California Department of Education

Executive Office

SBE-002 (REV. 11/2017)

memo-imab-adad-aug22item01

# **MEMORANDUM**

**DATE:** August 19, 2022

**TO:** MEMBERS, State Board of Education

**FROM:** TONY THURMOND, State Superintendent of Public Instruction

**SUBJECT:** California Alternate Assessment for Science Operational Field Test: Standard Setting Plan.

## Summary of Key Issues

This Information Memorandum provides a summary of activities related to the August 2021 California Alternate Assessment (CAA) for Science standard setting workshop. During this virtual workshop, California special education and science educators collaborated to develop recommended threshold scores for the CAA for Science. The 2020–21 administration of the assessment was intended to be the first operational field test; however, the number of students who completed testing was insufficient to reflect a “typical” instructional year that would provide the necessary impact data to establish the recommended threshold scores. Therefore, the finalization of the threshold scores was postponed until the more robust impact data from the 2021–22 CAA for Science administration could be used.

In September 2022, the California Department of Education (CDE) will bring the proposed State Superintendent of Public Instruction’s threshold scores to the California State Board of Education (SBE) for approval. If the proposed threshold scores receive the SBE’s approval, the Student Score Reports (SSRs) for the 2021–22 CAA for Science operational field test administration will be available to local educational agencies (LEAs) in December 2022, and summary results will be released shortly thereafter.

## CAA for Science Standard Setting Process

Standard setting is an empirical process that uses a panel of experts to recommend threshold scores for reporting student achievement levels. Using a consensus-building process, panelists make judgments that define what a student “can do” at the entry to each achievement level. The threshold scores are then used to include achievement level information, with test results, to students and parents.

Testing contractor, ETS, conducted the CAA for Science standard setting workshop August 3–5, 2021, while CDE staff observed. The 28 educators who served as panelists formed three panels, one for each grade level (i.e., grade five, eight, and high school) in which the assessment is administered. The panelists included a diverse group of educators, representing expertise in the California Next Generation Science Standards (CA NGSS), the Science Connectors (i.e., the alternate science standards), and working with students with the most significant cognitive disabilities.

During the standard setting, the general and range achievement level descriptors (ALDs) were used to determine threshold scores. Typically, four types of ALDs (General ALDs, Range ALDs, Threshold ALDs, and Reporting ALDs) are created for an assessment during separate development phases. General ALDs give a broad description of the three CAA for Science achievement levels (i.e., Level 3—Understanding; Level 2—Foundational Understanding; and Level 1—Limited Understanding). Range ALDs provide a detailed description of what students at each achievement level know and can do in the subject area. For the CAA for Science standard setting, the educator panelists used the Range ALDs to develop the threshold ALDs, which describe what students can do at the beginning of Levels 2 and 3. Using those descriptions, the panelists made judgments about each item, and those judgments resulted in panel recommendations for threshold scores at Level 2 and Level 3.

**Approaches Used**

For the standard setting workshop, ETS adhered to the guidelines and best practices proposed in standard setting literature, using two widely accepted standard setting approaches—the modified Angoff method and the extended Angoff method. This process was conducted over two rounds of judgments and included feedback and discussion between the rounds. The panel-recommended threshold scores were based on the panel average. The final panel-recommended threshold scores for Level 2 and Level 3 were based on the bank of items used to assemble CAA for Science test forms.

After the second round of individual panel judgments, representatives from each of the three panels assembled for a vertical articulation meeting to discuss recommendations across the grade levels and grade band. The vertical articulation panel discussed the development of the threshold student definitions to identify commonalities or differences across grade levels and grade band. The purpose of this meeting was to review the work of the panels for all grades and grade spans; no discrepancies in the process were noted.

Following the August 2021 standard setting workshop, ETS reported the panelists’ recommendations to the CDE for review. After the impact data from the 2021–22 CAA for Science operational field test was received, ETS and CDE psychometricians completed a review of those recommendations. In addition, selected California Assessment of Student Performance and Progress Technical Advisory Group members have reviewed the recommendations. These reviews have informed the State Superintendent of Public Instruction’s recommended threshold scores, which are to be presented to the SBE in September 2022 for approval.

Attachment 1, “California Assessment of Student Performance and Progress California Alternate Assessment for Science Standard Setting Plan,” provides detailed information on the standard setting process used in the standard setting workshop.

## Key Activities in the Development of the CAA for Science

Table 1 provides a timeline of key activities in the development of the CAA for Science.

