# State of Texas Assessments of Academic Readiness (STAAR®) Alternate 2 Standard Setting Report

## STAAR Alternate 2 Grades 3–8 Reading Language Arts (RLA)

June 2023

Pearson

Version 4.1

# Table of Contents

# Executive Summary

This report describes the standard setting process for the State of Texas Assessments (STAAR®) Alternate 2 Reading Language Arts (RLA) assessments for grades 3–8. A summary of the results is also provided.

## Standard Setting Process and Results

Performance levels are used to classify and describe student performance on an assessment. To classify student performance into the different performance levels, performance level descriptors (PLDs) and cut scores are generally required. The PLDs define what students at each performance level should know and be able to do within each content area, and cut scores represent the lowest boundary of each performance level on the scale. The process of recommending performance standards for the STAAR Alternate 2 RLA assessments is based on a nationally recognized best practice for standard settings. The standard setting methodology used is a modification of the well-known Angoff method (Angoff, 1971). Results and details of the process are presented in the following sections.

### Performance Level Descriptors

A multi-step iterative process was used to develop, review, and approve the PLDs. Prior to the standard setting meeting, content staff from the Texas Education Agency (TEA) created a draft set of PLDs representing requirements for classifying student performance on the STAAR Alternate 2 RLA assessments. The draft PLDs were reviewed by a panel of teachers from across the state who provided feedback and recommended revisions for TEA to consider when finalizing the PLDs. The performance levels for the STAAR Alternate 2 RLA assessments are as follows:

- Level 3: *Accomplished Academic Performance*
- Level 2: *Satisfactory Academic Performance*
- Level 1: *Developing Academic Performance*

### Standard Setting Meeting

A standard setting meeting was convened from June 12–16, 2023, to recommend cut scores for the STAAR Alternate 2 RLA grades 3–8 assessments. The panels for grades 3, 5, 6, and 8 met on June 12–13, while the panels for grades 4 and 7 met on June 14–15. Vertical articulation occurred on June 16. The committees were composed of 7–12 panelists who were educators with relevant content area teaching experience. They were a mix of general education teachers and special education teachers who had experience with the STAAR Alternate 2 population. The committee members were selected to provide content expertise and represent diverse state geographic regions, gender, ethnicity, educational experience, community size, and community socioeconomic status.

The Modified Angoff (Angoff, 1971) standard setting method is a content- and item-based method that leads panelists through a standardized process in which they consider student expectations, as defined by the PLDs, and the individual items that could be administered to students to recommend cut scores for each performance level. The standardized process was used by the committees for each content area.

To begin Day 1, the panelists participated in a general session where they were told why new performance standards were needed and given an overview of the Modified Angoff method. The panelists then moved to assigned breakout groups for each grade level. The committees first reviewed the test design and blueprint, followed by the "Experience the Assessment" activity that allowed them to consider the knowledge and skills needed to respond to each item. During this time, panelists discussed the STAAR Alternate 2 population and the administration and response modes available for the assessment and watched a video demonstrating the student test administration. Next, they reviewed the PLDs to gain a common understanding of the expectations for the performance levels and then narrowed the focus to key knowledge and skills at the borderlines that separated the performance levels. They worked in small groups to create specific descriptions of the knowledge and skills expected of students who just barely enter a performance level.

After discussion and general agreement about the borderline descriptions, the panelists were trained on the standard setting method and the judgment process that was to be applied during the remainder of the meeting. They were taught to review each item and the borderline descriptions and to consider the following question for each performance level for each multi-point (polytomously scored) item:

> *"How many points would a student with performance at the borderline of the level likely earn if they answered the question?"*

For the purposes of the standard setting, "likely" was defined as two out of three students who were at the borderline of the performance level and who correctly answered the item.

The panelists then engaged in a practice judgment activity using sample items where they discussed the process and results to clarify their understanding of the judgment task. They then began the three rounds of judgments. Following Rounds 1 and 2, the panelists reviewed their individual cut score recommendations and the group's cut score agreement and took part in a whole-group discussion of the items with the greatest level of disagreement. Panelists with the different judgment ratings provided a rationale for their decision to develop a common understanding across the group of expectations for being classified into each performance level. After Rounds 2 and 3, the panelists were presented with impact data showing the distribution of students falling into each performance level given the median cut score of the entire committee.

**Benchmark Reasonable Ranges**

Although there were changes to test design and administration mode, the policy expectations that trended in student distribution across the performance levels did not vary significantly from previous administrations. To assist in aligning the resulting standard setting cut score recommendations with the academic expectations defined in the PLDs while also maintaining similar impact data to previous administrations, benchmark values were established as reasonable ranges.

Table 1 presents the benchmark reasonable ranges presented to the panelists. The benchmark values represented a reasonable range for each performance level of every STAAR Alternate 2 RLA assessment based on the cut scores from the previous test design. Benchmark reasonable ranges were shared with panelists as part of the feedback data after Rounds 1 and 2. Placing the cut score recommendation within the reasonable ranges was not a requirement, but panelists were asked to provide a content-based rationale for placements outside the reasonable range.

**Table 1. Benchmark Reasonable Ranges (Raw Score Points)**

| Grade | *Satisfactory* | *Accomplished* |
|-------|-------------|--------------|
| 3 | 19–25 | 33–38 |
| 4 | 19–25 | 33–38 |
| 5 | 19–26 | 33–38 |
| 6 | 19–26 | 32–36 |
| 7 | 19–25 | 31–37 |
| 8 | 18–24 | 31–36 |

**Results**

Similar to Rounds 1 and 2, panelists made their judgments at the test level instead of the item level in Round 3. After Round 3, final recommended cut scores were computed, and panelists were shown their individual test-level judgments. Panelists also reviewed the group median judgment for each performance level and verified that the median judgments were within the reasonable ranges. The median Round 3 cut score of each performance level for each committee was used as the recommended cut score. Table 2 presents the recommended cut scores for the STAAR Alternate 2 RLA assessments.

**Table 2. Standard Setting Recommendations**

| Grade | Max. Score | *Satisfactory* Cut | *Satisfactory* Range | *Accomplished* Cut | *Accomplished* Range |
|-------|-----------|----------------|------------------|------------------|--------------------|
| 3 | 40 | 20 | 19–25 | 35 | 33–38 |
| 4 | 40 | 21 | 19–25 | 35 | 33–38 |
| 5 | 40 | 23 | 19–26 | 34 | 33–38 |
| 6 | 40 | 25 | 19–26 | 33 | 32–36 |
| 7 | 40 | 21 | 19–25 | 31 | 31–37 |
| 8 | 40 | 19 | 18–24 | 31 | 31–36 |

After Round 3, the panelists completed an evaluation of the standard setting process and their confidence in their recommended cut scores. Overall, the panelists understood the standard setting process and were confident in their recommendations.

**Vertical Articulation**

As a final step in the standard setting process, selected panelists from each grade-level committee were convened in a vertical articulation panel to evaluate the reasonableness of the cut score recommendations from the standard setting committees. The facilitator first led a content-focused discussion in which the panelists were asked to identify similarities and differences in performance expectations between grades. Next, panelists were shown impact data for each grade level based on the Round 3 cut score recommendations from each committee. The panelists evaluated the degree to which the impact data met their expectations, which was guided by their knowledge of the content and students.

The final cut scores from the individual standard setting committees were accompanied by recommended ranges for each performance level based on the Round 3 cut scores. The point estimate was the median cut score recommendation from the panelists. The range around the point estimate was defined by the first quartile (Q1) and third quartile (Q3) from the panelist cut score recommendations, with Q1 and Q3 representing the lower and upper bounds of the range, respectively. The recommended range essentially represented the variation in panelist cut score recommendations from the Round 3 judgments. Table 3 presents the recommended cut scores for the STAAR Alternate 2 RLA assessments provided by the articulation committee.

**Table 3. Vertical Articulation Recommendations**

| Grade | Max. Score | *Satisfactory* | *Accomplished* |
|-------|-----------|----------------|----------------|
| 3 | 40 | 20 | 35 |
| 4 | 40 | 22 | 35 |
| 5 | 40 | 23 | 35 |
| 6 | 40 | 23 | 34 |
| 7 | 40 | 22 | 34 |
| 8 | 40 | 23 | 34 |

Table 4 presents the impact data that denotes the percentage of students who took the STAAR Alternate 2 RLA assessment during the spring 2023 administration who would be classified into each performance level based on the recommended cut scores resulting from the articulation process. The percentage of students in a performance level is not directly comparable across grades as the student population is different for each assessment. Performance levels from different tests are not comparable because the cut scores for these tests are criterion-referenced (i.e., they are based on content-specific expectations of what students should know and be able to do).

**Table 4. Impact Data from the Vertical Articulation Recommendations**

| Grade | Developing | Satisfactory | Accomplished |
|-------|-----------|--------------|--------------|
| 3 | 10% | 74% | 16% |
| 4 | 10% | 69% | 21% |
| 5 | 11% | 62% | 27% |
| 6 | 9% | 59% | 32% |
| 7 | 7% | 63% | 30% |
| 8 | 8% | 58% | 34% |

**TEA Reasonableness Review**

TEA reviewed the resulting standard setting recommendations during a reasonableness review to examine the cut scores with an additional perspective of policy expectation and historical trends in student performance. This review incorporated a review of the impact data from the spring 2023 administration of the assessments, the reasonable ranges for the cut scores, and the committee-recommended cut score ranges with a focus on honoring the work of the standard setting committees while establishing performance levels that would work for the assessment program. The decision by TEA from this review was to keep the cut score recommendations from the vertical articulation committee presented in Table 3.

**Final Approval**

Mike Morath, the Commissioner of Education at TEA, reviewed and approved the performance level cut scores for the STAAR Alternate 2 RLA assessments on June 22, 2023.

# Chapter 1 – Overview of the Standard Setting Process

This chapter provides an overview of the standard setting process used for the State of Texas Assessments (STAAR®) Alternate 2 Reading Language Arts (RLA) grades 3–8 assessments and includes the following sections:

- Goals of the Standard Setting Meeting
- Performance Levels
- Standard Setting Process

## Goals of the Standard Setting Meeting

Once students are administered an assessment, various groups such as students, parents, educators, administrators, and policymakers want to know how the students performed on the assessment and how to interpret their performance. By establishing performance levels associated with different student performance on the assessment, a frame of reference is developed for interpreting student scores. Establishing the level of achievement on an assessment required for classification into each performance level is one of the most critical steps in developing an assessment program.

For criterion- and standards-based assessments, achievement on the assessment is compared to a set of predefined content standards. The standards communicated within the *TEKS Curriculum Framework for STAAR Alternate 2* define a set of knowledge and skills that the students taking the assessment are expected to demonstrate upon completion of each grade. The established cut scores represent the level of competence students are expected to demonstrate on the assessment to be classified into each performance level.

## Performance Levels

Federal statute requires that any statewide assessment used for accountability purposes includes at least three performance levels.[1] The performance levels relate student performance on the STAAR Alternate 2 RLA assessments directly to what students are expected to learn based on the standards in the *TEKS Curriculum Framework for STAAR Alternate 2*. The following performance levels delineate the knowledge and skills for which students are able to demonstrate mastery:

- Level 3: *Accomplished Academic Performance*
- Level 2: *Satisfactory Academic Performance*
- Level 1: *Developing Academic Performance*

---

[1] Every Student Succeeds Act (ESSA), Pub. L. No. 114–95, Stat. 1802 (2015). See SEC. 1111, (b), (1), (A). https://congress.gov/114/plaws/publ95/PLAW-114publ95.pdf

## Standard Setting Process

The recommendations by the standard setting committees represent the level of competence students are expected to demonstrate to be classified into each performance level. To establish the performance levels for each assessment, the Modified Angoff method (Angoff, 1971) was used to guide panelists as they determined their performance level cut score recommendations. This standard setting procedure is a systematic method for combining various considerations into the process for recommending cut scores for the different performance levels, including content standards and educator judgments regarding what students should know (based on the *TEKS Curriculum Framework for STAAR Alternate 2*) and be able to demonstrate at each performance level. The following steps were used for the standard setting process:

- *Pre-meeting development*—In anticipation of the standard setting meetings, the PLDs were reviewed, panelist materials were developed, the Pearson standard setting website was prepared, facilitator presentation materials were created, and data analysis sources and procedures were developed.
- *Standard setting meetings*—Committees of panelists referenced the PLDs to make recommendations for cut scores that define the different performance levels for each assessment.
- *Post-meeting*—The recommended cut scores for each assessment were submitted to TEA for approval or modification.

The subsequent chapters describe the specific procedures and activities that occurred during each of these steps.

# Chapter 2 – Pre-meeting Development

This chapter provides an overview of the work that was completed prior to the standard setting meetings and includes the following sections:

- Performance Level Descriptors
- Pearson Standard Setting Website
- Development of Panelist Materials
- Development of Presentation Materials
- Facilitator Training
- Preparation for Data Analysis During the Meetings

## Performance Level Descriptors

Performance level descriptors (PLDs) are statements that articulate the knowledge and skills that students classified into a particular performance level should be able to do to demonstrate competency at that performance level. The use of a well-defined set of PLDs is critical to ensuring the validity of the standard setting process. All assessments within STAAR Alternate 2 have three performance levels, as indicated in Chapter 1. The PLDs are associated with the performance levels in the following ways:

- *Performance levels* indicate a student's level of competency of the standards defined in the *TEKS Curriculum Framework for STAAR Alternate 2* through classification of their achievement on an assessment for a specific content area.
- *PLDs* indicate the knowledge and skills students must demonstrate to show competency within each content area to be classified into each performance level.
- *Cut scores* partition the test scale and represent the minimum test score that a student must earn on an assessment to be classified into a performance level.

The *TEKS Curriculum Framework for STAAR Alternate 2* provides a foundation for the development of the PLDs. In developing the PLDs, descriptors were written for each reporting category associated with the respective content area for each STAAR Alternate 2 RLA performance level. The knowledge and skills described at each performance level are cumulative, assuming students at a given performance level would be able to demonstrate competency at each preceding performance level for the same reporting category.

## Pearson Standard Setting Website

The Pearson standard setting website was the online platform for the meeting pre-work, standard setting facilitation, and panelist judgment collection throughout the process. The website was built using Moodle—an online, open-source collaboration and learning tool that has been successfully used for previous standard settings, including the Partnership for Assessment of Readiness for College and Careers (PARCC), National Assessment of

Educational Progress (NAEP), Indiana (ISTEP+), Massachusetts (Next-Generation MCAS), and Kentucky (Science) standard setting meetings. Each panelist was given a unique user identification and password that provided secure access to the website. Panelist access was restricted to sections of the website associated with their specific committee.

Even though the STAAR Alternate 2 RLA assessments are delivered on paper and paper test forms were used during the standard setting meeting, panelists provided their judgments on the Pearson standard setting website. The panelists accessed the website using a computer provided by Pearson set up specifically for this meeting. The facilitator provided training to all panelists on the use of the standard setting website and any additional guidance and instruction needed throughout the meeting.

### Development of Panelist Materials

The Pearson standard setting team worked with TEA to develop the materials used by panelists during the meeting. Because the standard setting website was used as a tool for facilitation, a specific website was developed for each committee. When appropriate, documents were presented online through the website. Table 5 presents a list of the materials developed for panelists and their mode of presentation.

**Table 5. Materials Prepared for Panelists**

| Panelist Material | Paper | Online |
|---|:---:|:---:|
| Meeting agenda | ✓ | ✓ |
| Panelist information survey | | ✓ |
| Non-disclosure agreement | | ✓ |
| *TEKS Standards* | | ✓ |
| Curriculum framework | | ✓ |
| Student test book | ✓ | |
| Teacher administration book | ✓ | |
| Image cards | ✓ | |
| "Experience the Assessment" response form | ✓ | |
| Test form item map/answer key | | ✓ |
| Practice judgment items | ✓ | |
| Practice judgment record sheet | ✓ | |
| Practice judgment survey | | ✓ |
| Practice judgment form test map/answer key | | ✓ |
| Judgment items | ✓ | |
| Judgment round record sheet | ✓ | |
| Judgment round surveys | | ✓ |
| Performance level descriptors | ✓ | ✓ |
| Borderline descriptions | ✓ | |
| Process evaluations | | ✓ |

Development of the materials and standard setting website started with the creation of templates for each resource that were reviewed and approved by TEA. Using the approved templates, the resources were created for each committee by the Pearson standard setting team. TEA reviewed the committee-specific documents and resources before they were finalized for publication for the meetings.

## Development of Presentation Materials

Customized Microsoft PowerPoint presentations were developed to guide facilitators through the presentation of information and materials throughout the meetings. TEA reviewed and provided suggested edits to the presentations that were resolved by the Pearson standard setting team. The following PowerPoint presentations were created:

- General session overview
- Breakout sessions

Presentation notes that coincide with the PowerPoint slides were developed for each presentation to guide facilitators. The notes provided information for each breakout meeting, including procedural steps, talking points, definitions to explain concepts to panelists, answers to commonly asked questions, and specific materials to distribute to panelists during the meeting.

## Facilitator Training

The facilitators underwent an extensive training program to facilitate the standard setting meeting that included the following:

- *STAAR Alternate 2 assessments*—The facilitators were provided an overview of the STAAR Alternate 2 assessment program, including the test design, item types, scoring rules, performance levels, and scaling design.
- *Use of the Pearson standard setting website*—Because the standard setting website was used as a facilitation tool, facilitators needed to be familiar with the platform. The website outlined a framework for each facilitator to follow and provided the panelists with defined and limited access. Specific guidelines for modeling the website and providing access to panelists were discussed.
- *Standard setting process*—The facilitators participated in a walkthrough of the agenda with a focus on time management, use of the online platform, and communicating feedback information.
- *Training slides and presentation notes*—The facilitators reviewed the training slides that included notes with guidance for facilitators to follow throughout the presentation, including when specific language was to be used. The use of presentation slides and notes ensured that each committee was facilitated using the same protocol to maintain standardization of the process across meetings.

## Preparation for Data Analysis during the Meetings

Pearson analysts developed programs to generate all feedback reports needed during the standard setting meeting, including the following feedback after each judgment round:

- *Individual panelist feedback*—The individual panelists' judgments for each performance level and the resulting individual cut score recommendations to ensure that they were recorded accurately
- *Committee-level feedback*—A summary of judgments from all panelists, including frequency distributions of judgments for each performance level and the mean and median cut scores (given to facilitators and TEA and presented to the panelists using tables and histograms in the PowerPoint slides)

The analysis programs created for the standard setting meetings used panelists' judgment data from each round. Analysts downloaded the panelists' judgments from the standard setting website at the conclusion of each judgment round. Each panelist's set of judgments was summed to determine an expected test-level raw score for each performance level. The analysis program completed the computation for each panelist and calculated summary statistics for the committee, including the median cut scores that were considered the committee cut score recommendations. Between judgment rounds, the estimated performance level cut score and ranges from the judgment process were presented so panelists could compare their content judgments to those from the process.

