State of Texas Assessments of Academic Readiness (STAAR®) Alternate 2 End-of-Course (EOC) Standard Setting Report

STAAR Alternate 2 EOC Assessments – English I and English II

February 2023

Pearson

Version 4.1

Table of Contents

Executive Summary

This report describes the Standard Setting process for the State of Texas Assessments of Academic Readiness (STAAR®) Alternate 2 End-of-Course (EOC) assessments in English I and English II. A summary of the results is also provided.

STAAR Alternate 2 EOC 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 illustrate 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 EOC assessments is based on national best practice for standard settings. The standard setting methodology used was 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 in developing, reviewing, and approving 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 EOC assessments for English I and English II. The draft PLDs were reviewed by a panel of teachers from across the state, which provided feedback and recommended revisions for TEA to consider when finalizing the PLDs. The performance levels for the STAAR Alternate 2 EOC 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 on February 13 and 14, 2023, to recommend cut scores for the STAAR Alternate 2 EOC assessments in English I and English II. The committees were composed of 11–14 panelists who were educators with relevant subject area teaching experience. The panelists were selected to provide content expertise during the meeting and to 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 procedure. The panelists then moved to assigned breakout groups for each course. The committees first reviewed the test design and test blueprint. STAAR Alternate 2 panelists had a more detailed discussion of how the assessment is administered and scored, along with a discussion of the population of students. The panelists then began the "Experience the Assessment" activity that allowed them to view the items as a student would and to consider the knowledge and skills needed to respond to each item. Each committee then reviewed the PLDs to gain a common understanding of the expectations for the performance levels and narrowed the focus to key knowledge and skills at the borderlines separating the performance levels. They worked in small groups to create 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 performance descriptions and consider the following question for each performance level:

> *"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 at the borderline of the performance level correctly answering the item.

The panelists then engaged in a practice judgment activity using sample items, discussing the process and results to clarify their understanding of the judgment task, before completing three rounds of individual judgments. Following Rounds 1 and 2, the panelists reviewed their individual cut score recommendations and the panelists' performance level cut score agreement and took part in a whole-group discussion of items from each performance level that displayed the greatest level of disagreement in the range of item judgments. Panelists with different judgment ratings on each of item were asked to provide a rationale for their decision to develop a common understanding across the group of expectations for being classified into each performance level.

Due to score reporting timeline requirements, standard setting meetings had to occur prior to the administration to establish the performance standards for the new STAAR Alternate 2 EOC assessments that would support rapid reporting of results after the administration. Thus, no impact data were available to inform the standard setting outcome. As a surrogate for the impact data, benchmark reasonable ranges were used.

Benchmark Reasonable Ranges

Although there were changes to test design and administration mode, there are policy expectations that trends in student distribution across the performance levels would 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 each STAAR Alternate 2 EOC assessment based on the performance level cut scores obtained 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 placement outside the range.

The benchmark reasonable ranges were created by mapping the performance level cuts 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 measures (CSEM) 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 of one performance level range 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, so ranges indicated subsequently greater expectations.

Table 1 presents the benchmark reasonable ranges presented to the panelists.

Table 1. Benchmark Reasonable Ranges (Raw Score Points)

Course	*Satisfactory*	*Accomplished*
English I	18–24	30–36
English II	18–24	29–35

Results

During Round 3, panelists made their judgments at the test level instead of at the item level, as was done in Rounds 1 and 2. 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 EOC assessments.

Table 2. Standard Setting Recommendations

Course	Max. Score	*Satisfactory* Cut	*Satisfactory* Range	*Accomplished* Cut	*Accomplished* Range
English I	40	19	18–23	31	30–32
English II	40	21	18–23	31	29–33

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 about their recommendations.

TEA Reasonableness Review

To support the rapid reporting of results, the standard setting was conducted prior to the administration of the STAAR Alternate 2 EOC assessments so the TEA commissioner could review and approve the cut score recommendations in time to support the reporting of student performance on these assessments shortly after the administration window ends. The standard setting process did not include the presentation of student performance on the assessment due to when the meeting occurred. As part of the process for validating the reporting scale for the STAAR Alternate 2 EOC assessments, TEA reviewed the distribution of student classification across the performance levels based on the approved cut score recommendations from the standard setting meeting.

TEA reviewed the recommendations from the standard setting committees in a reasonableness review to examine the performance level cut score recommendations 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 2022 administration, reasonable ranges for the cut scores, and the committee-recommended cut score ranges. The focus was on honoring the work of the standard setting committees while establishing performance levels that would work for the assessment program. Table 3 presents the final cut scores for the STAAR Alternate 2 EOC assessments.

Table 3. Final Performance Level Cut Scores

Course	Max. Score	*Satisfactory*	*Accomplished*
English I	40	21	32
English II	40	21	33

Final Approval

Mike Morath, the Commissioner of Education at TEA, reviewed and approved the final performance level cut scores for the STAAR Alternate 2 EOC assessments on February 27, 2023.

Chapter 1 – Overview of the Standard Setting Process

This chapter provides an overview of the standard setting process used for State of Texas Assessments (STAAR®) Alternate 2 End-of-Course (EOC) assessments in English I and English II and includes the following sections:

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

Goals of the Standard Setting Meeting

Once an assessment is administered, various groups such as students, parents, educators, administrators, and policymakers want to know how the students performed on the assessment and how to interpret that 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 a critical step in developing an assessment program.

