California Department of Education

Executive Office

SBE-002 (REV. 11/2017)

memo-pptb-adad-aug19item02

# **MEMORANDUM**

**DATE:** August 16, 2019

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

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

**SUBJECT:** Updates on the California Spanish Assessment Standard Setting Plan

## Summary of Key Issues

In September 2019, the California Department of Education (CDE) will bring proposed threshold scores for the California Spanish Assessment (CSA) to the California State Board of Education (SBE) for adoption. Threshold scores determine the “entry” or “exit” points between the three levels of achievement––Limited, Moderate, and High––on the CSA corresponding to the achievement level descriptors approved by the SBE in November 2017. This Memorandum provides a summary of related activities conducted prior to the CSA standard setting workshop.

**Purpose of the Standard Setting**

In August 2016, the SBE approved the high-level test design of the CSA to measure a student’s competency in Spanish reading language arts in grades three through eight and high school for the purposes of the following:

* Providing student-level data in Spanish competency
* Providing aggregate data that may be used for evaluating the implementation of Spanish language arts programs at the local level
* Providing a high school measure suitable to be used, in part, for the State Seal of Biliteracy

In November 2017, in preparation for the SBE’s approval of the CSA threshold scores, the SBE approved the CSA general achievement level descriptors (ALDs), which describe the expectations at each level; the expectations give meaning and context to scale scores by describing the knowledge and skills students must demonstrate to achieve each level (<https://www.cde.ca.gov/be/ag/ag/yr17/agenda201711.asp>). These general ALDs informed the development of content-specific or range ALDs, which aid in the interpretation of the threshold scores that distinguish between the three achievement levels for each grade or grade span for each domain of the assessment.

In preparation for the 2019 CSA standard setting workshop, Educational Testing Service (ETS) conducted a meeting in July 2018 with 23 educators to provide them with a thorough explanation of the draft range ALDs. The role of the educators was to provide ETS with feedback and recommended edits on the draft range ALDs. Those recommended changes to the draft ALDs were used in the August 2019 CSA standard setting workshop.

Prior to the standard setting workshop, a standard setting plan was submitted to the CDE for review and approval. For detailed information on the standard setting plan, see Attachment 1.

ETS will convene a CSA standard setting workshop August 6–9, 2019, to evaluate the degree to which the threshold scores and achievement levels of the CSA accurately distinguish between levels of the students’ Spanish competency. The standard setting plan recommends that 60 California educators representing all regions of the state form the group of participants. The selected participants will rely on their extensive experience in working with students learning Spanish as well as their familiarity with the Common Core State Standards en Espaňol for this work. This standard setting educator panel will recommend threshold scores based on their professional judgment. The threshold scores will allow for students to be placed, on the basis of their performance, in one of three achievement levels, by grade or grade span.

Following the standard setting workshop, ETS will report the educator panel-recommended threshold scores to the CDE for review. The review will be conducted by psychometricians from the CDE and select California Assessment of Student Performance and Progress Technical Advisory Group members and will inform the recommended threshold scores.

**Next Steps**

The State Superintendent of Public Instruction’s recommendations for the CSA threshold scores will be presented to the SBE for consideration in September 2019. Upon approval by the SBE, threshold scores and the associated achievement levels will then be reported on the student score reports for individual students from the 2018–19 CSA administration.

## Attachment(s)

Attachment 1: California Assessment of Student Performance and Progress California Spanish Assessment Standard Setting Plan (19 Pages)

****

# **California Assessment of Student Performance and Progress: California Spanish Assessment Standard Setting Plan**

**July 9, 2019**

Prepared by



Educational Testing Service

660 Rosedale Road, Princeton, NJ 08541

**Contents**

[Background 3](#_Toc13572747)

[Purpose and General Description of the Standard-Setting Process 3](#_Toc13572748)

[Panelists 4](#_Toc13572749)

[Standard-Setting Materials 5](#_Toc13572750)

[Standard Setting Process 6](#_Toc13572751)

[Test Familiarization 7](#_Toc13572752)

