



STANDARD SETTING TECHNICAL REPORT FOR THE CALIFORNIA SPANISH ASSESSMENT

**Prepared for the California Department of Education by
Educational Testing Service**

Presented November 14, 2019



Measuring the Power of Learning.™

Table of Contents

Chapter 1: Introduction	1
Purpose and General Description of the Standard Setting Workshops	2
Chapter 2: Method	3
Bookmark Method Standard Setting.....	3
Standard Setting Panels.....	3
Chapter 3: Materials	7
Ordered Item Booklet	7
Item Map	8
Evaluation Forms	8
Chapter 4: Process	9
Pre-meeting Preparation	9
General Session Training.....	10
Test Familiarization	10
Borderline Student Definitions	10
Table Leader Training	12
Bookmark Judgment Training.....	13
Feedback and Discussion	14
Cross-Grade Articulation for the Overall Score	14
Chapter 5: Results	16
Six Types of Data Tables	16
Impact Data	16
CSA Recommended Score Results	17
Incorporating Additional Considerations in Setting Threshold Scores.....	25
Evaluation of the Standard Setting Process	26
Evaluation Results from the CSA Standard Setting Final Evaluations	26
Chapter 6: Conclusion	38
References	39
Appendix 1: Attachments	40
Attachment A: Panelist Invitation to Participate and Pre-workshop Assignment	40
Attachment B: Final Borderline Student Definitions.....	44
CSA Borderline Student Definitions Grade Three	44
CSA Borderline Student Definitions Grade Four	45
CSA Borderline Student Definitions Grade Five.....	46
CSA Borderline Student Definitions Grade Six.....	47
CSA Borderline Student Definitions Grade Seven.....	48
CSA Borderline Student Definitions Grade Eight.....	49
CSA Borderline Student Definitions High School	50
Attachment C: Sample Item Map.....	52
Attachment D: Evaluation Forms.....	53
Training Evaluation Form	53
Final Evaluation.....	55
Attachment E: Nondisclosure Agreement Form	58
Appendix 2: Conditional Standard Error of Measurement	59
Appendix 3: Cross-Grade Articulation Data	62

List of Tables

Acronyms and Initialisms Used in the <i>Standard Setting Technical Report for the California Spanish Assessment</i>	v
Table 1. Panel Configuration	4
Table 2. Panelist Gender	5
Table 3. Panelist Primary Ethnicity/Race.....	5
Table 4. Geographical Region of Panelists	5
Table 5. Panelist Years Experience Teaching Bilingual Classes.....	6
Table 6. Panelist Subject(s) Currently Teaching.....	6
Table 7. Cross-Grade Consistency by Panel.....	12
Table 8. Median Bookmarks at the End of Each Round: Grade Three.....	17
Table 9. Median Bookmarks at the End of Each Round: Grade Four.....	17
Table 10. Median Bookmarks at the End of Each Round: Grade Five	17
Table 11. Median Bookmarks at the End of Each Round: Grade Six	17
Table 12. Median Bookmarks at the End of Each Round: Grade Seven	18
Table 13. Median Bookmarks at the End of Each Round: Grade Eight.....	18
Table 14. Median Bookmarks at the End of Each Round: High School Grade Span.....	18
Table 15. SEJs in Bookmark Placement by Round: Grade Three	19
Table 16. SEJs in Bookmark Placement by Round: Grade Four	19
Table 17. SEJs in Bookmark Placement by Round: Grade Five.....	19
Table 18. SEJs in Bookmark Placement by Round: Grade Six	19
Table 19. SEJs in Bookmark Placements by Round: Grade Seven	19
Table 20. SEJs in Bookmark Placements by Round: Grade Eight	20
Table 21. SEJs in Bookmark Placements by Round: High School Grade Span.....	20
Table 22. Distribution of 2018–19 Students Based on Round 2 Recommendations: Grade Three.....	20
Table 23. Distribution of 2018–19 Students Based on Round 2 Recommendations: Grade Four.....	21
Table 24. Distribution of 2018–19 Students Based on Round 2 Recommendations: Grade Five	21
Table 25. Distribution of 2018–19 Students Based on Round 2 Recommendations: Grade Six	21
Table 26. Distribution of 2018–19 Students Based on Round 2 Recommendations: Grade Seven	22
Table 27. Distribution of 2018–19 Students Based on Round 2 Recommendations: Grade Eight.....	22
Table 28. Distribution of 2018–19 Students Based on Round 2 Recommendations: High School Grade Span	22
Table 29. Distribution of 2018–19 Students Based on Round 3 Recommendations: Grade Three.....	23
Table 30. Distribution of 2018–19 Students Based on Round 3 Recommendations: Grade Four.....	23
Table 31. Distribution of 2018–19 Students Based on Round 3 Recommendations: Grade Five	23

Table 32. Distribution of 2018–19 Students Based on Round 3 Recommendations: Grade Six	24
Table 33. Distribution of 2018–19 Students Based on Round 3 Recommendations: Grade Seven	24
Table 34. Distribution of 2018–19 Students Based on Round 3 Recommendations: Grade Eight	24
Table 35. Distribution of 2018–19 Students Based on Round 3 Recommendations: High School Grade Span	25
Table 36. Final Evaluation Grades Three Through Four on the Usefulness of the Materials.....	27
Table 37. Final Evaluation Grades Three Through Four on the Influence of the Process Components	27
Table 38. Final Evaluation Grades Three Through Four on Timing.....	28
Table 39. Final Evaluation Grades Three Through Four on the Appropriateness of the Final Recommendations.....	29
Table 40. Final Evaluation Grades Three Through Four on Support of the Recommendations	29
Table 41. Final Evaluation Grades Five Through Six on the Usefulness of the Materials	29
Table 42. Final Evaluation Grades Five Through Six on the Influence of the Process Components	30
Table 43. Final Evaluation Grades Five Through Six on Timing.....	31
Table 44. Final Evaluation Grades Five Through Six on the Appropriateness of the Final Recommendations	31
Table 45. Final Evaluation Grades Five Through Six on Support of the Recommendations	31
Table 46. Final Evaluation Grades Seven Through Eight on the Usefulness of the Materials.....	32
Table 47. Final Evaluation Grades Seven Through Eight on the Influence of the Process Components	32
Table 48. Final Evaluation Grades Seven Through Eight on Timing	33
Table 49. Final Evaluation Grades Seven Through Eight on the Appropriateness of the Final Recommendations.....	34
Table 50. Final Evaluation Grades Seven Through Eight on Support of the Recommendations	34
Table 51. Final Evaluation High School on the Usefulness of the Materials.....	34
Table 52. Final Evaluation High School on the Influence of the Process Components	35
Table 53. Final Evaluation High School on Timing	36
Table 54. Final Evaluation High School on the Appropriateness of the Recommendations	36
Table 55. Final Evaluation High School on Support of the Recommendations.....	36
Table 1.1. CSA Sample Item Map	52
Table 2.1. Projected Percentage of 2018–19 Students at or Above the Recommended Threshold Score, +/-1 CSEM and +/-2 CSEMs for Grade Three.....	59
Table 2.2. Projected Percentage of 2018–19 Students at or Above the Recommended Threshold Score, +/-1 CSEM and +/-2 CSEMs for Grade Four.....	59
Table 2.3. Projected Percentage of 2018–19 Students at or Above the Recommended Threshold Score, +/-1 CSEM and +/-2 CSEMs for Grade Five	60

Table 2.4. Projected Percentage of 2018–19 Students at or Above the Recommended Threshold Score, +/-1 CSEM and +/-2 CSEMs for Grade Six	60
Table 2.5. Projected Percentage of 2018–19 Students at or Above the Recommended Threshold Score, +/-1 CSEM and +/-2 CSEMs for Grade Seven	60
Table 2.6. Projected Percentage of 2018–19 Students at or Above the Recommended Threshold Score, +/-1 CSEM and +/-2 CSEMs for Grade Eight.....	61
Table 2.7. Projected Percentage of 2018–19 Students at or Above the Recommended Threshold Score, +/-1 CSEM and +/-2 CSEMs for High School.....	61
Table 3.1. Percentage at Each Level Across Grade or Grade Span	62

List of Figures

Figure 1. Borderline Student Definitions.....	42
Figure 2. Percentage at each level across grade or grade span	62

Acronyms and Initialisms Used in the *Standard Setting Technical Report for the California Spanish Assessment*

Abbreviation	Definition
CDE	California Department of Education
CSEM	conditional standard error of measurement
ETS	Educational Testing Service
OIB	ordered item booklet
ALD	achievement level descriptor
SBE	State Board of Education
SEJ	standard error of judgment
SS Scale	Standard Setting Scale
SSPI	State Superintendent of Public Instruction

Chapter 1: Introduction

The California Spanish Assessment (CSA) is an optional assessment aligned with the Common Core State Standards en Español (CCSSeE) (Council of Chief State School Officers, California Department of Education [CDE], and San Diego County Office of Education, 2012). The purpose of the assessment is to measure students' Spanish reading language arts proficiency, and the results 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 proficiency in Spanish; competencies at the highest level are expected to be comparable to what is expected for the State Seal of Biliteracy.

The first operational administration of the CSA 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 CSA score reports. The computer-based, optional assessment is administered at 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 items using a variety of item types, including multiple-selection multiple choice, drag and drop, selection in-text, and drop-down lists.

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 ([CDE, 2017](#)). On July 18–19, 2018, 21 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 were used to inform the standard setting process. Threshold scores were to be developed to allow for score reporting.

To develop threshold-score recommendations aligned to the score-reporting hierarchy ([CDE, 2017](#)), Educational Testing Service (ETS) conducted standard setting workshops, in Sacramento, California, for the CSA on August 6–9, 2019. A standard setting plan was approved by the CDE on January 23, 2018, in preparation for the standard setting workshop.

This document provides the following information:

- The purpose of the standard setting workshops and a discussion of the work conducted prior to the workshop
- An overview of the standard setting methods implemented, including discussions of the Bookmark method used to develop the overall score thresholds
- A description of the panels and materials used in the approach, the process before and during the workshop, and a discussion of the data training
- The results, including summary data from the panel judgments and evaluations by the panelists

Purpose and General Description of the Standard Setting Workshops

The purpose of standard setting for the CSA was to collect recommendations for the CSA threshold scores for the CDE to review, with final approval by the SBE. For each grade and grade span assessment, there are three performance levels (Level 1 through Level 3). A threshold score defines the beginning of a higher level of performance or achievement. Therefore, panelists defined borderline students and made judgments for Level 2 and Level 3.

A review of the standard setting literature supports the need for attention to best practices (Brandon, 2004; Hambleton & Pitoniak, 2006; Tannenbaum & Katz, 2013), which include the following:

- A careful selection of panel members
- A sufficient number of panel members to represent varying perspectives and provide for replication
- Sufficient time devoted to developing a common understanding of the assessment domain
- Adequate training of panel members
- Development of a description of each performance level
- Multiple rounds of judgments
- The inclusion of data, where appropriate, to inform judgments

The approach used in this study, specifically, the Bookmark standard setting method (e.g., Karatonis & Sireci, 2006; Lewis, Mitzel, Mercado, & Schulz, 2012; Mitzel, Lewis, Patz, & Green, 2001), adheres to these guidelines.

Chapter 2: Method

Chapters 2 and 3 describe the following:

- Descriptions of the Bookmark method of standard setting
- Standard setting panels
- Materials used in the workshop

Bookmark Method Standard Setting

In this method, test items were ordered based on item parameters estimated from field test data from easiest to most difficult and were presented in a booklet known as an ordered item booklet (OIB). The task of each panelist was to place a “bookmark” in the OIB at the threshold of each performance level. The bookmark differentiates item content that a student with just enough Spanish language proficiency to be performing at a defined performance level would likely know from item content that the student would not likely know. Two bookmarks were placed to identify the beginning of Level 2 and Level 3.

To make judgments and place bookmarks in the OIB, panelists reviewed each item in the OIB, in sequence, and considered 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 were instructed to use the item map. The relative differences in item difficulty on the item map allowed panelists to more easily evaluate the difference in difficulty between adjacent items in the OIB.

Each panelist placed the Level 2 bookmark on the first item encountered in the OIB that the panelist believed the “borderline Level 2 student” would most likely not be able to answer correctly. This placement indicated that the items beyond that point were too difficult for that borderline student. Each panelist, starting at the Level 2 bookmark, then continued with a review of each item and placed 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 item response theory (IRT) model used to order the items. Panelists were 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). 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., refer to Cizek, 2012, p. 135). Using RP67 for item ordering and instructing panelists to think about a two-thirds likelihood, which is easily understood, provided an alignment between the instructions and the analytical model.

Standard Setting Panels

A diverse group, representative of Spanish-language educators in California, was recruited to participate as panelists in the standard setting sessions. In recruiting panelists, the goal was to include a representative group of California educators who were 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 was important to include teachers working with these

Chapter 2: Method

students as those educators provided a perspective on learning goals for the students taking the CSA, as well as students' progress toward Spanish reading language arts proficiency.

