cut score for Level 2: <i>Approached Expectations</i> is	–	1	11	–	1
The recommended cut score for Level 3: <i>Met Expectations</i> is	–	–	11	1	1
The recommended cut score for Level 4: <i>Met Expectations</i> is	–	1	8	2	2

**Process Evaluation #3 (Final Process Evaluation Survey)**

The purpose of this evaluation is to collect information about your experience in recommending cut scores associated with the achievement levels (performance level descriptors) for science. Your opinions provide an important part of our evaluation of this meeting.

**1. Please indicate below the degree to which you agree with each of the following statements.**

## Grade 5

Answer Option	Strongly Disagree	Disagree	Agree	Strongly Agree
I understood the purpose of the standard setting educator meeting	–	–	2	10
The instructions and explanations provided by the facilitators were clear	–	2	8	2
I had a solid understanding of what the test was intended to measure	–	–	5	7
I understand how the PLDs relate to the CAS	–	–	–	12
The training on the standard setting method gave me the information I needed to complete my assignment	–	–	7	5
The presentation of the feedback provided was adequate	1	2	5	4
The number of judgement rounds was adequate	–	–	10	2
The facilitators led the group through the standard setting process without imposing their ideas about where cut scores should be	–	–	7	5
I based my Rounds 2 and 3 judgments on the PLDs	–	–	5	7
The opportunity to make more than one round of judgments helped me to be more confident about my final ratings	–	1	2	9
I felt engaged in the process	–	–	4	8
I was comfortable sharing my ideas with the other panelists during the discussions	–	–	5	7
Table and group discussions were open and honest	–	–	4	8
My opinions and judgments were treated respectfully by the facilitators	–	1	6	5
My opinions and judgments were treated respectfully by my fellow panelists	–	–	4	8
I would be comfortable defending this process to my peers	–	–	7	5
Based on the CAS, I am confident this standard setting process will produce appropriate cut scores	–	–	8	4
I would be comfortable defending the final recommended cut scores to my peers	–	–	6	6
I had the opportunity to ask questions about the cut scores and how they will be used	–	2	6	4
I had the opportunity to ask questions about the process of making cut score recommendations	–	–	7	5

## Grade 8

Answer Option	Strongly Disagree	Disagree	Agree	Strongly Agree
I understood the purpose of the standard setting educator meeting	–	–	1	14
The instructions and explanations provided by the facilitators were clear	3	3	4	5
I had a solid understanding of what the test was intended to measure	–	1	2	12
I understand how the PLDs relate to the CAS	–	–	3	12
The training on the standard setting method gave me the information I needed to complete my assignment	–	3	6	6
The presentation of the feedback provided was adequate	–	3	4	8
The number of judgement rounds was adequate	–	1	6	8

Appendix F: Evaluation Survey Results

Answer Option	Strongly Disagree	Disagree	Agree	Strongly Agree
The facilitators led the group through the standard setting process without imposing their ideas about where cut scores should be	–	3	2	10
I based my Rounds 2 and 3 judgments on the PLDs	–	–	4	11
The opportunity to make more than one round of judgments helped me to be more confident about my final ratings	–	1	3	11
I felt engaged in the process	–	1	2	12
I was comfortable sharing my ideas with the other panelists during the discussions	–	1	5	9
Table and group discussions were open and honest	–	1	4	10
My opinions and judgments were treated respectfully by the facilitators	–	–	3	12
My opinions and judgments were treated respectfully by my fellow panelists	3	2	3	7
I would be comfortable defending this process to my peers	–	2	3	10
Based on the CAS, I am confident this standard setting process will produce appropriate cut scores	–	–	5	10
I would be comfortable defending the final recommended cut scores to my peers	–	–	5	10
I had the opportunity to ask questions about the cut scores and how they will be used	–	–	6	9
I had the opportunity to ask questions about the process of making cut score recommendations	–	–	6	9

Grade 11

Answer Option	Strongly Disagree	Disagree	Agree	Strongly Agree
I understood the purpose of the standard setting educator meeting	–	–	6	6
The instructions and explanations provided by the facilitators were clear	–	1	9	2
I had a solid understanding of what the test was intended to measure	–	1	6	5
I understand how the PLDs relate to the CAS	–	–	7	5
The training on the standard setting method gave me the information I needed to complete my assignment	–	1	10	1
The presentation of the feedback provided was adequate	–	–	11	1
The number of judgement rounds was adequate	–	–	11	1
The facilitators led the group through the standard setting process without imposing their ideas about where cut scores should be	–	–	7	5
I based my Rounds 2 and 3 judgments on the PLDs	–	–	7	5
The opportunity to make more than one round of judgments helped me to be more confident about my final ratings	–	–	7	5
I felt engaged in the process	–	–	6	6
I was comfortable sharing my ideas with the other panelists during the discussions	–	–	6	6
Table and group discussions were open and honest	–	–	8	4
My opinions and judgments were treated respectfully by the facilitators	–	–	5	7
My opinions and judgments were treated respectfully by my fellow panelists	–	2	4	6
I would be comfortable defending this process to my peers	–	2	8	2
Based on the CAS, I am confident this standard setting process will produce appropriate cut scores	–	2	10	–
I would be comfortable defending the final recommended cut scores to my peers	–	1	11	–
I had the opportunity to ask questions about the cut scores and how they will be used	–	1	8	3
I had the opportunity to ask questions about the process of making cut score recommendations	–	–	9	3