[Defining the Borderline Student 7](#_Toc13572753)

[Standard Setting Methodology 8](#_Toc13572754)

[Bookmark Standard Setting 9](#_Toc13572755)

[Feedback and Discussion 10](#_Toc13572756)

[Recommendations and Technical Report 11](#_Toc13572757)

[Staffing, Logistics, and Security of Panel Meetings 12](#_Toc13572758)

[Appendix A. Sample Agenda 13](#_Toc13572759)

[Day 1 13](#_Toc13572760)

[Day 2 14](#_Toc13572761)

[Day 3 15](#_Toc13572762)

[Day 4 16](#_Toc13572763)

[Appendix B. Sample Item Map 17](#_Toc13572764)

[Appendix C. Sample Judgment Forms 18](#_Toc13572765)

[References 19](#_Toc13572766)

## Background

The first operational administration of the California Spanish Assessment (CSA), which is aligned with the *Common Core State Standards en Español* (CCSS en Español) (Council of Chief State School Officers, California Department of Education [CDE], and San Diego County Office of Education, 2012), occurred April through July, 2019. Standard setting is required so that threshold scores and achievement levels will be available for the fall 2019 release of the CSA score reports. The purpose of the CSA is to measure students’ Spanish language arts proficiency. The student-level data provided by this assessment can be used to evaluate the implementation of Spanish language arts programs at the local level. High school students with Spanish and English biliteracy skills can take the CSA to demonstrate their proficiency in Spanish; competencies at the highest level are expected to be comparable to what is expected for the State Seal of Biliteracy.

The computer-based, optional CSA is administered in six grades (three through eight) and the high school grade span (grades nine through twelve). It is presented in Spanish, includes technology-enhanced items, and is accessible to students with disabilities. It assesses three Spanish language arts domains (Reading, Writing Mechanics, and Listening) with 52 multiple-choice items.

The achievement level descriptors (ALDs) describe what students should be able to do at each level. The general, or policy, ALDs were approved by the California State Board of Education (SBE) in November 2017. On July 18 and 19, 2018, twenty-one California educators convened in Sacramento to review and provide input on the descriptions of the Spanish reading/language arts knowledge and skills necessary for students in grades three through eight and high school to be placed into one of three achievement levels. These range ALDs will be used to inform the standard setting process. Threshold scores will be developed to allow for score reporting. The plan presented in this document is for the CSA standard setting scheduled for August 2019.

## Purpose and General Description of the Standard-Setting Process

The purpose of standard setting for the CSA is to collect recommendations for the CSA threshold scores. These recommendations will be reviewed, along with additional data, by the CDE, and the final approval of the threshold scores will be made by the SBE in September 2019. The purpose, general process, logistics, security, and staffing for the CSA standard setting are described herein. The approach used in this process adheres to the guidelines and best practices recommended in the standard setting literature; specifically, the Bookmark standard-setting method (e.g., Karatonis and Sireci, 2006; Lewis, Mitzel, Mercado, and Schulz, 2012; Mitzel, Lewis, Patz, and Green, 2001) will be applied to the assessment. A walk-through of the process will be conducted for the CDE prior to the workshop by Patricia Baron, Ed.D, the standard setting director at Educational Testing Service (ETS).

The CSA standard setting workshop will be held August 6, 2019 through August 9, 2019, at the Sacramento County Office of Education (SCOE) in Mather, California. Panelists will make three rounds of judgments, with feedback and discussion occurring between each round. The last day of the workshop will consist of a cross-grade articulation workshop by representatives from each panel who instruct students at each grade level assessed.

### Panelists

A diverse group, representative of Spanish-language educators in California, will be recruited to participate as panelists in the standard setting sessions. In recruiting panelists, the goal will be to include a representative group of California educators who are familiar with the CCSS en Español and who have experience in the education of students in grades three through twelve who will take the CSA. It is important to include teachers working with these students, as they will provide a perspective on learning goals for the students taking the CSA as well as students’ progress toward Spanish language proficiency.