# Chapter 3 – Standard Setting Meetings

This chapter provides details about the standard setting meeting process and includes the following sections:

- Purpose of the Standard Setting Meetings
- Committee Composition
- Facilitators and Staff
- Materials
- Procedure
- Standard Setting Meeting Proceedings
- Recommended Cut Scores from Standard Setting Committees

## Purpose of the Standard Setting Meetings

Standard setting is based, to a large degree, on the judgment of educators. Committees of educators make expert recommendations about the level of performance expected for each performance level based on their experience with different groups of students and knowledge of the assessed content. A specific process, or standard setting method, is used to capture the educator judgments and to translate these into cut scores for the performance levels. The purpose of the STAAR Alternate 2 RLA standard setting meetings was to gather expert cut score recommendations from educators across the state of Texas to define the performance levels of each STAAR Alternate 2 RLA assessment.

Student performance on the STAAR Alternate 2 RLA assessments is classified into one of three performance levels. Each standard setting committee was asked to recommend two cut scores to define the boundaries between the performance levels. These recommended cut scores represent the performance on each assessment that a student would need to meet or exceed to be classified into a performance level.

## Committee Composition

One committee was created for each STAAR Alternate 2 RLA assessment. Individuals in each meeting included three distinct groups, as illustrated in Figure 1:

- Meeting facilitators
- Committee panelists
- Observers and staff

**Figure 1. General Room Setup for the Meetings**
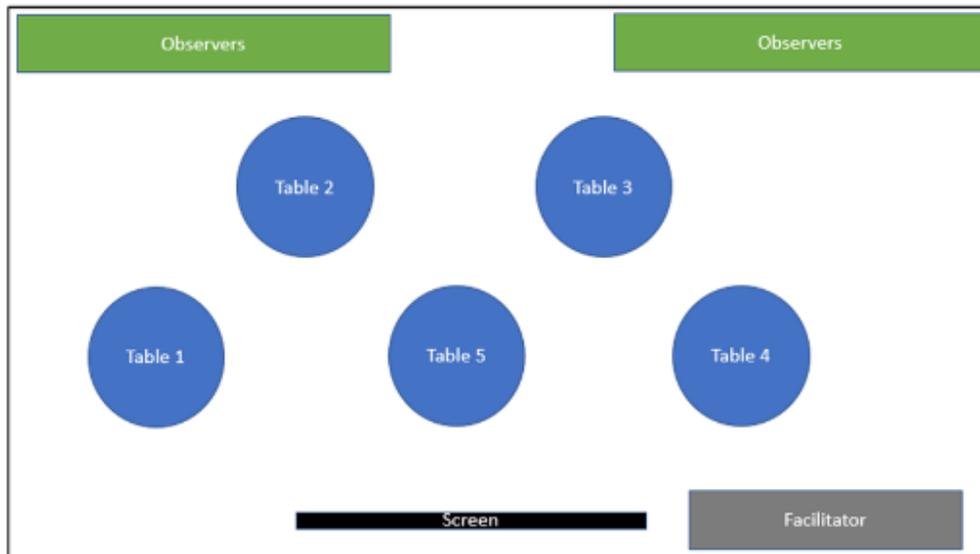


TEA selected the panelists to represent statewide educators with relevant content knowledge and experience with a variety of student groups. When setting cut scores, it is important to obtain the very best judgments from people in the best possible position to make those judgments. To meet this goal, panelists had the following qualifications:

- Were subject matter experts well-versed in the *TEKS Curriculum Framework Standards for STAAR Alternate 2*
- Understood the student special education population
- Were able to estimate item difficulty
- Understood the instructional environment
- Appreciated the consequences of the standards
- Were representative of key stakeholder groups

The six STAAR Alternate 2 RLA panels had 7–12 panelists, with a total of 62 panelists. To ensure that the panelists were subject-matter experts with grade-level expertise, recruited educators had experience in the grade for which performance standards were being established. The panelists in each committee were assigned to table groups such that each table was balanced in terms of the various demographic considerations.

Prior to the standard setting meeting, one panelist at each table was selected as the table leader for the group. The table leader was someone who had demonstrated leadership at previous educator committees (e.g., data review, content review) or someone known by TEA to be a suitable candidate for this role. The table leader assisted the facilitator in maintaining appropriate discussions among the panelists, distributing and collecting materials, maintaining established security measures, and performing other duties as deemed appropriate by the facilitator.

## Facilitators and Staff

TEA and Pearson collaborated to conduct the standard setting meetings, working in facilitative and observational roles but not contributing to the cut score recommendations during the meeting.

### Facilitators

The lead facilitator was Eric L. Moyer, Ph.D., from Pearson. Each breakout committee meeting was led by a process facilitator with knowledge and experience facilitating standard setting meetings. The facilitator was responsible for ensuring that appropriate processes were followed throughout all phases of the meeting and verifying that panelists had a solid understanding of the tasks they were being asked to complete. Content experts from Pearson and TEA were also available as observers to help answer content and policy questions that arose during the meeting. Prior to the meeting, a staffing plan was provided to TEA that communicated the psychometric, content, and support staff required to attend each committee meeting. Table 6 presents the process and content facilitators for each standard setting committee.

**Table 6. Process and Content Facilitators**

| Committee | Process Facilitator | Content Facilitator |
|-----------|---------------------|---------------------|
| Grade 3 | Jaime Kavanaugh | Kathleen McDonnold |
| Grade 4 | Jaime Kavanaugh | Kathleen McDonnold |
| Grade 5 | Kshawna Askew | Kathleen McDonnold |
| Grade 6 | Ross Markle | Megan Carlin |
| Grade 7 | Russell Keglovits | Megan Carlin |
| Grade 8 | Russell Keglovits | Megan Carlin |

### Observers

Observers did not participate in the standard setting process. The individuals that attended as observers consisted of TEA staff, vendor staff (Pearson and Cambium Assessment), content experts, and any selected evaluators. The purpose of observation was to allow individuals to experience the standard setting process and, in some cases, provide feedback. Observers were invited to attend the meeting by TEA. The number of observers in a committee meeting was kept to a maximum of 1–2 individuals so the panelists did not feel overwhelmed.

### Data Analysts

Four data analysts performed all analyses for the standard setting committees. Michelle Anderson and Toby Parker were the onsite data analysts, and Jack Kissell and Adam Bouchard were the offsite data analysts. During the meeting, the analysts collected panelist judgment data, performed independent analysis to verify analysis results, and prepared the feedback presented to the panelists.

**TEA Staff**

TEA staff members attended the standard setting meeting to observe the process, answer assessment and curriculum questions, and address policy questions. TEA also monitored the cut score recommendations for each performance level throughout the standard setting meetings. TEA was represented at the meeting by Chris Rozunick, Director, Assessment Development Division, and Dr. Mi-Suk Shim, Director of Psychometric Services. Additional TEA staff assisted these individuals in monitoring the standard setting meeting, including content specialists and psychometricians.

## Materials

Extensive materials were needed for the successful implementation of the standard setting meeting. This section describes the development of meeting materials.

**Pearson Standard Setting Website**

The Pearson standard setting website is the online platform for the meeting pre-work, facilitating the standard setting meeting, and collecting panelist judgments throughout the standard setting process. The website provided panelists access to the standard setting meeting materials and tools to record their judgments. Figure 2 presents an example.

**Figure 2. Example Website Interface**

Even though the assessment was paper based, the panelists provided their judgments on the Pearson standard setting website. During the meeting, panelists accessed the website using a computer provided by Pearson and set up specifically for this meeting. The facilitator provided training to all panelists on the use of the standard setting website and any additional guidance and instruction needed throughout the meeting.

**Committee Panelist Folders**

In addition to the online resources provided through the website, panelists were given a meeting folder to organize the following hard-copy materials used throughout the meeting:

- Meeting agenda
- PLDs
- "Experience the Assessment" activity response form
- Practice judgment record form
- Rounds 1, 2, and 3 judgment record form

The panelist folders were prepared in advance of the standard setting meetings. Panelists were required to check in at the start of each day and to return their folders and check out at the end of each day. Panelists were provided additional materials throughout the meeting, which they were instructed to insert into their folders.

**Computers**

Each panelist was provided a laptop computer in their meeting room to access the online resources through the Pearson standard setting website. Panelists were also provided an external monitor so they could navigate the online materials with ease. Panelists were seated in table groups in pod configuration to provide each panelist with enough space to work with the computer and folder materials. The panelists used Google Chrome to access the standard setting website, which was programmed with a whitelist of websites to restrict the panelists' use of the computers during the meeting.

## Procedure

The Modified Angoff method (Angoff, 1971) was used during the standard setting meeting to assist panelists in recommending performance level cut scores for each assessment. This standard setting procedure operated as both a content- and item-based method that led panelists through a standardized process in which they considered student expectations, as defined by the PLDs, and the knowledge and skills measured by the individual items administered to students to make judgments about student performance on each item.

For Rounds 1 and 2, the panelists made item-level judgments, as was typical of this method. The set of judgments made by panelists was used to determine both individual and committee cut score recommendations for each performance level. Panelists completed three rounds of item judgments. Between the item judgment rounds, the panelists were provided feedback information such as data relative to panelist agreement, student performance on the items, and student performance on the overall test.

During Round 3, the panelists were presented the following question for the Level 2 (*Satisfactory*) cut score:

> *"How many points would a student performing at the borderline of the specific performance level likely earn if they answered all the questions? This would be a number between 1 and 40."*

For the Level 3 (*Accomplished*) cut score, the panelists were presented the following question:

> *"How many points would a student performing at the borderline of the specific performance level likely earn if they answered all the questions? This would be a number between 1 and 40 and greater than your recommended cut score for Satisfactory."*

## Standard Setting Meeting Proceedings

The standard setting meetings were conducted across five days from June 12–16, 2023. Appendix D presents a complete agenda for the meetings, and Table 7 presents an overview.

**Table 7. Overview of Agenda**

| Session | Activity |
|---|---|
| Pre-Work | Complete Pre-Meeting Activities, Including a Review of PLDs |
| General Session | Welcome and Overview of STAAR Alternate 2 RLA Assessments |
| | Overview of Standard Setting Process |
| Breakout Session | Introductions and Process Overview |
| | "Experience the Assessment" Activity |
| | Review of PLDs |
| | Development of Borderline Descriptions |
| | Standard Setting Training |
| | Practice Judgment Activity and Discussion |
| | Round 1 Recommendations |
| | Discussion of Round 1 Results |
| | Round 2 Recommendations |
| | Discussion of Round 2 Recommendations |
| | Round 3 Recommendations |
| | Closing Remarks and Final Evaluation |

**Pre-Work**

The recruited committee panelists were registered into the Pearson standard setting website one to two weeks prior to the meeting. In an email from the website, panelists were provided with their unique user ID, a temporary password, and a link to the website. When panelists first logged in, they were required to create a unique, strong password consisting of at least eight characters, including at least one lowercase letter, one uppercase letter, one number, and one symbol.

Once panelists logged into the website, they had limited access to certain materials for their assigned committee, as this occurred before the first day of the standard setting meeting. Panelists were asked to complete a set of tasks as pre-work at a convenient time prior to attending the standard setting meeting. The required pre-work ensured that panelists were oriented and trained to perform each step of the process during the meeting. Completion of meeting pre-work also maximized the efficiency of time usage during the meetings. Pre-work activities included the following:

- Panelist information survey
- Review resource materials, including PLDs
- Standard setting training videos

A training video was included that gave a brief overview of the purpose of standard setting, what would happen at the meeting, and the role of a panelist. The standard setting website provided panelists access to the materials and activities for the pre-work, and panelist completion of the pre-work was monitored through the website. Follow-up emails were sent to panelists several days prior to the standard setting meetings to remind them to complete the pre-work if they had not done so already.

**General Session**

During the opening general session, panelists were presented an overview of the STAAR Alternate 2 RLA assessment program and the standard setting process. This information was critical for all panelists to begin the process with a common understanding of the assessment program and their role in setting performance standards. The overview included the following:

- Goals and rationale
- Legislative requirements
- Stakes for the students and teachers
- Uses for state and federal accountability purposes
- Introductions of key staff

An overview of the standard setting process, including a description of the Modified Angoff method (Angoff, 1971), was presented by the lead psychometrician from Pearson to emphasize that committees were making recommendations for other groups, including policymakers, to review and use to determine the final performance level cut scores.

**Breakout Session**

After the general session, panelists moved into their assigned committees in breakout sessions for the remainder of the meeting. Each committee was responsible for providing recommendations for cut scores for each performance level for the assessment associated with the committee during the activities described below.

**Introductions and Overview**

To begin the breakout session, the individuals in the room—facilitator, panelists, and observers—introduced themselves. After introductions, the facilitator distributed the meeting folders and reviewed the materials in the folder, the use of the website, and how those resources were to be used during the standard setting process. The panelists had an opportunity to ask questions before proceeding.

**"Experience the Assessment" Activity**

The panelists were given an overview of the overall test design and item types that appeared on the spring 2023 STAAR Alternate 2 RLA assessment for their grade-level committee. Panelists then reviewed a sample of test items that students took during the spring 2023 administration. The panelists were given time to review the sample assessment and encouraged to think from a student's perspective and take notes of the specific knowledge and skills a student would need to correctly respond to the item.

During this activity, panelists had the opportunity to score their responses to the items. This allowed panelists to understand the scoring rules for the different types of items included on the test, which also provided a good reference point for the judgment tasks that came later in the process. The panelists were trained in the specific scoring rules used for the test. Content specialists from Pearson and test development specialists from TEA were available to assist in the presentation and training on the scoring of items.

The amount of time given to panelists to complete the "Experience the Assessment" activity was less than that given to students, as it was assumed that the content experts would not need the same allotted time as students to complete the assessment. If panelists did not complete the test in the allotted time, they still had an opportunity to review items during the judgment tasks.

**Borderline Descriptions**

An essential component to the Modified Angoff standard setting method (Angoff, 1971) is the development of borderline descriptions to ensure that panelists developed a common understanding of the minimum level of knowledge and skills required to be classified into each performance level. To begin the activity, panelists reviewed the PLDs associated with their committee's grade-level STAAR Alternate 2 RLA assessment. The panelists were informed that the PLDs provided a snapshot of the typical characteristics of each performance level, including the breadth and depth of knowledge and skills demonstrated by students within the performance level.

To complete the activity, panelists were asked to consider the knowledge and skills of students performing at the borderline (i.e., a student that is "just barely" past the point of entry for that performance level). Because the assessments have three performance levels, panelists were asked to develop borderline descriptions for two performance levels (Level 2: *Satisfactory* and Level 3: *Accomplished*). Panelists were led through the following multi-step process to develop the borderline descriptions:

- *Step 1*—The facilitator modeled the creation of one or two borderline descriptions for the Level 2 performance level with the entire committee to create a framework for the activity.
- *Step 2*—The panelists worked within their table groups to review the PLDs for the Level 2 performance level. Each table group created a set of descriptions that identified the key characteristics of student performance at the borderline of Level 2. Panelists were asked to consider the following questions:
    - What would a student with performance "just barely" at Level 2 be able to do with respect to the PLDs?
    - What differentiates student performance at the borderline of Level 2 relative to a student in the middle or upper end of this level?
    - What differentiates student performance at the borderline of Level 2 relative to the upper end of Level 2?
- *Step 3*—The facilitator collected the Level 2 borderline descriptions from each group into a single document. The collected descriptions were then reviewed with the whole group for consistency in expectations. Additional edits or clarifications were made as needed.
- *Step 4*—The process was repeated for the Level 3 performance level, with panelists working in their table groups to craft borderline descriptions, followed by a whole-group review and discussion of the entire set of borderline descriptions to ensure coherence and an appropriate progression of knowledge and skills across performance levels.

The result of the whole-group discussion was a list of borderline descriptions for each performance level that was printed and provided to each panelist as a reference for the subsequent activities. The resulting borderline descriptions were not official documents and will not be published beyond the standard setting meeting. The goal of the borderline description activity was to help panelists develop a common understanding of the characteristics of performance at the borderline of each performance level.

**Item Judgment Process Training**

The panelists were provided thorough training on the steps used to make their recommendations. The Modified Angoff method was "sensitive to both the questions on the test and to the knowledge, skills, and abilities of the examinees at each transition point" (Plake & Cizek, 2012, pg. 190). For the STAAR Alternate 2 RLA assessments, the Modified Angoff method was extended to support judgments with polytomously scored items, where multiple score points were possible through partial-credit scoring. Panelists reviewed each item and answered the following question:

> *"How many points would a student performing at the borderline of the [specific] performance level likely earn if he or she answered the question?"*

Significant time was spent describing the thought process the panelists should go through using each part of the question. Discussion focused on the following parts of the question:

- *"How many points…"*—Rather than recording "yes" or "no" judgments, panelists recorded the number of points for an item.
- *"… would…"*—When considering expected student performance on an item, the panelists needed to consider how a student *would* perform rather than how they *should* perform. Where "should" was an aspirational expectation, "would" was a more realistic expectation of student performance on the item.
- *"… a student performing at the borderline of the [specific] performance level…"*—The panelists referenced the borderline descriptions for the performance level to determine how a student performing at the borderline would be expected to perform.
- *"… likely earn if he or she answered the question?"*—*Likely* was defined as two out of three times, or 67%. To help panelists understand this, the facilitator asked them to think about three students performing at the borderline of a performance level for a specific point value, starting with 1 point. If panelists believed two out of three students performing at the borderline would earn a specific number of points, they entered that number of points for that question. If not, they considered whether two out of three students performing at the borderline would earn the next lower point value for the question. If so, that value would be recorded. If not, the process would continue until a point value to which two out of three students performing at the borderline would earn was found. Zero was a possible point value.