For the CSA, there were four panels of educators (refer to [Table 1](#)). The targeted number of panelists was 15 per panel, for a total of 60 educators. Although 57 panelists completed the standard setting process, one panelist who served on the panel for grades seven and eight needed to leave due to a family emergency. She was not able to complete the demographic information on the final evaluation. Therefore, 56 panelists are represented in [Table 2](#) through [Table 6](#).

Table 1. Panel Configuration

Panel	Total
Grades 3–4	15
Grades 5–6	15
Grades 7–8	13
High school	14
Total	57

The CSA panels were assembled in test-specific panel rooms for much of the standard setting work. For three panels, the panelists worked on two tests, assessing adjacent grades, as shown in [Table 1](#). Educators teaching students in each of the relevant grade levels served on those panels. Educators teaching across the high school grade span served on the high school panel. This design supported consistency in the panelists' judgments as well as consistency in the definition of the threshold ALDs (described later in this plan). Panelists were seated at three tables, with five educators at each table. The composition of each panel included the following as criteria for selection:

- Educators who were teaching Spanish-language learners, in the grade level(s) assigned to the panel
- Educators who were teaching students who would take the CSA
- Educators who were familiar with the Common Core State Standards en Español

Final decisions regarding the panelists' selection were made by the CDE. After the panelists were approved and notified, travel arrangements were made. Panelists were required to sign a security agreement acknowledging the confidentiality of the materials used in the standard setting and prohibiting the removal of the materials from the meeting area.

For every panel, one member was identified as a table leader at each table, for a total of three per panel. The table leader's responsibility was to help keep discussions on track at the table, report interim discussions to the room, and collect materials at the table. Table leaders were advised of the role during the first day and joined the lead facilitator for table-leader training prior to bookmark judgments.

Because standard setting is based on expert judgment—informed by performance data—it is important that panelists collectively reflect the diversity of the educators working with students who take the assessment. Special efforts were made to assemble panels that were representative of the geographic and socioeconomic diversity of California in general, and the CSA educator population in particular. [Table 2](#) provides the distribution of the panel by gender; all panels included at least two male educators.

Table 2. Panelist Gender

Gender	Grades 3–4	Grades 5–6	Grades 7–8	High School	Total
Female	13	10	8	11	42
Male	2	5	4	3	14

[Table 3](#) provides the educators’ self-reported ethnic or racial background. The largest group represented was Hispanic (n = 40). All panelists responded to the question, “What is your primary ethnicity/race?”

Table 3. Panelist Primary Ethnicity/Race

Ethnicity/Race	Grades 3–4	Grades 5–6	Grades 7–8	High School	Total
American Indian/Alaska Native	0	0	0	1	1
Asian	0	0	0	0	0
Black or African American	0	1	0	0	1
Filipino	0	0	0	0	0
Hispanic or Latino	11	13	7	9	40
Native Hawaiian or Other Pacific Islander	0	0	0	0	0
White	1	1	2	1	5
Two or More Races	3	0	3	3	9

[Table 4](#) presents the regions in California in which the panelists were teaching. A majority of the educators indicated the southern region of California. One panelist from the Northern region of the state who served on the panel for the CSA grades seven and eight did not complete the final evaluation and is not included in the other tables describing the panels.

Table 4. Geographical Region of Panelists

Region	Grades 3–4	Grades 5–6	Grades 7–8	High School	Total
Northern	4	3	1	2	10
Central	2	0	2	0	4
Southern	9	12	10	12	43

[Table 5](#) presents the teaching experience in each panel and across the standard setting workshop by the number of years taught. A majority of the educators indicated having more than 10 years of experience teaching bilingual classes.

Table 5. Panelist Years Experience Teaching Bilingual Classes

Experience	Grades 3–4	Grades 5–6	Grades 7–8	High School	Total
0 years	0	2	1	3	6
1–3 years	2	2	1	3	8
4–6 years	1	2	1	2	6
7–10 years	3	2	0	1	6
10+ years	9	7	9	5	30

Each educator was asked what subject or subjects that educator currently teaches; multiple responses were permitted. [Table 6](#) shows that most panels included educators with experience teaching English, all subjects, and subjects other than the options listed.

Table 6. Panelist Subject(s) Currently Teaching

Subject	Grades 3–4	Grades 5–6	Grades 7–8	High School	Total
All Subjects	6	6	0	0	12
Mathematics	2	0	1	1	4
Science	3	1	1	1	6
Social Studies	4	1	2	0	7
English	3	3	7	3	16
Other	5	6	9	11	31

Chapter 3: Materials

Prior to the standard setting workshop, panel members were provided with a letter describing the purpose and procedures of the standard setting workshop, along with the following materials:

- A pre-workshop assignment specific to the individual educator’s panel assignment
- A notetaking form for the assignment
- A link to the training tests
- General and Range ALDs for the tests the panelists would be reviewing

At the standard setting workshop, panelists received training materials and a set of operational materials. Items were kept secure by assigning panelists an individual identification number and giving material marked with the same number. Each panelist was asked to sign a nondisclosure agreement, check the material out and in each day, and accept responsibility for controlling all documents labeled with the assigned identification number. The Sacramento County Office of Education and ETS staff monitored each room to ensure that materials remained in the rooms, and that no room was left unattended when unlocked.

The test forms reviewed during the standard setting were the operational test forms developed for the first operational test administration, which occurred in April 2019. The set of operational materials included the following:

- A printed version of the CSA and the answer key with scoring rules for 2-point items
- An OIB with corresponding passage/script book
- An item map
- A judgment recording form
- Training materials based on the online training test: A printed training test form, an OIB, a passage/script book, an item map
- A training evaluation form

Panelists developed borderline student definitions in the workshop; refer to [attachment B](#) in [appendix 1](#). The OIB, student responses, and item map are described more fully in the next subsections.

Ordered Item Booklet

For each item, the page of the OIB showed the item, the possible responses, and the correct answer or rubric score. For the items associated with a passage, a separate passage booklet was included allowing panelists to reference items associated with a passage.

The OIB contained one item per page. Items were ordered from least difficult to most difficult based on a response probability of 0.67 employed in the item response theory model (Mitzel, Lewis, Patz & Green, 2001). Items with a maximum score of 1 appeared once. Items with a maximum score of 2 appeared twice, representing each of the response probabilities of earning 1 point or 2 points. Scores of zero did not appear in the OIB.

Item Map

The item map is a summary document displaying relevant information regarding each item in the CSA (such as the key, the corresponding passage or script references, the maximum number of points, and a difficulty value). This information was presented in the same order as the order in the corresponding OIB—from least to most difficult, based on a response probability of 0.67. The item map also listed the original order from the test form as well as relative item-difficulty information on a temporary scale, called the standard setting scale (SS Scale). The SS Scale looked similar to the reporting scale but was only to be used during standard setting to aid panelists' understanding of the relative difficulty of the items in the OIB. Items on the CSA included 1-point and 2-point items; item scores were indicated on the item map. Two-point items appeared twice in the OIB and item map. A score of 1+ represented a score of 1 on a 2-point item, and a score of 2+ represented a score of 2 on a 2-point item. Refer to [attachment C](#) in [appendix 1](#) for a sample item map.

Evaluation Forms

It was important to collect information from the panelists to document procedural validity (Cizek & Bunch, 2007; Hambleton & Pitoniak, 2006). Panelists received evaluation forms at two points during the process to gauge the panelists' understanding and gather other information ([attachment D](#) in [appendix 1](#) for the evaluation forms). Evaluations included questions about the following:

- Training
- Understanding the tasks
- The influence of different aspects of the standard setting process
- Panelists' beliefs about the final recommended threshold scores

Because ETS was interested in knowing, as soon as possible, if panelists were not satisfied with the level of training received, the evaluation form was given to the panelists at the end of the training to gauge the panelists' current understanding of the process and comfort level with the tasks to be performed. The responses on the evaluation forms were analyzed immediately by the panel facilitators so any tasks or materials that appeared to be unclear could be reviewed. In all four panels, no additional review and discussion was needed; panelists indicated comfort with the process and a readiness to proceed. An overview of the results obtained through the evaluation forms is included in the [Results](#) section of this report.

Chapter 4: Process

This section of the report describes what occurred prior to and during the standard setting workshop.

The workshop process began with a welcome provided by the CDE and a general overview and initial training provided by the ETS standard setting director. At the conclusion of the general session, panel facilitators experienced in working with educators in standard setting provided in-depth training and practice on the method in each panel room.

Panelists then completed the following steps:

- Worked in grade- or grade-span-specific panel rooms
- Took the test for which standards were being set (refer to the [Test Familiarization](#) subsection)
- Defined the borderline students (refer to the [Borderline Student Definitions](#) subsection)
- Received training (including practice)
- Placed bookmarks to indicate threshold scores in an OIB over the course of three rounds of judgments. Feedback and discussion took place after each round of judgment (refer to the [Feedback and Discussion](#) subsection)

The final round of judgments was made on the overall score. Panelists reviewed the individual judgments provided by the group, discussed the range of judgments in the panel, and then reviewed the impact data (i.e., the percent of students who would be classified into each of four performance levels based on Round 2 judgments and the overall score). After feedback and discussion, including the impact data, panelists made Round 3 judgments for final recommended threshold scores. At the conclusion of the workshop, the results were shared with the panelists in each panel and with the CDE. During the articulation workshop, a subset of panelists assembled to consider cross-grade articulation (grade three through the high school grade span) of the threshold scores. While reviewing the borderline student definitions used in standard setting, the panel discussed the rationales and judgments in each panel and the Round 3 recommendations across the five grades and the high school grade span.

Pre-meeting Preparation

Prior to the standard setting, a pre-workshop assignment, consisting of two parts, was given to the panelists approximately two weeks before the in-person workshop. For the first part, the panelists were provided with a link to the CSA training test on the CDE website and asked to take the training test for the grade level (or high school grade span) the panelists were scheduled to work with first. This allowed the panelists to become familiar with the task types on the CSA; the same training test was used during the practice round of judgments (refer to the [Bookmark Judgment Training](#) subsection). The second part of the assignment included familiarization with the range ALDs and access to a link to the Common Core State Standards en Español and the general ALDs. Panelists were 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 Level 2 and Level 3, and bring those notes to the

standard setting workshop. The pre-workshop assignment is shown in [attachment A](#) in [appendix 1](#).

General Session Training

Panelists were trained in various aspects of the process throughout the course of the workshop; training was often followed immediately by doing the task addressed in the training. On the first day, a general orientation session was held for the entire group where the need for threshold scores was explained. Because this was the first administration of the CSA, panelists were invited to ask questions during the general session, and staff from the CDE and ETS were available throughout the process to answer questions about the test, the policies surrounding the test, and the standard setting procedures.

Dr. Wanda Swiggett, ETS Standard Setting Director, introduced the Bookmark method for setting threshold scores and presented the agenda and expectations for panel members' participation. Dr. Swiggett then continued the general session with initial training on the method, after which panelists moved into grade-specific groups, where the panel facilitators continued with training and guided the panelists through the rest of the standard setting activities.

Test Familiarization

For the CSA, there were four panels (refer to [Table 1](#)). Each panel began the test familiarization with one of the two tests assigned to that panel (i.e., the CSA test for grades four, five, or eight, respectively). The high school panel only needed to become familiar with the one CSA test for high school. During the test-familiarization process, an ETS assessment development (AD) specialist presented the computer-administered assessment to the room. The panelists, using paper forms, followed along and recorded responses to the items. Audio prompts were played to the entire room and passages were slowly scrolled through for the entire room to read independently.

After reviewing the test, panelists checked the individual responses against the answer key. The panelists then discussed the content measured, what they thought might be particularly challenging for students, and what might be less difficult. The goal of this activity was for panelists to begin to think about and articulate the perceived general difficulty of the tested content for students.

Borderline Student Definitions

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 worked in small groups to define borderline students. The process to arrive at borderline student definitions involved 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 was expected. For the CSA, two definitions were needed for two thresholds—the Level 2 borderline and Level 3 borderline student definitions. Panels worked first on the Level 3 borderline, because this is the point at which students are 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 referred to the range ALDs that describe the full range for each of the three levels. After the facilitator familiarized the panel with the task, panelists worked as a whole group to describe one aspect of the Level 3 borderline student. By working on one bullet as a whole group, the process was modeled and the facilitator provided guidance to ensure that the focus was on the differentiation between Level 2 and Level 3 ALDs. The panel then worked in two small groups to complete the borderline student definitions for Level 3.

In small groups, the panelists started by utilizing notes from the pre-workshop assignments, the group's knowledge of the standards and the ALDs for the panel's assigned grade, and the panelists' knowledge of the students who were administered the CSA. Panelists then began describing the skills and knowledge required of a borderline student based on an understanding of what the test is assessing. This work was done first at the table level, where panelists listed the major components that defined the borderline Level 3 student. One panelist at each table served as scribe for the table. The next step was to share the descriptions from each table; a full-room discussion occurred to reach consensus on definitions for each borderline student definition.