For the CSA, there will be four panels of educators ( Table 1). The targeted number of panelists is 15 per panel, for a total of 60 educators.

 Table 1. Panel Configuration

| **Panel** | **Test(s)** |
| --- | --- |
| A | Grades 3 and 4 |
| B | Grades 5 and 6 |
| C | Grades 7 and 8 |
| D | High school\*  |

 \*Grades 9–12.

The CSA panels will be assembled in test-specific panel rooms for much of the standard setting work. For three panels, the panelists will work on two tests, assessing adjacent grades, as shown in Table 1. Educators teaching students in each of the relevant grade levels will serve on those panels. Educators teaching across the high school grade span will serve on the high school panel. This design supports consistency in the panelists’ judgments as well as consistency in the definition of the threshold ALDs (described later in this plan). Panelists will sit at three tables, with five educators at each table. ETS recommends that the composition of each panel include the following as criteria for selection. Educators who are teaching Spanish-language learners, in the grade level(s) assigned to the panel;

* Educators who are teaching students who would take the CSA
* Educators who are familiar with the CCSS en Español

The decision on the panelists’ selection will be made by the CDE. After the panelists are approved, they will be notified, and travel arrangements will be made. 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 removal of the materials from the meeting area.

## Standard-Setting Materials

The following list of materials will be provided to panelists for each CSA grade level assessed:

* Workshop agenda (Appendix A)
* Familiarization materials––CSA test book (on paper) and answer key with scoring rules for 2-point items
* Judgment materials––passage book with scripts for audio prompts, ordered item booklet (OIB), and corresponding item map
* Impact data––data based on the spring 2019 administration of the CSA
* Forms––practice judgment form for the Bookmark method, operational judgment forms, training evaluation form, and final evaluation form

**Familiarization materials.** Panelists will use these materials to become familiar with the test content. They will “take the test,” receive the key, and self-score. Operational test forms will be used for all grades and grade spans. During the test-familiarization process, the computer-administered version of the test will be displayed as panelists follow along using the printed materials, make notes, and respond to the questions.

**OIBs.** Items will be ordered from least difficult to most difficult, based on a response probability of 0.67 employed in the item response theory (IRT) model (Mitzel, Lewis, Patz and Green, 2001). Items with a maximum score of 1 will appear once. Items with a maximum score of 2 will appear twice, representing each of the response probabilities of earning 1 point or 2 points. Scores of 0 will not appear in the OIB.

**Passage books.** Some CSA items are a part of a set, in that more than one item corresponds to a single prompt. The prompt could be a written passage or an audio prompt. The passage book will have each written passage printed once and the scripts for each audio passage printed once. The item map will provide information for the panelists on how to find the appropriate passage or audio script for each item in the OIB.

**Item map** The item map (Appendix B) will provide item information (such as a key, corresponding passage references, and the maximum number of points) for every item on the CSA. This information will be presented in the same order as that of the corresponding OIB—from least to most difficult, based on a response probability of 0.67. The item map also will list the original order from the test form as well as relative item-difficulty information on a temporary scale called the standard setting scale. The standard setting scale will look similar to the reporting scale but will be used only during standard setting to aid panelists’ understanding of the relative difficulty of the items in the OIB.

**Impact data.** After the second and third rounds of judgment, panelists will learn the percentage of students’ scores from the spring 2019 CSA administration that would fall into each of the three achievement levels. The impact data will be based on the panel-recommended score for Round 2 and Round 3 judgments.

**Judgment and evaluation forms.** During the practice round, the panelists will complete their judgments on a practice judgment form, which will be nearly identical to the operational judgment form (Appendix C). The boxes in which panelists write their judgments (the OIB page numbers) will be shaded for the rounds and levels that are not part of the practice round. Following the practice round, the panelists will be asked to complete an evaluation of the training on how to make their bookmark judgments, which will be their initial evaluation. At the conclusion of the workshop, panelists will be asked to complete a final evaluation of the entire process.