**Table 1. Timelines of Key Development Activities—CAA for Science**

| **Date(s)** | **Activity** |
| --- | --- |
| July 2016 | SBE approves high-level test design. |
| January 2018 | SBE approves blueprint and general ALDs. |
| September 8, 2020, through July 15, 2021 | CAA for Science operational field test (year one) is administered statewide.  |
| August 3 through August 5, 2021 | ETS conducts virtual CAA for Science standard setting workshop, with California special education and science educators as panelists. |
| September 7, 2021, through July 15, 2022 | CAA for Science operational field test (year two) is administered statewide.  |
| May 2022 | ETS confirms the number students tested meets the minimum amount needed for psychometrics analyses. |
| September 2022 | SBE considers for approval the CAA for Science threshold scores. |
| December 2022(tentative) | ETS provides CAA for Science SSRs to LEAs for distribution to parents (contingent on approval of threshold scores in September 2022). |

### **Next Steps for CAA for Science Threshold Scores**

In September 2022, the CDE will bring the State Superintendent of Public Instruction’s recommended CAA for Science threshold scores to the SBE for approval. If the SBE approves these recommended threshold scores, LEAs will receive electronic SSRs to send to parents/guardians and students. In addition to sending the SSRs, LEAs will be able to access student results in a downloadable student-level data file in the Test Operations Management System. Once available, the statewide-, county-, LEA-, and school-level summary results by student groups will be displayed on the Test Results for California's Assessments website at <https://caaspp-elpac.cde.ca.gov/caaspp/>.

## Attachment(s)

* Attachment 1: California Alternate Assessment for Science Standard Setting Plan (18 Pages)



**California Assessment of Student Performance and Progress**

**California Alternate Assessment for Science Standard Setting Plan**

**Prepared for the California Department of Education by ETS**

**Presented June 15, 2021**



**Contract No. CN150012**

**Contents**

[**California Alternate Assessment for Science Standard Setting Plan: Background 3**](#_Toc74054953)

[**Standard Setting Process 4**](#_Toc74054954)

[General Description of the Approach 4](#_Toc74054956)

[Educator Panelists 4](#_Toc74054957)

[Preworkshop Activities 5](#_Toc74054958)

[Workshop Overview 5](#_Toc74054959)

[Test Familiarization 6](#_Toc74054960)

[Threshold Student Definitions 6](#_Toc74054961)

[**Standard Setting Methodology 8**](#_Toc74054962)

[Item Scoring, Judgments, and Rating Scales 8](#_Toc74054963)

[Modified and Extended Angoff Judgments 8](#_Toc74054964)

[Feedback and Discussion 10](#_Toc74054965)

[**Standard Setting Materials 12**](#_Toc74054966)

[CAA for Science General and Range ALDs 12](#_Toc74054967)

[Judgment Materials 12](#_Toc74054968)

[Evaluation Forms 12](#_Toc74054969)

[**Recommendations and Technical Report 13**](#_Toc74054970)

[**Staffing, Logistics, and Security of Panel Meetings 14**](#_Toc74054971)

[Appendix A: General Achievement Level Descriptors **15**](#_Toc74054972)

[Appendix B: Draft Agenda **16**](#_Toc74054973)

[**References 18**](#_Toc74054977)

**California Alternate Assessment for Science Standard Setting Plan: Background**

The plan presented here describes the process that testing contractor ETS will use to conduct standard setting for the California Alternate Assessment (CAA) for Science. That process includes a range of activities, starting with the selection of standard setting methodology, proceeding with the recruitment and selection of a panel of California educators to serve as standard setting panelists, and ending with implementation and documentation of the standard setting workshop. At the conclusion of the standard setting workshop, California educator recommendations for threshold scores will be presented to the California Department of Education (CDE) for review. Subsequently, once student performance data on the CAA for Science is available, these recommendations will be presented to the California State Board of Education (SBE) at the SBE meeting in September 2022 along with the State Superintendent of Public Instruction’s (SSPI’s) recommended threshold scores.

The CAA for Science is designed for students with the most significant cognitive disabilities and measures what these students know and can do in science. It is aligned with the alternate science standards, also known as the Science Core Content Connectors (Science Connectors). The Science Connectors are derived from the California Next Generation Science Standards (CA NGSS). This assessment is administered to eligible students in grade levels five and eight and once in high school (i.e., grade ten, eleven, or twelve).