The facilitators instructed panelists to limit the definitions of their borderline students to a brief description that would allow decisions to be made about the student's performance on the test. It was also pointed out to the panelists that the documents provided were for use during the process; perfect language was not necessary. Rather, the goal was to capture the essence of the knowledge and skills of each borderline student. Each room reached agreement on the descriptions of the borderline students. The descriptions were used by the panels as working definitions in the standard setting process. The borderline student definitions are provided in [attachment B](#) in [appendix 1](#).

After the "borderline Level 3 student" definition was drafted, two pairs of two panels working on adjacent CSA grade-level or grade-span tests met to discuss the drafts, provide feedback to each other, and finalize the definitions. These discussions and this work focused 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 reconvened in the panel rooms to complete the "borderline Level 2 student" definition, working to ensure its consistency with the Level 3 definition. [Table 7](#) illustrates the manner in which cross-grade consistency was considered during the workshop.

Each panel, with the exception of the high school panel, completed the standard setting process on two CSA grade-level tests. After completing the process for the first grade level of the CSA, the panel began the entire process again with the second assessment. The grades five and six panel and the grades seven and eight panel met again to consider cross-grade

consistency when creating the borderline student definitions for their second CSA (for grades six and seven, respectively), as illustrated in [Table 7](#).

Table 7. Cross-Grade Consistency by Panel

Borderline Student Definition Process	Grades three through four	Grades five through six	Grades seven through eight	High School
Draft Level 3 (within-panel)	Grade four definition	Grade five definition	Grade eight definition	High school definition
Complete Level 3 (across-panels)	Grade four consistent with grade five	Grade five consistent with grade four	Grade eight consistent with high school	High school, consistent with grade eight
Draft, then complete Level 2 (within-panel)	Grade four Level 2, consistent with grade four, Level 3	Grade five Level 2, consistent with grade five, Level 3	Grade eight Level 2, consistent with grade eight, Level 3	High school, Level 2, consistent with high school, Level 3
Draft Level 3 (within-panel)	Grade three consistent with grade four	Grade six consistent with grade five	Grade seven consistent with grade eight	NA
Complete Level 3 (across-panels)	NA	Grade six consistent with grade seven	Grade seven consistent with grade six	NA
Draft, then complete Level 2 (within-panel)	Grade three, Level 2, consistent with grade three, Level 3	Grade six, Level 2, consistent with grade six, Level 3	Grade seven, Level 2, consistent with grade seven, Level 3	NA

Table Leader Training

ETS trained the table leaders in a half-hour session during breakfast on Day 2 of each week. The training began with a description of a table leader’s role and responsibilities. The table leaders received instruction on the following tasks:

- Helping to control secure materials
- Notifying the facilitator of any difficulties during discussions
- Coordinating the review of the OIB
- Collecting and checking all rating forms for completeness and accuracy
- Taking notes and presenting a summary of the Round 1 table discussion
- Sharing feedback data with the panelists at the table

Bookmark Judgment Training

Panelists were trained in the Bookmark method, practiced making judgments, and discussed rationales for judgments, prior to the start of actual standard setting, as described next. After training, panelists completed a training evaluation form confirming the panelists' understanding and readiness to proceed.

Panelists were provided the opportunity to practice on items for the Bookmark standard setting method prior to the start of the actual standard setting. As part of the training, the facilitator asked the panelists to discuss the rationales behind the judgments. The facilitator guided this instructional discussion and provided clarity on the procedure as needed. After practice for each method, each panelist was asked to complete a training evaluation form indicating the extent to which the training in the procedure and materials had been clear and whether or not the panelist was ready to proceed. The evaluation forms were reviewed, and any retraining needs were addressed. Panelists placed the Round 1 operational judgments independently.

The first activity for each panelist was to read independently the consensus borderline student definitions. Panelists then reviewed the items in the OIB and discussed with others at the table what makes each item more difficult than the previous item in the OIB, taking into consideration all aspects of the item (e.g., prompt and options). During this review, panelists were instructed to answer and discuss two questions:

1. What do these items measure?
2. What makes this item more difficult than the previous item?

The table leaders facilitated this discussion; panel facilitators monitored each table. The purpose of this exercise was for the panelists to gain a common understanding about the knowledge and skills assessed by these items. This stage is considered essential to placing the first round of bookmarks. At this point, however, panelists were cautioned not to discuss the placement of the bookmark, but instead, to focus only on comparisons of the content of the items.

Panelists were then asked to practice placing a bookmark using the borderline Level 2 student description, placing only the first bookmark. Because this was “practice,” participants were told to place a bookmark for the Level 2 threshold score only. Panelists were reminded to place a bookmark on the first item the borderline student was *not* likely to answer correctly. Note that “not likely” was conceived of in terms of the “two-thirds rule” described previously.

Facilitators provided instructions on the use of the item map, which provides student performance data about item difficulty. The item map included a column called “Standard-Setting Scale,” which indicated the difference in difficulty across items. The relative differences in item difficulty on the item map allowed panelists to more easily evaluate the difference in difficulty between adjacent items in the OIB. Facilitators also reminded panelists that gaps in the relative item difficulty did not represent where bookmarks should be placed and that the training should be followed when making judgments are made. A sample item map is provided in [attachment C](#) in [appendix 1](#).

Facilitators were available during the practice task to answer questions. When the panelists were comfortable with the process, practice materials were returned and the training forms completed.

Feedback and Discussion

The purpose of feedback and discussion is to allow panelists to hear the rationales of the other panelists, to both receive empirical information about item performance and student performance and to arrive at a mutual understanding of the expectations of the borderline students on this test. The process of judgment, feedback, and discussion is repeated over the four-day period until all threshold scores are collected.

Feedback was given to the panelists after Round 1 judgments were collected and summarized. The table-level feedback provided an opportunity for the panelists to discuss in a small-group setting the range of judgments and rationales for why individual judgments were made. The panelists were provided with the median and range of the panel judgments for the table, and then discussed the judgments in table-level groups. As part of the discussion of individual rationales, panelists were 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 allowed panelists to get a better sense of how significant the differences were among the judgments, thereby aiding discussion. Table leaders took notes on the themes of the table-level discussion as the panelists discussed the items from the lowest to highest OIB item judgments. Then panelists made independent Round 2 judgments, which were collected and summarized.

After making Round 2 independent judgments, but before receiving feedback on those judgments, the table leaders provided a report to the room of the table's post-Round 1 table-level discussions. After all questions had been answered, the tables commenced Round 2 table-level feedback. Once the table-level discussions had concluded, the room received room-level feedback from Round 2 judgments. Results were 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 discussed the data, the student performance data showing the impact, or consequence data, of the Round 2 judgments was presented. This impact data was based on spring 2018–19 CSA student performance, and the feedback showed “what percentage of students would fall into each level based on Round 2 judgments.” After the room-level discussions, panelists were asked to undertake table-level discussions again.

Once all discussions had concluded, panelists were asked to make a final round of judgments. The results from the Round 3 judgments were considered the final threshold score recommendations from the standard setting panel. Panelists reviewed Round 3 feedback and responded to a final evaluation form, which was confidential.

Cross-Grade Articulation for the Overall Score

The last step in the standard-setting workshop involved a subset of panelists from the four panel rooms—three representatives from each panel were recruited to attend the cross-grade articulation meeting. The cross-grade articulation meeting was held on August 9, 2019.

The goal of this meeting was to ask panelists to consider the score recommendations by considering feedback and data across the seven sets of threshold score recommendations. Panelists were provided with the borderline student definitions across all panels. The panelists were asked to review the definitions for the assigned grade levels or grade span,

along with the two adjacent grade levels or grade span. For example, the educators from panel B (for grades five and six) reviewed the definitions for grades three through eight. (During the standard setting, these panelists had already reviewed the Level 3 definitions for grades four and seven.)

The panel facilitator asked the panelists to share the rationales and the discussions that occurred in each panel. Panelists next reviewed the impact data for all seven sets of threshold scores (refer to [appendix 2](#)). The panel facilitator and the CDE psychometric staff were available to provide further explanation about data and clarification on the limitations of the data.

Chapter 5: Results

This section describes the results from the workshop, which include the rounds of bookmark judgments, the student impact data, and an evaluation of the process based on questionnaires completed by the panelists.

Data for each grade and the high school grade span is presented in this section, in [appendix 2](#), and in [appendix 3](#). Six types of tables are presented; a general description of the six types follows:

Six Types of Data Tables

1. **Median bookmark judgments, by round.** Median threshold scores for each round of judgment are presented in the metric of panelists' judgments, i.e., its location in the OIB. The range of bookmark values is from one to the number of possible points in the test; the test includes one-point and two-point items.
2. **Standard Error of Judgment (SEJ), by round.** SEJs for each round of judgment are presented in the metric of panelists' judgments, OIB locations.
3. **Projected distribution of 2018–19 students based on Round 2 median threshold scores.** This data represents the impact data that was presented to the panelists after Round 2. Refer to [Table 22](#) through [Table 28](#).
4. **Projected distribution of 2018–19 students based on Round 3 median threshold scores.** This data represents the impact data that was presented to the panelists after their final round of judgments. Refer to [Table 29](#) through [Table 35](#).
5. **Projected percentage of 2018–19 students at or above the recommended threshold score, +/-1 conditional standard errors of measurement (CSEM), and +/-2 CSEM.** Refer to [appendix 2](#).
6. **Cross-grade articulation data.** Table and figure of the impact data based on Round 3 median threshold judgments at each achievement level across grades and the high school grade span. Refer to [appendix 3](#).

Impact Data

The impact data provided in this report is based on the scores of students who took the CSA during the first operational administration in 2018–19. Impact data is derived using item response theory, which provides a scale score equivalent to the theta values associated with the recommended threshold scores.

The SBE reviewed both the panel recommendations and the State Superintendent of Public Instruction's (SSPI's) recommendations and approved reporting of three score ranges for CSA. Following SBE approval, a reporting scale was developed. The scaling process provides a reported scale for Student Score Reports only; the scaling process is not part of standard setting, and the details are not provided in this report. In the tables that follow, the recommended scale-score threshold scores and CSEMs are based on the working scale, or SS Scale equivalent of the median Round 3 threshold scores, as described previously.

CSA Recommended Score Results

[Table 8](#) through [Table 14](#) display the median bookmark threshold scores after each round for each grade and the high school grade span. The median was calculated for each table and for the room. The tables show how panelists moved the bookmarks across rounds. Lower numbers represent bookmark placements earlier in the OIB, indicating a threshold score on a less difficult item that translates to a lower threshold score. Higher numbers translate to a higher threshold score; a higher threshold score means that more is required for a student to be included in the level.

In general, the threshold scores changed by only one or two bookmarks from one round to the next, especially between Round 2 and Round 3. The most notable difference is in the high school panel's judgments. The scores rose and lowered across the rounds by multiple bookmarks.

[Table 8](#) shows the bookmark threshold scores for grade three.

Table 8. Median Bookmarks at the End of Each Round: Grade Three

Level	Round 1	Round 2	Round 3
Level 2	15	15	15
Level 3	35	34	35

[Table 9](#) shows the bookmark threshold scores for grade four.

Table 9. Median Bookmarks at the End of Each Round: Grade Four

Level	Round 1	Round 2	Round 3
Level 2	19	22	20
Level 3	41	38	38

[Table 10](#) shows the bookmark threshold scores for grade five.

Table 10. Median Bookmarks at the End of Each Round: Grade Five

Level	Round 1	Round 2	Round 3
Level 2	15	17	17
Level 3	39	36	36

[Table 11](#) shows the bookmark threshold scores for grade six.

Table 11. Median Bookmarks at the End of Each Round: Grade Six

Level	Round 1	Round 2	Round 3
Level 2	17	20	20
Level 3	37	37	39

[Table 12](#) shows the bookmark threshold scores for grade seven.

Table 12. Median Bookmarks at the End of Each Round: Grade Seven

Level	Round 1	Round 2	Round 3
Level 2	18.5	18	18
Level 3	44	47	46

[Table 13](#) shows the bookmark threshold scores for grade eight.

Table 13. Median Bookmarks at the End of Each Round: Grade Eight

Level	Round 1	Round 2	Round 3
Level 2	18	22	22
Level 3	38	44	44

[Table 14](#) shows the bookmark threshold scores for the high school grade span.

Table 14. Median Bookmarks at the End of Each Round: High School Grade Span

Level	Round 1	Round 2	Round 3
Level 2	17.5	23.5	19
Level 3	38	43	40

[Table 15](#) through [Table 21](#) represent the room SEJs after each round for each grade and the high school grade span. The SEJs are provided in the bookmark, or OIB, metric. The SEJ is calculated by multiplying 1.25 by the Round 3 standard error of the mean, which is a research-based estimate of the standard error of the median (refer to, for example, MacCann & Stanley, 2004) and is one way of estimating the reliability of the judgments. The SEJ indicates how close the threshold score is likely to be to the score of other panels of educators similar in composition to the current panel and similarly trained in the same standard setting method. A comparable panel's threshold score would be within one SEJ of the current threshold score 68 percent of the time and within two SEJs 96 percent of the time.¹

Lower numbers from Round 1 to Round 2 indicate the convergence of panelists' judgments across tables over rounds during the bookmark process. Ideally, the SEJ should decrease across rounds; although, occasionally, the introduction of impact data results in the SEJ increasing, as panelists have different reactions to the normative data.