## Standard Setting Process

A pre-workshop assignment, consisting of two parts, will be given to the panelists approximately two weeks before the in-person workshop. For the first part, the panelists will be provided with a link to the CSA practice test on the CDE website and asked to take the practice test for their grade level to become familiar with the task types on the CSA. The second part of the assignment will include familiarization with the rangeALDs for one grade level or grade span and a link to the CCSS en Español and the general ALDs. Panelists will be asked to consider the expectations of a student in each of the achievement levels, take notes about the knowledge and skills of students *at the beginning* of levels 2 and 3, and bring those notes to the standard setting workshop.

Once on-site at the workshop, all panelists will attend a general session that will include an overview of the CSA and the Bookmark standard setting procedure. At the conclusion of the general session, panel facilitators experienced in working with educators in standard setting will provide in-depth training and practice on the method in each panel room. Panelists will then complete three rounds of judgments. Feedback and discussion will take place after each round of judgment (refer to the [Feedback and Discussion](#_Feedback_and_Discussion) section).

For every panel and at every table, one member will be identified as a table leader, for a total of three per panel. The table leader’s responsibility will be to help keep discussions on track at the table, report interim discussions to the room, and collect materials at the table. Table leaders will be advised of their role during the first day and will join the lead facilitator for table-leader training prior to bookmark judgments.[[1]](#footnote-2)

### Test Familiarization

Immediately following the general training session, panelists will break into their assigned panels associated with the test(s) for which they will be setting standards. For the CSA, there will be four panels (refer to Table 1). During the test-familiarization process, an assessment development (AD) content specialist will present the computer-administered assessment to the room. The panelists, using their paper forms, will follow along and record their responses to the items. After independently reviewing the test, panelists will check their responses against the answer key. They will then discuss the content measured, what they think might be particularly challenging for students, and what might be less difficult. The goal of this activity is for panelists to begin to think about and articulate their perception of the general difficulty of the tested content for students.

Once the panelists are familiar with the content of the assessment, they will begin discussing the pre-workshop assignment, including articulation of the knowledge and skills necessary for students to reach achievement level 2 and level 3 for one grade. The focus in each room will be on the grade level or grade span assessed by the CSA (refer to Table 1); however, the range ALDs for the adjacent grade level or grade span will be available to all panelists. These ALDs will be provided to allow the panels to have a clear understanding of the progression of expectations across grades. In addition, these materials, as well as impact data, will be provided on day four, during the vertical articulation workshop.

### Defining the Borderline Student

Developing definitions of borderline students is a critical component of any standard setting workshop. For each grade and for the high school grade span, panelists will work in small groups to define borderline students. The process to arrive at borderline student definitions will involve small-group discussions and the development of draft borderline-student definitions, followed by a whole-panel discussion of the draft definitions in order to reach a panel consensus of what is expected. For the CSA, two definitions will be needed for two thresholds—the level 2 borderline and level 3 borderline student definitions. Panels will work first on the level 3 borderline, because this is the point at which students will be classified as *on track* for demonstrating competencies consistent with those needed to receive the State Seal of Biliteracy.

For each grade or grade span of the CSA, panelists will refer to the range ALDs that describe the full range for each of the three levels. Beginning with level 3, borderline-student definitions will be drafted for the student just entering level 3, the “Borderline Level 3 student.” ETS facilitators will instruct panelists to limit the definitions of their borderline students to a brief description that will allow them to make decisions about the student’s performance on the test.

After the “Borderline Level 3 student” definition is drafted, two pairs of two panels working on adjacent CSA grade-level or grade-span tests will meet to discuss their drafts, provide feedback to each other, and finalize the definitions. Their discussions and work will focus on cross-grade consistency of the ALDs and the description of the borderline student for level 3. After completing the “Borderline Level 3 student” definition, the panels will reconvene in their panel rooms to complete the “Borderline Level 2 student” definition, working to ensure its consistency with the level 3 definition. Table 2 illustrates the manner in which cross-grade consistency will be considered during the workshop.