The general achievement level descriptors (ALDs) for the CAA for Science were approved by the SBE in January 2018 (refer to [appendix A](#_Appendix_A:_General))*.* General ALDs describe the expectations for student achievement at each reporting level across all grades tested. From March 2 through 4, 2021, 11 California educators convened for remote meetings to review and provide input on the draft descriptions of the range ALDs. The range ALDs are aligned with the SBE-approved general ALDs and are used to describe, in detail, what students at each achievement level know and can do in science. They are used to distinguish between the three CAA for Science achievement levels and aid in the interpretation of threshold scores for each grade level or grade band assessment. Range ALDs are important because they define the knowledge, skills, and abilities required at each achievement level; describe the Science Connectors as they apply to threshold scores; and give standardized meaning to scores or score ranges. That is, they provide the detail required in the standard setting process.

Utilizing the range ALDs, California educator panelists will develop and use threshold ALDs during the standard setting process. Panelists will describe what a student is expected to be able to know and do at the beginning of levels 2 and 3, thus distinguishing between the three achievement levels. These ALDs will be used to make standard setting judgments. Two threshold scores allow three achievement levels to be reported. Achievement levels for the CAA for Science are as follows:

Level 3—Understanding

Level 2—Foundational Understanding

Level 1—Limited Understanding

The reporting ALDs describe the meaning of the levels on score reports and are anticipated to be finalized in September 2022, following the SBE’s approval of the threshold scores.

**Standard Setting Process**

The standard setting process for the CAA for Science will result in recommendations for the CAA for Science threshold scores, which are the minimum scores at Level 2 and Level 3. The CDE will review these recommendations, along with additional data, and present them for approval to the SBE in September 2022.

**General Description of the Approach**

The approach used in this workshop adheres to the guidelines and best practices recommended in the standard setting literature. Specifically, the modified Angoff and extended Angoff standard setting methods (Cizek and Bunch 2007; Hambleton and Plake 1995) will be implemented.

The virtual workshop will be facilitated using Zoom conferencing. The panelists will make two rounds of judgments, with feedback and discussion occurring via Zoom between the rounds. Because of the low volume of test takers in the 2021 administration of the CAA for Science, the impact data often presented after Round 2 will not be available during the standard setting workshop. The workshop will include a vertical articulation discussion after Round 2. That discussion will include representatives from each grade level or grade band. This will allow for cross-panel understanding of how the process has worked in each room, from the development of the threshold ALDs to the discussion of the assessment to the sharing of the in-panel discussions of the recommendations across the grade levels and grade bands.

**Educator Panelists**

A group of special education and science educators, representative of the diversity of California students in grades five and eight and high school, will be recruited to participate as panelists in the standard setting workshop. Recruitment efforts will be made through various educator outreach opportunities—such as the California Association for Science Educators listserv and item writing workshops—and educator trainings as well as through other communication sources, such as the CAASPP Opportunities to Get Involved web page at <https://www.caaspp.org/getinvolved/getinvolved.html>. These recruitment efforts and resources also are highlighted in the Assessment Spotlight email updates (archived at <https://www.cde.ca.gov/ta/tg/sa/assessmentspotlight.asp>) and on social media (e.g., the CDE Twitter account for the California Statewide Assessment System at <https://twitter.com/CDEAssessments>).

In recruiting panelists, the goal will be to include a range of panelists who have familiarity with instructing students with the most significant cognitive disabilities, experience in administering the CAA for Science, and familiarity with the Science Connectors. An important element in the recruiting efforts is the inclusion of teachers who are working with these students, as they will provide perspectives that include experience teaching the standards and understanding the students’ knowledge and skills.

Each panelist will be assigned to one of three panels, each panel focusing on one grade level or grade band: grade five, grade eight, or high school. The targeted number of panelists from this population of educators is 10 per panel, for a total of 30 educators as shown in [table 1](#Table1).

Table 1. Educator Panel Recruitment Goals

|  |  |
| --- | --- |
| **Panel** | **Number of Participants** |
| Grade five | 10 |
| Grade eight | 10 |
| High school | 10 |
| Total | 30 |

The CDE will select the educators who will participate, and then will determine their panel assignments at the standard setting. The panelists will be notified after they are approved, and SCOE will assist with communicating to the selected educators about logistics and reimbursement for substitutes or honoraria. The panelists will be required to sign a security agreement notifying them of the confidentiality of the materials used in the standard setting and prohibiting the use or disclosure of the meeting materials after the workshop.

**Preworkshop Activities**

The process for the CAA for Science will include preworkshop training on the remote meeting tools (e.g., technology and devices needed) and a preworkshop assignment.