The standard setting training included an orientation to the following components and how each was used during the process:

- *Standard setting website*—Provided access to the items used in the judgment activity and the judgment survey, where panelists recorded their individual judgments for each item and performance level
- *Operational test items*—A set of items that represented the operational test administered to students, shown in test booklets in the order they were administered during the operational test
- *Test map*—A summary of the items in the test form that included the following information:
    - Item position from the order of presentation
    - Item scoring key and item type
    - Maximum number of possible points for each item
    - *TEKS Curriculum Framework for STAAR Alternate 2* strand and essence statement(s) aligned to each item
- *Judgment record form*—The form on which panelists recorded their judgments in addition to the standard setting website for each judgment round

Panelists reviewed each item and made a judgment for each borderline performance level, starting with Level 2. Because student performance on an item was expected to increase or stay the same as the performance level increases, panelists were trained to check their judgments for expected patterns across performance levels. This training included multiple examples with different judgment patterns that were reviewed with panelists to assist them in their understanding of the judgment task. The examples included responses that followed and did not follow the expected judgment patterns, as well as floor and ceiling patterns in the judgments. The panelists' judgment data were analyzed to ensure that the judgment pattern was reasonable (i.e., that the judgment increased or remained the same with increases in the performance levels). Any panelist who provided judgment patterns that were not reasonable was removed from the analysis and indicated for additional instruction or review of the process by the facilitator.

**Practice Judgment Activity**

At the end of the training session, panelists practiced making judgments prior to beginning the actual judgment rounds. The goals of this activity were as follows:

- Give panelists experience reviewing and making judgments for different item types
- Familiarize panelists with the paper judgment record sheet and judgment survey in the standard setting website
- Build panelists' confidence in their understanding of the task to be completed

A subset of items was selected for the practice judgment activity. Items were either publicly available or a subset of the items that panelists would review during the actual judgment rounds. The practice activity included a range of item types, item difficulty, and scoring types.

Following the practice judgments, facilitators interactively showed item-level results through the standard setting website, including the percentage of panelists that selected each point value for each performance level. The facilitator walked through the judgment materials for the first few items to ensure that panelists knew where to locate key information when making their judgments. The group also discussed a few practice items to better understand that various judgments were possible. Panelists were reminded to refer to the borderline descriptions and other key considerations when making judgments. Finally, the facilitator demonstrated how the judgments were used to calculate individual and committee cut score recommendations.

**Item Judgment Rounds**

After receiving training on the standard setting process, the panelists participated in three rounds of judgments. Prior to making judgments for each round, panelists responded to a survey to indicate their readiness to participate in the standard setting activity and confirm their understanding of the task. The facilitator answered questions as needed about the upcoming activity. Panelists were not able to start the judgment survey until they answered "yes" to each of the following readiness questions:

- Do you understand your task for the judgment activity? (Rounds 1, 2, and 3)
- Are you ready to begin the judgment activity? (Rounds 1, 2, and 3)
- Do you understand the feedback data provided? (Rounds 2 and 3)

During Rounds 1 and 2, panelists independently made judgments for each item. Starting with the first item, the panelists made their judgment for the Level 2 performance level based on the borderline descriptions and the knowledge, skills, and abilities required by the item. The panelists then made judgments about the same item for the Level 3 performance level and continued the same process until all items were completed.

Judgments were recorded on the website using the judgment survey for the specific round. Panelists were provided a paper judgment sheet so they could keep a record of their judgments for each round. Once the panelists had completed their judgments for each item, they submitted their online judgment survey for analysis.

During Round 3, panelists independently completed judgments for the entire test form. As part of the Round 2 judgment feedback, panelists were provided with the sum of their individual item judgments as a reference point for the Round 3 judgments. Panelists made separate judgments for the number of points a student with performance at the borderline of each level would likely earn across all items on the test form.

After all panelists completed the judgment activity for the round, the statistical analysts from Pearson conducted an analysis of the data, applied quality control checks, and created feedback data for the panelists.

**Feedback and Discussion**

After each judgment round, the panelists were given feedback that was based on their current cut score recommendations, the recommendations of others in the committee, and relevant information from actual student results on the assessment. Feedback data included the following information about panelists' cut scores for each performance level:

- *Individual cut scores*—Item judgments for each performance level were summed to obtain a cut score for each level. The panelists were presented with their recommended cut score for each performance level, along with all their item judgments for each level.
- *Committee cut score recommendations and statistics*—Committee-level recommendations for each performance level were the median cut scores across all panelists. The committee panelists were presented with the committee-level cut score recommendations and summary statistics (minimum, maximum, median, mean, Q1, and Q3) for each performance level.
- *Panelist agreement data*—Bar graphs showed the frequency of individual cut score recommendations for each performance level and across adjacent performance levels.
- *Item-level judgment agreement across panelists*—The distribution of individual judgments for each item and performance level (i.e., item-level agreement) were provided.
- *Cut scores*—The estimated cut score for each performance level were also provided.

Table 8 summarizes the feedback information that was introduced after each judgment round. Before each round of feedback discussion, panelists were given guidance regarding the independence of their judgments. They were instructed to listen to other panelists and consider the rationales given for their judgments but to not feel pressured to change their judgments to reach a consensus.

**Table 8. Feedback Data by Round**

| Feedback Data | Round 1 | Round 2 | Round 3 |
|---|:---:|:---:|:---:|
| Panelist Judgments | ✓ | ✓ | |
| Panelist Agreement Data | ✓ | ✓ | |
| Individual Cut Scores | ✓ | ✓ | ✓ |
| Committee Cut Scores | ✓ | ✓ | ✓ |
| Panelist Agreement Data | ✓ | ✓ | |
| Benchmark Ranges | ✓ | ✓ | |

**Process Evaluation**

The validity of standard setting outcomes relies partially on the procedural validity of the meeting. Evidence of the procedural validity was gathered through evaluation surveys administered during the standard setting at specific points throughout the process, including after the practice judgment activity and after Round 3. The purpose of the evaluation surveys was to determine the perceived effectiveness of the standard setting meeting, including the panelists' understanding of the process, their comfort with the overall process, and their agreement level with the results. The evaluations were delivered through the standard setting website, and the results were aggregated and included in this technical report in Appendix I.

**Closing**

As part of the closing process, panelists returned all materials and documents used during the standard setting meeting. The panelists were instructed about the process that followed the standard setting meeting and how their cut score recommendations would be used.

**Benchmark Reasonable Ranges**

Although there were changes to test design and administration mode, the policy expectations that trended in student distribution across the performance levels did not vary significantly from previous administrations. To assist in aligning the resulting standard setting cut score recommendations with the academic expectations defined in the PLDs while also maintaining similar impact data to previous administrations, benchmark values were established as reasonable ranges.

The benchmark values represented a reasonable range for each performance level of every STAAR Alternate 2 RLA assessment based on the cut scores from the previous test design. Benchmark reasonable ranges were shared with panelists as part of the feedback data after Rounds 1 and 2. Placing the cut score recommendation within the reasonable ranges was not a requirement, but panelists were asked to provide a content-based rationale for placements outside the reasonable range.

Table 9 presents the benchmark reasonable ranges presented to the panelists. The benchmark reasonable ranges were created by mapping the performance level cut scores from the previous administration onto the spring 2023 administration standard setting form and determining ranges around each performance level. Specifically, the raw cut scores from the spring 2022 raw score look-up table, along with the associated theta values and conditional standard error of measurements (CSEMs), were determined for each performance level of each assessment. The CSEM was used to create a reasonable range around the cut score for each performance level. The reasonable range values on the spring 2022 theta scale were matched with the nearest theta values on the raw score look-up table from the 2023 pre-equated raw score look-up table for each assessment. The raw scores associated with the reasonable range of theta values from the spring 2023 pre-equated raw score look-up table were used to establish the benchmark reasonable ranges. If the raw score values associated with the maximum range of a performance level was greater than the minimum of the range of the next performance level, the minimum of the performance level range was increased to be one raw score greater than the maximum of the range of the previous performance level. Thus, the ranges indicated subsequently greater expectations.

**Table 9. Benchmark Reasonable Ranges (Raw Score Points)**

| Grade | *Satisfactory* | *Accomplished* |
|:-----:|:--------------:|:--------------:|
| 3 | 19–25 | 33–38 |
| 4 | 19–25 | 33–38 |
| 5 | 19–26 | 33–38 |
| 6 | 19–26 | 32–36 |
| 7 | 19–25 | 31–37 |
| 8 | 18–24 | 31–36 |

## Recommended Cut Scores from Standard Setting Committees

During the standard setting meeting, it was expected that there would be variation between panelists' cut score recommendations for each performance level. To determine a single cut score recommendation for a performance level for a committee, the cut score recommendations for the performance level were analyzed across panelists. Specifically, the median cut score from a set of panelists' cut score recommendations was used to determine the recommended cut score for a performance level for the committee. The recommendation resulting from Round 3 became the committee's recommendation for each performance level. Table 10 presents the recommended raw cut scores for each performance level based on the Round 3 recommendations for each grade.

**Table 10. Cut Score Recommendations from Standard Setting Committees**

| Grade | Max. Score | *Satisfactory* Cut Score | *Accomplished* Cut Score |
|-------|-----------|--------------------------|--------------------------|
| 3 | 40 | 20 | 35 |
| 4 | 40 | 21 | 35 |
| 5 | 40 | 23 | 34 |
| 6 | 40 | 25 | 33 |
| 7 | 40 | 21 | 31 |
| 8 | 40 | 19 | 31 |

Appendix F presents the committee-recommended cut scores for each performance level by round (represented as raw scores), Appendix G presents the recommended cut score summary statistics for each performance level by round, and Appendix H presents the panelists' judgment agreement data by performance level for judgment Rounds 1–3 for each committee.

# Chapter 4 – Post-Standard Setting

This chapter provides details about the work completed after the standard setting committee meetings and includes the following sections:

- Vertical Articulation
- Linear Scaling Process
- TEA Reasonableness Review
- Final Approval

## Vertical Articulation

After the Round 3 recommendations were finalized, select members of each grade-level committee became part of the vertical articulation committee. The lead psychometrician of the standard setting meeting, Eric Moyer, Ph.D., was the facilitator for the vertical articulation meeting.

The purpose of the articulation meeting was to evaluate the reasonableness of the cut score recommendations from the standard setting committees. The recommendations from the standard setting committees were made with a specific focus on the respective grade level for a committee, whereas the focus of the articulation committee was to view the cut score recommendations across grade levels (grades 3–8) to evaluate whether they resulted in a cohesive assessment program. The vertical articulation panelists were guided through the process in which they reviewed the recommendations from the standard setting committees and recommended changes if necessary, which resulted in a set of recommended cut scores from the vertical articulation committee.

### Meeting Process

The vertical articulation process involved the following steps:

- PLD cross-grade review activity
- Review and discussion of the cross-grade impact data
- Review and recommendation to recommended cut scores

The meeting began with an introduction to the articulation process. Panelists were told they would have the opportunity to review the recommended cut scores from the standard setting meetings across the grade levels to ensure that the recommendations represented a cohesive assessment program. Moreover, panelists were informed that the focus of the standard setting meetings was primarily on the content related to the grade represented by their committee, whereas the articulation meeting would emphasize reviewing the recommendations across the grade levels from a policy perspective.

To start the vertical articulation process, the panelists independently reviewed the PLDs across grades, looking for differences or similarities in student expectation across grades that could be used to explain the articulation of student impact across grades. The panelists then discussed the PLDs in table groups, discussing what their expectation would be of the articulation of the impact data across grades to establish a content-based expectation for the impact data across grades.

The panelists were then presented the grade-level impact data that reflected the results from Round 3 of all the standard setting committees. The groups discussed how the results looked across grade levels based on their initial expectations. Based on their expectations of student impact relative to their review of the PLDs, the panelists investigated changes to the recommended cut scores from Round 3 using an interactive spreadsheet accessed through the standard setting website.

The interactive spreadsheet allowed panelists to investigate possible changes to the cut scores from their committee by adjusting the current cut scores and simultaneously viewing the change to the impact data. The panelists were instructed to investigate changes to the recommended cut scores if they felt that the pattern of the impact data across grades was inconsistent with what they expected based on their review of the PLDs and their understanding of a cohesive assessment program. The changes were made directly at the cut score level. The range of individual panelist's cut score recommendations from Round 3 was used as a guide when evaluating a reasonable amount of changes to make. The panelists were aware of the need to honor the work that the standard setting committees had done and were judicious in making changes.

The committee could recommend changes to cut scores for performance levels for the grades that they determined had inconsistent results compared to their expectations of student performance across the grades. A recommended change to a cut score was entered into a master interactive spreadsheet by the facilitator for the entire committee to view. The committee viewed one recommended change at a time then discussed and either accepted or rejected the change as a group. This process was repeated until all recommended changes were discussed and the vertical articulation committee agreed with the entire set of cut score recommendations across all grades.

Table 11 presents the changes made to the recommended cut scores from the vertical articulation committee, and Table 12 presents the impact data for each grade based on the recommended cut scores from the vertical articulation process.

**Table 11. Vertical Articulation Cut Score Recommendations**

| Grade | Max. Score | Satisfactory | Accomplished |
|-------|-----------|--------------|--------------|
| 3 | 40 | 20 | 35 |
| 4 | 40 | 22 | 35 |
| 5 | 40 | 23 | 35 |
| 6 | 40 | 23 | 34 |
| 7 | 40 | 22 | 34 |
| 8 | 40 | 23 | 34 |

**Table 12. Impact Data for Vertical Articulation Cut Scores**

| Grade | Developing | Satisfactory | Accomplished |
|-------|-----------|--------------|--------------|
| 3 | 10% | 74% | 16% |
| 4 | 10% | 69% | 21% |
| 5 | 11% | 62% | 27% |
| 6 | 9% | 59% | 32% |
| 7 | 7% | 63% | 30% |
| 8 | 8% | 58% | 34% |

**Process Evaluation Survey**

At the end of the vertical articulation meeting, panelists completed an evaluation survey on the website about each panelists' experience in the vertical articulation meeting. The evaluation asked panelists to provide feedback on the following information:

- The level of success of the various meeting components
- The usefulness of the activities conducted during the meeting
- The adequacy of the various meeting components
- The level of support the panelists had in setting the recommended cut scores for each performance level across all grades
- The confidence panelists had in the recommended cut score recommendations

The panelists could also provide any additional information concerning their evaluation of the process of the vertical articulation meeting through an open-response question. Appendix I presents the results of the vertical articulation evaluation survey.

## Linear Scaling Process

The recommendations from the standard setting committees were cut scores in terms of raw scores on the test. Student results are not reported as raw scores because the overall difficulty of tests may change from year to year, so the results are not comparable across years. To address this, student results on the STAAR Alternate 2 RLA assessments are reported using scale scores that are comparable across administration years.

After the standard setting meeting, it was determined that the STAAR Alternate 2 RLA reporting scale would have the following properties across all grades:

- The lowest obtainable scale score (LOSS) was set at 100.
- The highest obtainable scale score (HOSS) was not set.
- The cut score for Level 2: *Satisfactory* would be set at 300 for every grade.

The reporting scale was set using the cut score for Level 2: *Satisfactory*. While the cut scores were defined with the same scale score cut scores for Level 2: *Satisfactory* across the grades, they are not identical, and direct comparisons through averaging and aggregation across grades should not be made without study and/or statistical adjustments.

## TEA Reasonableness Review

TEA examined the performance level cut score recommendations with an additional perspective of policy expectation and historical trends in student performance. This reasonableness review incorporated a review of the impact data from the spring 2023 administration of the assessments, reasonable ranges for the cut scores, and the committee-recommended cut score ranges. The reviewers focused on honoring the work of the standard setting committees while establishing performance levels that would work for the assessment program. The decision by TEA from this review was to keep the cut score recommendations from the vertical articulation committee. Table 13 presents the final raw cut scores and the final scale score cuts on the item response theory (IRT) scale following the TEA reasonableness review, including the *a* and *b* scaling constants.

## Table 13. Final Recommended Cut Scores on the IRT Scale

| Grade | Satisfactory (Raw Score) | Accomplished (Raw Score) | Satisfactory (IRT) | Accomplished (IRT) | A (Slope) | B (Intercept) |
|---|---|---|---|---|---|---|
| 3 | 20 | 35 | 300 | 388 | 51.7409 | 300.6002 |
| 4 | 22 | 35 | 300 | 380 | 52.5281 | 289.7045 |
| 5 | 23 | 35 | 300 | 374 | 52.1646 | 285.3261 |
| 6 | 23 | 34 | 300 | 370 | 52.7711 | 284.9813 |
| 7 | 22 | 34 | 300 | 378 | 53.4243 | 290.3676 |
| 8 | 23 | 34 | 300 | 371 | 50.2019 | 283.8651 |

*Note*. The lowest obtainable scale score (LOSS) was set at 100, and no highest obtainable scale score (HOSS) was set.

## Final Approval

Mike Morath, the Commissioner of Education at TEA, approved the performance level cut scores for the STAAR Alternate 2 RLA grades 3–8 assessments on June 22, 2023.

# Chapter 5 – Evidence of Procedural Validity of the Standard Setting Process

This chapter details various evidence for the validity of the process used during the standard setting meetings and includes the following sections:

- Committee Representation
- Committee Training
- Panelists' Perceived Validity of the Workshop

## Committee Representation

As part of the standard setting evaluation, panelists completed a demographic survey that collected information about their background relevant to educational experience. Appendix C presents the results of the self-reported demographic characteristics of the panelists.

Panelists provided their current professional position (Table C.1) and their number of years teaching the grade related to their standard setting committee (Table C.3). Most panelists in each committee were teachers in grades K–12, and most panelists (in five of the six committees) had more than 10 years of experience. The panelists' experience included teaching different populations of students, as displayed in Table C.4. Most panelists in each committee had experience teaching general education, self-contained special education, mainstream special education, and emergent bilingual (EB) students.

All panelists were working in school districts at the time of the meeting, as presented in Table C.9. The panelists that worked within school districts represented the various types of districts across the state, including size, type, and socioeconomic status. Nevertheless, across the six committees, most panelists taught in districts that had either a low or moderate socioeconomic status.

## Committee Training

During the standard setting meeting, it was essential that panelists understood how to make judgments as part of the Modified Angoff method. The training on the standard setting methodology was provided during the general session and in the individual standard setting committees and was standardized across committees through the PowerPoint training slides.

Panelists completed a practice judgment round to implement the standard setting methodology without consequence, including making judgments within the standard setting website. During the practice activity, the panelists reviewed a reduced set of items and provided judgments for Level 2: *Satisfactory* and Level 3: *Accomplished*.