Each panel, with the exception of the high school panel, will complete the standard setting process on two CSA grade-level tests. After completing the process for the first CSA, the panel will begin the entire process again with the second assessment. The work the panels will do for cross-grade consistency will continue when they create the borderline student definitions for the second CSA grade level, as illustrated in Table 2.

Table 2. Cross-Grade Consistency, by Panel

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Borderline Student Definition Process** | **Panel A** | **Panel B** | **Panel C** | **Panel D** |
| Draft level 3 (within panel)  | Grade 4 definition | Grade 5 definition | Grade 8 definition | High school definition |
| Complete level 3 (across panels) | Grade 4 consistent with grade 5 | Grade 5 consistent with grade 4 | Grade 8 consistent with HS | High school, consistent with grade 8 |
| Draft, and then complete level 2 (within panel) | Grade 4, level 2, consistent with grade 4, level 3 | Grade 5, level 2, consistent with grade 5, level 3 | Grade 8, level 2, consistent with grade 8, level 3 | High school, level 2, consistent with high school, level 3 |
| Draft level 3 (within panel) | Grade 3 consistent with grade 4 | Grade 6 consistent with grade 5 | Grade 7 consistent with grade 8 | NA |
| Complete level 3 (across panels) | NA | Grade 6 consistent with grade 7 | Grade 7 consistent with grade 6 | NA |
| Draft, and then complete level 2 (within panel) | Grade 3, level 2, consistent with grade 3, level 3 | Grade 6, level 2, consistent with grade 6, level 3 | Grade 7, level 2, consistent with grade 7, level 3 | NA |

## Standard Setting Methodology

Panelists will be trained in, and have an opportunity to practice, making judgments prior to the start of actual standard setting rounds of judgments, as described in the next section. After training, panelists will be asked to sign a training evaluation form confirming their understanding and readiness to proceed; additional training will be provided as needed. Panelists will make three rounds of judgments for each of the two achievement levels. Round 1 judgments will be made independently, without discussion. Panelists will then discuss their independent judgments with the other panelists at their respective tables. The feedback received from the Round 1 judgment data and Round 1 table-level discussion will inform Round 2 judgments. Panelists will then engage in another round of feedback based on the Round 2 data, having table-level and room-level discussions before making their third and final round of judgments (Round 3). After each round, panelists’ judgments will be collected, analyzed, and summarized. Feedback and discussions are described in detail next.

Each panel will be seated at three tables of five educators to facilitate discussion. This table format will provide an environment more conducive to panelists sharing their opinions and rationales, as some panelists may be less inclined to speak or have less opportunity to be heard in a large group. The table format also will increase the likelihood of alignment among panelists when considering the achievement expectations.

During table-level discussion, each educator will review the range of the table members’ recommendations and participate in a discussion of the rationales for their bookmark judgments. Table recommendations will be reviewed, discussed, and then aggregated across the tables. This will allow an analysis of the variability across tables and can be considered a type of replication. A room-level discussion will follow, during which members will share their individual perspectives, as well as themes from the table-level conversation, with the rest of the panel. This process provides an opportunity for the varied perspectives of the experts serving on the panel to be heard.

### Bookmark Standard Setting

To make judgments and place bookmarks in the OIB, panelists will review each item in the OIB, in sequence, and consider whether a student at the beginning of level 2, known as the “Borderline Level 2 student”, would most likely be able to correctly respond to the item (in order to earn 1 or 2 points, as represented in the OIB). Panelists also will be instructed to use the item map. The relative differences in item difficulty on the item map will allow panelists to more easily evaluate the difference in difficulty between adjacent items in the OIB.

Each panelist will place the level 2 bookmark on the first item encountered in the OIB that the panelist believes the “Borderline Level 2 student,” would most likely not be able to answer correctly. This placement indicates that the items beyond that point are too difficult for that borderline student. Each panelist, starting at the level 2 bookmark, will then continue with a review of each item and place a bookmark on the first item in the OIB that the “Borderline Level 3 student” would most likely not be able to answer correctly (the item that likely exceeds the content understanding of the “Borderline Level 3 student”).