**Remote Meeting Preparation**

The training on remote meeting tools will take place two weeks prior to the workshop. Panelists will receive instructions on how to prepare for the remote meeting and will be invited to a practice session during which ETS staff will provide a walk-through of the features of the software that will be used, offer technical assistance, and answer any questions.

**Standard Setting Preparation**

Approximately one week prior to the standard setting, a preworkshop assignment consisting of two parts will be sent to the panelists. For this assignment, the panelists will be asked to take the CAA for Science practice test to become familiar with the item types, task types, and scoring rubrics. The panelists will also be asked to review the general ALDs and the grade-specific range ALDs aligned with their panel assignment. Panelists will be asked to consider the expected performance of a student at each of the achievement levels, take notes about the knowledge and skills of students *at the beginning* of Level 3, and bring those notes to the standard setting workshop.

**Workshop Overview**

During the first day of the workshop, all panelists will participate in a general session led by the standard setting director. The general session will include an overview of the CAA for Science and the standard setting process. At the conclusion of the general session, the panelists will break into separate grade-level or grade-band panel rooms via Zoom breakouts. In each of three panel rooms, an ETS facilitator, experienced in working with educators in standard setting, will present the test to familiarize the panelists with the performance tasks and provide in-depth training and practice on each step in the standard setting process. Each step in the process will begin with training and an evaluation of the training before proceeding.

The panelists will complete two rounds of judgments. Feedback and discussion will take place after each round. Each panel will complete the standard setting process on one grade-level or grade-band assessment. During the last day of the workshop, representatives from each panel will assemble for a vertical articulation meeting to discuss recommendations across the grade levels and grade band. Refer to the [Vertical Articulation Discussion](#_Vertical_Articulation_Discussion) section for more information.

**Test Familiarization**

Once in grade-level or grade-band panels, the panelists will become familiar with the CAA for Science items for that grade and discuss the test content before making any judgments.

The CAA for Science is administered online to each student individually by a test examiner. So panelists can review all items administered in 2020–21 that will be evaluated in the standard setting process, an ETS assessment expert will facilitate a testing session with the panel. Each item will be displayed on the screen, and text from the *Directions for Administration* (*DFAs*) will be read to the panelists. The panelists will follow the administration on their own screens and independently record the answers to the test questions, with no key provided, in the comments section of the Content Review Tool (CRT). After the panelists are administered all of the items, they will be provided with instructions on how to use the answer keys and asked to self-score their tests. The purpose of taking the test is to allow the panelists to familiarize themselves with the content and the difficulty of the items on the test.

Answer keys differ by task type. Specifically, some items include one or more selected responses, while others require an interaction with a technology-enhanced item (TEI) (e.g., when students must drag and drop an image into a blank to complete a diagram or check a cell in a grid). The panelists will be provided with rubrics when needed for TEIs as well as the scoring rules for items that have multiple parts, requiring different response types (e.g., when students must respond to selected-response and grid parts of a single item).

Once the test familiarization is complete, the panelists will discuss the demands of the items, what content is measured by the test, and what might be challenging for the students who take the CAA for Science. ETS assessment specialists will be available to respond to any questions the panelists may have related to the task types or scoring.

**Threshold Student Definitions**

Developing a common understanding of expectations of students at the threshold of each achievement level is a critical component of any standard setting workshop. The threshold student is that student who has just sufficient knowledge and skills to be included in a higher achievement level, compared with the highest performing student in the next lower achievement level. The process to arrive at threshold student definitions will involve small-group discussions and the development of a draft of the threshold student definitions, followed by a whole-panel discussion of the draft definitions, to reach a panel consensus of what is expected.

For the CAA for Science, two definitions will be needed for two thresholds—the Level 2 and Level 3 threshold student definitions. Panels will work first on the Level 3 threshold because this is the point at which students will be classified as having *understanding,* demonstrating that they can apply their knowledge and skills to problems aligned with the Science Connectors.

The panelists will refer to the range ALDs that describe the full range for three levels. The Level 3 threshold student will be defined by considering what is expected of students in Level 2 (compared with expectations in Level 3) and describing what the student just entering Level 3—the Level 3 threshold student—can do beyond what the highest-performing student in Level 2 can do. ETS facilitators will instruct panelists on the process and will provide a model, in the form of a few bullets, to exemplify the level of detail expected in a threshold student definition. The panelists will be asked to review and discuss the exemplar and modify it as needed. Instructions to the panelists will include a goal to limit the definitions of their threshold students to a sufficient, but not all-encompassing, description.