**2. If you answered Disagree or Strongly Disagree to any of the above, please give us feedback on what could have been done differently to change this outcome.**

**3. When the recommended cut scores are taken to the State Board, would you be interested in representing your fellow panelists as part of the presentation team?**

Panel	Yes	No
Grade 5	7	5
Grade 8	11	4
Grade 11	7	5

**4. Please indicate how influential each of the factors were in completing your Rounds 2 and 3 judgments.**

#### Grade 5

Answer Option	Not Influential	Influential	Very Influential
Completing the Experience the Assessment activity	1	2	9
My understanding of the PLDs	–	–	12
My perception of the difficulty of the items	–	6	6
Committee statistics provided after Round 1	–	5	7
The discussion after Round 1	–	2	10
Committee statistics provided after Round 2	–	6	6
Discussion after Round 2	–	5	7
My experience with students in my classroom	–	8	4

#### Grade 8

Answer Option	Not Influential	Influential	Very Influential
Completing the Experience the Assessment activity	2	5	8
My understanding of the PLDs	–	3	12
My perception of the difficulty of the items	–	8	7
Committee statistics provided after Round 1	1	13	1
The discussion after Round 1	–	4	11
Committee statistics provided after Round 2	–	10	5
Discussion after Round 2	–	5	10
My experience with students in my classroom	2	7	6

#### Grade 11

Answer Option	Not Influential	Influential	Very Influential
Completing the Experience the Assessment activity	2	7	3
My understanding of the PLDs	–	2	10
My perception of the difficulty of the items	1	8	3
Committee statistics provided after Round 1	–	9	3
The discussion after Round 1	2	6	4
Committee statistics provided after Round 2	–	10	2
Discussion after Round 2	–	9	3
My experience with students in my classroom	1	6	5

**5. Select the option that best reflects your opinion about the level of usefulness of the various components of the meeting in which you participated. The activities were designed to help you both understand the process and be supportive of the recommendations made by the committee.**

## Grade 5

Answer Option	Not Useful	Somewhat Useful	Useful	Very Useful
Completing the Experience the Assessment Activity	1	–	3	8
Practicing the procedure with real items prior to beginning the actual rating task	–	–	2	10
Referencing the PLDs	–	–	–	12
Reviewing data after Round 1	–	–	6	6
Discussion after Round 1	–	1	1	10
Reviewing data after Round 2	–	–	4	8
Discussion after Round 2	–	1	2	9

## Grade 8

Answer Option	Not Useful	Somewhat Useful	Useful	Very Useful
Completing the Experience the Assessment Activity	1	–	4	10
Practicing the procedure with real items prior to beginning the actual rating task	–	2	3	10
Referencing the PLDs	–	–	1	14
Reviewing data after Round 1	–	1	3	11
Discussion after Round 1	–	–	3	12
Reviewing data after Round 2	–	–	3	12
Discussion after Round 2	–	1	1	13

## Grade 11

Answer Option	Not Useful	Somewhat Useful	Useful	Very Useful
Completing the Experience the Assessment Activity	–	2	7	3
Practicing the procedure with real items prior to beginning the actual rating task	–	–	7	5
Referencing the PLDs	–	–	3	9
Reviewing data after Round 1	–	1	7	4
Discussion after Round 1	–	1	4	7
Reviewing data after Round 2	–	1	8	3
Discussion after Round 2	–	–	6	6



**6. How adequate was the time dedicated to each of the following activities of Day 2?**

## Grade 5

Answer Option	Too Little Time	About Right	Too Much Time
Completing the Experience the Assessment activity	–	10	2
Reviewing the PLDs	2	9	1
Training on the judgment task before Round 1	2	8	2
Round 1 of the judgment task	4	6	2
Review of data and discussion after Round 1	3	6	3
Round 2 of the judgment task	3	9	–
Review of data and discussion after Round 2	2	10	–
Discussion after Round 2	3	9	–
Round 3 of the judgment task	–	12	–
Presentation and discussion of Round 3 Final recommendations	1	11	–
Final review of PLDs	–	12	–