After the practice round, the facilitator led a whole-group discussion to identify and respond to any questions or issues panelists encountered while implementing the standard setting process. Before each judgment round, panelists responded to a readiness survey that asked whether they were prepared to make their judgments. Panelists were not able to continue to the judgment survey unless they answered "yes" to both readiness questions. They were encouraged to ask the facilitator questions if they responded "no" to either question.

At various points during the standard setting meeting, panelists completed a process evaluation survey to record their impressions of the effectiveness of the materials and methods employed throughout the process. Figure 3 presents the results of the evaluation survey across grade-level committees for several questions related to the training on the standard setting process. Appendix J presents the results of these process evaluations for each individual committee.

**Figure 3. Evaluation Results on Standard Setting Process Training Activities**

| Introduction to the standard setting process | Rating | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|---|
| | Not successful | – | – | – | – | – | – |
| | Partially successful | – | – | – | – | – | – |
| | Successful | 4 | 4 | – | 2 | 2 | 5 |
| | Very Successful | 8 | 7 | 10 | 7 | 6 | 7 |

| Practice exercise for the standard setting procedures | Rating | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|---|
| | Not successful | – | – | – | – | – | – |
| | Partially successful | – | – | – | 1 | – | – |
| | Successful | 3 | 4 | 3 | 3 | 2 | 4 |
| | Very Successful | 9 | 7 | 7 | 5 | 6 | 8 |

| Training provided on the standard setting process | Rating | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|---|
| | Not adequate | – | – | – | – | – | – |
| | Somewhat adequate | – | – | 1 | – | – | – |
| | Adequate | 5 | 2 | 2 | 4 | 2 | 5 |
| | More than adequate | 7 | 9 | 7 | 5 | 6 | 7 |

Panelists were asked about the effectiveness of the training they received on the standard setting process. One question asked panelists to rate the level of success of the initial introduction to the standard setting process during the general session. Overall, the panelists perceived the initial introduction to the standard setting process as successful, with all panelists for each grade responding that it was either *Successful* or *Very Successful*. However, for one panelist in the grade 6 committee, a similar favorable balance of responses held for the practice judgment activity was also provided for the standard setting process. One panelist in the grade 5 committee thought that the training on the standard setting process in the breakout groups was only *Somewhat Adequate*, but most panelists perceived that the training provided on the standard setting process was either *Adequate* or *More than Adequate*.

## Panelists' Perceived Validity of the Workshop

Panelists communicated their perceived validity of the standard setting process and the recommended cut scores as part of the standard setting process evaluation. Evaluations are important evidence for establishing the validity of recommended cut scores for the performance levels.

**Panelist Evaluations**

The panelists were satisfied with their recommendations and with the overall standard setting process. As part of the process evaluation from each committee, the panelists indicated their confidence that the PLDs were reasonable for each performance level. Figure 4 presents the results of the evaluation survey across committees and indicates that the PLDs were reasonable for each of the performance levels. Appendix I presents the results for all evaluation survey questions.

**Figure 4. Evaluation Results on Reasonableness of the PLDs for Each Performance Level**

| Satisfactory | Rating | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|---|
| | Not confident | – | – | – | – | – | – |
| | Somewhat confident | – | – | – | – | – | – |
| | Confident | 2 | 2 | 6 | 1 | 1 | 3 |
| | Very Confident | 10 | 9 | 5 | 8 | 6 | 9 |

| Accomplished | Rating | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|---|
| | Not confident | – | – | – | – | – | – |
| | Somewhat confident | – | – | – | – | – | – |
| | Confident | 3 | 2 | 5 | 1 | 1 | 3 |
| | Very Confident | 9 | 9 | 6 | 8 | 6 | 9 |

All panelists were either *Confident* or *Very Confident* that the PLDs were reasonable for both performance levels. The grade 5 committee panelists were divided between *Confident* and *Very Confident*, whereas most panelists on the other five committees were *Very Confident* in the PLDs. Given that most panelists were either *Confident or Very Confident* in the PLDs for all grade levels and performance levels, these responses provide evidence that, overall, the panelists perceived the PLDs as providing reasonable expectations for each performance level.

The panelists also indicated their confidence in the cut scores recommended by the standard setting committees. Figure 5 presents the results of the evaluation survey across committees for their confidence in the recommended cut scores. Appendix I presents the results for all evaluation survey questions.

**Figure 5. Evaluation Results on Reasonableness of Cut Scores for Each Performance Level**

| Satisfactory | Rating | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|---|
| | Not confident | – | – | – | – | – | – |
| | Somewhat confident | – | – | – | – | – | – |
| | Confident | 1 | 1 | 5 | 1 | 1 | 3 |
| | Very Confident | 11 | 10 | 6 | 8 | 6 | 9 |

| Accomplished | Rating | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|---|
| | Not confident | – | – | – | – | – | 1 |
| | Somewhat confident | – | – | – | – | – | – |
| | Confident | 3 | 1 | 4 | 1 | 1 | 4 |
| | Very Confident | 9 | 10 | 7 | 8 | 6 | 7 |

The confidence in the *Satisfactory* cut score at all grades was either *Confident* or *Very Confident*, although the confidence was not consistent among the grade 5 panelists who were again split between *Confident* and *Very Confident*. All but one panelist (in the grade 8 committee) had some level of confidence in the *Accomplished* cut score. Overall, this feedback from the standard setting panelists provides evidence for the validity of the cut score recommendations for each performance level.

# References

Angoff, W. H. (1971). Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed, pp. 508–600). American Council on Education.

Plake, B. S., & Cizek, G. J. (2012). Variations on a theme. In G. J. Cizek (Ed.), *Setting performance standards* (2nd ed.). Routledge.

# Appendix A – Performance Level Descriptors

**State of Texas Assessments of Academic Readiness (STAAR®) Alternate 2**
**Performance Level Descriptors**
**Grade 3**

| Performance Level Descriptors |
|---|
| **When reading texts of increasing complexity,\* students achieving Accomplished Performance can do the following:** |
| • Make inferences and use text evidence to support understanding of informational and narrative text, with adult assistance |
| • Describe the main events in literary texts |
| • Identify the author's purpose for writing text |
| • Revise drafts by adding details using pictures or words |
| **When reading texts of increasing complexity,\* students achieving Satisfactory Performance can do the following:** |
| • Use illustrations and texts to clarify word meanings |
| • Evaluate details to determine what is most important |
| • Describe main characters in literary texts |
| • Identify declarative sentences that begin with a capital letter and end with a period |
| **When reading texts of increasing complexity,\* students achieving Developing Performance can do the following:** |
| • Identify and use words that name categories such as colors, shapes, actions, and locations, with adult assistance |
| • Synthesize information to create new understanding, with adult assistance |
| • Identify main characters in literary texts, with adult assistance |
| • Identify capital letters, with adult assistance |

\*Text complexity increases from grade to grade. Texts can become increasingly complex because the vocabulary/use of language may be more varied and challenging; sentence structures may be more varied and sophisticated; the topic/content may be less familiar or more challenging; and relationships among ideas may require more reasoning and inferential thinking. This applies to text that is read aloud to the student, as well as text the student may be able to read independently.

**State of Texas Assessments of Academic Readiness (STAAR®) Alternate 2**
**Performance Level Descriptors**
**Grade 4**

| Performance Level Descriptors |
|---|
| **When reading texts of increasing complexity,\* students achieving Accomplished Performance can do the following:** |
| • Use text to hear or read to clarify the meaning of unfamiliar words |
| • Make inferences and use text evidence to support understanding of informational and narrative text, with adult assistance |
| • Identify the problem and resolution and the importance of setting in literary texts |
| • Identify the author's purpose for writing text |
| • Identify a complete sentence |
| • Revise drafts by adding details using pictures or words |
| **When reading texts of increasing complexity,\* students achieving Satisfactory Performance can do the following:** |
| • Use illustrations in text to hear or read to clarify meaning of unfamiliar words |
| • Evaluate details to determine what is most important |
| • Identify elements of the plot including setting and main events in literary texts |
| • Identify punctuation marks including period, question mark, and exclamation point |
| **When reading texts of increasing complexity,\* students achieving Developing Performance can do the following:** |
| • Identify and use words that name categories such as textures, sequences, and positions, with adult assistance |
| • Synthesize information to create new understanding, with adult assistance |
| • Identify the main characters and setting in literary text, with adult assistance |
| • Identify a capital letter at the beginning of a sentence, with adult assistance |

\*Text complexity increases from grade to grade. Texts can become increasingly complex because the vocabulary/use of language may be more varied and challenging; sentence structures may be more varied and sophisticated; the topic/content may be less familiar or more challenging; and relationships among ideas may require more reasoning and inferential thinking. This applies to text that is read aloud to the student as well as text the student may be able to read independently.

**State of Texas Assessments of Academic Readiness (STAAR®) Alternate 2**
**Performance Level Descriptors**
**Grade 5**

| Performance Level Descriptors |
|---|
| **When reading texts of increasing complexity,\* students achieving Accomplished Performance can do the following:** |
| • Use context to identify the meaning of antonyms, synonyms, and homographs |
| • Make inferences and use text evidence to support understanding of informational and narrative text, with adult assistance |
| • Identify topics in informational text and themes in literary text |
| • Identify the author's purpose for writing text |
| • Use punctuation marks at the end of declarative and interrogative sentences |
| • Revise drafts by adding or deleting words or sentences |
| **When reading texts of increasing complexity,\* students achieving Satisfactory Performance can do the following:** |
| • Use context to clarify the meaning of unfamiliar words |
| • Evaluate details to determine what is most important |
| • Identify plot elements including sequence of events, conflict, and resolution |
| • Use capital letters in the days of the week and months of the year |
| • Revise drafts by adding details using pictures or words |
| **When reading texts of increasing complexity,\* students achieving Developing Performance can do the following:** |
| • Use illustrations and text to identify or demonstrate word meanings, with adult assistance |
| • Synthesize information to create new understanding, with adult assistance |
| • Identify the main characters, main events, and setting, with adult assistance |
| • Identify punctuation marks including periods, question marks, and exclamation points, with adult assistance |

*Text complexity increases from grade to grade. Texts can become increasingly complex because the vocabulary/use of language may be more varied and challenging; sentence structures may be more varied and sophisticated; the topic/content may be less familiar or more challenging; and relationships among ideas may require more reasoning and inferential thinking. This applies to text that is read aloud to the student as well as text the student may be able to read independently.

**State of Texas Assessments of Academic Readiness (STAAR®) Alternate 2**
**Performance Level Descriptors**
**Grade 6**

| Performance Level Descriptors |
|---|
| **When reading texts of increasing complexity,\* students achieving Accomplished Performance can do the following:** <br>• Use context to determine the meaning of unfamiliar words and multiple-meaning words<br>• Make inferences and use text evidence to support understanding of informational and narrative text, with adult assistance<br>• Discuss topics in informational texts and themes in narrative texts<br>• Identify the author's purpose for writing text<br>• Use coordinating conjunctions to form compound subjects and predicates<br>• Revise drafts by adding or deleting words or sentences |
| **When reading texts of increasing complexity,\* students achieving Satisfactory Performance can do the following:** <br>• Use context within and beyond a sentence to determine the meaning of unfamiliar words<br>• Evaluate details to determine what is most important<br>• Describe plot elements including sequence of events, conflict, and resolution<br>• Revise drafts by adding details using pictures or words |
| **When reading texts of increasing complexity,\* students achieving Developing Performance can do the following:** <br>• Use illustrations and texts to clarify word meanings, with adult assistance<br>• Synthesize information to create new understanding, with adult assistance<br>• Identify characters, events, or settings, with adult assistance<br>• Identify the capital letter at the beginning of sentences and the pronoun "I," with adult assistance<br>• Revise drafts by adding details using pictures or words, with adult assistance |

\*Text complexity increases from grade to grade. Texts can become increasingly complex because the vocabulary/use of language may be more varied and challenging; sentence structures may be more varied and sophisticated; the topic/content may be less familiar or more challenging; and relationships among ideas may require more reasoning and inferential thinking. This applies to text that is read aloud to the student as well as text the student may be able to read independently.

**State of Texas Assessments of Academic Readiness (STAAR®) Alternate 2**
**Performance Level Descriptors**
**Grade 7**

| Performance Level Descriptors |
|---|
| **When reading texts of increasing complexity,\* students achieving Accomplished Performance can do the following:** <br> • Use context to determine the meaning of unfamiliar words and multiple-meaning words <br> • Make inferences and use text evidence to support understanding of informational and narrative text, with adult assistance <br> • Discuss topics in informational texts and themes in narrative texts <br> • Identify the author's purpose for writing text <br> • Use coordinating conjunctions to form compound subjects, predicates, and sentences <br> • Revise drafts by adding and deleting ideas to improve sentence structure and word choice |
| **When reading texts of increasing complexity,\* students achieving Satisfactory Performance can do the following:** <br> • Use context within and beyond a sentence to determine the meaning of unfamiliar words <br> • Evaluate details to determine what is most important <br> • Describe plot elements including sequence of events, conflict, and resolution <br> • Revise drafts by adding or deleting sentences for improved clarity |
| **When reading texts of increasing complexity,\* students achieving Developing Performance can do the following:** <br> • Use illustrations and texts to clarify word meanings, with adult assistance <br> • Synthesize information to create new understanding, with adult assistance <br> • Identify characters, events, or settings, with adult assistance <br> • Identify capitalization in days of the week and months of the year, with adult assistance <br> • Revise drafts by adding details using pictures or words, with adult assistance |

\*Text complexity increases from grade to grade. Texts can become increasingly complex because the vocabulary/use of language may be more varied and challenging; sentence structures may be more varied and sophisticated; the topic/content may be less familiar or more challenging; and relationships among ideas may require more reasoning and inferential thinking. This applies to text that is read aloud to the student as well as text the student may be able to read independently.

**State of Texas Assessments of Academic Readiness (STAAR®) Alternate 2**
**Performance Level Descriptors**
**Grade 8**

| Performance Level Descriptors |
|---|
| **When reading texts of increasing complexity,\* students achieving Accomplished Performance can do the following:** |
| • Use context to determine the relevant meaning of unfamiliar words or multiple-meaning words |
| • Make inferences and use text evidence to support understanding of informational and narrative text, with adult assistance |
| • Discuss topics in informational texts and themes in narrative texts |
| • Identify the author's purpose for writing text |
| • Use coordinating conjunctions to form compound subjects, predicates, and sentences |
| • Revise drafts by adding and deleting ideas to improve sentence structure and word choice |
| **When reading texts of increasing complexity,\* students achieving Satisfactory Performance can do the following:** |
| • Use context within and beyond a sentence to determine the meaning of unfamiliar words |
| • Evaluate details to determine what is most important |
| • Describe plot elements including sequence of events, conflict, and resolution |
| • Revise drafts by adding or deleting sentences for improved clarity |
| **When reading texts of increasing complexity,\* students achieving Developing Performance can do the following:** |
| • Use illustrations and texts to clarify word meanings, with adult assistance |
| • Synthesize information to create new understanding, with adult assistance |
| • Identify characters, events, or settings, with adult assistance |
| • Identify singular and plural nouns, with adult assistance |
| • Revise drafts by adding details using pictures or words, with adult assistance |

\*Text complexity increases from grade to grade. Texts can become increasingly complex because the vocabulary/use of language may be more varied and challenging; sentence structures may be more varied and sophisticated; the topic/content may be less familiar or more challenging; and relationships among ideas may require more reasoning and inferential thinking. This applies to text that is read aloud to the student as well as text the student may be able to read independently.

# Appendix B – Panelist Meeting Materials

This appendix presents examples of the materials provided to the standard setting panelists. Because the materials contained secure information, that information has been redacted from the examples in this appendix. The following materials are also not provided in the appendix:

- *Test form*—This was presented on paper to panelists as it was presented to students during the spring 2023 administration.
- *Practice item judgment set*—This was presented on paper to panelists as it was presented to students during the spring 2023 administration.

**Panelist Agendas**

# STAAR Alternate 2 RLA Assessments
# Standard Setting Meeting
# June 2023

## Grades 3, 5, 6, and 8

Day 1—June 12

| | |
|---|---|
| *8:30 am* | *General Session* |
| | Welcome |
| | Overview of STAAR Alternate 2 RLA Assessments |
| | Standard Setting Overview |
| | |
| *9:50 am* | *Break* |
| | |
| *10:00 am* | *Breakout Sessions* |
| | Welcome and Introductions |
| | Assessment Overview |
| | Experience the Assessment Activity |
| | |
| *11:30 am* | *Lunch (Q & A with TEA)* |
| | Performance Level Descriptors Discussion |
| | Borderline Descriptor Development |
| | |
| *2:15 pm* | *Break* |
| | Borderline Descriptor Development (con't) |
| | Standard Setting Training |
| | Practice Judgment Activity and Discussion |
| | |
| *5:00 pm* | *End-of-Day* |

## Day 2—June 13

*8:30 am*          *Breakout Session*
                       Welcome
                       Standard Setting Review
                       Round 1 Judgments

*10:30 am*        *Break*
                       Round 1 Judgment Feedback and Discussion

*11:30 am*        *Lunch*
                       Round 1 Judgment Feedback and Discussion (con't)
                       Round 2 Judgments

*2:00 pm*          *Break*
                       Round 2 Judgment Feedback and Discussion
                       Round 3 Judgments

*4:15 pm*          *Break*
                       Round 3 Judgment Discussion and Next Steps

*5:00 pm*          *End-of-Day*

# STAAR Alternate 2 RLA Assessments
# Standard Setting Meeting
# June 2023

## Grades 4 and 7

**Day 3—June 14**

| | | |
|---|---|---|
| *8:30 am* | *Breakout Sessions* | |
| | | Welcome and Introductions |
| | | Assessment Overview |
| | | |
| *9:45 am* | *Break* | |
| | | Experience the Assessment Activity |
| | | PLD Overview and Individual Review |
| | | |
| *11:30 am* | *Lunch* | |
| | | Performance Level Descriptors Discussion |
| | | Borderline Descriptor Development |
| | | |
| *2:00 pm* | *Break* | |
| | | Borderline Descriptor Development (con't) |
| | | Standard Setting Training Review |
| | | Round 1 Judgments |
| | | |
| *5:00 pm* | *End-of-Day* | |

## Day 4—June 15

*8:30 am*              *Breakout Session*

> Welcome
> Standard Setting Process Review
> Round 1 Judgment Feedback and Discussion

*10:30 am*             *Break*

> Round 2 Judgments

*11:45 am*             *Lunch*

> Round 2 Judgment Feedback and Discussion

*2:00 pm*              *Break*

> Round 3 Judgments

*3:00 pm*              *Break*

> Round 3 Judgment Discussion and Next Steps

*4:30 pm*              *End-of-Day*

**Non-Disclosure Agreement**

State of Texas                                         Texas Education Agency

County of _____               Texas Student Assessment Program

## PERSONAL OATH OF SECURITY AND CONFIDENTIALITY

I, _____, do solemnly swear, or affirm, that I will
                     (Print Full Name)
faithfully execute the duty imposed upon me by Sections 39.030 and 39.0303 of the Texas
Education Code (TEC) to insure the security of the assessment instruments and achievement tests,
and by my oath or affirmation do agree to safeguard the confidentiality of all assessment
instruments, assessment instrument items, or achievement tests.