**Standard Setting Methodology**

This section describes the methodology whereby the panelists will make the standard setting judgments.

Following the completion of the threshold student definitions and before the start of official standard setting rounds of judgments, the panelists will be trained in the judgment-making procedure. During the training, the panelists will review a small set of practice items, and then make practice judgments on judgment forms (i.e., Excel spreadsheets). ETS staff members will be available to help panelists as needed. A variety of items and judgment types will be included in the training and practice. The main purpose of the training is to familiarize the panelists with how to make judgments on the judgment forms. The training also will serve as a reminder to the panelists to use the threshold student definitions when making judgments. The panel facilitator will provide feedback on the panel’s judgments, and then the panelists will discuss the rationales they used for making those judgments.

In this exercise, the standard setting panel facilitator will listen to the participants’ rationales and comments and ensure that they are consistent with the training. If the facilitators notice that the comments indicate any misconceptions about how to make judgments for standard setting, they will redirect the panelists. After the training session, the panelists will be asked to complete a training evaluation through which their understanding and readiness to proceed in the standard setting workshop can be confirmed. Additional training will be provided to participants as needed. Having completed this training, the panelists will be directed by the facilitator to make judgments on the operational items.

For each of the achievement levels, the panelists will make two rounds of judgments. The first round of judgments, Round 1, will be made independently and without discussion. Panelists’ judgments made in Round 2 will include consideration of judgment data from Round 1 as well as panelist discussions of the data and their rationales, as described in the [Feedback and Discussion](#_Feedback_and_Discussion) section. The steps in this process, including training, practice, judgments, and feedback, can also be found in the agenda (refer to [appendix B](#_Appendix_B:_Sample).)

**Item Scoring, Judgments, and Rating Scales**

The CAA for Science includes multiple item types, which solicit a variety of response types. The assessment includes rubric-scored 2-point items and dichotomously scored 1-point items as well as some 2-point items that are scored as two dichotomously scored 1-point items. One important goal in standard setting is to reduce the cognitive complexity of making judgments; instructions and judgments will be more intuitive for the panelists if the standard setting judgments are aligned with the scoring rules. Specifically, the CAA for Science standard setting process will include both the modified Angoff judgment procedure (Brandon 2004; Hambleton and Pitoniak 2006) and the extended Angoff procedure (Cizek and Bunch 2007; Hambleton and Plake 1995). Using these two judgment procedures allows panelist judgments to align with the scoring of the item. Rubrics and scoring rules will be provided to the panelists, along with the answer keys, as described in next sections.

**Modified and Extended Angoff Judgments**

For all 1-point items, the modified Angoff method will be implemented. There are two types of 2-point items, and they will be treated differently in the standard setting process. For the 2‑point composite items that are scored as two 1-point dichotomously scored items, the modified Angoff method will be implemented. For all other 2-point items, the extended Angoff method will be implemented. One-point items include discrete item types, such as selected-response items or TEIs (e.g., grid, match). The 2-point items include grid, multiple-choice multiple-select, match, and composite items, with scoring rules that indicate how a student will obtain a score of 0, 1, or 2. During the standard setting workshop, the panelists will review three performance tasks per grade (one per domain). This is aligned with the operational blueprint (CDE 2018).

**Modified Angoff**

The modified Angoff method is a probability-based standard setting method. For 1-point items, each panelist will judge the item on the likelihood that the threshold student would answer the item correctly. Panelists will make their judgments using a rating scale between 0 and 1 (i.e., 0, .05, .10, .20, .30, .40, .50, .60, .70, .80, .90, .95, 1). If an item is judged as difficult for the threshold student, the panelist would select a lower value, meaning that the probability the threshold student would answer the item correctly is low. If another item is judged as being easier for the threshold student, a higher value would be selected, indicating it is more likely that the threshold student would answer the item correctly.

The panelists will receive the consensus version of the threshold student definitions and will be instructed to refer to those definitions in comparison with the demands of the item for each judgment. They will be advised to approach the modified Angoff judgment-making process in two stages. First, they will consider what is the most likely probability range that the threshold student would answer the question correctly. The facilitator will encourage the panelists to consider the following general guidelines to guide their decisions:

* Items in the 0 to .30 range are those that the threshold student would have a low chance of answering correctly.
* Items in the .40 to .60 range are those that the threshold student would have a moderate chance of answering correctly.
* Items in the .70 to 1 range are those that the threshold student would have a high chance of answering correctly.