In the Bookmark method, the definition of “most likely” is related to the IRT model used to order the items. Thus, panelists will be instructed to think of “most likely” as having a two-thirds likelihood of answering the item correctly (in order to earn 1 or 2 points, as represented in the OIB). As described earlier, the item ordering in the OIB for CSA standard setting is based on a response probability of 0.67 (RP67) as recommended by research (e.g., see Cizek, 2012, p.135). Using RP67 for item ordering and instructing panelists to think about a two-thirds likelihood, which is easily understood, will provide an alignment between the instructions and the analytical model.

Panelists will receive training, have an opportunity to practice making bookmark judgments, and respond to the training evaluation form. Next, panelists will be asked to place their first judgment independently (the Round 1 judgment). The instructions to the panelists for the operational judgments will be as follows:

1. Focus on Level 2 first.
2. Review the *borderline student* definition and refer to the ALDs as needed.
3. Review the first item and identify the knowledge and competencies required to respond successfully to the item. Continue to the next item.
4. Place a bookmark on the first item that exceeds the knowledge and skills of the Borderline Level 2 student.
5. Focus on Level 3.
6. Continue with the next item in the OIB and repeat steps 2–4 for the Borderline Level 3 student.

After placing the bookmarks, panelists will record the bookmark page, or OIB number, for each threshold score. Judgments will be summarized and discussed prior to the next round of judgments.

### Feedback and Discussion

The purpose of feedback and discussion is to allow panelists to hear the rationales of the other panelists, to receive empirical information about item performance and student performance, and to arrive at a mutual understanding of the expectations of borderline students’ performance on this test. The process of judgment, feedback, and discussion will be repeated over the entire standard setting workshop until all threshold scores are set.

Feedback will be given to the panelists after Round 1 judgments are collected and summarized. The table-level feedback will provide an opportunity for the panelists to discuss in a small-group setting the range of judgments and rationales for why they made their judgments. The panelists will be provided with the median and range of the panel judgments for the table, and they will then discuss their judgments in table-level groups. As part of the discussion of their rationales, panelists will be instructed to reflect on the borderline student definitions, the knowledge and skills required to correctly respond to the items, and the relative difficulties between the items as shown on the item map. These relative difficulties will allow panelists to get a better sense of how significant the differences are among their judgments, thereby aiding discussion. Table leaders will take notes on the themes of the table-level discussion as the panelists discuss the items from the lowest to highest OIB item judgments. Then panelists will make independent Round 2 judgments, which will be collected and summarized.

After making Round 2 independent judgments, but before receiving feedback on those judgments, the table leaders will provide a report to the room of their post-Round 1 table-level discussions. After all questions have been answered, the tables will commence their Round 2 table-level feedback. Once the table-level discussions have concluded, the room will receive room-level feedback from Round 2 judgments. Results will be projected in each panel room, including summary statistics of the panel’s threshold scores: the panel average (median), minimum, maximum, and range of judgments. After the panelists have discussed the data, the student performance data showing the impact, or consequence data, of the Round 2 judgments will be presented. This impact data will be based on spring 2019 CSA student performance, and the feedback will show “what percentage of students would fall into each level based on these decisions.” After the room-level discussions, panelists will be asked to undertake their table-level discussions again.

Once all discussions have concluded, panelists will be asked to make their final round of judgments. The results from the Round 3 judgments will be considered the final threshold score recommendations from the standard setting panel. Panelists will review Round 3 feedback and respond to a final evaluation form, which will be confidential.

## Recommendations and Technical Report

ETS will deliver the recommended threshold scores and the data files containing score distributions for grades three through eight and the high school grade span (nine through twelve) to the CDE on August 12, 2019. In addition to the recommended threshold score tables typically designed for presentation to the SBE, additional tables may be developed by the CDE and data verified by ETS. Further discussion between ETS and the CDE to define the composition of the tables and a timeline for delivery will be required prior to the standard setting.

ETS will produce and deliver the final technical report for the standard setting by October 11, 2019. The technical report will contain a description of the process used to set standards, a description of the panelists’ qualifications, the results presented during the standard setting process, and statistical information related to the threshold score judgments: two standard errors of judgment and two standard errors of measurement above and below the panel-recommended threshold score.