The highest SEJ occurred during the Round 1, Level 3 judgments for most panels, except for the grade three CSA and the grade seven CSA. However, for most panels, the Level 3 SEJ was at its lowest by Round 3. For each grade, the Level 2 SEJ was at its lowest by Round 3,

¹ Probabilities assume normality of the sample; sampling theory suggests that, as the size of the group increases, the distribution gets closer to normal.

with the exception of the high school CSA; the Round 3 SEJ remained the same as the Round 2 SEJ.

[Table 15](#) provides the room SEJs for grade three.

Table 15. SEJs in Bookmark Placement by Round: Grade Three

Level	Round 1	Round 2	Round 3
Level 2	1.5	1.7	1.3
Level 3	1.2	1.0	1.0

[Table 16](#) provides similar information for grade four.

Table 16. SEJs in Bookmark Placement by Round: Grade Four

Level	Round 1	Round 2	Round 3
Level 2	2.8	2.3	1.1
Level 3	4.3	1.5	1.2

[Table 17](#) provides similar information for grade five.

Table 17. SEJs in Bookmark Placement by Round: Grade Five

Level	Round 1	Round 2	Round 3
Level 2	1.8	1.4	1.2
Level 3	3.2	2.0	1.9

[Table 18](#) provides similar information for grade six.

Table 18. SEJs in Bookmark Placement by Round: Grade Six

Level	Round 1	Round 2	Round 3
Level 2	1.6	1.5	1.1
Level 3	3.2	2.9	2.4

[Table 19](#) provides similar information for grade seven.

Table 19. SEJs in Bookmark Placements by Round: Grade Seven

Level	Round 1	Round 2	Round 3
Level 2	2.9	1.9	1.0
Level 3	2.1	1.2	1.0

[Table 20](#) provides similar information for grade eight.

Table 20. SEJs in Bookmark Placements by Round: Grade Eight

Level	Round 1	Round 2	Round 3
Level 2	2.0	1.9	1.1
Level 3	4.1	3.5	2.5

[Table 21](#) provides similar information for high school.

Table 21. SEJs in Bookmark Placements by Round: High School Grade Span

Level	Round 1	Round 2	Round 3
Level 2	1.8	1.5	1.5
Level 3	2.9	1.4	2.0

Panelists viewed the scores associated with the impact data as bookmark judgments. The impact data in [Table 22](#) through [Table 28](#) show the percent of students who would be placed in each performance level based on the Round 2 median threshold score recommendations shown as scale scores. These impact data are based on the 2018–19 operational administration of the CSA data and may differ from the percent of students in each level in future test administrations. These impact data were presented to the panels as feedback to consider when making Round 3 judgments.

[Table 22](#) provides the Round 2 impact data for grade three.

Table 22. Distribution of 2018–19 Students Based on Round 2 Recommendations: Grade Three

Performance Level	Threshold Overall Scale Score	Percentage
Level 1	-	63.4
Level 2	404	27.7
Level 3	416	8.9

[Table 23](#) provides similar data for grade four.

Table 23. Distribution of 2018–19 Students Based on Round 2 Recommendations: Grade Four

Performance Level	Threshold Overall Scale Score	Percentage
Level 1	-	66.6
Level 2	404.5	22.3
Level 3	416	11.1

[Table 24](#) provides similar data for grade five.

Table 24. Distribution of 2018–19 Students Based on Round 2 Recommendations: Grade Five

Performance Level	Threshold Overall Scale Score	Percentage
Level 1	-	56.9
Level 2	402	35.6
Level 3	417	7.5

[Table 25](#) six provides similar data for grade six.

Table 25. Distribution of 2018–19 Students Based on Round 2 Recommendations: Grade Six

Performance Level	Threshold Overall Scale Score	Percentage
Level 1	-	52.8
Level 2	400.5	34.8
Level 3	413	12.4

Chapter 5: Results

[Table 26](#) provides similar data for grade seven.

Table 26. Distribution of 2018–19 Students Based on Round 2 Recommendations: Grade Seven

Performance Level	Threshold Overall Scale Score	Percentage
Level 1	-	68.9
Level 2	405	29.2
Level 3	421	1.9

[Table 27](#) provides similar data for grade eight.

Table 27. Distribution of 2018–19 Students Based on Round 2 Recommendations: Grade Eight

Performance Level	Threshold Overall Scale Score	Percentage
Level 1	-	68.8
Level 2	405	26.7
Level 3	418	4.5

[Table 28](#) provides similar data for high school.

Table 28. Distribution of 2018–19 Students Based on Round 2 Recommendations: High School Grade Span

Performance Level	Threshold Overall Scale Score	Percentage
Level 1	-	83.4
Level 2	410	13.4
Level 3	419	3.2

[Table 29](#) through [Table 35](#) present Round 3 results, based on consideration of the median recommended threshold scores applied to the 2018–19 operational performance data. This impact data was presented to each panel for each grade or grade span. The panels used this final recommendation when responding to the last questions in the final evaluation form.

[Table 29](#) provides the Round 3 impact data for grade three.

Table 29. Distribution of 2018–19 Students Based on Round 3 Recommendations: Grade Three

Performance Level	Threshold Overall Scale Score	Percentage
Level 1	-	63.4
Level 2	404	27.7
Level 3	416.5	8.9

[Table 30](#) provides the Round 3 impact data for grade four.

Table 30. Distribution of 2018–19 Students Based on Round 3 Recommendations: Grade Four

Performance Level	Threshold Overall Scale Score	Percentage
Level 1	-	63.9
Level 2	403.5	25.0
Level 3	416	11.1

[Table 31](#) provides the Round 3 impact data for grade five.

Table 31. Distribution of 2018–19 Students Based on Round 3 Recommendations: Grade Five

Performance Level	Threshold Overall Scale Score	Percentage
Level 1	-	56.9
Level 2	402	35.6
Level 3	417	7.5

Chapter 5: Results

[Table 32](#) provides the Round 3 impact data for grade six.

Table 32. Distribution of 2018–19 Students Based on Round 3 Recommendations: Grade Six

Performance Level	Threshold Overall Scale Score	Percentage
Level 1	-	52.8
Level 2	400.5	37.0
Level 3	414	10.2

[Table 33](#) provides the Round 3 impact data for grade seven.

Table 33. Distribution of 2018–19 Students Based on Round 3 Recommendations: Grade Seven

Performance Level	Threshold Overall Scale Score	Percentage
Level 1	-	68.9
Level 2	405	29.2
Level 3	420.5	1.9

[Table 34](#) provides the Round 3 impact data for grade eight.

Table 34. Distribution of 2018–19 Students Based on Round 3 Recommendations: Grade Eight

Performance Level	Threshold Overall Scale Score	Percentage
Level 1	-	68.8
Level 2	405	26.7
Level 3	418	4.5

[Table 35](#) provides the Round 3 impact data for high school.

Table 35. Distribution of 2018–19 Students Based on Round 3 Recommendations: High School Grade Span

Performance Level	Threshold Overall Scale Score	Percentage
Level 1	.	70.7
Level 2	406	25.1
Level 3	417	4.2

The data displayed in [appendix 2](#) present the final-round threshold-score recommendations converted to rounded scale scores and the CSEM at each recommended threshold score. Every test has measurement error, and the CSEM presents the error surrounding one particular score, the recommended threshold score. The CSEM is a way to take into consideration the reliability of test scores. More specifically, this statistic is an indication of the degree of uncertainty at each scale score and is sometimes used as guidance when evaluating the appropriateness of threshold scores. The CSEM is the same across both levels for each CSA grade level and for Level 2 of the high school grade span. For Level 3 of the high school grade span, the CSEM is slightly higher.

Incorporating Additional Considerations in Setting Threshold Scores

In standard setting, policymakers sometimes wish to reduce the number of examinees who fall below the panel-recommended threshold scores due to random error. In addition to measurement error metrics (e.g., SEM, SEJ), policymakers should consider the likelihood of classification error; that is, when adjusting a threshold score, policymakers should consider whether it is more important to minimize a false-positive decision or to minimize a false-negative decision.

A false-positive decision occurs when a test taker’s score suggests one level of knowledge and skills, but the student’s actual level is lower (i.e., the student does not possess the required skills). A false-negative decision occurs when a test taker’s score suggests that the student does not possess the required skills, but that student nevertheless actually does possess those skills.

In order to reduce the number of false negatives, policymakers will decide to lower the threshold score(s). On the other hand, policymakers may desire to reduce the number of test takers who attain a score above the recommended threshold score because of random error at each level in order to reduce the number of false positives and thus raise the threshold score(s).

Raising threshold scores reduces false positives but increases false negatives; the reverse occurs when threshold scores are lowered. Policymakers need to consider which decision error to minimize; it is not possible to eliminate both types of decision errors simultaneously.

Evaluation of the Standard Setting Process

Each panelist was asked at two points over the course of the workshop to rate the following:

1. The panelist's understanding of the process
2. The usefulness of different training exercises
3. The influence of various factors on the bookmark placements

Panelists' ratings were collected using evaluation forms. The purpose of the first evaluation form, completed prior to the placement of bookmarks was to provide an early check on the level of panelists' understanding and to identify any areas of confusion. Assessing the level of clarity prior to beginning the judgment process is essential to validating the overall standard setting process. The final evaluation form contained additional questions used to analyze the whole process, including the following:

- The training
- The usefulness of materials and procedures
- The influence of ALDs and borderline student definitions
- Individual and group perceptions
- The student performance data
- The discussion

Results from the evaluation forms are panel-based and are specific to each panel. Any comparisons across panels should acknowledge the independence of the panels.

Evaluation Results from the CSA Standard Setting Final Evaluations

Although 57 panelists completed the standard setting process (refer to [Table 1](#)), one panelist, who served on the panel for grades seven and eight, needed to leave due to a family emergency. She was not able to complete the demographic information on the final evaluation. Therefore, 56 panelists are represented in [Table 36](#) through [Table 55](#), which provide the results of final evaluations. The evaluation forms are in [appendix 1: attachment D](#). The results provide information about panelists' thoughts as to the usefulness and influence of the materials and aspects of the four-day process. They also detail the panelists' stated beliefs as to the appropriateness of the threshold-score recommendations and whether the panelists individually support them.

[Table 36](#) provides the evaluation results on the usefulness of materials for grades three through four.

Table 36. Final Evaluation Grades Three Through Four on the Usefulness of the Materials

How <i>useful</i> was each of the following materials or procedures in completing the standard setting process?	Not at All Useful N	Not at All Useful %	Somewhat Useful N	Somewhat Useful %	Very Useful N	Very Useful %
Completing the pre-work assignment	0	0	6	40	9	60
Taking the test before making judgments	0	0	0	0	15	100
Defining the borderline students	0	0	0	0	15	100
Reviewing the organization of the ordered item booklet	0	0	1	7	14	93
Information in the item map	1	7	2	13	12	80
Practicing the procedure	0	0	1	7	14	93
Group discussions	0	0	2	13	13	87
Impact information (percent of students in each achievement level)	0	0	5	33	10	67

[Table 37](#) provides the evaluation results on the influence of the process components for grades three through four.

Table 37. Final Evaluation Grades Three Through Four on the Influence of the Process Components

How <i>influential</i> was each of the following in making your judgments?	Not at All Influential N	Not at All Influential %	Somewhat Influential N	Somewhat Influential %	Very Influential N	Very Influential %
Achievement level descriptors	0	0	1	7	14	93
Borderline student definitions	0	0	1	7	14	93
My perception of the difficulty of the items and tasks	0	0	7	47	8	53
My experience with the students	0	0	5	33	10	67

Table 37 (continuation)

How <i>influential</i> was each of the following in making your judgments?	Not at All Influential N	Not at All Influential %	Somewhat Influential N	Somewhat Influential %	Very Influential N	Very Influential %
Group discussions	0	0	7	47	8	53
Judgments and rationales of other panelists	0	0	4	27	11	73
Percent of students in each achievement level	1	7	5	33	9	60
My sense of what students need to know in each achievement level	0	0	2	13	13	87

[Table 38](#) provides the final evaluation on timing for grades three through four.

Table 38. Final Evaluation Grades Three Through Four on Timing

How appropriate was the <i>amount of time</i> you were given to complete the different components of the process?	Too Little Time N	Too Little Time %	About Right N	About Right %	Too Much Time N	Too Much Time %
Training in the Bookmark procedure	1	7	11	73	3	20
Test familiarization	0	0	11	73	4	27
Reviewing the ordered item booklet	0	0	13	87	2	13
Placing bookmark judgments	0	0	11	73	4	27
Group discussion	3	20	9	60	3	20

[Table 39](#) provides the final evaluation for grades three through four on the appropriateness of the final recommendations.