Second, the panelists will decide how to refine their judgment within the range. For example, if a panelist thinks that there is a high chance that the threshold student would answer the question correctly, the initial decision will be in the .70 to 1 range. The second decision for the panelist will be to judge whether the likelihood of the threshold student answering it correctly is .70, .80, .90, .95, or 1.

The panelists will be asked to make two judgments for each item. The overall instructions will include a reminder tha when making Level 2 (L2) and then Level 3 (L3) judgments, it is expected that each judgment value must be at least the same or higher than the previously assigned value for each item. For example, if the threshold L2 judgment is .30, then the threshold L3 judgment must be .30 or higher. Note that the judgment forms on which the panelists will enter their judgments are Excel spreadsheets; the filtering feature will require that the panelists’ recorded judgments are the same value as, or higher than, the level before. That is, if the panelists enter an L3 judgment that is lower than the L2 judgment previously entered, they will receive a warning message.

**Extended Angoff**

An extended Angoffmethod (Cizek and Bunch 2007; Hambleton and Plake 1995) will be used for the rubric-scored 2-point items. For these items, panelists will decide on the assigned score value that would most likely be earned by the threshold student for each item scored as 0, 1, or 2. The panelists will be asked to first review the definition of the threshold student, and then to review the item and its scoring rubric. The rubric, or scoring rule, defines the quality of the evidence that would merit a response earning a particular score. The scoring rules for these 2-point items describe which responses are required to achieve one point and what responses are required to achieve two points.

During this review, each panelist will independently consider the level of knowledge and skill required to respond to the item as well as which response would earn a particular score as defined by the rubric. Each panelist will decide on the score most likely to be earned by each threshold student from the possible values a student can earn. The panelists will be reminded to refer to the knowledge and skills of the threshold student definition and the scoring rules and not to expect the two levels to match the two possible scores. For the two judgments—L2 and L3—the higher level should have the same or higher expectation.

**Feedback and Discussion**

The purposes of feedback and discussion are to allow the panelists to see the range of judgments across the panel, hear the rationales of the other panelists, and arrive at a mutual understanding of the expectations of threshold students’ performance on the test. The facilitators will encourage discussion, and the process of judgment, feedback, and discussion will be repeated during the entire standard setting workshop until all threshold score recommendations have been collected.

**Round 1 Feedback and Discussion and Round 2 Judgments**

The feedback and discussion from the Round 1 judgment data will inform Round 2 judgments. As part of the post-Round 1 feedback, panelists will receive their own judgments in the context of the range of judgments across the panel, and the facilitator will share feedback on the similarities and differences of the panel judgments on the CAA for Science items.

The feedback on panelists’ judgments on 1-point items is displayed first as a list showing the range of individual total score judgments, from low to high, and then as a summary table with the panel’s mean, minimum, and maximum judgment for each level. The judgments are also displayed in three ranges, that is, the number of panelists rating an item from 0 to .30, the number of panelists rating an item from .40 to .60, and the number of panelists rating an item from .70 to 1. This helps the panelists to see where there is agreement; it is also consistent with the instructions on how to consider the full range of 0 to 1 when making judgments.

Feedback for item judgments based on extended Angoff judgments will be displayed as 0, 1, and 2. For both types of judgments—modified Angoff and extended Angoff—items for which two-thirds, or 67 percent, of panelists were in the same range for a level will be considered as “close to agreement.” Items with a more widespread distribution of judgments will be targeted for discussion. However, if any panelist would like to hear and provide feedback on any item, regardless of the degree of agreement, data will be available for a discussion.

In addition, if the empirical data from the 2020–21 CAA for Science administration indicates a meaningful discrepancy compared with the panel judgments, those items will be discussed. Given that the sample size of test takers is small because of the novel coronavirus disease 2019 pandemic, the item difficulty data will be used in consultation with the CDE and considered as additional information only. It will not be displayed as a table of percent correct to the panel in order to limit misinterpretation.

The panelists will be encouraged to discuss their judgments and rationales. Items will be discussed in batches of 10. During discussion and at the end of each batch of 10, the panelists will make changes on their own judgment forms by entering independent Round 2 judgments for any judgments they want to change.

**Round 2 Feedback and Final Evaluations of the Process**

After the Round 2 data has been analyzed, panel results will again be projected in each panel room—in this case the summary statistics of the panel’s threshold scores for Round 1 and Round 2. This allows the panelists to see what changed between Round 1 and their final Round 2 judgments. After the panel has discussed this data, panelists will be invited to complete the confidential final evaluation form, which asks about both the standard setting process and the level of confidence in the final recommendations.