This oath or affirmation is intended by me to extend to any meeting or portion of meetings held
pursuant to TEC Section 39.030 or other applicable law, in which assessment instruments or
assessment instrument items are discussed. I acknowledge that failure to abide by this, my
oath or affirmation, will make me subject to the maximum criminal and professional penalties that
can be imposed by law. Penalties involved include:

- a permanent reprimand affixed to the face of all Texas Teacher Certificates
  and other educator credentials,
- a one-year suspension of all Texas Teacher Certificates and other education
  credentials,
- a permanent cancellation of all Texas Teacher Certificates and other education
  credentials, and
- a Class C misdemeanor.

As a testament to this oath, I affix my signature below:

Executed this _____ day of _____, 20____ .

_____       _____
(School Name/Organization Affiliation)       (Signature)

_____       _____
(Work Address)       (Home Address)

_____       _____
(City and Zip Code)       (City and Zip Code)

_____       _____
(Telephone Number)       (Telephone Number)

**Experience the Assessment Notes Sheet**

Only grade 3 is presented as an example.

# Texas STAAR Alternate 2
# Standard Setting Meeting
# June 2023

## Experience the Assessment
## Notes Sheet
## Grade 3

| Sequence | |
|---|---|
| 1 | |
| 2 | |
| 3 | |
| 4 | |
| 5 | |
| 6 | |
| 7 | |

| 8 | |
|----|----|
| 9 | |
| 10 | |
| 11 | |
| 12 | |
| 13 | |
| 14 | |
| 15 | |
| 16 | |
| 17 | |
| 18 | |
| 19 | |
| 20 | |

**Item Judgment Round Record Form**

Only grade 3 is presented as an example.

**Panelist Name: _____**

# Texas STAAR Alternate 2
# Standard Setting Meeting
# June 2023

## Judgment Rounds Record Sheet
## Grade 3

*"How many points would a student with performance at the borderline of the level likely earn if they answered the question?"*

| | | Judgement Round | | | |
| --- | --- | --- | --- | --- | --- |
| | | Round 1 | | Round 2 | |
| Seq. | Item Code | Satisfactory | Accomplished | Satisfactory | Accomplished |
| 1 | | | | | |
| 2 | | | | | |
| 3 | | | | | |
| 4 | | | | | |
| 5 | | | | | |

| Seq. | Item Code | Judgement Round | | | |
| --- | --- | --- | --- | --- | --- |
| | | Round 1 | | Round 2 | |
| | | Satisfactory | Accomplished | Satisfactory | Accomplished |
| 6 | | | | | |
| 7 | | | | | |
| 8 | | | | | |
| 9 | | | | | |
| 10 | | | | | |
| 11 | | | | | |
| 12 | | | | | |
| 13 | | | | | |
| 14 | | | | | |
| 15 | | | | | |
| 16 | | | | | |
| 17 | | | | | |
| 18 | | | | | |
| 19 | | | | | |
| 20 | | | | | |

# Judgment Round 3 Record Sheet

*How many points would a student performing at the borderline of the specific performance level likely earn if they answered all the questions?*

| Level | Total Points |
|---|---|
| Satisfactory Performance Level | |
| Accomplished Performance Level | |

**Item Judgment Survey**

The survey for only the first two items is shown, and the key response is obscured for security.

*Rounds 1 and 2 Judgments*



*Round 3 Judgments*

**Process Evaluation #1**

## State of Texas Assessments of Academic Readiness (STAAR) Alternate 2 Standard Setting Meeting

### Process Evaluation Survey #1

### Alternate 2 RLA Grade 3

The purpose of this evaluation is to collect information about your experience in recommending cut scores associated with the achievement levels for the STAAR Alternate 2 assessments. Your opinions provide an important part of our evaluation of this meeting.

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

| | Not Successful | Partially Successful | Successful | Very Successful |
|---|---|---|---|---|
| Overview of the STAAR Alternate 2 assessments | ● | ○ | ○ | ○ |
| Introduction to the standard setting process | ● | ○ | ○ | ○ |
| Experiencing the actual assessment | ● | ○ | ○ | ○ |
| Discussion of the scoring of items on the assessment | ● | ○ | ○ | ○ |
| Discussion of performance level descriptors (PLDs) | ● | ○ | ○ | ○ |
| Overview of the standard setting procedure | ● | ○ | ○ | ○ |
| Practice exercise for the standard setting procedure | ● | ○ | ○ | ○ |

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

| | Very Useful | Useful | Somewhat Useful | Not Useful |
|---|---|---|---|---|
| Performance Level Descriptors (PLDs) | ● | ○ | ○ | ○ |
| Borderline Description Development | ● | ○ | ○ | ○ |
| Standard Setting Training | ● | ○ | ○ | ○ |

How adequate were the following elements of the session?

| | Not Adequate | Somewhat Adequate | Adequate | More Than Adequate |
|---|---|---|---|---|
| Training provided on the standard-setting process | ● | ○ | ○ | ○ |
| Amount of time spent training | ● | ○ | ○ | ○ |
| Total amount of time to discuss the PLDs | ● | ○ | ○ | ○ |
| Total amount of time to develop the borderline descriptions | ● | ○ | ○ | ○ |
| Total amount of time to discuss the practice judgments | ● | ○ | ○ | ○ |

**Process Evaluation #2**

## State of Texas Assessments of Academic Readiness (STAAR) Alternate 2 Standard Setting Meeting

### Process Evaluation Survey #2
### STAAR Alternate 2 RLA Grade 3

The purpose of this evaluation is to collect information about your experience with the activities of the standard setting meeting to this point. Your opinions are an important part of our evaluation of this meeting.

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

|  | | Not Successful | Partially Successful | Successful | Very Successful |
|---|---|---|---|---|---|
| Judgment rounds | ● | ○ | ○ | ○ | ○ |
| Judgment round feedback - committee-level statistics | ● | ○ | ○ | ○ | ○ |
| Judgment round feedback - panelist agreement data | ● | ○ | ○ | ○ | ○ |
| Judgment round feedback - impact data | ● | ○ | ○ | ○ | ○ |
| Discussions after each round | ● | ○ | ○ | ○ | ○ |

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

|  | | Very Useful | Useful | Somewhat Useful | Not Useful |
|---|---|---|---|---|---|
| Committee-level statistics after Round 2 | ● | ○ | ○ | ○ | ○ |
| Panelist agreement data provided after Round 1 | ● | ○ | ○ | ○ | ○ |
| Panelist agreement data provided after Round 2 | ● | ○ | ○ | ○ | ○ |
| Impact data after Round 2 | ● | ○ | ○ | ○ | ○ |
| Discussion after each judgment round | ● | ○ | ○ | ○ | ○ |

How adequate were the following elements of the session?

|  | | Not Adequate | Somewhat Adequate | Adequate | More Than Adequate |
|---|---|---|---|---|---|
| Amount of time to make judgments | ● | ○ | ○ | ○ | ○ |
| Visual presentation of the feedback provided | ● | ○ | ○ | ○ | ○ |
| Number of judgment rounds | ● | ○ | ○ | ○ | ○ |

In applying the standard-setting method, you were asked to recommend cut scores (separating three performance levels) for student performance on STAAR Alternate 2 assessments.

How confident do you feel that the Performance Level Descriptors (PLDs) for STAAR Alternate RLA Grade 3 are reasonable for each student performance level?

|  | Not Confident | Somewhat Confident | Confident | Very Confident |
|---|---|---|---|---|
| Satisfactory | ○ | ○ | ○ | ○ |
| Accomplished | ○ | ○ | ○ | ○ |

In applying the standard-setting method, you were asked to recommend cut scores (separating three performance levels) for student performance on STAAR Alternate 2 assessments.

How confident do you feel that the recommended cut scores for STAAR Alternate RLA Grade 3 are reasonable for each student performance level?

|  | Not Confident | Somewhat Confident | Confident | Very Confident |
|---|---|---|---|---|
| Satisfactory | ○ | ○ | ○ | ○ |
| Accomplished | ○ | ○ | ○ | ○ |

How adequate were the following elements of the meeting?

|  | Not Adequate | Somewhat Adequate | Adequate | More Than Adequate |
|---|---|---|---|---|
| Facilities used for the meeting | ○ | ○ | ○ | ○ |
| Computers used during the meeting | ○ | ○ | ○ | ○ |
| Pearson standard setting website for accessing materials and making judgments | ○ | ○ | ○ | ○ |
| Content review system for viewing items | ○ | ○ | ○ | ○ |
| Materials provided in the folder | ○ | ○ | ○ | ○ |
| Work space in table groups during the meeting | ○ | ○ | ○ | ○ |

Did you have adequate opportunities during the session to do the following?

|  | Not Adequate | Somewhat Adequate | Adequate | More Than Adequate |
|---|---|---|---|---|
| Express your opinions about student performance levels | ○ | ○ | ○ | ○ |
| Ask questions about the cut score and how they will be used | ○ | ○ | ○ | ○ |
| Ask questions about the process of making cut score recommendations | ○ | ○ | ○ | ○ |
| Interact with your fellow panelists | ○ | ○ | ○ | ○ |

Do you believe your opinions and judgments were treated with respect by:

|  | Yes | Sometimes | No |
|---|---|---|---|
| Fellow panelists | ○ | ○ | ○ |
| Facilitators | ○ | ○ | ○ |

Please use the space below to provide any additional comments you have regarding the standard setting process, facilitators, materials, etc.

Path: p

# Appendix C – Committee Panelist Composition

**Table C.1. Panelist Position**

| Response Option | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|
| Teacher (K–12) | 7 | 4 | 5 | 6 | 7 | 9 |
| Teacher (Higher Ed.) | – | – | – | – | – | – |
| Administrator (School) | – | 1 | 1 | 1 | 1 | – |
| Administrator (District) | 1 | – | – | – | – | – |
| Instructional Specialist/Coach | – | – | – | – | – | – |
| Other Position | 2 | 4 | 3 | – | – | 2 |
| Total | 10 | 9 | 9 | 7 | 8 | 11 |

**Table C.2. Years of Total Teaching Experience**

| Response Option | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|
| None | – | – | – | – | – | – |
| 1 to 5 years | – | 1 | – | 1 | – | – |
| 6 to 10 years | 2 | 2 | 1 | 2 | 2 | 6 |
| 11 to 15 years | 2 | – | – | 3 | 2 | 1 |
| 16 to 20 years | 5 | 3 | 3 | 1 | 2 | 1 |
| More than 20 years | 1 | 3 | 5 | – | 2 | 3 |
| Total | 10 | 9 | 9 | 7 | 8 | 11 |

**Table C.3. Years of Experience Teaching STAAR Alt2 RLA at this Grade Level**

| Response Option | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|
| None | 4 | 4 | 4 | 2 | 3 | 5 |
| 1 to 5 years | 4 | 3 | 1 | 5 | 4 | 5 |
| 6 to 10 years | 1 | 2 | 3 | – | – | – |
| 11 to 15 years | 1 | – | – | – | 1 | 1 |
| 16 to 20 years | – | – | 1 | – | – | – |
| More than 20 years | – | – | – | – | – | – |
| Total | 10 | 9 | 9 | 7 | 8 | 11 |

**Table C.4. Experience Teaching Student Populations (Check all that apply)**

| Response Option | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|
| Mainstream special education | 7 | 6 | 7 | 6 | 7 | 10 |
| Self-contained special education | 8 | 6 | 5 | 4 | 6 | 7 |
| Emergent Bilinguals (EBs) | 10 | 8 | 7 | 5 | 4 | 7 |
| General education instruction | 8 | 8 | 7 | 5 | 7 | 9 |
| Vocational technical instruction | 1 | 1 | 2 | 1 | – | 2 |

**Table C.5. Highest Degree Completed**

| Response Option | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|
| Bachelor's degree (B.A., B.S.) | 3 | 3 | 4 | 3 | 1 | 2 |
| Master's degree (M.A., M.S.) | 6 | 5 | 5 | 4 | 6 | 7 |
| Doctoral degree (Ph.D., Ed.D.) | 1 | 1 | – | – | 1 | 2 |
| Total | 10 | 9 | 9 | 7 | 8 | 11 |

**Table C.6. Demographic: Gender**

| Response Option | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|
| Female | 9 | 8 | 9 | 7 | 7 | 9 |
| Male | 1 | 1 | – | – | 1 | 1 |
| Other/No answer | – | – | – | – | – | – |

**Table C.7. Demographic: Ethnicity**

| Response Option | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|
| Hispanic or Latino | 5 | 2 | 2 | 1 | 3 | 3 |
| Not Hispanic or Latino | 5 | 7 | 7 | 6 | 4 | 6 |
| No answer | – | – | – | – | – | – |

**Table C.8. Demographic: Race**

| Response Option | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|
| Native American or Alaskan Native | 1 | – | – | – | – | – |
| Asian | – | – | – | – | 1 | 1 |
| Black or African American | – | – | 2 | 2 | 3 | 3 |
| Native Hawaiian or Pacific Islander | – | – | – | 1 | – | – |
| White | 9 | 9 | 6 | 4 | 4 | 6 |
| No answer | – | – | – | – | – | – |

**Table C.9. Currently Work in a School District**

| Response Option | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|
| Yes | 10 | 9 | 9 | 7 | 8 | 11 |
| No (Higher Ed) | – | – | – | – | – | – |

**Table C.10. Size of School District**

| Response Option | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|
| Small | 5 | 5 | 4 | – | – | 1 |
| Medium | 2 | 2 | 4 | 5 | 5 | 5 |
| Large | 3 | 2 | 1 | 2 | 3 | 5 |

**Table C.11. Type of School District**

| Response Option | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|
| Rural | 2 | 3 | 3 | 2 | 1 | 4 |
| Metropolitan/Urban | 2 | – | 1 | 2 | 2 | 1 |
| Suburban | 6 | 6 | 5 | 3 | 5 | 6 |

**Table C.12. Socioeconomic Status of School District**

| Response Option | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|
| Low | 3 | 4 | 4 | 3 | 6 | 8 |
| Moderate | 4 | 5 | 4 | 4 | 2 | 3 |
| High | 3 | – | 1 | – | – | – |

## Appendix D – Standard Setting Meeting Agenda

<div align="center">

STAAR Alternate 2 RLA Assessments
Standard Setting Meeting
June 2023

Facilitator Agenda

</div>

Day 1: June 12

| Start Time | End Time | |
|---|---|---|
| *General Session* | | |
| 8:30 am | 8:45 am | Orientation and Security |
| 8:45 am | 9:15 am | Assessment Overview |
| 9:15 am | 9:50 am | Standard Setting Overview |
| 9:50 am | 10:00 am | *Break* |

| | | |
|---|---|---|
| *Breakout Sessions* (Grades 3, 5, 6, and 8) | | |
| 10:00 am | 10:15 am | Welcome and Orientation |
| 10:15 am | 10:30 am | Assessment Overview |
| 10:30 am | 11:30 am | Experience the Assessment |
| 11:30 am | 12:15 pm | *Lunch* |
| | | |
| 12:15 pm | 12:45 pm | PLD Overview and Discussion |
| 12:45 pm | 1:15 pm | Borderline Description Training and Modeling |
| 1:15 pm | 1:45 pm | Borderline Description Development—Level 2 Group Work |
| 1:45 pm | 2:15 pm | Borderline Description Development—Level 2 Whole-Group Discussion |
| 2:15 pm | 2:30 pm | *Break* |
| 2:30 pm | 3:00 pm | Borderline Description Development—Level 3 Group Work |
| 3:00 pm | 3:30 pm | Borderline Description Development—Level 3 Whole-Group Discussion |
| 3:30 pm | 4:00 pm | Standard Setting Training and Practice Judgments |
| 4:00 pm | 5:00 pm | Practice Judgment and Discussion |
| | 5:00 pm | *End of Day* |

<u>Day 2 : June 13</u>
Start Time       End Time
*Breakout Sessions* (Grades 3, 5, 6, and 8)

| Start Time | End Time | |
|---|---|---|
| 8:30 am | 8:45 am | Welcome and Review |
| 8:45 am | 9:00 am | Standard Setting Process Review |
| 9:00 am | 10:30 am | Round 1 Judgments |
| 10:30 am | 11:00 am | *Break (Data Analysis)* |
| 11:00 am | 11:30 am | Round 1 Judgment Feedback and Discussion |
| 11:30 am | 12:15 pm | *Lunch* |
| 12:15 pm | 1:00 pm | Round 1 Judgment Feedback and Discussion (cont.) |
| 1:00 pm | 2:00 pm | Round 2 Judgments |
| 2:00 pm | 2:30 pm | *Break (Data Analysis)* |
| 2:30 pm | 3:30 pm | Round 2 Judgment Feedback and Discussion |
| 3:30 pm | 4:15 pm | Round 3 Judgments |
| 4:15 pm | 4:30 pm | *Break (Data Analysis)* |
| 4:30 pm | 5:00 pm | Round 3 Discussion and Next Steps |
| | 5:00 pm | *End of Day* |

<u>Day 3: June 14</u>
Start Time     End Time
*Breakout Sessions* (Grades 4 and 7)

| Start Time | End Time | |
|---|---|---|
| 8:30 am | 8:50 am | Welcome and Orientation |
| 8:50 am | 9:30 am | Assessment Overview |
| 9:30 am | 10:00 am | *Break* |
| 10:00 am | 10:45 am | Experience the Assessment |
| 10:45 am | 11:30 am | PLD Overview and Individual Review |
| 11:30 am | 12:15 pm | *Lunch* |
| 12:15 pm | 12:45 pm | PLD Group Discussion |
| 12:45 pm | 1:00 pm | Borderline Description Training Review |
| 1:00 pm | 1:30 pm | Borderline Description Development—Level 2 Group Work |
| 1:30 pm | 2:00 pm | Borderline Description Development—Level 2 Whole-Group Discussion |
| 2:00 pm | 2:15 pm | *Break* |
| 2:15 pm | 2:45 pm | Borderline Description Development—Level 3 Group Work |
| 2:45 pm | 3:15 pm | Borderline Description Development—Level 3 Whole-Group Discussion |
| 3:15 pm | 3:30 pm | Standard Setting Training Review |
| 3:30 pm | 5:00 pm | Round 1 Judgments |
| | 5:00 pm | End of Day |

<u>Day 4 : June 15</u>

Start Time        End Time

*Breakout Sessions* (Grades 4 and 7)

| Start Time | End Time | |
|---|---|---|
| 8:30 am | 8:45 am | Welcome and Review |
| 8:45 am | 9:00 am | Standard Setting Process Review |
| 9:00 am | 10:30 am | Round 1 Judgment Feedback and Discussion |
| 10:30 am | 10:45 am | *Break* |
| 10:45 am | 11:45 am | Round 2 Judgments |
| 11:45 am | 12:45 pm | *Lunch (Data Analysis)* |
| 12:45 pm | 2:00 pm | Round 2 Judgment Feedback and Discussion |
| 2:00 pm | 2:15 pm | *Break* |
| 2:15 pm | 3:00 pm | Round 3 Judgments |
| 3:00 pm | 3:30 pm | *Break (Data Analysis)* |
| 3:30 pm | 4:30 pm | Round 3 Discussion and Next Steps |
| | 4:30 pm | *End of Day* |

Day 5: June 16

*Vertical Articulation (Reading Language Arts Grades 3–8)*

| | | |
|---|---|---|
| *8:30 am* | 9:00 am | Welcome and Overview of Vertical Articulation |
| 9:00 am | 10:00 am | Review of PLDs across Grade-Bands |
| | | Individual Review of PLDs |
| | | Discussion of Performance Expectations |
| | | |
| 10:00 am | 10:10 am | *Break* |
| | | |
| 10:10 am | 11:00 pm | Vertical Articulation Discussion |
| | | Review of Cross-Grade Impact Data |
| | | |
| 11:00 am | Noon | Overall Performance Level Rules Discussion |
| | | |
| Noon | | *End of Day* |

# Appendix E – Examples of Feedback Data

The following are examples of feedback data provided to panelists after each judgment round.