Table 39. Final Evaluation Grades Three Through Four on the Appropriateness of the Final Recommendations

	Too Low N	Too Low %	About Right N	About Right %	Too High N	Too High %
Do you believe that the final recommended threshold score for entering each of the performance levels is too low, about right, or too high?						
Level 2	1	7	12	80	2	13
Level 3	0	0	13	87	2	13

[Table 40](#) provides the final evaluation for grades three through four on support of the recommendations.

Table 40. Final Evaluation Grades Three Through Four on Support of the Recommendations

Do you support the final recommendations of the panel?	Yes N	Yes %	No N	No %
Do you support the final recommendations of the panel?	15	100	0	0

[Table 41](#) provides the evaluation results on the usefulness of materials for grades five through six.

Table 41. Final Evaluation Grades Five Through Six on the Usefulness of the Materials

	Not at All Useful N	Not at All Useful %	Somewhat Useful N	Somewhat Useful %	Very Useful N	Very Useful %
How <i>useful</i> was each of the following materials or procedures in completing the standard setting process?						
Completing the pre-work assignment	1	7	9	60	5	33
Taking the test before making judgments	0	0	3	20	12	80
Defining the borderline students	0	0	4	27	11	73
Reviewing the organization of the ordered item booklet	0	0	3	20	12	80
Information in the item map	0	0	3	20	12	80

Table 41 (continuation)

How <i>useful</i> was each of the following materials or procedures in completing the standard setting process?	Not at All Useful <i>N</i>	Not at All Useful %	Somewhat Useful <i>N</i>	Somewhat Useful %	Very Useful <i>N</i>	Very Useful %
Practicing the procedure	0	0	5	33	10	67
Group discussions	1	7	2	13	12	80
Impact information (percent of students in each achievement level)	0	0	3	20	12	80

[Table 42](#) provides the evaluation results on the influence of the process components for grades five through six.

Table 42. Final Evaluation Grades Five Through Six on the Influence of the Process Components

How <i>influential</i> was each of the following in making your judgments?	Not at All Influential <i>N</i>	Not at All Influential %	Somewhat Influential <i>N</i>	Somewhat Influential %	Very Influential <i>N</i>	Very Influential %
Achievement level descriptors	0	0	3	20	12	80
Borderline student definitions	0	0	6	40	9	60
My perception of the difficulty of the items and tasks	0	0	12	80	3	20
My experience with the students	0	0	8	53	7	47
Group discussions	1	7	7	47	7	47
Judgments and rationales of other panelists	0	0	10	67	5	33
Percent of students in each achievement level	4	27	7	47	4	27
My sense of what students need to know to be in each achievement level	0	0	7	47	8	53

[Table 43](#) provides the final evaluation on timing for grades five through six.

Table 43. Final Evaluation Grades Five Through Six on Timing

How appropriate was the <i>amount of time</i> you were given to complete the different components of the process?	Too Little Time N	Too Little Time %	About Right N	About Right %	Too Much Time N	Too Much Time %
Training in the Bookmark procedure	0	0	13	87	2	13
Test familiarization	2	13	11	73	2	13
Reviewing the ordered item booklet	0	0	14	93	1	7
Placing bookmark judgments	0	0	12	80	3	20
Group discussion	0	0	11	73	4	27

[Table 44](#) provides the final evaluation for grades five through six on the appropriateness of the final recommendations.

Table 44. Final Evaluation Grades Five Through Six on the Appropriateness of the Final Recommendations

Do you believe that the final recommended threshold score for entering each of the performance levels is too low, about right, or too high?	Too Low N	Too Low %	About Right N	About Right %	Too High N	Too High %
Level 2	0	0	15	100	0	0
Level 3	2	13	13	87	0	0

[Table 45](#) provides the final evaluation for grades five through six on support of the recommendations.

Table 45. Final Evaluation Grades Five Through Six on Support of the Recommendations

Do you support the final recommendations of the panel?	Yes N	Yes %	No N	No %
Do you support the final recommendations of the panel?	15	100	0	0

Chapter 5: Results

[Table 46](#) provides the evaluation results on the usefulness of materials for grades seven through eight.

Table 46. Final Evaluation Grades Seven Through Eight on the Usefulness of the Materials

	Not at All Useful <i>N</i>	Not at All Useful %	Somewhat Useful <i>N</i>	Somewhat Useful %	Very Useful <i>N</i>	Very Useful %
How <i>useful</i> was each of the following materials or procedures in completing the standard setting process?						
Completing the pre-work assignment	0	0	7	58	5	42
Taking the test before making judgments	0	0	2	17	10	83
Defining the borderline students	0	0	2	17	10	83
Reviewing the organization of the ordered item booklet	0	0	0	0	12	100
Information in the item map	0	0	2	17	10	83
Practicing the procedure	0	0	2	17	10	83
Group discussions	0	0	0	0	12	100
Impact information (percent of students in each achievement level)	0	0	2	17	10	83

[Table 47](#) provides the evaluation results on the influence of the process components for grades seven through eight.

Table 47. Final Evaluation Grades Seven Through Eight on the Influence of the Process Components

	Not at All Influential <i>N</i>	Not at All Influential %	Somewhat Influential <i>N</i>	Somewhat Influential %	Very Influential <i>N</i>	Very Influential %
How <i>influential</i> was each of the following in making your judgments?						
Achievement level descriptors	1	8	2	17	9	75
Borderline student definitions	0	0	1	8	11	92
My perception of the difficulty of the items and tasks	0	0	2	17	10	83

Table 47 (continuation)

How <i>influential</i> was each of the following in making your judgments?	Not at All Influential N	Not at All Influential %	Somewhat Influential N	Somewhat Influential %	Very Influential N	Very Influential %
My experience with the students	1	8	2	17	9	75
Group discussions	0	0	4	33	8	67
Judgments and rationales of other panelists	0	0	4	33	8	67
Percent of students in each achievement level	1	8	8	67	3	25
My sense of what students need to know to be in each achievement level	0	0	3	25	9	75

[Table 48](#) provides the final evaluation on timing for grades seven through eight.

Table 48. Final Evaluation Grades Seven Through Eight on Timing

How appropriate was the <i>amount of time</i> you were given to complete the different components of the process?	Too Little Time N	Too Little Time %	About Right N	About Right %	Too Much Time N	Too Much Time %
Training in the Bookmark procedure	0	0	10	83	2	17
Test familiarization	2	17	10	83	0	0
Reviewing the ordered item booklet	0	0	11	92	1	8
Placing bookmark judgments	0	0	9	75	3	25
Group discussion	0	0	10	83	2	17

Chapter 5: Results

[Table 49](#) provides the final evaluation for grades seven through eight on the appropriateness of the final recommendations.

Table 49. Final Evaluation Grades Seven Through Eight on the Appropriateness of the Final Recommendations

	Too Low <i>N</i>	Too Low %	About Right <i>N</i>	About Right %	Too High <i>N</i>	Too High %
Do you believe that the final recommended threshold score for entering each of the performance levels is too low, about right, or too high?						
Level 2	0	0	12	100	0	0
Level 3	0	0	12	100	0	0

[Table 50](#) provides the final evaluation for grades seven through eight on support of the recommendations.

Table 50. Final Evaluation Grades Seven Through Eight on Support of the Recommendations

Do you support the final recommendations of the panel?	Yes <i>N</i>	Yes %	No <i>N</i>	No %
Do you support the final recommendations of the panel?	12	100	0	0

[Table 51](#) provides the evaluation results on the usefulness of materials for the high school grade span.

Table 51. Final Evaluation High School on the Usefulness of the Materials

How <i>useful</i> was each of the following materials or procedures in completing the standard setting process?	Not at All Useful <i>N</i>	Not at All Useful %	Somewhat Useful <i>N</i>	Somewhat Useful %	Very Useful <i>N</i>	Very Useful %
Completing the pre-work assignment	0	0	4	29	10	71
Taking the test before making judgments	1	7	3	21	10	71
Defining the borderline students	0	0	1	7	13	93
Reviewing the organization of the ordered item booklet	0	0	1	7	13	93
Information in the item map	0	0	2	14	12	86

Table 51 (continuation)

How <i>useful</i> was each of the following materials or procedures in completing the standard setting process?	Not at All Useful <i>N</i>	Not at All Useful %	Somewhat Useful <i>N</i>	Somewhat Useful %	Very Useful <i>N</i>	Very Useful %
Practicing the procedure	0	0	0	0	14	100
Group discussions	0	0	0	0	14	100
Impact information (percent of students in each achievement level)	0	0	1	7	13	93

[Table 52](#) provides the evaluation results on the influence of the process components for high school.

Table 52. Final Evaluation High School on the Influence of the Process Components

How <i>influential</i> was each of the following in making your judgments?	Not at All Influential <i>N</i>	Not at All Influential %	Somewhat Influential <i>N</i>	Somewhat Influential %	Very Influential <i>N</i>	Very Influential %
Achievement level descriptors	0	0	3	21	11	79
Borderline student definitions	0	0	3	21	11	79
My perception of the difficulty of the items and tasks	0	0	2	14	12	86
My experience with the students	0	0	1	7	13	93
Group discussions	0	0	0	0	14	100
Judgments and rationales of other panelists	0	0	3	21	11	79
Percent of students in each achievement level	1	7	4	29	9	64
My sense of what students need to know to be in each achievement level	0	0	1	7	13	93

Chapter 5: Results

[Table 53](#) provides the final evaluation on timing for high school.

Table 53. Final Evaluation High School on Timing

How appropriate was the <i>amount of time</i> you were given to complete the different components of the process?	Too Little Time N	Too Little Time %	About Right N	About Right %	Too Much Time N	Too Much Time %
Training in the Bookmark procedure	0	0	13	93	1	7
Test familiarization	7	50	5	36	2	14
Reviewing the ordered item booklet	1	7	12	86	1	7
Placing bookmark judgments	1	7	12	86	1	7
Group discussion	0	0	13	93	1	7

[Table 54](#) provides the final evaluation for high school on the appropriateness of the final recommendations.

Table 54. Final Evaluation High School on the Appropriateness of the Recommendations

Do you believe that the final recommended threshold score for entering each of the performance levels is too low, about right, or too high?	Too Low N	Too Low %	About Right N	About Right %	Too High N	Too High %
Level 2	0	0	13	93	1	7
Level 3	1	7	13	93	0	0

[Table 55](#) provides the final evaluation for high school on support of the recommendations.

Table 55. Final Evaluation High School on Support of the Recommendations

Do you support the final recommendations of the committee?	Yes N	Yes %	No N	No %
Do you support the final recommendations of the committee?	14	100	0	0

There were no panelists who indicated on the initial evaluation that additional training or review was needed in the Bookmark method. On the final evaluation, the majority of panelists indicated having a clear understanding of the standard setting process and indicated that the materials and processes were somewhat or very useful. Panelists overall indicated that most

of the process materials, data, and discussion were somewhat or very influential. In some panels, one panelist indicated that one aspect was not personally influential (e.g., responded that “the percent of students in each performance level was not at all influential”).

Panelists’ responses to the questions about the appropriate amount of time allowed for each step was somewhat variable. In most panels, one or more of the panelists indicated that there was either too little or too much time allotted to some aspect of the process. For example, in the grades six through eight panel, one panelist indicated there was too little time for making bookmark judgments, and three panelists indicated there was too much time for group discussion. In the grades nine through twelve panel, one panelist indicated there was too little time for reviewing the OIB, and one panelist indicated there was too much time for placing bookmark judgments. Experience indicates that variability in panelists’ sense of the training and process is expected and dependent on the characteristics and interactions of the panel.

Panelists indicated the independent judgments on the standard-setting forms and were given another opportunity to provide opinions when asked in the final evaluation if the recommended threshold scores were too low, about right, or too high, based on the Round 3 panel median judgments. Generally, a majority of the panelists were comfortable with the threshold-score recommendations, indicating that the recommended threshold scores were “about right,” and 100% of panelists across all panels supported the threshold score recommendations.

Chapter 6: Conclusion

At the request of the CDE, ETS conducted standard-setting workshops for the CSA for grades three through eight and the high school grade span (grades nine through twelve) on August 6–9, 2019, using the Bookmark method.

The process was implemented as planned: Three rounds of judgments with feedback and discussion were completed, and evidence of internal procedural validity was collected via the panelists' evaluations.

The results of the evaluations indicated that the panelists understood the process and the tasks the panelists were asked to complete, found the instructions easy to follow and the training and materials sufficient and clear, and had adequate time to complete the various tasks. In all panels, the majority of panelists judged the final recommended threshold scores to be appropriate (not too high or too low). At the end of the workshop, a cross-grade articulation panel met to review the process and the final threshold-score panel recommendations. There was general consensus that the process was reasonable, educators were engaged and thoughtful, and the threshold scores were reasonably articulated across the six grades and the high school grade span.

Immediately following the workshop, preliminary results were provided to the CDE in the form of recommended threshold scores for each performance level for all six grades and the high school grade span. Data files were provided to the CDE on August 13, 2019. The final standard setting report presented here provides details about panelists, materials, and processes that were not included in the preliminary results table.