## Grade 8

Answer Option	Too Little Time	About Right	Too Much Time
Completing the Experience the Assessment activity	–	12	3
Reviewing the PLDs	2	10	3
Training on the judgment task before Round 1	1	6	8
Round 1 of the judgment task	5	9	1
Review of data and discussion after Round 1	3	9	3
Round 2 of the judgment task	5	8	2
Review of data and discussion after Round 2	2	7	6
Discussion after Round 2	5	5	5
Round 3 of the judgment task	2	12	1
Presentation and discussion of Round 3 Final recommendations	3	10	2
Final review of PLDs	2	13	–

## Grade 11

Answer Option	Too Little Time	About Right	Too Much Time
Completing the Experience the Assessment activity	–	11	1
Reviewing the PLDs	3	8	1
Training on the judgment task before Round 1	–	8	4
Round 1 of the judgment task	–	12	–
Review of data and discussion after Round 1	5	7	–
Round 2 of the judgment task	–	12	–
Review of data and discussion after Round 2	2	10	–
Discussion after Round 2	3	8	1
Round 3 of the judgment task	–	9	3
Presentation and discussion of Round 3 Final recommendations	1	11	–
Final review of PLDs	2	10	–

### 7. How confident do you feel that the Performance Level Descriptors (PLDs) for Science are reasonable for each performance level?

#### Grade 5

Answer Option	Not Confident	Somewhat Confident	Confident	Very Confident
Level 1: <i>Did Not Yet Meet Expectations</i>	–	–	5	7
Level 2: <i>Approached Expectations</i>	–	–	4	8
Level 3: <i>Met Expectations</i>	–	–	5	7
Level 4: <i>Exceeded Expectations</i>	–	1	5	6

#### Grade 8

Answer Option	Not Confident	Somewhat Confident	Confident	Very Confident
Level 1: <i>Did Not Yet Meet Expectations</i>	–	–	7	8
Level 2: <i>Approached Expectations</i>	–	–	7	8
Level 3: <i>Met Expectations</i>	–	2	6	7
Level 4: <i>Exceeded Expectations</i>	–	5	6	4

#### Grade 11

Answer Option	Not Confident	Somewhat Confident	Confident	Very Confident
Level 1: <i>Did Not Yet Meet Expectations</i>	1	3	5	3
Level 2: <i>Approached Expectations</i>	–	3	9	–
Level 3: <i>Met Expectations</i>	1	3	6	2
Level 4: <i>Exceeded Expectations</i>	2	6	2	2

### 8. How adequate were the following elements of the two-day meeting?

#### Grade 5

Answer Option	Not Adequate	Somewhat Adequate	Adequate	More than Adequate
Facilities used for the general session	–	–	3	9
Facilities used for the breakout session	–	–	3	9
Computers used during the meetings	–	–	6	6
Moodle site for accessing materials and making judgements	–	–	4	8
Materials provided in the folder	–	–	3	9
Workspace in table groups during the meeting	–	–	8	4

Appendix F: Evaluation Survey Results

Grade 8

Answer Option	Not Adequate	Somewhat Adequate	Adequate	More than Adequate
Facilities used for the general session	1	2	2	10
Facilities used for the breakout session	–	3	5	7
Computers used during the meetings	–	–	5	10
Moodle site for accessing materials and making judgements	–	–	5	10
Materials provided in the folder	–	–	6	9
Workspace in table groups during the meeting	1	1	6	7

Grade 11

Answer Option	Not Adequate	Somewhat Adequate	Adequate	More than Adequate
Facilities used for the general session	–	1	6	5
Facilities used for the breakout session	–	4	5	3
Computers used during the meetings	–	1	7	4
Moodle site for accessing materials and making judgements	–	1	9	2
Materials provided in the folder	–	–	9	3
Workspace in table groups during the meeting	1	2	7	2

### Appendix G: Panelist Cut Score Agreement

This appendix presents the items in the OIS selected by panelists as their cut score recommendation for each performance level by round (Rounds 2 and 3 only) and the number of panelists who selected each cut score item. Please note that the tables only show the items that were selected as cuts and not the full range of items in the OIS available for each assessment.