**Vertical Articulation Discussion**

Once all final evaluations have been completed, a subset of panelists across the three grade-level or grade-band panel rooms will participate in a vertical articulation discussion. The standard setting director will provide information to the full group on the process in each room and solicit feedback from the panelists. The vertical articulation panel will share notable aspects of the discussions their panels had during the development of the threshold student definitions and about the rationales about judgments. The purpose is to share commonalities and differences across the grade levels and grade band. Because the three grade-level or grade-band assessments are not equated, are not for adjacent grades, and are not on the same scale, direct comparisons cannot be made using raw score threshold scores. Therefore, during vertical articulation, the focus will be on the expectations defined by the threshold student definitions and the rationales for judgments on the items in and across grades.

**Standard Setting Materials**

The panelists for each CAA for Science grade-level or grade-band assessment will be provided with the following materials:

* Workshop agenda ([appendix B](#_Appendix_B:_Draft))
* Test materials in the CRT
* Operational field test items administered in 2020–21
* Answer key with scoring rules and rubrics
* *DFAs*
* CAA for Science general and range ALDs
* Judgment forms for electronic data collection and feedback (Excel spreadsheets)
* Training materials (e.g., practice judgment forms and feedback)
* Standard setting evaluations launched in Zoom
* Training evaluations
* Final evaluation

The manner in which these materials are used is described more fully in the [Standard Setting Process](#_Standard_Setting_Process) and [Standard Setting Methodology](#_Standard_Setting_Methodology) sections.

**CAA for Science General and Range ALDs**

The general ALDs provide the meaning of achievement levels across grades. Range ALDs define the achievement levels specific to each grade level and science domain.

**Judgment Materials**

During the training round, the panelists will make their judgments on a small number of items from the CAA for Science practice test. A variety of task types will be provided during the training round to familiarize the panelists with the types of judgments they will make. The training process, which includes learning how to make judgments on Excel spreadsheets, is described in the [Standard Setting Methodology](#_Standard_Setting_Methodology) section. During the operational rounds of judgments, panelists will again enter their judgments on the judgment form.

**Evaluation Forms**

Following the training on the purpose and process for developing the threshold student definitions, and after the training on the judgment and feedback process, the panelists will be asked to complete a training evaluation. They will be able to indicate their understanding or ask questions anonymously. As needed, the facilitator will review aspects of the training. At the conclusion of the workshop, panelists will be asked to complete a final evaluation of the entire process.

**Recommendations and Technical Report**

ETS will deliver the recommended threshold scores for grades five and eight and the high school grade band to the CDE one week after the standard setting workshop is complete. The data files containing score distributions for grades five and eight and the high school grade band will be provided to the CDE after the 2021–22 CAA for Science administration. Also provided will be the recommended threshold score tables, developed by applying the 2021–22 student performance data, which will be designed for presentation to the SBE. ETS understands that additional tables presenting the SSPI’s recommendations may be developed by the CDE and verified by data analysts at ETS.

ETS will produce and deliver the final technical report for the standard setting based on the standard setting panel judgments by October 8, 2022. The technical report will contain a description of the process used to set the standards, a description of the panelists’ qualifications, results presented during the standard setting process, and statistical information related to the threshold score judgments (i.e., two standard errors of judgment above and below the panel-recommended threshold score).

**Staffing, Logistics, and Security of Panel Meetings**

The CAA for Science standard setting workshop will be held from August 3 to August 5, 2021, using Zoom. Patricia Baron, Ed.D., the standard setting director at ETS, will conduct a walk-through of the process for the CDE on or before July 27, 2021.

To allow the standard setting meetings to run smoothly, all groups will be led by trained and experienced standard setting facilitators who will conduct the training, facilitate the process, and keep the discussions on track. The panel facilitators will respond to any standard setting process questions.

Dr. Baron will lead the introductory training session and will oversee the workshop process. In addition, ETS will provide two assessment development specialists, who will be available to respond to questions about items and other materials, and a psychometrician, who will provide information to the CDE about the data from the 2020–21 administration. For the duration of the workshop, ETS will also provide two data managers familiar with the judgment data collection and remote tools and one process specialist well versed in the CRT. Representatives of ETS’s program management staff, Jackson Brown and Kelly Bolton, also will attend the sessions and be available to the CDE as needed.

All logistics concerns will be addressed by SCOE. ETS understands that CDE staff will be present during the standard setting sessions to hear discussions, observe the process, and address any policy-level issues, as appropriate. Security of all confidential material will be the responsibility of ETS and SCOE staff.