**Individual Item-Level Judgments**

The feedback data were provided the panelists with the actual item-level judgments that were recorded in Moodle. The data were provided so that the panelists could check that the system recorded the judgments correctly. Panelist names and the unique identification numbers (UINs) have been redacted for security purposes in this example.

**Alternate RLA Grade 3—Individual Rating—Round 1**

Table=1 Name=

| SeqN | UIN | L2 | L3 |
|------|-----|----|----|
| 1ALT |  | 2 | 2 |
| 2ALT |  | 2 | 2 |
| 3ALT |  | 0 | 1 |
| 4ALT |  | 0 | 1 |
| 5ALT |  | 2 | 2 |
| 6ALT |  | 1 | 2 |
| 7ALT |  | 1 | 2 |
| 8ALT |  | 0 | 1 |
| 9ALT |  | 0 | 1 |
| 10ALT |  | 0 | 1 |
| 11ALT |  | 0 | 1 |
| 12ALT |  | 1 | 2 |
| 13ALT |  | 1 | 2 |
| 14ALT |  | 1 | 2 |
| 15ALT |  | 0 | 1 |
| 16ALT |  | 0 | 1 |
| 17ALT |  | 2 | 2 |
| 18ALT |  | 2 | 2 |
| 19ALT |  | 1 | 2 |
| 20ALT |  | 2 | 2 |

**Individual Test-Level Recommendation**

This provided the panelists with the recommendations for test-level cut scores based on their item judgments for the *Satisfactory* and *Accomplished* performance levels.

**Alternate RLA Grade 3—Individual Cut Scores—Round 1**

Table=1 Name=

| L2 Raw Score | L3 Raw Score |
|---|---|
| 18 | 32 |

**Overall Test-Level Recommendations**

This provided the panelists with the aggregate test-level recommendation based on the individual panelists in the committee, including the number of panelists, the mean recommendation, the median recommendation, the minimum and maximum recommendation, and the first and third quartiles for each performance level.

**Alternate RLA Grade 3 Round 1 Summary Statistics—Overall**

|  | N | Mean | Median | Min. | Max. | Q1 | Q3 |
|---|---|---|---|---|---|---|---|
| L2 Raw Score | 12 | 20.33 | 19.00 | 15.00 | 31.00 | 18.50 | 21.50 |
| L3 Raw Score | 12 | 31.08 | 31.00 | 26.00 | 40.00 | 28.50 | 32.50 |

**Item-level Judgment Agreement**

This provided the panelists with item-level judgment distributions for the committees for each item. Additionally, for each performance level, the items with the greatest level of judgment disagreement were identified.

**Alternate RLA Grade 3 Round 1 Round 1 Level L2**

| SeqNo | UIN | Max. Points | 0 | 1 | 2 |
|-------|-----|-------------|-----|-----|-----|
| 1ALT | | 2 | . | 17% | 83% |
| 2ALT | | 2 | . | 58% | 42% |
| 3ALT | | 2 | 42% | 42% | 17% |
| 4ALT | | 2 | 67% | 33% | . |
| 5ALT | | 2 | . | 17% | 83% |
| 6ALT | | 2 | . | 83% | 17% |
| 7ALT | | 2 | 8% | 92% | . |
| 8ALT | | 2 | 83% | 8% | 8% |
| 9ALT | | 2 | 8% | 33% | 58% |
| 10ALT | | 2 | 17% | 58% | 25% |
| 11ALT | | 2 | 75% | 17% | 8% |
| 12ALT | | 2 | 58% | 33% | 8% |
| 13ALT | | 2 | . | 33% | 67% |
| 14ALT | | 2 | 25% | 58% | 17% |
| 15ALT | | 2 | 42% | 42% | 17% |
| 16ALT | | 2 | 58% | 33% | 8% |
| 17ALT | | 2 | . | 17% | 83% |
| 18ALT | | 2 | 8% | 8% | 83% |
| 19ALT | | 2 | 33% | 58% | 8% |
| 20ALT | | 2 | 83% | 8% | 8% |

**Test-Level Panelist Recommendation Agreement**

This feedback was presented to panelists by the facilitator. The facilitator presented bar graphs displaying the distribution of panelists' recommendations for the cut score, by raw score, for each performance level (i.e., Level 2: *Satisfactory* and Level 3: *Accomplished*). Graphs displaying consecutive performance levels on the scale graph were also presented. This figure shows the panelists' agreement for Level 2: *Satisfactory* (L2) and Level 3: Accomplished for Round 1 for Grade 8.
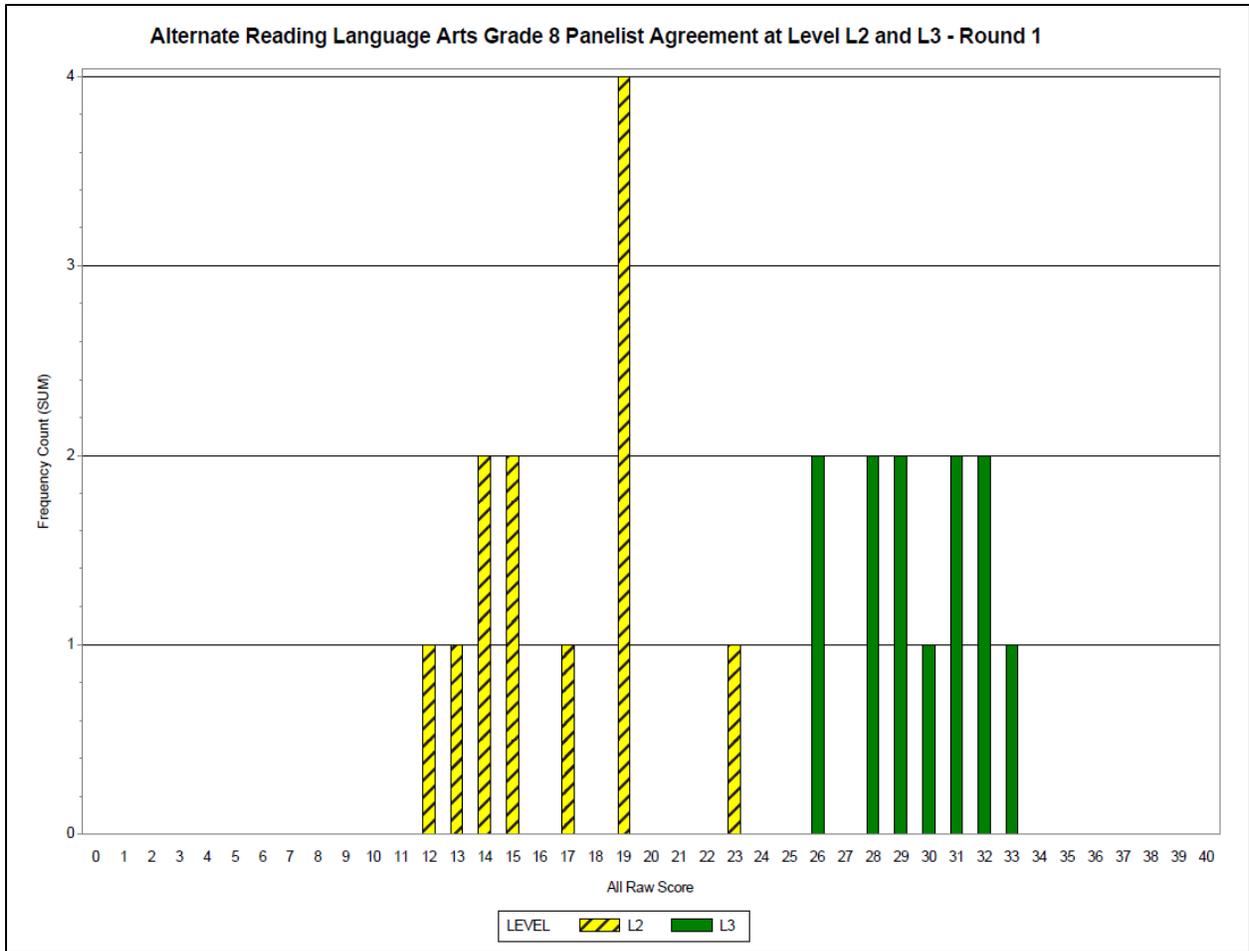
# Appendix F – Committee-Recommended Cut Scores by Round

**Table F.1. Committee-Recommended Cut Scores by Round, Grade 3**

| Performance Level | Max. Score | Round 1 | Round 2 | Round 3 |
|---|---|---|---|---|
| Level 2: *Satisfactory* | 40 | 19 | 20 | 20 |
| Level 3: *Accomplished* | 40 | 31 | 31 | 35 |

**Table F.2. Committee-Recommended Cut Scores by Round, Grade 4**

| Performance Level | Max. Score | Round 1 | Round 2 | Round 3 |
|---|---|---|---|---|
| Level 2: *Satisfactory* | 40 | 20 | 20 | 21 |
| Level 3: *Accomplished* | 40 | 32 | 33 | 35 |

**Table F.3. Committee-Recommended Cut Scores by Round, Grade 5**

| Performance Level | Max. Score | Round 1 | Round 2 | Round 3 |
|---|---|---|---|---|
| Level 2: *Satisfactory* | 40 | 20 | 22 | 23 |
| Level 3: *Accomplished* | 40 | 31 | 33 | 34 |

**Table F.4. Committee-Recommended Cut Scores by Round, Grade 6**

| Performance Level | Max. Score | Round 1 | Round 2 | Round 3 |
|---|---|---|---|---|
| Level 2: *Satisfactory* | 40 | 23 | 25 | 25 |
| Level 3: *Accomplished* | 40 | 32 | 34 | 33 |

**Table F.5. Committee-Recommended Cut Scores by Round, Grade 7**

| Performance Level | Max. Score | Round 1 | Round 2 | Round 3 |
|---|---|---|---|---|
| Level 2: *Satisfactory* | 40 | 17 | 21 | 21 |
| Level 3: *Accomplished* | 40 | 31 | 31 | 31 |

**Table F.6. Committee-Recommended Cut Scores by Round, Grade 8**

| Performance Level | Max. Score | Round 1 | Round 2 | Round 3 |
|---|---|---|---|---|
| Level 2: *Satisfactory* | 40 | 16 | 19 | 19 |
| Level 3: *Accomplished* | 40 | 30 | 31 | 31 |

# Appendix G – Recommended Cut Score Summary Statistics

**Table G.1. Recommended Cut Score Summary Statistics, Grade 3**

| Round | Statistic | *Satisfactory* | *Accomplished* |
|---|---|---|---|
| 1 | Mean | 20.33 | 31.08 |
| | Minimum | 15.0 | 26.0 |
| | Q1 | 18.5 | 28.5 |
| | **Median** | **19.0** | **31.0** |
| | Q3 | 21.5 | 32.5 |
| | Maximum | 31.0 | 40.0 |
| 2 | Mean | 19.67 | 31.58 |
| | Minimum | 10.0 | 28.0 |
| | Q1 | 17.5 | 30.0 |
| | **Median** | **20.0** | **31.0** |
| | Q3 | 22.5 | 32.0 |
| | Maximum | 26.0 | 40.0 |
| 3 | Mean | 20.70 | 33.60 |
| | Minimum | 14 | 27 |
| | Q1 | 19 | 34 |
| | **Median** | **20** | **35** |
| | Q3 | 21 | 35 |
| | Maximum | 31 | 37 |

**Table G.2. Recommended Cut Score Summary Statistics, Grade 4**

| Round | Statistic | *Satisfactory* | *Accomplished* |
|---|---|---|---|
| 1 | Mean | 20.00 | 31.91 |
| | Minimum | 16.0 | 26.0 |
| | Q1 | 19.0 | 30.0 |
| | **Median** | **20.0** | **32.0** |
| | Q3 | 21.0 | 34.0 |
| | Maximum | 26.0 | 35.0 |
| 2 | Mean | 21.09 | 34.00 |
| | Minimum | 19.0 | 32.0 |
| | Q1 | 20.0 | 33.0 |
| | **Median** | **20.0** | **33.0** |
| | Q3 | 23.0 | 35.0 |
| | Maximum | 25.0 | 36.0 |
| 3 | Mean | 21.60 | 34.60 |
| | Minimum | 20 | 33 |
| | Q1 | 20 | 34 |
| | **Median** | **21** | **35** |
| | Q3 | 23 | 35 |
| | Maximum | 25 | 37 |

**Table G.3. Recommended Cut Score Summary Statistics, Grade 5**

| Round | Statistic | *Satisfactory* | *Accomplished* |
|---|---|---|---|
| 1 | Mean | 19.55 | 32.36 |
| | Minimum | 9.0 | 28.0 |
| | Q1 | 16.0 | 29.0 |
| | **Median** | **20.0** | **31.0** |
| | Q3 | 23.0 | 35.0 |
| | Maximum | 31.0 | 39.0 |
| 2 | Mean | 21.00 | 32.09 |
| | Minimum | 18.0 | 29.0 |
| | Q1 | 19.0 | 30.0 |
| | **Median** | **22.0** | **33.0** |
| | Q3 | 23.0 | 34.0 |
| | Maximum | 26.0 | 34.0 |
| 3 | Mean | 23.50 | 34.10 |
| | Minimum | 21 | 32 |
| | Q1 | 22 | 33 |
| | **Median** | **23** | **34** |
| | Q3 | 26 | 35 |
| | Maximum | 27 | 37 |

**Table G.4. Recommended Cut Score Summary Statistics, Grade 6**

| Round | Statistic | *Satisfactory* | *Accomplished* |
|---|---|---|---|
| 1 | Mean | 21.67 | 32.22 |
| | Minimum | 13.0 | 28.0 |
| | Q1 | 19.0 | 32.0 |
| | **Median** | **23.0** | **32.0** |
| | Q3 | 26.0 | 34.0 |
| | Maximum | 27.0 | 35.0 |
| 2 | Mean | 24.89 | 34.22 |
| | Minimum | 19.0 | 32.0 |
| | Q1 | 24.0 | 33.0 |
| | **Median** | **25.0** | **34.0** |
| | Q3 | 26.0 | 35.0 |
| | Maximum | 30.0 | 40.0 |
| 3 | Mean | 24.30 | 33.30 |
| | Minimum | 21 | 32 |
| | Q1 | 23 | 32 |
| | **Median** | **25** | **33** |
| | Q3 | 26 | 34 |
| | Maximum | 27 | 35 |

**Table G.5. Recommended Cut Score Summary Statistics, Grade 7**

| Round | Statistic | Satisfactory | Accomplished |
|---|---|---|---|
| 1 | Mean | 19.25 | 30.75 |
| | Minimum | 14.0 | 28.0 |
| | Q1 | 16.0 | 28.0 |
| | **Median** | **17.0** | **31.0** |
| | Q3 | 23.5 | 33.5 |
| | Maximum | 27.0 | 34.0 |
| 2 | Mean | 20.29 | 30.14 |
| | Minimum | 17.0 | 28.0 |
| | Q1 | 18.0 | 29.0 |
| | **Median** | **21.0** | **31.0** |
| | Q3 | 22.0 | 31.0 |
| | Maximum | 25.0 | 31.0 |
| 3 | Mean | 20.00 | 30.30 |
| | Minimum | 17 | 27 |
| | Q1 | 18 | 29 |
| | **Median** | **21** | **31** |
| | Q3 | 21 | 31 |
| | Maximum | 24 | 33 |

**Table G.6. Recommended Cut Score Summary Statistics, Grade 8**

| Round | Statistic | Satisfactory | Accomplished |
|---|---|---|---|
| 1 | Mean | 16.58 | 29.58 |
| | Minimum | 12.0 | 26.0 |
| | Q1 | 14.0 | 28.0 |
| | **Median** | **16.0** | **30.0** |
| | Q3 | 19.0 | 31.5 |
| | Maximum | 23.0 | 33.0 |
| 2 | Mean | 18.00 | 30.75 |
| | Minimum | 15.0 | 28.0 |
| | Q1 | 16.5 | 30.0 |
| | **Median** | **19.0** | **31.0** |
| | Q3 | 19.0 | 32.0 |
| | Maximum | 21.0 | 33.0 |
| 3 | Mean | 18.80 | 31.20 |
| | Minimum | 17 | 29 |
| | Q1 | 19 | 30 |
| | **Median** | **19** | **31** |
| | Q3 | 19 | 32 |
| | Maximum | 20 | 34 |

# Appendix H – Test-Level Panelist Judgment Agreement

This appendix presents the raw score cuts selected by panelists for each proficiency level by round and the number of panelists who selected each cut score. Please note that the tables only show the raw score cuts that were selected and not the full range of raw scores available for each assessment.