References

- Brandon, P. R. (2004). Conclusions about frequently studied modified Angoff standard setting topics. *Applied Measurement in Education*, 17(1), 59–88.
- California Department of Education. (2012). *English language development standards: kindergarten through grade 12*. Retrieved from <http://www.cde.ca.gov/sp/el/er/documents/eldstndpublication14.pdf>
- California Department of Education. (2017, November 8). *California State Board of Education: Final minutes: November 8–9, 2017*. Retrieved from <https://www.cde.ca.gov/be/mt/ms/documents/finalminutes0809nov2017.docx>
- Cizek, G. J., & Bunch, M. B. (2007). *Standard setting: A guide to establishing and evaluating performance standards on tests*. Thousand Oaks, CA: Sage Publications.
- Hambleton, R. K., & Pitoniak, M. J. (2006). Setting performance standards. In R. L. Brennan (Ed.), *Educational Measurement* (4th ed., 433–470). Westport, CT: Praeger.
- Lewis, D. M., Green, D. R., Mitzel, H. C., Baum, K., & Patz, R. J. (1999). *The bookmark standard setting procedure: Methodology and recent implications*. Manuscript under review.
- MacCann, R. G. & Stanley, G. (2004). Estimating the standard error of the judging in a modified-Angoff standards setting procedure. *Practical Assessment, Research, and Evaluation* 9(5). Retrieved from <http://pareonline.net/getvn.asp?v=9&n=5>.
- Mitzel, H. C., Lewis, D. M., Patz, R. J., & Green, D. R. (2001). The bookmark procedure: Psychological perspectives. In G. J. Cizek (Ed.), *Setting performance standards: concepts, methods, and perspectives*. 249–281. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Tannenbaum, R. J. & Katz, I. R. (2013). Standard setting. In K. F. Geisinger (Ed.), *APA handbook of testing and assessment in psychology*, (Vol. 3, pp. 455–77). Washington, DC: American Psychological Association.

Appendix 1: Attachments

Attachment A: Panelist Invitation to Participate and Pre-workshop Assignment

Dear Standard Setting Panelist,

Thank you once again for agreeing to serve as a member of a standard setting panel for the California Spanish Assessment (CSA). You have been selected because you have the appropriate expertise to make the necessary recommendations, you know the *California Common Core State Standards en Español*, you are familiar with the CSA, and you are working with students who have taken or will take the CSA. You have been assigned to a panel that will work on one or two grades. Panelists represent regions across the state and include educators working with students in the six CSA grades and the high school grade span. Four panels of educators will work on the CSA for grades three and four, grades five and six, grades seven and eight, and the high school grade span. **Your grade(s) and panel assignment** are at the top of your notetaking form.

During the standard setting workshop, you will work with your fellow panelists and participate in training for the procedure to develop threshold scores that define the entry into two of the three CSA achievement levels. Standard setting facilitators and assessment specialists from Educational Testing Service will guide you through the process, and the California Department of Education (CDE) will be present to answer any policy questions you may have. The results of the workshop will be presented to the CDE, and the California State Board of Education will make the final decision concerning the threshold scores.

An important part of your work will be to define the knowledge and skills at the entry point of the achievement levels. To help you become familiar with the expectations for the CSA, we have attached the achievement level descriptors (ALDs). The general ALDs are listed in the first row of the document; they are the same for every grade and grade span. The rest of the document lists the range ALDs, which are grade or grade-span specific.

In order to help you prepare for the workshop, we have a two-part task we invite you to complete. The purpose of the first part of the task is for you to become familiar with the item types on the CSA. The second part of the task will give you some familiarity with another important resource for setting standards.

Part 1: It will be very helpful for you to take the CSA training test for the grade(s) to which you are assigned. The website link in this email will allow you to access the training test. We are also sending you a link to the scoring guide, which includes details about the items, as well as the item scoring guide and answer key. To take the training test, select this link: <http://www.caaspp.org/practice-and-training/index.html>.

Once you open that link:

1. Select “Student Interface Practice and Training Tests.”
2. Select “Sign In” (at the bottom of the page).

3. Select your grade (three, four, five, eight, or high school). (Please note that for grades three through five, the training test is the same.)
4. Scroll to the bottom and select the “California Spanish Assessment (CSA) Training Test.”
5. Select “Seleccionar” without making any changes.
6. Play the video and select “Sí, pude reproducir el video y su sonido,” then select “Continuar.”
7. Select “Comenzar la prueba ahora” and take the test.

To score yourself, select the link for your grade to score your training test.

- Links to the Scoring Guide for the training test:

Grades three through five (Spanish)—<http://www.caaspp.org/rsc/pdfs/CSA.training-scoring-guide-gr3-5-Espanol.pdf>

Grades three through five (English)—<http://www.caaspp.org/rsc/pdfs/CSA.training-scoring-guide-gr3-5-English.pdf>

Grades six through eight (Spanish)—<http://www.caaspp.org/rsc/pdfs/CSA.training-scoring-guide-gr6-8-Espanol.pdf>

Grades six through eight (English)—<http://www.caaspp.org/rsc/pdfs/CSA.training-scoring-guide-gr6-8-English.pdf>

High School (Spanish)—<http://www.caaspp.org/rsc/pdfs/CSA.training-scoring-guide-HS-Espanol.pdf>

High School (English)—<http://www.caaspp.org/rsc/pdfs/CSA.training-scoring-guide-HS-English.pdf>

Part 2: We also ask you to become familiar with the ALDs for one of your assigned grades. In order to help you prepare for the workshop, we have attached a notetaking form. The task described on the notetaking form will help you structure your thoughts as you read through the range ALDs. Please focus on the **one** grade listed at the top of the notetaking form (even if your panel will work on two grades). Please bring these printed notes with you to the standard setting workshop. You do not have to bring the ALDs; we will have printed ALDs as well as test materials for your reference at the workshop.

We have found that by completing this pre-workshop task, panelists feel more prepared at the workshop. If you have any questions or concerns before you arrive for the standard setting, please contact Jonathan Moyer at JMoyer@SCOE.net for any travel-related questions. You can contact me at WSwiggett@ets.org for questions regarding standard setting. Thank you in advance for your involvement in this very important work, and we look forward to seeing you in Sacramento.

CALIFORNIA SPANISH ASSESSMENT (CSA)

NOTETAKING TASK

The CSA achievement level descriptors (ALDs) reflect expected performance for a range of students at each achievement level. [Figure 1](#) represents students ordered according to the students' Spanish proficiency in each grade tested. Three performance levels are indicated. In Levels 2 and 3, the student at the beginning of each level is the **borderline student**. The Level 2 (in solid pink) borderline student has slightly more knowledge than the highest-performing student in Level 1 (in green plaid). There is no borderline student for Level 1 because any student whose performance is lower than the Level 2 borderline student is automatically in Level 1.

In this task, you will focus only on the ALDs for one grade level (even if your panel is working on two grades). Your task is to think about the Level 3 borderline student. The task on the following pages will allow you to become familiar with the ALDs and with the type of comparisons we will be making at the standard setting workshop.

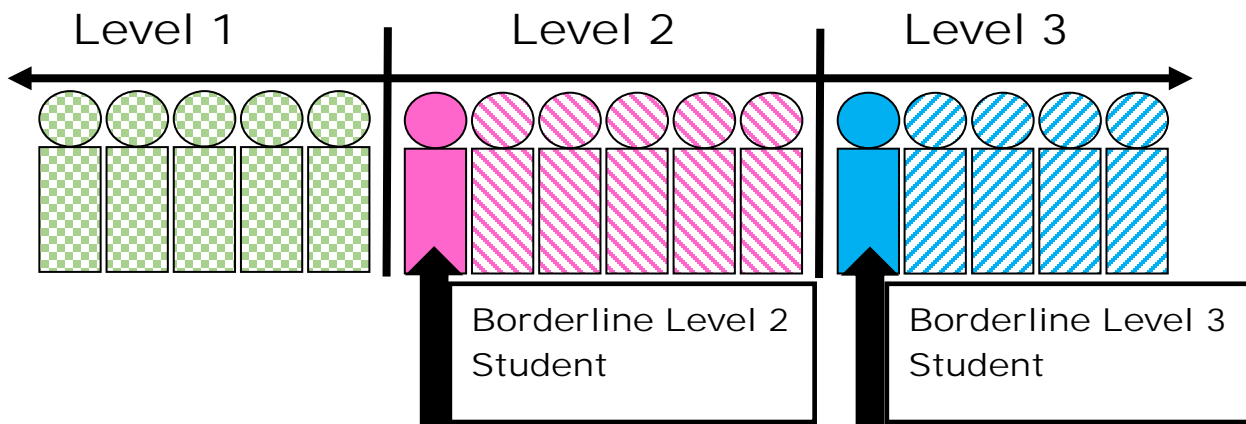


Figure 1. Borderline Student Definitions

Task:

Write down what you think is important; your notes will not be exhaustive.

Spanish Level 3 borderline student

For a student who *just barely* meets the requirements for **Level 3**:

Question	Reponse
1. What does the student know and what can the student do relative to the achievement level descriptors (ALDs)?	[Response]
2. What might the student not be able to do?	[Response]
3. How would you distinguish the student from the highest-performing Level 2 student?	[Response]

Attachment B: Final Borderline Student Definitions

CSA Borderline Student Definitions Grade Three

Borderline Level 2 Student: Reading

The grade three borderline Level 2 student can...

1. Ask and answer some common questions to demonstrate understanding and identify main ideas and some key details of a variety of written texts and genres.
2. Begin to demonstrate how text features contribute to the meaning of written text.
3. Begin to make sufficient comparisons of the information, themes, and literary elements across a variety of texts.
4. Begin to describe the significant aspects of the relationships among event, concepts, and procedures and how characters contribute to textual meaning.

Borderline Level 2 Student: Listening

The grade three borderline Level 2 student can...

1. Begin to determine some main ideas with supporting details in the apparent relationship of text, both visual and oral, using appropriate extrapolation.

Borderline Level 2 Student: Writing

The grade three borderline Level 2 student can...

1. Begin to apply sufficient phonics skills, including syllabication, word analysis, and Spanish language conventions and mechanics.
2. Begin to identify some introductions and concluding elements to a topic or a story.
3. Create agreement among parts of speech in simple sentences using regular and some commonplace irregular nouns and verbs.
4. Begin to apply knowledge of organizational text structures, including a logical sequence of events.

Borderline Level 3 Student: Reading

The grade three borderline Level 3 student can...

1. Ask and answer some challenging questions to demonstrate strong understanding and identify main ideas and key details of a variety of written texts and genres.
2. Often demonstrate how text features contribute to the meaning of written text.
3. Make some insightful comparisons of the information, themes, and literary elements across a variety of texts.
4. Begin to describe the complex aspects of the relationships among events, concepts, and procedures and how characters contribute to textual meaning.

Borderline Level 3 Student: Listening

The grade three borderline Level 3 student can...

1. Begin to determine some complex ideas with supporting details in the relationship of texts both visual and oral using some insightful extrapolation.

Borderline Level 3 Student: Writing

The grade three borderline Level 3 student can...

1. Begin to apply a variety of phonics skills, including syllabication, word analysis, and Spanish language conventions and mechanics.
2. Identify some atypical or unique introduction and concluding elements to a topic or a story.
3. Create agreement among parts of speech in simple, compound, and some complex sentences using regular and some irregular nouns and verbs.
4. Begin to apply strong knowledge of organizational text structures across a variety of genres.

CSA Borderline Student Definitions Grade Four

Borderline Level 2 Student: Reading

The grade four borderline Level 2 student can...

1. Determine some clear main ideas and themes by beginning to identify key details while summarizing a text.
2. Begin to explain some significant differences within literary works and genres using some commonplace terminology while distinguishing aspects of point of view used by authors, narrators, and characters.

Borderline Level 2 Student: Listening

The grade four borderline Level 2 student can...

1. Comprehend and paraphrase some main ideas while providing appropriate support.

Borderline Level 2 Student: Writing

The grade four borderline Level 2 student can...

1. Demonstrate some control of Spanish written text conventions, including mechanics.
2. Begin to identify an introduction and concluding statement to a topic or story, including details.
3. Create agreement among the parts of speech in simple and some compound sentences.

Borderline Level 3 Student: Reading

The grade four borderline Level 3 student can...

1. Determine some complex main ideas and themes by beginning to identify implicit details while summarizing a text.
2. Begin to explain some intricate differences within literary works/genres using relevant terminology while distinguishing aspects of the point of view used by authors, narrators, and/or characters.

Appendix 1: Attachments

Borderline Level 3 Student: Listening

The grade four borderline Level 3 student can...

1. Comprehend and paraphrase the main idea and identify some complex ideas while providing some insightful support.

Borderline Level 3 Student: Writing

The grade four borderline Level 3 student can....

1. Demonstrate some command of Spanish written text conventions, including mechanics.
2. Often identify an appropriate introduction and concluding statement to a topic or story, including details.
3. Create agreement among the parts of speech in simple, compound, and some complex sentences in meaningful texts.
4. Understand simple logical connections across parts of text in a variety of texts.

CSA Borderline Student Definitions Grade Five

Borderline Level 2 Student: Reading

The grade five borderline Level 2 student can....