**Appendix A: General Achievement Level Descriptors**

General ALDs are generic descriptors of student performance level expectations. They provide the range of measurement for the assessment and characterize results with regard to the defined standards. General ALDs apply across grade levels and are referenced when drafting the specific descriptors used in setting threshold scores.

The proposed CAA for Science general ALDs are presented in table A.1.

Table A.1. CAA for Science General ALDs

|  |  |
| --- | --- |
| **Level** | **Description** |
| 3 | Students at this level demonstrate **understanding** of core subject matter in the content area. They are actively working with adapted grade-level content that focuses on the essential knowledge and skills and may need occasional prompts and assistance to complete tasks and activities. |
| 2 | Students at this level demonstrate **foundational understanding** of core subject matter in the content area when provided with frequent prompts and supports. They are actively working with adapted grade-level content that focuses on the essential knowledge and skills and may frequently need supports to complete tasks and activities. |
| 1 | Students at this level demonstrate **limited understanding** of adapted grade-level content that focuses on much of the basic knowledge and skills, even with extensive supports. |

**Appendix B: Draft Agenda**

All times are approximate and may change depending on the panel’s progress.

**Day 1**

Table B.1. Day 1 Agenda

|  |  |
| --- | --- |
| **Time** | **Event** |
| 7:15 a.m. | Zoom session open |
| 7:30 a.m. | Welcome and general session |
| 8:15 a.m. | Morning break and move to panel room Zoom breakouts |
| 8:30 a.m. | Panel introductions and agenda for the day |
| 8:45 a.m. | Test familiarization, self-score |
| 10:30 a.m. | Morning break |
| 10:45 a.m. | Discuss content measured on the test |
| 11 a.m. | Training to draft threshold student definitions; complete training evaluation; begin drafting in small groups |
| 12:15 p.m. | Lunch break |
| 1 p.m. | Continue small group work on threshold student definitions |
| 1:45 p.m. | Reach consensus on threshold student definitions |
| 3:15 p.m. | Overview of Day 2 agenda; end of Day 1 |

**Day 2**

Table B.2. Day 2 Agenda

| **Time** | **Event** |
| --- | --- |
| 7:15 a.m. | Zoom session open |
| 7:30 a.m. | Goals for today |
| 7:45 a.m. | Complete training and practice on modified Angoff standard setting judgments on 1-point items; complete training evaluation 2 |
| 9:15 a.m. | Morning break |
| 9:30 a.m. | Complete Round 1 standard-setting judgments on 1-point items |
| Noon | Lunch break |
| 1 p.m. | Training on extended Angoff methods; training evaluation 2Complete Round 1 standard setting judgments on 2-point items |
| 2:30 p.m. | Overview of Day 3 agenda; end of Day 2 |

**Day 3**

Table B.3. Day 3 Agenda

|  |  |
| --- | --- |
| **Time** | **Event** |
| 7:15 a.m. | Zoom session open |
| 7:30 a.m. | Goals for today |
| 7:45 a.m. | Round 1 feedback and complete Round 2 judgments |
| 10:15 a.m. | Break  |
| 10:30 a.m. | Review final Round 2 recommendations; complete final evaluation. |
| 12:30 p.m. | Lunch break |
| 1:15 p.m. | Feedback on CAA for Science panel recommendations; complete final evaluation |
| 1:30 p.m. | Vertical articulation discussion |
| 3:30 p.m. | End of CAA for Science vertical articulation discussion |

**References**

Brandon, P. R. 2004. “Conclusions About Frequently Studied Modified Angoff Standard-Setting Topics.” *Applied Measurement in Education* 17: 59–88.

California Department of Education. 2018. “California Alternate Assessment for Science Blueprint.” Sacramento: California Department of Education. [https://www.cde.ca.gov/‌ta/tg/ca/documents/caascienceblueprint.docx](https://www.cde.ca.gov/ta/tg/ca/documents/caascienceblueprint.docx)

Cizek, G. J., and M. B. Bunch. 2007. *Standard Setting: A Guide to Establishing and Evaluating Performance Standards on Tests*. Thousand Oaks, CA: Sage.

Hambleton, R. K., and B. S Plake. 1995. “Using an Extended Angoff Procedure to Set Standards on Complex Performance Assessments.” *Applied Measurement in Education* 8: 41–55.

Hambleton, R. K., and M. J. Pitoniak. 2006.”Setting Performance Standards,” in *Educational Measurement,* edited byR. L. Brennan (4th ed., pp. 433–70). Westport, CT: Praeger.