**Table G.1. Panelist Cut Score Agreement—  
Grade 5, Round 2**

OIS Item	Level 2 Cut	Level 3 Cut	Level 4 Cut
8	1		
9	1		
10	1		
12	5		
13	3		
14	2		
15	1		
16	1		
17	1		
20		2	
21		1	
25		2	
30		2	
31		3	
32		3	
37		1	
41		2	
42			1
43			3
45			1
48			1
49			2
50			1
53			1
54			1
55			1
57			3
59			1

**Table G.2. Panelist Cut Score Agreement—  
Grade 5, Round 3**

OIS Item	Level 2 Cut	Level 3 Cut	Level 4 Cut
11	2		
12	7		
13	5		
14	1		
16	1		
21		2	
24		1	
28		1	
30		2	
31		2	
32		6	
37		1	
41		1	
43			4
46			2
49			1
50			2
52			1
54			3
55			1
57			2

**Table G.3. Panelist Cut Score Agreement—  
Grade 8, Round 2**

OIS Item	Level 2 Cut	Level 3 Cut	Level 4 Cut
7	1		
8	4		
10	1		
11	2		
12	4		
13	1		
14	1	1	
16	1	3	
20		1	
22		1	
23		1	
29		1	
31		1	
32		1	
33		3	
34		1	
37		1	
38			1
40			1
44			2
50			1
52			1
53			2
54			3
55			1
56			3

**Table G.4. Panelist Cut Score Agreement—  
Grade 8, Round 3**

OIS Item	Level 2 Cut	Level 3 Cut	Level 4 Cut
10	3		
11	3		
12	8		
14	1		
21		1	
23		1	
31		1	
32		4	
33		7	
34		1	
52			1
53			10
54			2
55			2

**Table G.5. Panelist Cut Score Agreement—  
Grade 11, Round 2**

OIS Item	Level 2 Cut	Level 3 Cut	Level 4 Cut
5	1		
6	3		
7	3		
9	1		
12	2		1
13	1		
14	1		
15	1		
16	1		
17		1	
20		1	
25	1		
26		1	
27		4	
28		3	
29		1	
30		2	
33		1	
34		1	
36			1
45			5
47			2
49			1
50			3
54			1
56			1

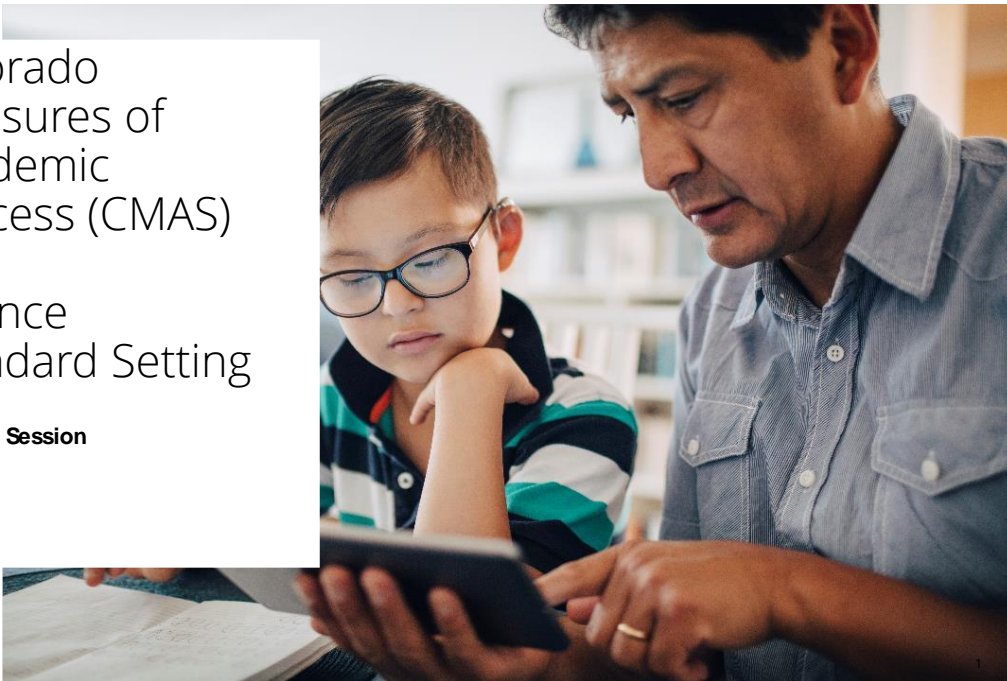
**Table G.6. Panelist Cut Score Agreement—  
Grade 11, Round 3**

OIS Item	Level 2 Cut	Level 3 Cut	Level 4 Cut
1	1		
4		1	
5	1		
6	2		
7	4		
8	1		
9	1		
10	3		
13	1		
15	1		
24		1	
25		1	
26		1	
27		5	
28		4	
30		1	1
34		1	
45			2
47			3
49			2
50			6
53			1


## Appendix H: PowerPoint Presentations

This appendix provides attachments to the general session PowerPoint presentation and the Grade 5 breakout session presentation as an example. Full copies of the presentations are accessible by double-clicking each slide in the Word document or by clicking on the attachments available in your PDF reader.

### General Session



Colorado Measures of Academic Success (CMAS)  
Science Standard Setting  
General Session



### Breakout Session—Grade 5



Colorado Measures of Academic Success (CMAS) Science Standard Setting  
Grade 5  
Day 1