**Table H.1. Panelist Agreement Data: Grade 3, Round 1**

| Raw Score | Satisfactory | Accomplished |
|---|---|---|
| 15 | 2 | |
| 18 | 1 | |
| 19 | 4 | |
| 20 | 1 | |
| 21 | 1 | |
| 22 | 1 | |
| 26 | 1 | 1 |
| 28 | | 2 |
| 29 | | 1 |
| 30 | | 2 |
| 31 | 1 | 1 |
| 32 | | 2 |
| 33 | | 1 |
| 34 | | 1 |
| 40 | | 1 |

**Table H.2. Panelist Agreement Data: Grade 3, Round 2**

| Raw Score | Satisfactory | Accomplished |
|---|---|---|
| 10 | 1 | |
| 15 | 1 | |
| 16 | 1 | |
| 19 | 3 | |
| 20 | 1 | |
| 22 | 2 | |
| 23 | 1 | |
| 25 | 1 | |
| 26 | 1 | |
| 28 | | 2 |
| 30 | | 3 |
| 31 | | 3 |
| 32 | | 2 |
| 36 | | 1 |
| 40 | | 1 |

**Table H.3. Panelist Agreement Data: Grade 3, Round 3**

| Raw Score | Satisfactory | Accomplished |
|---|---|---|
| 14 | 1 | |
| 18 | 1 | |
| 19 | 4 | |
| 20 | 3 | |
| 22 | 1 | |
| 27 | 1 | 1 |
| 28 | | 1 |
| 31 | 1 | |
| 33 | | 1 |
| 34 | | 3 |
| 35 | | 4 |
| 36 | | 1 |
| 37 | | 1 |

**Table H.4. Panelist Agreement Data: Grade 4, Round 1**

| Raw Score | Satisfactory | Accomplished |
|---|---|---|
| 16 | 1 | |
| 18 | 1 | |
| 19 | 3 | |
| 20 | 3 | |
| 21 | 1 | |
| 22 | 1 | |
| 26 | 1 | 1 |
| 29 | | 1 |
| 30 | | 1 |
| 32 | | 3 |
| 33 | | 1 |
| 34 | | 3 |
| 35 | | 1 |

**Table H.5. Panelist Agreement Data: Grade 4, Round 2**

| Raw Score | Satisfactory | Accomplished |
|---|---|---|
| 19 | 1 | |
| 20 | 5 | |
| 21 | 2 | |
| 23 | 2 | |
| 25 | 1 | |
| 32 | | 1 |
| 33 | | 5 |
| 35 | | 3 |
| 36 | | 2 |

**Table H.6. Panelist Agreement Data: Grade 4, Round 3**

| Raw Score | Satisfactory | Accomplished |
|---|---|---|
| 20 | 3 | |
| 21 | 3 | |
| 22 | 2 | |
| 23 | 2 | |
| 25 | 1 | |
| 33 | | 2 |
| 34 | | 3 |
| 35 | | 4 |
| 36 | | 1 |
| 37 | | 1 |

**Table H.7. Panelist Agreement Data: Grade 5, Round 1**

| Raw Score | Satisfactory | Accomplished |
|---|---|---|
| 9 | 1 | |
| 14 | 1 | |
| 16 | 1 | |
| 17 | 1 | |
| 19 | 1 | |
| 20 | 2 | |
| 22 | 1 | |
| 23 | 1 | |
| 24 | 1 | |
| 28 | | 1 |
| 29 | | 2 |
| 30 | | 1 |
| 31 | 1 | 2 |
| 34 | | 2 |
| 35 | | 1 |
| 36 | | 1 |
| 39 | | 1 |

**Table H.8. Panelist Agreement Data: Grade 5, Round 2**

| Raw Score | Satisfactory | Accomplished |
|---|---|---|
| 18 | 2 | |
| 19 | 3 | |
| 22 | 3 | |
| 23 | 2 | |
| 26 | 1 | |
| 29 | | 2 |
| 30 | | 1 |
| 32 | | 2 |

| Raw Score | Satisfactory | Accomplished |
|---|---|---|
| 33 | | 3 |
| 34 | | 3 |

**Table H.9. Panelist Agreement Data: Grade 5, Round 3**

| Raw Score | Satisfactory | Accomplished |
|---|---|---|
| 21 | 1 | |
| 22 | 4 | |
| 23 | 3 | |
| 26 | 1 | |
| 27 | 2 | |
| 32 | | 1 |
| 33 | | 3 |
| 34 | | 4 |
| 35 | | 1 |
| 36 | | 1 |
| 37 | | 1 |

**Table H.10. Panelist Agreement Data: Grade 6, Round 1**

| Raw Score | Satisfactory | Accomplished |
|---|---|---|
| 13 | 1 | |
| 15 | 1 | |
| 19 | 1 | |
| 20 | 1 | |
| 23 | 1 | |
| 25 | 1 | |
| 26 | 1 | |
| 27 | 2 | |
| 28 | | 1 |
| 29 | | 1 |
| 32 | | 3 |
| 33 | | 1 |
| 34 | | 1 |
| 35 | | 2 |

**Table H.11. Panelist Agreement Data: Grade 6, Round 2**

| Raw Score | Satisfactory | Accomplished |
|---|---|---|
| 19 | 1 | |
| 23 | 1 | |
| 24 | 2 | |
| 25 | 1 | |
| 26 | 2 | |
| 27 | 1 | |

| Raw Score | Satisfactory | Accomplished |
|---|---|---|
| 30 | 1 | |
| 32 | | 2 |
| 33 | | 2 |
| 34 | | 2 |
| 35 | | 2 |
| 40 | | 1 |

**Table H.12. Panelist Agreement Data: Grade 6, Round 3**

| Raw Score | Satisfactory | Accomplished |
|---|---|---|
| 21 | 1 | |
| 22 | 1 | |
| 23 | 1 | |
| 24 | 1 | |
| 25 | 2 | |
| 26 | 2 | |
| 27 | 1 | |
| 32 | | 3 |
| 33 | | 2 |
| 34 | | 2 |
| 35 | | 2 |

**Table H.13. Panelist Agreement Data: Grade 7, Round 1**

| Raw Score | Satisfactory | Accomplished |
|---|---|---|
| 14 | 1 | |
| 15 | 1 | |
| 17 | 3 | |
| 23 | 1 | |
| 24 | 1 | |
| 27 | 1 | |
| 28 | | 3 |
| 30 | | 1 |
| 31 | | 1 |
| 33 | | 1 |
| 34 | | 2 |

**Table H.14. Panelist Agreement Data: Grade 7, Round 2**

| Raw Score | Satisfactory | Accomplished |
|---|---|---|
| 17 | 1 | |
| 18 | 2 | |
| 21 | 2 | |
| 22 | 1 | |
| 25 | 1 | |

| Raw Score | Satisfactory | Accomplished |
|---|---|---|
| 28 | | 1 |
| 29 | | 1 |
| 30 | | 1 |
| 31 | | 4 |

**Table H.15. Panelist Agreement Data: Grade 7, Round 3**

| Raw Score | Satisfactory | Accomplished |
|---|---|---|
| 17 | 1 | |
| 18 | 2 | |
| 21 | 3 | |
| 24 | 1 | |
| 27 | | 1 |
| 29 | | 1 |
| 30 | | 1 |
| 31 | | 3 |
| 33 | | 1 |

**Table H.16. Panelist Agreement Data: Grade 8, Round 1**

| Raw Score | Satisfactory | Accomplished |
|---|---|---|
| 12 | 1 | |
| 13 | 1 | |
| 14 | 2 | |
| 15 | 2 | |
| 17 | 1 | |
| 19 | 4 | |
| 23 | 1 | |
| 26 | | 2 |
| 28 | | 2 |
| 29 | | 2 |
| 30 | | 1 |
| 31 | | 2 |
| 32 | | 2 |
| 33 | | 1 |

**Table H.17. Panelist Agreement Data: Grade 8, Round 2**

| Raw Score | Satisfactory | Accomplished |
|---|---|---|
| 15 | 2 | |
| 16 | 1 | |
| 17 | 1 | |
| 18 | 2 | |
| 19 | 4 | |
| 20 | 1 | |

| Raw Score | Satisfactory | Accomplished |
|---|---|---|
| 21 | 1 | |
| 28 | | 1 |
| 30 | | 5 |
| 31 | | 2 |
| 32 | | 3 |
| 33 | | 1 |

**Table H.18. Panelist Agreement Data: Grade 8, Round 3**

| Raw Score | Satisfactory | Accomplished |
|---|---|---|
| 17 | 1 | |
| 18 | 2 | |
| 19 | 7 | |
| 20 | 2 | |
| 29 | | 1 |
| 30 | | 3 |
| 31 | | 4 |
| 32 | | 2 |
| 33 | | 1 |
| 34 | | 1 |

# Appendix I – Panelist Evaluation Results

## Process Evaluation Survey #1

The purpose of this evaluation is to collect information about your experience with the activities of the standard setting meeting. Your opinions are an important part of our evaluation of this meeting. Select the option that best reflects your opinion about the level of success of the various components of the meeting in which you are participating. The activities are designed to help you both understand the process and be supportive of the recommendations made by the committee.

### Overview of the STAAR Alternate 2 RLA Assessment

| Rating | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|
| Not successful | – | – | – | – | – | – |
| Partially successful | 1 | – | – | – | – | – |
| Successful | 1 | 2 | 1 | 2 | 2 | 1 |
| Very Successful | 10 | 10 | 9 | 7 | 6 | 11 |

### Introduction to the Standard Setting Process

| Rating | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|
| Not successful | – | – | – | – | – | – |
| Partially successful | – | – | – | – | – | – |
| Successful | 4 | 4 | – | 2 | 2 | 5 |
| Very Successful | 8 | 7 | 10 | 7 | 6 | 7 |

### Experiencing the Actual Assessment

| Rating | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|
| Not successful | – | – | – | – | – | – |
| Partially successful | – | – | – | – | – | – |
| Successful | 4 | 4 | 1 | 2 | 2 | 2 |
| Very Successful | 8 | 7 | 9 | 7 | 6 | 10 |

### Discussion of Scoring Items on the Assessment

| Rating | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|
| Not successful | – | – | – | – | – | – |
| Partially successful | 1 | – | – | – | – | – |
| Successful | 2 | 2 | 4 | 4 | 3 | 2 |
| Very Successful | 9 | 9 | 6 | 5 | 5 | 10 |

**Discussion of Performance Level Descriptors**

| Rating | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|
| Not successful | – | – | – | – | – | – |
| Partially successful | – | – | – | – | – | – |
| Successful | 1 | 2 | 2 | 1 | 2 | 3 |
| Very Successful | 11 | 9 | 8 | 8 | 6 | 9 |

**Overview of the Standard Setting Procedure**

| Rating | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|
| Not successful | – | – | – | – | – | – |
| Partially successful | – | – | – | – | – | – |
| Successful | 3 | 4 | – | 4 | 2 | 3 |
| Very Successful | 9 | 7 | 10 | 5 | 6 | 9 |

**Practice Exercise for the Standard Setting Procedure**

| Rating | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|
| Not successful | – | – | – | – | – | – |
| Partially successful | – | – | – | 1 | – | – |
| Successful | 3 | 4 | 3 | 3 | 2 | 4 |
| Very Successful | 9 | 7 | 7 | 5 | 6 | 8 |

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

**Performance Level Descriptors**

| Rating | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|
| Very useful | 8 | 9 | 9 | 7 | 7 | 10 |
| Useful | 2 | 2 | 1 | 1 | 1 | 2 |
| Somewhat Useful | – | – | – | – | – | – |
| Not Useful | 2 | – | – | 1 | – | – |

**Borderline Description Development**

| Rating | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|
| Very useful | 8 | 9 | 7 | 8 | 7 | 9 |
| Useful | 2 | 2 | 3 | – | 1 | 2 |
| Somewhat Useful | – | – | – | – | – | 1 |
| Not Useful | 2 | – | – | 1 | – | – |

**Standard Setting Training**

| Rating | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|
| Very useful | 7 | 8 | 8 | 7 | 7 | 9 |
| Useful | 3 | 3 | 2 | 1 | 1 | 3 |
| Somewhat Useful | – | – | – | – | – | – |
| Not Useful | 2 | – | – | 1 | – | – |

How adequate were the following elements of the session?

**Training Provided on the Standard Setting Process**

| Rating | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|
| Not adequate | – | – | – | – | – | – |
| Somewhat adequate | – | – | 1 | – | – | – |
| Adequate | 5 | 2 | 2 | 4 | 2 | 5 |
| More than adequate | 7 | 9 | 7 | 5 | 6 | 7 |

**Amount of Time Spent Training**

| Rating | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|
| Not adequate | – | – | – | – | – | – |
| Somewhat adequate | – | – | – | – | – | – |
| Adequate | 7 | 2 | 1 | 2 | 2 | 5 |
| More than adequate | 5 | 9 | 9 | 7 | 6 | 7 |

**Total Amount of Time to Discuss the Performance Level Descriptors**

| Rating | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|
| Not adequate | – | – | – | – | – | – |
| Somewhat adequate | – | – | – | – | – | – |
| Adequate | 6 | 1 | 2 | 3 | 2 | 6 |
| More than adequate | 6 | 10 | 8 | 6 | 6 | 6 |

**Total Amount of Time to Develop the Borderline Descriptions**

| Rating | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|
| Not adequate | – | – | – | – | – | – |
| Somewhat adequate | – | – | – | 1 | – | – |
| Adequate | 3 | 4 | 3 | 3 | 2 | 4 |
| More than adequate | 9 | 7 | 7 | 5 | 6 | 8 |

**Total Amount of Time to Discuss the Practice Judgments**

| Rating | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|
| Not adequate | – | – | 1 | – | – | – |
| Somewhat adequate | – | – | – | – | – | 1 |
| Adequate | 7 | 3 | 3 | 4 | 2 | 4 |
| More than adequate | 5 | 8 | 6 | 5 | 6 | 7 |

## Process Evaluation Survey #2

The purpose of this evaluation is to collect information about your experience with the activities of the standard setting meeting. Your opinions are an important part of our evaluation of this meeting. Select the option that best reflects your opinion about the level of success of the various components of the meeting in which you participated. The activities are designed to help you both understand the process and be supportive of the recommendations made by the committee.

### Judgment Rounds

| Rating | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|
| Not successful | – | – | – | – | – | – |
| Partially successful | – | – | – | – | – | – |
| Successful | 3 | 1 | 4 | 1 | – | 4 |
| Very Successful | 9 | 10 | 7 | 8 | 7 | 8 |

### Judgment Round Feedback—Committee-Level Statistics

| Rating | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|
| Not successful | – | – | – | – | – | – |
| Partially successful | – | – | – | – | – | – |
| Successful | 3 | 2 | 3 | 1 | – | 3 |
| Very Successful | 9 | 9 | 8 | 8 | 7 | 9 |

### Judgment Round Feedback—Panelist Agreement Data

| Rating | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|
| Not successful | – | – | – | – | – | – |
| Partially successful | – | – | – | – | – | – |
| Successful | 4 | 1 | 3 | 1 | – | 3 |
| Very Successful | 8 | 10 | 8 | 8 | 7 | 9 |

### Judgment Round Feedback—Impact Data

| Rating | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|
| Not successful | – | – | – | – | – | – |
| Partially successful | – | – | – | – | – | – |
| Successful | 4 | 1 | 2 | 1 | – | 3 |
| Very Successful | 8 | 10 | 9 | 8 | 7 | 9 |

### Discussions after Each Round

| Rating | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|
| Not successful | – | – | – | – | – | – |
| Partially successful | – | – | – | – | – | – |
| Successful | 1 | – | 2 | 1 | – | 3 |
| Very Successful | 11 | 11 | 9 | 8 | 7 | 9 |

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

**Committee-Level Statistics after Round 2**

| Rating | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|
| Very useful | 8 | 9 | 7 | 8 | 6 | 9 |
| Useful | 3 | 2 | 3 | – | 1 | 3 |
| Somewhat Useful | – | – | – | – | – | – |
| Not Useful | 1 | – | 1 | 1 | – | – |

**Panelist Agreement Data Provided after Round 1**

| Rating | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|
| Very useful | 6 | 11 | 9 | 8 | 6 | 8 |
| Useful | 5 | – | 1 | – | 1 | 4 |
| Somewhat Useful | – | – | – | – | – | – |
| Not Useful | 1 | – | 1 | 1 | – | – |

**Panelist Agreement Data Provided after Round 2**

| Rating | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|
| Very useful | 7 | 11 | 9 | 8 | 6 | 8 |
| Useful | 4 | – | 1 | – | 1 | 4 |
| Somewhat Useful | – | – | – | – | – | – |
| Not Useful | 1 | – | 1 | 1 | – | – |

**Impact Data after Round 2**

| Rating | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|
| Very useful | 8 | 10 | 9 | 8 | 6 | 10 |
| Useful | 3 | 1 | 1 | – | 1 | 2 |
| Somewhat Useful | – | – | – | – | – | – |
| Not Useful | 1 | – | 1 | 1 | – | – |

**Discussion after Each Judgment Round**

| Rating | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|
| Very useful | 8 | 11 | 7 | 8 | 6 | 9 |
| Useful | 3 | – | 3 | – | 1 | 3 |
| Somewhat Useful | – | – | – | – | – | – |
| Not Useful | 1 | – | 1 | 1 | – | – |

How adequate were the following elements of the session?

**Amount of Time to Make Judgments**

| Rating | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|
| Not adequate | – | – | – | – | – | – |
| Somewhat adequate | – | – | – | – | – | – |
| Adequate | 3 | 1 | 1 | 2 | 1 | 3 |
| More than adequate | 9 | 10 | 10 | 7 | 6 | 9 |

**Visual Presentation of the Feedback Provided**

| Rating | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|
| Not adequate | – | – | – | – | – | – |
| Somewhat adequate | – | – | – | – | – | – |
| Adequate | 5 | 1 | – | 1 | 1 | 4 |
| More than adequate | 7 | 10 | 11 | 8 | 6 | 8 |

**Number of Judgment Rounds**

| Rating | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|
| Not adequate | – | – | – | – | – | – |
| Somewhat adequate | – | – | – | – | – | – |
| Adequate | 6 | 2 | 1 | 2 | 1 | 5 |
| More than adequate | 6 | 9 | 10 | 7 | 6 | 7 |

In applying the standard setting method, you were asked to recommend cut scores (separating three performance levels) for student performance on STAAR Alternate 2 assessments.