## Staffing, Logistics, and Security of Panel Meetings

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

Wanda Swiggett, Ph.D., the lead facilitator, will lead the introductory training session and the table-leader training and will oversee the workshop process. In addition, ETS will provide three assessment development specialists who will be available to respond to questions about items and other materials. ETS also will provide a data analyst experienced in the bookmark process, and a lead CSA psychometrician, Likun Hao, Ph.D., for the duration of the workshop. ETS program management staff Zulma Torres, Yoon Mee Byun, and Marques Barron also will attend the session and be available to the CDE as needed.

All logistics and panelists’ travel 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.

Panelists will be provided with materials on the first day at the time of registration and as needed during the four-day process. All of the secure materials will have the panelists’ ID numbers printed on them and will be checked out and in as they are distributed and collected, respectively. In addition, panelists will sign in each day on the sign-in sheet next to their printed names and ID numbers. At the end of the process, ETS staff will collect and destroy all confidential materials.

## Appendix A. Sample Agenda

### Day 1

|  |  |
| --- | --- |
| **Time\*** | **Event** |
| 7 a.m. | Registration and breakfast |
| 8 a.m. | General session with all four panels |
| 9:30 a.m. | Morning break—reconvene in four separate panel rooms |
| 9:45 a.m. | Introductions of panelists and goals for today |
| 10 a.m. | Test familiarization and discussion |
| Noon | Lunch |
| 1 p.m. | Review achievement level descriptors (ALDs) and draft borderline student definition for level 3 |
| 2 p.m. | Afternoon break |
| 2:15 p.m. | Cross-panel consistency meetings in two meeting rooms (panels 3/4 and 5/6 meet together; panels 7/8 and HS meet together) |
| 3 p.m. | Reconvene in four separate panel rooms and complete borderline student definition for level 2 |
| 5 p.m. | End of day 1 |

### Day 2[[2]](#footnote-3)

|  |  |
| --- | --- |
| **Time\*** | **Event** |
| 7 a.m. | Breakfast (table leader training at 7:50) |
| 8 a.m. | Review and discussion of standard setting materials  |
| 9 a.m. | Standard setting training and practice |
| 10 a.m. | Round 1 bookmark judgments |
| 11 a.m. | Morning break |
| 11:15 a.m. | Post-round 1 feedback and discussion; then place Round 2 bookmark judgments |
| 12:30 p.m. | Lunch |
| 1:30 p.m. | Post-round 2 feedback and discussion; then place Round 3 bookmark judgments |
| 2:30 p.m. | Afternoon break |
| 2:45 p.m. | Test familiarization and discussion |
| 4 p.m. | Review ALDs and draft borderline student definition for level 3 |
| 5 p.m. | Finish day 2 |

### Day 3[[3]](#footnote-4)

|  |  |
| --- | --- |
| **Time\*** | **Event** |
| 7 a.m. | Breakfast |
| 8 a.m. | Cross-panel consistency meeting for panels 5/6 and 7/8 |
| 9:30 a.m. | Panels 5/6 and 7/8 reconvene in separate panel rooms; all panels complete borderline student definition for level 2 |
| 10:45 a.m. | Morning break |
| 11 a.m. | Review and discussion of standard-setting materials  |
| Noon | Round 1 bookmark judgments |
| 12:30 p.m. | Lunch |
| 1:30 p.m. | Post-round 1 feedback and discussion; then place Round 2 bookmark judgments |
| 3 p.m. | Afternoon break |
| 3:15 p.m. | Post-round 2 feedback and discussion; then place Round 3 bookmark judgments |
| 4 p.m. | Afternoon break |
| 4:15 p.m. | Final evaluation; remarks from the California Department of Education |
| 5 p.m. | End of day 3 |