1. Analyze and interpret various texts by determining clear main ideas and details, comparing or contrasting story elements, text features, and explaining relationships with some clarity and evidence.

Borderline Level 2 Student: Listening

The grade five borderline Level 2 student can....

1. Comprehend spoken Spanish by summarizing and describing ideas and providing some appropriate support.

Borderline Level 2 Student: Writing

The grade five borderline Level 2 student can....

1. Utilize general vocabulary and language structures to demonstrate some control of Spanish and to communicate ideas.

Borderline Level 2 Student: Reading and Writing

The grade five borderline Level 2 student can....

1. Begin to demonstrate domain-specific use of language to enhance their craft/structure by connecting ideas, adding some details, and use of appropriate linking words when reading and writing.

Borderline Level 3 Student: Reading

The grade five borderline Level 3 student can....

1. Analyze and interpret various texts by determining main ideas, details, comparing and contrasting story elements, text features and explaining relationships with some complexity, evidence and inference.

Borderline Level 3 Student: Listening

The grade five borderline Level 3 student can...

1. Comprehend spoken Spanish by summarizing and describing ideas with some complexity and inference and providing strong support.

Borderline Level 3 Student: Writing

The grade five borderline Level 3 student can...

1. Use grammatical and language functions to demonstrate grade-level appropriateness in various elements of writing.

Borderline Level 3 Student: Reading and Writing

The grade five borderline Level 3 student can...

1. Begin to demonstrate a varied use of language to enhance their craft/structure when reading and writing.

CSA Borderline Student Definitions Grade Six

Borderline Level 2 Student: Reading

The grade six borderline Level 2 student can...

1. Refer to details or evidence to analyze and interpret a variety of texts and genres by evaluating evidence of some clear claims, at times determining complex ideas and themes, explaining how texts contribute to theme, authors point of view, and synthesizing information presented in more than one medium as well as sometimes clarifying commonplace figurative language and meaning of words.

Borderline Level 2 Student: Listening

The grade six borderline Level 2 student can...

1. Comprehend spoken Spanish in a range of contexts by summarizing some clear ideas, aspects of topic, text, or issue and demonstrate comprehension of arguments and claims providing general support.

Borderline Level 2 Student: Writing

The grade six borderline Level 2 student can...

1. Use standard Spanish, with basic and some commonplace words, to demonstrate an understanding and use of how, at times, word choice affects academic tone and formal style.
2. Demonstrate basic use of Spanish conventions and manage differences between academic Spanish and English by explaining the general functions of the parts of speech and create agreement in some grade-level sentences.
3. Identify a common introduction to a story thesis statement and group related information according to organizational structures; at times use of sensory details to convey information with an appropriate conclusion.

Appendix 1: Attachments

Borderline Level 3 Student: Reading

The grade six borderline Level 3 student can...

1. Read, analyze, and interpret a variety of texts and genres through Spanish by evaluating evidence of some implicit claims, determining complex ideas and themes, explaining how texts contribute to theme, author's point of view, and synthesizing information presented in more than one medium, as well as clarifying figurative language with some insight.

Borderline Level 3 Student: Listening

The grade six borderline Level 3 student can...

1. Comprehend spoken Spanish in a range of contexts by summarizing some complex ideas and intricate aspects of topic, text, or issue; and demonstrating some nuanced comprehension of arguments and claims providing insightful support.

Borderline Level 3 Student: Writing

The grade six borderline Level 3 student can...

1. Use standard orthography in commonplace and some challenging words to demonstrate an understanding of how word choice affects academic tone and formal style.
2. Demonstrate the use of Spanish conventions and manage a variety of differences between academic Spanish and English by explaining the function of the parts of speech and creating agreement in simple compound and some complex sentences.
3. Identify an introduction to a story thesis statement and group information according to various organizational structures, connect details, support claims with specific evidence and reasoning, using linking words and sensory details to convey information with an appropriate conclusion.

CSA Borderline Student Definitions Grade Seven

Borderline Level 2 Student: Reading

The grade seven borderline Level 2 student can...

1. Determine some aspects of an author's purpose and how the author compares/contrasts various points of view.
2. Sometimes identify a summary of the text based on details and themes presented visually and orally.
3. Determine how a portion of the text contributes to the whole and fits into the structure.
4. Sometimes refer to evident details, reasons, or evidence to analyze what a text says, and begin to evaluate the evidence and reasoning of apparent claims and arguments from a variety of appropriate texts.

Borderline Level 2 Student: Writing

The grade seven borderline Level 2 student can...

1. Choose language that, at times, relates ideas precisely and concisely.
2. Identify and revise sentences and sentence elements using sufficient grade-level writing and grammar conventions, including academic tone and formal style.

Borderline Level 2 Student: Integrated Skills

The grade seven borderline Level 2 student can...

1. At times make sufficient comparisons, while sometimes contrasting two authors' presentations by synthesizing information presented in more than one medium.

Borderline Level 3 Student: Reading

The grade seven borderline Level 3 student can...

1. Sometimes determine the nuances of an author's purpose and how the author compares/contrasts various points of view.
2. Evaluate how a portion of the text contributes to the whole and fits into the structure.
3. Sometimes identify a focused summary of the text based on implicit details and complex themes presented visually and orally.
4. Sometimes refer to implicit details, reasons, and evidence to analyze what a text says, and begin to evaluate the evidence and reasoning of implicit claims and arguments from a variety of appropriate texts.

Borderline Level 3 Student: Writing

The grade seven borderline Level 3 student can...

1. Choose language that, at times, expresses ideas precisely and concisely.
2. Identify and revise sentences and sentence elements using a wide variety of writing/grammar conventions, including academic tone and formal style.

Borderline Level 3 Student: Integrated skills

The grade seven borderline Level 3 student can...

1. Sometimes make insightful comparisons and contrast two authors' presentations by synthesizing information presented in more than one medium.

CSA Borderline Student Definitions Grade Eight

Borderline Level 2 Grade 8 Student

The grade eight borderline Level 2 student can...

1. Develop a text by adding some relevant details, textual evidence, or examples, and begin to control Spanish conventions and rhetorical elements (e.g. voice, mood, tenses).
2. Sometimes identify clear details and apply clear evidence and occasionally provide solid support from one or more texts.
3. Sometimes make sufficient observations and appropriate evaluations by comparing and contrasting in order to synthesize written information.

Appendix 1: Attachments

4. Identify and/or revise sentences and sentence elements with respect to standard writing/grammar conventions, including academic tone and formal style.
5. Determine and sometimes analyze central idea(s), supporting details, and an author's purpose in literary and informational texts (presented orally and in text/print).
6. Sometimes make appropriate analysis of two authors' conflicting accounts on the same topic identifying how texts disagree on matters of fact or interpretation by using key details to identify a summary.

Borderline Level 3 Grade 8 Student

The grade eight borderline Level 3 student can...

1. Sometimes analyze and evaluate complex idea(s) and supporting details in literary and informational texts (presented orally and in print).
2. Develop a text by adding some connecting details, textual evidence, clarifying examples, and manage a wide variety of rhetorical elements (e.g. voice, mood, and verb tenses).
3. Sometimes make insightful analysis of two authors' conflicting accounts on the same topic identifying how texts disagree on matters of fact or interpretation by using key details to identify a focused summary that is objective.
4. At times refers to implicit details, implicit reasons, and implicit evidence and occasionally provides strong support by quoting from a variety of texts.
5. Make insightful observations by comparing and contrasting in order to synthesize written information.
6. Determine complex central ideas and complex themes in various texts and can sometimes determine the nuances of an author's purpose.
7. Recognize and/or revise challenging sentences and sentence elements with respect to standard writing/grammar conventions including academic tone and formal style.
8. Know and apply a wide variety of strategies to clarify the meaning of words, and sometimes figurative language and figures of speech, with moderate success.

CSA Borderline Student Definitions High School

Borderline Level 2 High School Student

The high school borderline Level 2 student can...

1. Recognize explicit details, reasons, or evidence to analyze text.
2. Understand the author's purpose, point of view, and style to analyze some clear central ideas and themes in a variety of texts.
3. Identify some details that support the central idea and analyze their development in the text (presented orally and in print).
4. Demonstrate some comprehension of point of view, purpose arguments, and claims in a text read aloud.
5. Make revisions to commonplace sentences by applying basic academic Spanish conventions that demonstrate some understanding of Spanish syntax and grammar.

6. Demonstrate an ability to synthesize information based on some commonplace organizational structures presented in one or more medium.
7. Make obvious connections between sources and identify when information is left open-ended or uncertain.
8. Distinguish between simple claims and obvious fallacies in a variety of texts.

Borderline Level 3 High School Student

The high school borderline Level 3 student can...

1. Compare and contrast texts citing evidence to describe subtleties and make insightful observations.
2. Refer to some implicit details, reasons, or evidence to analyze a text.
3. Understand the author's purpose and style to analyze complex and central themes.
4. Identify some complex aspects of how an author's choices impact events, ideas, reasoning, or some complex characters.
5. Analyze and evaluate central idea(s) and supporting details in literary and informational texts (presented orally and in print).
6. Analyze an author's point of view and make some connections between sources.
7. Distinguish between facts and invalid information in a variety of texts.
8. Link information, group ideas, and summarize information based on details from the text.
9. Make revisions to sentences of varying degrees of difficulty by applying academic Spanish conventions such as syntax, grammar, and word choice.

Attachment C: Sample Item Map

The passage titles were deleted to protect the security of the CSA. The item map in [Table 1.1](#) provides information about items in the ordered item booklet (OIB). The Standard-Setting Scale (SS Scale) column represents relative item difficulty and is not the CSA score-reporting scale.

Table 1.1. CSA Sample Item Map

OIB Sequence	Original Item Order	Measure	Passage or Script Title	Key	Pts1	SS Scale	Item ID
1	41	Reading	Confidential Title	3, 1, 2	1 of 2	161	854612
2	15	Listening	Confidential Title	2, 3, 1	1 of 1	209	752410
3	48	Writing	Confidential Title	1 or 3	1 of 2	210	570650
4	41	Reading	Confidential Title	2	2 of 2	216	854612
5	1	Reading	Confidential Title	4	1 of 1	218	145381
6	20	Listening	Confidential Title	2	1 of 1	230	851451
7	24	Listening	Confidential Title	3 or 1	1 of 2	236	486513
8	12	Reading	N/A	3	1 of 1	245	458545
9	30	Writing	N/A	1 and 3	2 of 2	298	570650
10	18	Writing	Confidential Title	1	1 of 2	305	894735

Attachment D: Evaluation Forms

Training Evaluation Form

California Spanish Assessment

Training Evaluation Form

The purpose of this evaluation form is to obtain your feedback about the training you have received on the standard setting process. Your feedback will provide a basis for determining what to review before you begin making your standard setting judgments.

Please indicate the degree to which you agree with each statement using the scale given (Strongly Agree, Agree, Disagree, Strongly Disagree). Please choose only one response for each statement.

Statement	Strongly Agree	Agree	Disagree	Strongly Disagree
I understand the purpose of this workshop.	[Response]	[Response]	[Response]	[Response]
The large-group facilitator explained things clearly.	[Response]	[Response]	[Response]	[Response]
The panel facilitator explained things clearly.	[Response]	[Response]	[Response]	[Response]
I understand the purpose of the ALDs in this process.	[Response]	[Response]	[Response]	[Response]
I understand what is meant by the borderline student.	[Response]	[Response]	[Response]	[Response]
I understand what the ordered item booklet is.	[Response]	[Response]	[Response]	[Response]
I understand the information presented in the item map.	[Response]	[Response]	[Response]	[Response]
The training in the Bookmark method seems adequate to give me the information I need to complete my assignment.	[Response]	[Response]	[Response]	[Response]
I understand how to record my bookmark placements.	[Response]	[Response]	[Response]	[Response]
I am ready to place my first bookmark for the test.	[Response]	[Response]	[Response]	[Response]

If you checked "Disagree" or "Strongly Disagree" for any of the statements on the previous page, please indicate what additional information or explanation you need.

Appendix 1: Attachments

Have you participated in a Bookmark or other standard setting workshop before today?
No Yes

Did you participate in the CSA ALD workshop last year, held at the ETS Sacramento office?
No Yes

By signing this form, I state that I am ready to proceed with the process.

Signature_[RESPONSE]_____

Date_[RESPONSE]

Final Evaluation

California Spanish Assessment Standard Setting Final Evaluation

The purpose of the final evaluation form is to obtain your feedback about the standard setting process overall. Your feedback will provide a basis for evaluating the training, methods, and materials in the standard setting process. Your responses will be anonymous; no individuals will be identified. Please select one unless otherwise indicated.

Gender

Female Male Non-binary

Race

American Indian/Alaskan Native Asian Black or African American Filipino

Hispanic or Latino Native Hawaiian or Pacific Islander

White Two or more races

Grade(s) you currently teach (check all that apply)

3 4 5 6 7 8 9–12

Other

What subjects do you currently teach in Spanish? (Check all that apply.)

All Subjects Math Science Social Studies Reading/language Arts Other

How many years of experience do you have teaching bilingual classes?