How confident do you feel that the Performance Level Descriptors (PLDs) for STAAR Alternate 2 RLA were reasonable for each student performance level?

**Level 2—Satisfactory**

| Rating | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|
| Not confident | – | – | – | – | – | – |
| Somewhat confident | – | – | – | – | – | – |
| Confident | 2 | 2 | 6 | 1 | 1 | 3 |
| Very Confident | 10 | 9 | 5 | 8 | 6 | 9 |

**Level 3—Accomplished**

| Rating | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|
| Not confident | – | – | – | – | – | – |
| Somewhat confident | – | – | – | – | – | – |
| Confident | 3 | 2 | 5 | 1 | 1 | 3 |
| Very Confident | 9 | 9 | 6 | 8 | 6 | 9 |

How confident do you feel that the recommended cut scores for STAAR Alternate 2 RLA were reasonable for each student performance level?

### Level 2—Satisfactory

| Rating | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|
| Not confident | – | – | – | – | – | – |
| Somewhat confident | – | – | – | – | – | – |
| Confident | 1 | 1 | 5 | 1 | 1 | 3 |
| Very Confident | 11 | 10 | 6 | 8 | 6 | 9 |

### Level 3—Accomplished

| Rating | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|
| Not confident | – | – | – | – | – | 1 |
| Somewhat confident | – | – | – | – | – | – |
| Confident | 3 | 1 | 4 | 1 | 1 | 4 |
| Very Confident | 9 | 10 | 7 | 8 | 6 | 7 |

How adequate were the following elements of the session?

### Facilities Used for the Meeting

| Rating | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|
| Not adequate | – | – | – | – | – | – |
| Somewhat adequate | – | – | – | – | – | – |
| Adequate | 4 | 3 | 2 | 1 | 1 | 4 |
| More than adequate | 8 | 8 | 9 | 8 | 6 | 8 |

### Computers Used during the Meeting

| Rating | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|
| Not adequate | – | – | – | – | – | – |
| Somewhat adequate | – | – | 1 | – | – | – |
| Adequate | 2 | 3 | 2 | 2 | 1 | 5 |
| More than adequate | 10 | 8 | 8 | 7 | 6 | 7 |

### Pearson Standard Setting Website for Accessing Materials and Making Judgments

| Rating | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|
| Not adequate | – | – | – | – | – | – |
| Somewhat adequate | – | – | – | – | – | – |
| Adequate | 3 | 1 | 5 | 1 | 1 | 5 |
| More than adequate | 9 | 10 | 6 | 8 | 6 | 7 |

**Content Review System for Viewing Items**

| Rating | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|
| Not adequate | – | – | – | – | – | – |
| Somewhat adequate | – | – | – | – | – | – |
| Adequate | 1 | 3 | 1 | 1 | 1 | 4 |
| More than adequate | 11 | 8 | 10 | 8 | 6 | 8 |

**Materials Provided in the Folder**

| Rating | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|
| Not adequate | – | – | – | – | – | – |
| Somewhat adequate | – | – | – | – | – | – |
| Adequate | 1 | 1 | 2 | 1 | 1 | 4 |
| More than adequate | 11 | 10 | 9 | 8 | 6 | 8 |

**Workspace in Table Groups during the Meeting**

| Rating | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|
| Not adequate | – | – | – | – | – | – |
| Somewhat adequate | – | – | – | – | – | – |
| Adequate | 2 | 2 | 1 | 1 | 1 | 3 |
| More than adequate | 10 | 9 | 10 | 8 | 6 | 9 |

Did you have adequate opportunities during the session to do the following?

**Express Your Opinions about Student Performance Levels**

| Rating | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|
| Not adequate | – | – | – | – | – | – |
| Somewhat adequate | – | – | – | – | – | – |
| Adequate | 1 | – | 1 | 1 | 1 | 3 |
| More than adequate | 11 | 11 | 10 | 8 | 6 | 9 |

**Ask Questions about the Cut Score and How They Will Be Used**

| Rating | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|
| Not adequate | – | – | – | – | – | – |
| Somewhat adequate | – | – | – | – | – | – |
| Adequate | 1 | – | 1 | 1 | 1 | 3 |
| More than adequate | 11 | 11 | 10 | 8 | 6 | 9 |

**Ask Questions about the Process of Making Cut Score Recommendations**

| Rating | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|
| Not adequate | – | – | – | – | – | – |
| Somewhat adequate | – | – | – | – | – | – |
| Adequate | 1 | – | 1 | 1 | 1 | 3 |
| More than adequate | 11 | 11 | 10 | 8 | 6 | 9 |

**Interact with Your Fellow Panelists**

| Rating | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|
| Not adequate | – | – | – | – | – | – |
| Somewhat adequate | – | – | 1 | – | – | – |
| Adequate | 1 | – | – | 1 | 1 | 2 |
| More than adequate | 11 | 11 | 10 | 8 | 6 | 10 |

Do you believe your opinions and judgments were treated with respect by the following people?

**Fellow Panelists**

| Rating | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|
| Yes | 11 | 11 | 10 | 8 | 7 | 12 |
| Sometimes | – | – | 1 | 1 | – | – |
| No | 1 | – | – | – | – | – |

**Facilitators**

| Rating | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|
| Yes | 11 | 11 | 11 | 9 | 7 | 12 |
| Sometimes | – | – | – | – | – | – |
| No | 1 | – | – | – | – | – |

Please use this table to provide any additional comments you have regarding the standard setting process, facilitators, materials, and any other relevant topic.

Grade 3:

- Mrs. Kavanaugh was a fantastic leader and gave us plenty of time to have discussions and never pressured us to make a decision based on what "should" be done but rather what we felt was the best based on the PLD's. This was a wonderful experience and a great process.
- This group was amazing and I love that we were able to talk thing out and take information and build on that and really get to what all our students understanding. It was very well organized and ran very well. I would love to do more of this. It is well needed.
- The gradual release of the process made it accessible to all panelists. The amount of time for discussion felt short but very productive!
- I enjoyed being able to participate in this process.
- I have been here on 3 different occasions and by far this has been one of the smoothest sessions. Everything seemed well plan, the flow of the session was smooth and the facilitator just kept the group going. Thank you
- I have enjoyed the opportunity to be a part of this process and hope to be able to participate in others in the future.
- The information was very helpful.

- Thank you for this opportunity. It helped to give me a better understanding of how the exam is scored and evaluated. It was also eye opening to know which additional skills I need to teach my students when looking at a particular TEK
- The standard setting process was difficult to understand at first, however working on it throughout the day it became easier. The facilitators did a great job.
- Thank you very much for this informative & collaborative session! This was my first time participating in this and I appreciate the opportunity to do so!
- Very informative. I am glad I was able to be part of the process,
- Having the discussions after every judging round helped so much with reviewing each test item as well the printed materials that were provided for us to take notes. I really appreciate the time to explain the process of how to do each round. Thank you so much and hope to attend to more sessions like this :)

Grade 4:

- Jamie Kavanaugh was a fantastic facilitator and very professional.
- With having going through this process the second time I think it would be great idea if possible to make a small video about what the overall process is and what it all details look like I think that might help.
- The Chromebooks were still challenging, but the 2nd screen was very useful.
- This was an enjoyable process. It was much easier the second time around. It was smart to use those did the previous grade levels and had the experience.
- I was very happy with the conversations and respect around the process. I think that nothing really needs to be changed and everything was organized and the facilitators were wonderful. Thank you for the opportunity, I learned so much about the standards and the process surrounding the creation of the test.
- I think this process has been very enlightening. This experience will help me to better serve students and teachers going forward with this new testing model.
- It was a great learning experience being part of the Standard Setting Committee. Facilitators and all supporting team from Pearson and TEA were a big part of the teaching and guidance that made our job in completing our task for the week easier! Thank you so much!
- I truly enjoyed this experience and appreciate the opportunity.
- I have a better understanding with the process in regards to standard setting. This was a great learning experience.
- We were excellently accommodated, computers, time, materials, it was a great and very informative time lots of great discussion and feedback. Hope to come back again very informative and helpful.
- Thank you so much for the opportunity to be part of this committee! It was a great learning experience and like always I learn something new. Thank you so much and hope to see you in a future standard setting committee soon!

Grade 5:

- Ms. Kshawna did an awesome job facilitating the session! Great to see Kathleen and Ms. Angie

- Our facilitator was very helpful in explaining the process and helping us accomplish the task. She would summarize what a person's thoughts were so that we could better understand each other's comments. The people from Pearson were helpful also explaining the thinking used for different items.
- I really wanted to have the student data in front of me, because it really did impact what I was still not committed to as far as scores.
- Having to download the same file 15 times just to see it was not helpful, nor was seeing the same file name for the first part of the files "output alternate Reading Language Arts Grade... Round - finally, something I can use to help me understand what file it is, but Chromebook's previews cut all that off...very frustrating!
- The facilitator and other representatives in our session were wonderful!
- Kshawna did a great Job facilitating the meetings. The support from TEA and Pearson representatives was very helpful to us.
- Big thanks for the whole team!
- I felt that this experience provided me with a deeper understanding of how cut scores are developed. I appreciate this opportunity to grow my craft as an educator.
- Thank you for the opportunity to assist on the STAAR Alt 2 panel. It was a learning experience and eye-opener to this test.
- Minor request-- more gluten free/dairy free snack options. Thank you so much!!! I enjoyed it.
- It was a very interesting and informative.
- My cut wishes I would have left my accomplished at 32 instead of bumping to 33, but I do think this will still be successful. The Kshawna was wonderful and the panelists were wonderful and knowledgeable. Thank you and I look forward to coming again.
- I love working with Kshawna Askew.

Grade 6:

- thanks for opportunity to share my opinions and hope to come be on the next one. Ross was nice.
- Russ was very knowledgeable and courteous and very nice to work with for the past two days!!
- Facilitators were amazingly helpful. The entire process was explained thoroughly.
- Very well organized and facilitated. I would definitely volunteer to serve on another panel.
- The feedback and the discussions were very helpful.
- Ross was very helpful and great the entire time. The group was amazing as well and worked well together.
- Excellent Facilitator.
- Ross was great!
- Ross was a wonderful facilitator and explained the information very well. He should definitely be invited back for future sessions.

Grade 7:

- The knowledge and experience that was gained through this process was very useful.
- The standard setting process has been a great, I have learned a lot of valuable information.

- It was a great meeting and the fellow panelists were very open and objective with data and conversations.
- Great group for this round, lots of thoughtful conversations.
- Russ was an awesome facilitator and very knowledgeable about the standard setting process!

Grade 8:

- Great training. Thank you for the opportunity.
- This was an eye-opening process that I am grateful that I was able to participate in this session. Now, I can take the details about the standard setting process back to my district and help my teachers understand how these cut scores are determined. In addition, I really appreciated our facilitator and his ability to run an effective session. Thank you for having us!
- Great process to determine cut rate scores. I gained a wealth of information.
- Participating in the standard setting process was definitely a learning experience. It provides an opportunity to learn and collaborate with other colleagues.
- I felt that the process was really helpful in making the cutoff scores. That all materials were good for us to used and make decision what needed to be done.
- This was thoroughly planned and it was quite enlightening.
- The process was a learning experience and I truly enjoyed the work that was completed. Looking at all relevant information was crucial to the final determination in order to address the required expectations of all students with cognitive disabilities in the state of Texas. I look forward to working on other data analysis in the future.
- Being this was my first experience as an Educator, I have learned a lot of valuable information. Everything is very well set up and explained. I want to do more testing item help.
- I think things went relatively well. Since many of us did not know what to expect, it was a little confusing and unorganized to begin with. I believe having a general overview of steps required and possibly some examples of what it should look like would be beneficial for some.
- This was a great experience and I learned a lot. I enjoyed the productive and healthy debate and being able to express my reasoning in the scores I had. The materials provided were useful, however we did have to have the TEA representative provide us with some additional data from the previous year on impact that was not readily available.
- This meeting was extremely beneficial. However, I do not agree with the accomplished cut scores being so low. The students that are taking this test have IQs that are typically below 70. When a student is answering level 3 & 4 questions, those TEKS are very close to grade level. This then means that if they are scoring very high, maybe this test and/or the placement is not correct for them.
- I feel the process was great, in general; however, there seemed to be a lot of wait time. I believe it was for data to be returned to our group, but I felt like some time was wasted waiting on this data to be returned.

## Process Evaluation Vertical Articulation

The purpose of this evaluation is to collect information about your experience in participating in the vertical articulation meeting for the STAAR Alternate 2 RLA assessments. Your opinions provide an important part of our evaluation of this meeting.

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

**Discussion of Recommended Changes**

| Rating | Vertical Articulation |
|---|---|
| Not Successful | – |
| Partially Successful | – |
| Successful | – |
| Very Successful | 9 |

**Use of Interactive Vertical Articulation Spreadsheet**

| Rating | Vertical Articulation |
|---|---|
| Not Successful | – |
| Partially Successful | – |
| Successful | – |
| Very Successful | 9 |

**Review of the Cross-Grade Impact Data**

| Rating | Vertical Articulation |
|---|---|
| Not Successful | – |
| Partially Successful | – |
| Successful | – |
| Very Successful | 9 |

**Review of the Performance Level Descriptors**

| Rating | Vertical Articulation |
|---|---|
| Not Successful | – |
| Partially Successful | – |
| Successful | 1 |
| Very Successful | 8 |

**Intro to Vertical Articulation Process**

| Rating | Vertical Articulation |
|---|---|
| Not Successful | – |
| Partially Successful | – |
| Successful | – |
| Very Successful | 9 |

How adequate were the following elements of the session?

**Amount of Time Working with the Interactive Spreadsheet**

| Rating | Vertical Articulation |
|---|---|
| Not Adequate | – |
| Somewhat Adequate | – |
| Adequate | 1 |
| More than Adequate | 8 |

**Amount of Time Discussing the Impact Data**

| Rating | Vertical Articulation |
|---|---|
| Not Adequate | – |
| Somewhat Adequate | – |
| Adequate | 1 |
| More than Adequate | 8 |

**Amount of Time Spent Reviewing the PLDs**

| Rating | Vertical Articulation |
|---|---|
| Not Adequate | – |
| Somewhat Adequate | – |
| Adequate | 2 |
| More than Adequate | 7 |

During this standard setting meeting, which was the grade you initially worked with?

| Grade | Vertical Articulation |
|---|---|
| 3 | 2 |
| 5 | 2 |
| 6 | 3 |
| 8 | 2 |

How confident do you feel that the final cut score recommendations for STAAR Alternate 2 RLA for this grade represent appropriate levels of student performance?

**Accomplished**

| Rating | Vertical Articulation |
|---|---|
| Not Confident | – |
| Somewhat Confident | – |
| Confident | – |
| Very Confident | 9 |

**Satisfactory**

| Rating | Vertical Articulation |
|---|---|
| Not Confident | – |
| Somewhat Confident | – |
| Confident | – |
| Very Confident | 9 |

Did you participate in the whole week of standard setting?

| Rating | Vertical Articulation |
|---|---|
| Yes | 9 |
| No | – |

What was the second grade that you participated in during the standard setting meeting?

| Grade | Vertical Articulation |
|---|---|
| 4 | 4 |
| 7 | 5 |

How confident do you feel that the final cut score recommendations for STAAR Alternate 2 RLA for the second grade represent appropriate levels of student performance?

**Accomplished**

| Rating | Vertical Articulation |
|---|---|
| Not Confident | – |
| Somewhat Confident | 1 |
| Confident | – |
| Very Confident | 8 |

**Satisfactory**

| Rating | Vertical Articulation |
|---|---|
| Not Confident | – |
| Somewhat Confident | – |
| Confident | 1 |
| Very Confident | 8 |

# Appendix J – PowerPoint Presentations

This appendix presents a sampling of presentations from the general session and breakout sessions. Full copies of the presentations are accessible by clicking on the attachments available on the left margin of your PDF reader.

**General Session**
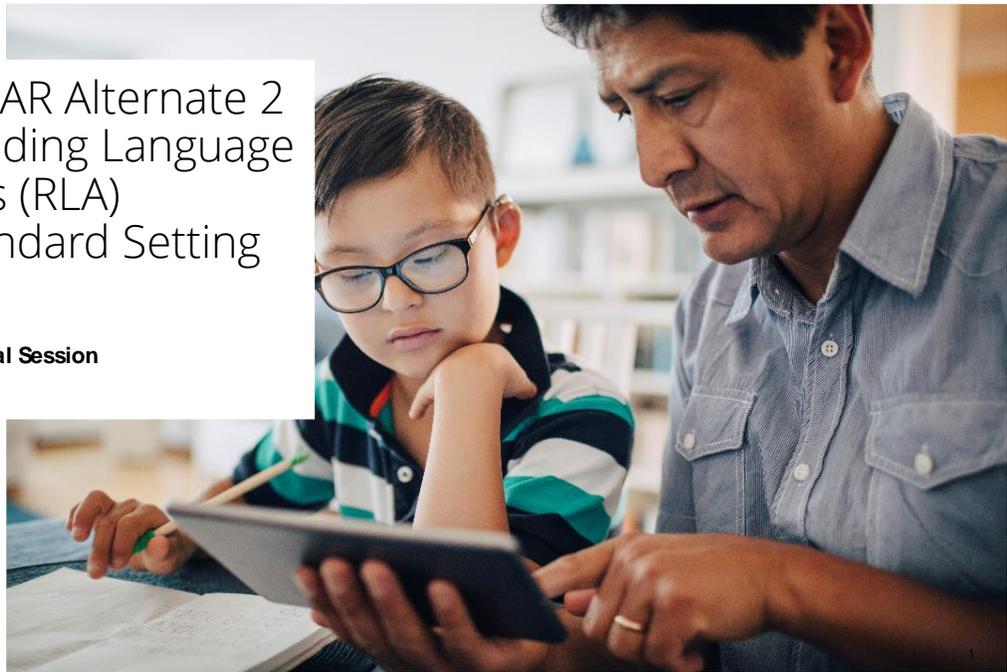


**Breakout Session – Grade 3**

**Vertical Articulation**



STAAR
Alternate 2
RLA Standard
Setting

**Vertical Articulation**