***Thank you for your time and contributions!***

### Day 4[[4]](#footnote-5)

|  |  |
| --- | --- |
| Time\* | Event |
| 7 a.m. | Breakfast |
| 8 a.m. | Introductions |
| 8:30 a.m. | Cross-grade articulation |
| Noon | Lunch and dismissal of panelists |

***Thank you for your time and contributions!***

## Appendix B. Sample Item Map

| **Ordered Item Number Sequence** | **Original****Item order** | **Measure** | **Passage/ Script Title** | **Key** | **Points1** | **Standard Setting Scale** | **Item ID** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 41 | Reading | Passage 1 | 3,1,2 | 1 of 2 | 161 | 854612 |
| 2 | 15 | Listening | Passage 2 | 2,3,1 | 1 of 1 | 209 | 752410 |
| 3 | 48 | Writing | ~ | 1 or 3 | 1 of 2 | 210 | 570650 |
| 4 | 41 | Reading | Passage 1 | 2 | 2 of 2 | 216 | 854612 |
| 5 | 1 | Reading | ~ | 4 | 1 of 1 | 218 | 145381 |
| 6 | 20 | Listening | Passage 2 | 2 | 1 of 1 | 230 | 851451 |
| 7 | 24 | Listening | Passage 2 | 3 or 1 | 1 of 2 | 236 | 486513 |
| 8 | 12 | Reading | Passage 3 | 3 | 1 of 1 | 245 | 458545 |
| 9 | 30 | Writing | ~ | 1 and 3 | 2 of 2 | 298 | 570650 |
| 10 | 18 | Writing | ~ | 1 | 1 of 2 | 305 | 894735 |

## Appendix C. Sample Judgment Forms

**Bookmark Recording Form**[[5]](#footnote-6)

**Panel member ID\_\_\_\_\_\_\_\_\_\_\_\_ Table \_\_\_\_\_\_\_\_\_\_\_**

**Grade or grade span (circle one): 3–4 5–6 7–8 HS**

Please place your bookmark, and then record the item number from the ordered item booklet (OIB) on this judgment form. This should be the first item in the OIB where the borderline student is *not* likely to be able to answer the item correctly.

| **Achievement Level** | **OIB Item #****Round 1** | **OIB Item #****Round 2** | **OIB Item #****Round 3** |
| --- | --- | --- | --- |
| **Level 2** |  |  |  |
| **Level 3** |  |  |  |
| **Panelist Initials** | \_\_\_\_\_\_ | \_\_\_\_\_\_ | \_\_\_\_\_\_ |

Please initial the bottom of each column to certify that these are your final judgments.

## References

Council of Chief State School Officers, California Department of Education, and San Diego County Office of Education. (2012). *Common Core State Standards in language arts and literacy in history/social studies, science, and technical subjects: Spanish language version.* Retrieved from <https://commoncore-espanol.sdcoe.net/CCSS-en-Espanol/SLA-Literacy>

Cizek, G. J. (Ed.). (2012). *Setting performance standards: foundations, methods, and innovations.* Abingdon-on-Thames, UK: Routledge.

Karatonis, A., and Sireci, S. G. (2006). The bookmark standard-setting method: a literature review. *Educational Measurement: Issues and Practice*, 25(1), 4–12.

Lewis, D. M., Mitzel, H. C., Mercado, R. L., and Schulz, E. M. (2012). The bookmark standard setting procedure. In G. J. Cizek (Ed.), *Setting performance standards: Foundations, methods, and innovations* (2nd ed., pp. 225-253). New York: Routledge.

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

1. Leader training will take place prior to the start of panel-room work on the second day of the workshop. Table leaders will be provided with instructions regarding their roles and responsibilities. [↑](#footnote-ref-2)
2. All times are approximate and may change based on the panel’s progress. [↑](#footnote-ref-3)
3. All times are approximate and may change based on the panel’s progress. [↑](#footnote-ref-4)
4. All times are approximate and may change based on the panel’s progress. [↑](#footnote-ref-5)
5. For the practice round, a nearly identical form will be used, but the Round 2 and Round 3 columns will be gray, as will the Level 3 box in the Round 1 column. [↑](#footnote-ref-6)