1–3 4–6 7–10 More than 10

How many years of experience do you have working with the California Common Core State Standards en Español?

Not Applicable 1–2 3–4 More than 4

Does your teaching experience include students from these populations? (Check all that apply.)

General education English learners Special education

Appendix 1: Attachments

Please choose only one response for each statement below.

1. How useful was each of the following materials or procedures in completing the standard setting?

Statement	Not at All Useful	Somewhat Useful	Very Useful
Completing the pre-work assignment	[Response]	[Response]	[Response]
Taking the test before making judgments	[Response]	[Response]	[Response]
Defining the borderline students	[Response]	[Response]	[Response]
Reviewing the organization of the ordered item booklet	[Response]	[Response]	[Response]
Information in the item map	[Response]	[Response]	[Response]
Practicing the procedures	[Response]	[Response]	[Response]
Group discussions	[Response]	[Response]	[Response]
Impact information (percent of students in each performance level)	[Response]	[Response]	[Response]

2. How influential was each of the following in making your judgments?

Statement	Not at All Influential	Somewhat Influential	Very Influential
Achievement level descriptors	[Response]	[Response]	[Response]
Borderline student definitions	[Response]	[Response]	[Response]
My perception of the difficulty of the items and tasks	[Response]	[Response]	[Response]
My experiences with the students	[Response]	[Response]	[Response]
Reviewing authentic student responses	[Response]	[Response]	[Response]
Group discussions	[Response]	[Response]	[Response]
Judgments and rationales of other panelists	[Response]	[Response]	[Response]
Percent of students in each performance level	[Response]	[Response]	[Response]
My sense of what students need to know to be in each achievement level	[Response]	[Response]	[Response]

3. How appropriate was the *amount of time* you were given to complete the different components of the process?

Statement	Too Little Time	About Right	Too much Time
Training in the Bookmark procedure	[Response]	[Response]	[Response]
Test familiarization	[Response]	[Response]	[Response]
Reviewing the ordered item booklet (OIB)	[Response]	[Response]	[Response]
Placing bookmark judgments	[Response]	[Response]	[Response]
Group discussion	[Response]	[Response]	[Response]

4. California Spanish Assessment threshold scores

Do you believe that the final recommended threshold scores for entering each of the achievement levels are too low, about right, or too high?

Level	Too Low	About Right	Too High
Level 2	[Response]	[Response]	[Response]
Level 3	[Response]	[Response]	[Response]

5. Do you support the final recommendations of the panel?

Yes No

Attachment E: Nondisclosure Agreement Form

Test security for the California Assessment of Student Performance and Progress (CAASPP) is of the utmost importance, and it is the California Department of Education’s obligation to ensure the security of all test materials. The nature and content of any test, test item, proposed or draft test item, or other secure assessment material, including but not limited to the specific language or the subject of test items or proposed or draft test items and any art such as drawings, graphs, tables and sketches, must not be divulged.

By signing below, you acknowledge and agree that the CAASPP test materials are highly secure and that the unauthorized disclosure of any test materials associated with the CAASPP could result in substantial monetary and nonmonetary costs to the State to replace the test and materials. You agree that your access to the CAASPP test items, proposed or draft test items, or any other test materials is only for the purpose of review as charged by your role as a member of this panel. You agree not to reproduce the tests or any questions within them, directly or indirectly, and not to reveal the nature or content of the test or test items to any other person other than those participating in this meeting.

I understand that the California Spanish Assessment passages and written items for the Item and Passage Review Meeting are classified as confidential. I understand that these materials cannot be posted publicly, sold, or reproduced. The materials included in this training set contain information copyrighted by the Regents of the University of California, the California Department of Education, and/or independent publishers. IN WITNESS WHEREOF, Participant has executed this Agreement.

Signature: [RESPONSE]_____

Print Name: [RESPONSE]_____

Affiliation/Organization: [RESPONSE]_____

Date: [RESPONSE]_____

Appendix 2: Conditional Standard Error of Measurement

[Table 2.1](#) through [Table 2.7](#) provide data on the scale-scores for each grade and high school grade span. Based on the panel recommendations for each threshold score, the recommended threshold scores were calculated using item response theory. The recommended threshold score is presented for each level—Level 2 and Level 3—as well as the percent of students at or above that median threshold score. The tables also display, for each level, a range around the recommended threshold score, specifically, +/-1 CSEM and +/-2 CSEM, and the percent of students who would be placed at or above each of those scores. These impact data are based on the 2018–19 operational administration of the CSA and may differ from the percent of students in each level in following administrations.

[Table 2.1](#) provides this scale score data for grade three.

Table 2.1. Projected Percentage of 2018–19 Students at or Above the Recommended Threshold Score, +/-1 CSEM and +/-2 CSEMs for Grade Three

Threshold	Level 2 Scale Score	Level 2 Percent at or Above	Level 3 Scale Score	Level 3 Percent at or Above
-2 CSEM	396	62.9	409	21.8
-1 CSEM	400	47.3	413	14.3
Panel Recommended	404	36.6	417	8.9
+1 CSEM	408	24.5	421	3.7
+2 CSEM	412	16.7	425	1.9

[Table 2.2](#) provides this scale score data for grade four.

Table 2.2. Projected Percentage of 2018–19 Students at or Above the Recommended Threshold Score, +/-1 CSEM and +/-2 CSEMs for Grade Four

Threshold	Level 2 Scale Score	Level 2 Percent at or Above	Level 3 Scale Score	Level 3 Percent at or Above
-2 CSEM	396	61.8	408	27.4
-1 CSEM	400	50.3	412	17.3
Panel Recommended	404	36.1	416	11.1
+1 CSEM	408	27.4	420	6.2
+2 CSEM	412	17.3	424	4.0

Appendix 2: Conditional Standard Error of Measurement

[Table 2.3](#) provides this scale score data for grade five.

Table 2.3. Projected Percentage of 2018–19 Students at or Above the Recommended Threshold Score, +/-1 CSEM and +/-2 CSEMs for Grade Five

Threshold	Level 2 Scale Score	Level 2 Percent at or Above	Level 3 Scale Score	Level 3 Percent at or Above
-2 CSEM	394	72.3	409	23.0
-1 CSEM	398	54.4	413	13.6
Panel Recommended	402	43.1	417	7.5
+1 CSEM	406	29.5	421	3.4
+2 CSEM	410	20.5	425	1.5

[Table 2.4](#) provides this scale score data for grade six.

Table 2.4. Projected Percentage of 2018–19 Students at or Above the Recommended Threshold Score, +/-1 CSEM and +/-2 CSEMs for Grade Six

Threshold	Level 2 Scale Score	Level 2 Percent at or Above	Level 3 Scale Score	Level 3 Percent at or Above
-2 CSEM	393	76.3	406	31.3
-1 CSEM	397	62.4	410	17.8
Panel Recommended	401	47.2	414	10.2
+1 CSEM	405	34.9	418	5.1
+2 CSEM	409	21.2	422	1.4

[Table 2.5](#) provides this scale score data for grade seven.

Table 2.5. Projected Percentage of 2018–19 Students at or Above the Recommended Threshold Score, +/-1 CSEM and +/-2 CSEMs for Grade Seven

Threshold	Level 2 Scale Score	Level 2 Percent at or Above	Level 3 Scale Score	Level 3 Percent at or Above
-2 CSEM	397	59.3	412	13.9
-1 CSEM	401	46.6	417	4.8
Panel Recommended	405	31.1	422	1.9
+1 CSEM	409	18.9	427	0.2
+2 CSEM	413	11.3	432	0.0

[Table 2.6](#) provides this scale score data for grade eight.

Table 2.6. Projected Percentage of 2018–19 Students at or Above the Recommended Threshold Score, +/-1 CSEM and +/-2 CSEMs for Grade Eight

Threshold	Level 2 Scale Score	Level 2 Percent at or Above	Level 3 Scale Score	Level 3 Percent at or Above
-2 CSEM	397	62.7	409	22.0
-1 CSEM	401	46.4	414	9.7
Panel Recommended	405	31.2	419	4.5
+1 CSEM	409	22.0	424	1.2
+2 CSEM	413	11.5	429	0.4

[Table 2.7](#) provides this scale score data for high school.

Table 2.7. Projected Percentage of 2018–19 Students at or Above the Recommended Threshold Score, +/-1 CSEM and +/-2 CSEMs for High School

Threshold	Level 2 Scale Score	Level 2 Percent at or Above	Level 3 Scale Score	Level 3 Percent at or Above
-2 CSEM	398	56.8	409	19.8
-1 CSEM	402	40.4	413	11.0
Panel Recommended	406	29.3	417	4.2
+1 CSEM	410	16.6	421	1.7
+2 CSEM	414	9.0	425	0.9

Appendix 3: Cross-Grade Articulation Data

The data in [Table 3.1](#) and [Figure 2](#) was available to support the cross-grade articulation workshop. During the workshop, the facilitator focused on having the panelists discuss the borderline student definitions and processes they used to create them. The panelists first reviewed the borderline student definitions for adjacent grades that they had not seen. Then, the reasonableness of those definitions was confirmed by the panelists. Panelists also engaged in discussion regarding the impact data and the reasonableness of the final scores in relation to the panelists' personal experiences with students. [Table 3.1](#) provides this data in tabular form.

Table 3.1. Percentage at Each Level Across Grade or Grade Span

Level	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	High School
Level 1	63.4	63.9	56.9	52.8	68.9	68.8	70.7
Level 2	27.7	25.0	35.6	37.0	29.2	26.7	25.1
Level 3	8.9	11.1	7.5	10.2	1.9	4.5	4.2

[Figure 2](#) provides this data in graph form.

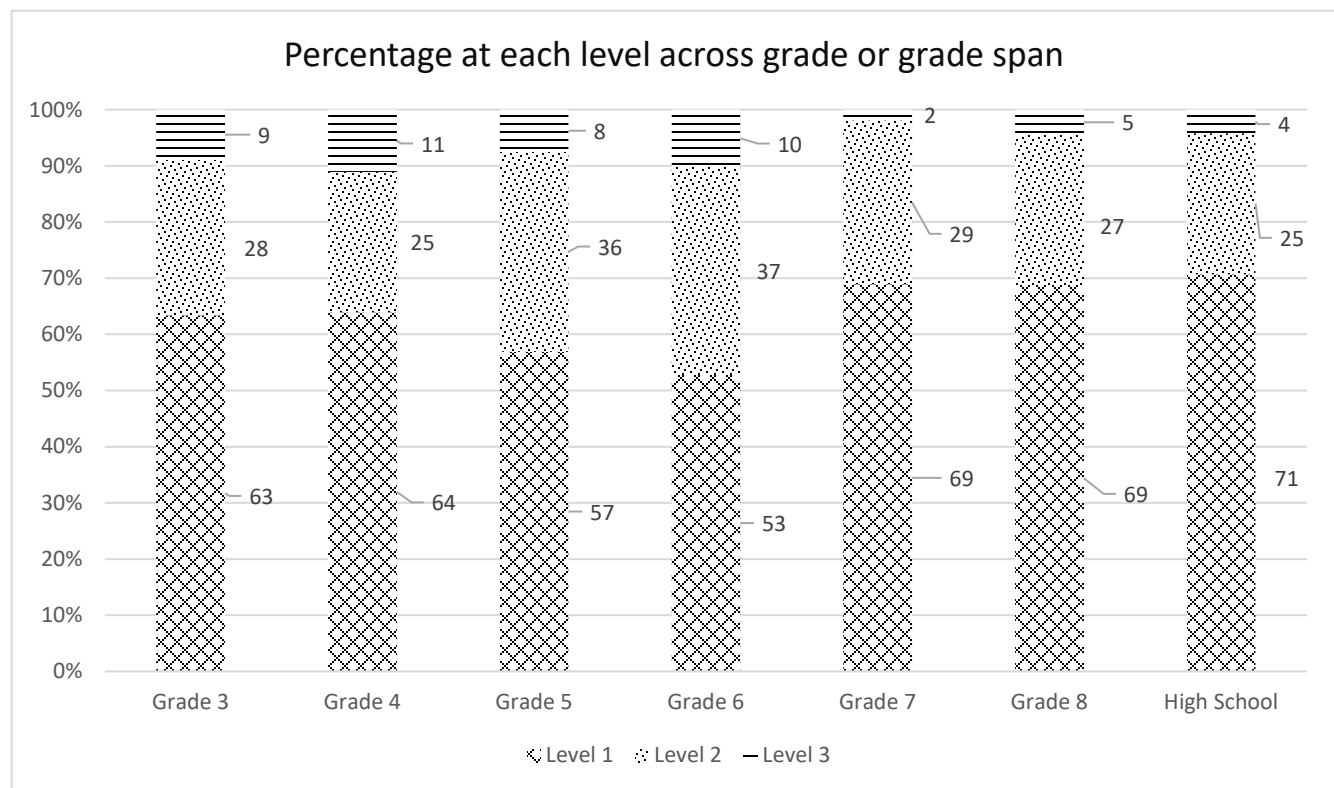


Figure 2. Percentage at each level across grade or grade span