For criterion standards-based assessments, achievement is compared to a set of predefined content standards. These standards, communicated within the *TEKS Curriculum Framework,* define a set of knowledge and skills the students taking the assessment are expected to demonstrate upon completion of each course. The cut scores established 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] These performance levels relate student performance on the STAAR Alternate 2 EOC assessments directly to what students are expected to learn based on the standards in the *TEKS Curriculum Framework* for STAAR Alternate 2. Student achievement on all STAAR Alternate 2 EOC assessments is classified into the following three performance levels that 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* 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, the panelist materials were developed, the Pearson standard setting website was prepared, and 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 during each step.

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

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 a given performance level. The use of a well-defined set of PLDs is critical to ensuring the validity of the standard setting process. All STAAR Alternate 2 EOC assessments have three performance levels, as indicated in Chapter 1. The PLDs are associated with the performance levels in the following way:

- *Performance levels* indicate a student's level of competency of the standards defined in the *TEKS Curriculum Framework* through classification of their achievement on a STAAR Alternate 2 EOC assessment.
- *Performance level descriptors* indicate the knowledge and skills expected of students to demonstrate competency in each specific 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 for each content area to be classified into a given 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 of the three STAAR Alternate 2 EOC performance levels. The knowledge and skills described at each performance level were cumulative, assuming students at a given performance level would be able to demonstrate competency at each of the preceding performance levels, for the same reporting category.

Pearson Standard Setting Website

The Pearson standard setting website is the online platform for meeting pre-work, facilitating the standard setting meeting, and collecting panelist judgments throughout the standard setting process. The website is built using Moodle, an online, open-source collaboration and learning tool that has been successfully used for previous standard setting meetings, 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 settings. 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 EOC assessments are delivered on paper and paper test forms were used during the standard setting, 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 and to ensure that all materials were accurate. Because the meetings used the standard setting website as a tool for facilitation, a specific website was developed for each committee. When appropriate, documents were presented online through the standard setting website. Table 4 presents a list of the materials developed for panelists and their mode of presentation.

Table 4. Materials Prepared for Panelists

Panelist Material	Paper	Online
Meeting agenda	✓	✓
Panelist information survey		✓
Non-disclosure agreement		✓
Texas Essential Knowledge and Skills (TEKS) Standards		✓
Curriculum framework		✓
Subject 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	✓	

Panelist Material	Paper	Online
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		✓

The process for developing materials and the standard setting website started with the creation of templates for each resource, which were reviewed and approved by TEA. Using the approved templates, the resources were then created for each committee meeting by the Pearson standard setting team. The committee-specific documents/resources were reviewed by TEA before being finalized for publication for the meetings.

Development of Presentation Materials

Customized PowerPoint presentations were developed to guide facilitators through the presentation of information and materials throughout the standard setting meetings. TEA had the opportunity to review and provide suggested edits to the presentations, which were resolved by the Pearson standard setting team. The following PowerPoint presentations were created for the standard setting meetings:

- General Session Overview
- Standard Setting Breakout Meeting

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 program of training to facilitate the STAAR Alternate 2 EOC standard setting meetings. Facilitator training included the following:

- *STAAR Alternate 2 EOC assessments*—The facilitators were provided an overview of the Texas STAAR Alternate 2 EOC assessment program, including the test design, item types, scoring rules, performance levels, and scaling design.

- *Use of the Pearson standard setting website*—Because the Pearson standard setting website was used as a facilitation tool during the meeting, facilitators needed to be familiar with the use of the platform. The website outlines a framework for each of the facilitators to follow and provides the standard setting 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 specific issues for these meetings, such as time management, use of the online platform, and communicating feedback information.
- *Training slides and presentation notes*—As part of the walkthrough of the standard setting process, facilitators reviewed the standard setting training slides. Notes in the slides were provided to facilitators with guidance 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, which was intended 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. For example, statistical analysts produced the following after each judgment round:

- *Individual panelist feedback*—The judgments of the panelists for each performance level, to ensure that they were recorded accurately, and the resulting individual cut score recommendations (provided to all panelists)
- *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 PowerPoint slides)

The analysis programs created for the standard setting meetings used panelists' judgment data from each round. Panelists' judgments were downloaded from the standard setting website by analysts 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, which 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 Standard Setting Meetings
- Committee Composition
- Facilitators and Staff
- Materials
- 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 EOC standard setting meetings was to gather expert cut score recommendations from educators across the state of Texas. These cut scores define the performance levels of each STAAR Alternate 2 EOC assessment for English I and English II.

Student performance on the STAAR Alternate 2 EOC assessments is classified into one of three performance levels. Each STAAR Alternate 2 standard setting committee was asked to recommend two cut scores to define the boundaries between the different 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 the specific performance level.

Committee Composition

One committee was convened for each STAAR Alternate 2 EOC course (English I and English II). Figure 1 exhibits the room setup for the meetings. Individuals in each meeting included three distinct groups:

- Meeting facilitators
- Committee panelists
- Observers and staff

Figure 1. General Room Setup for the Meetings
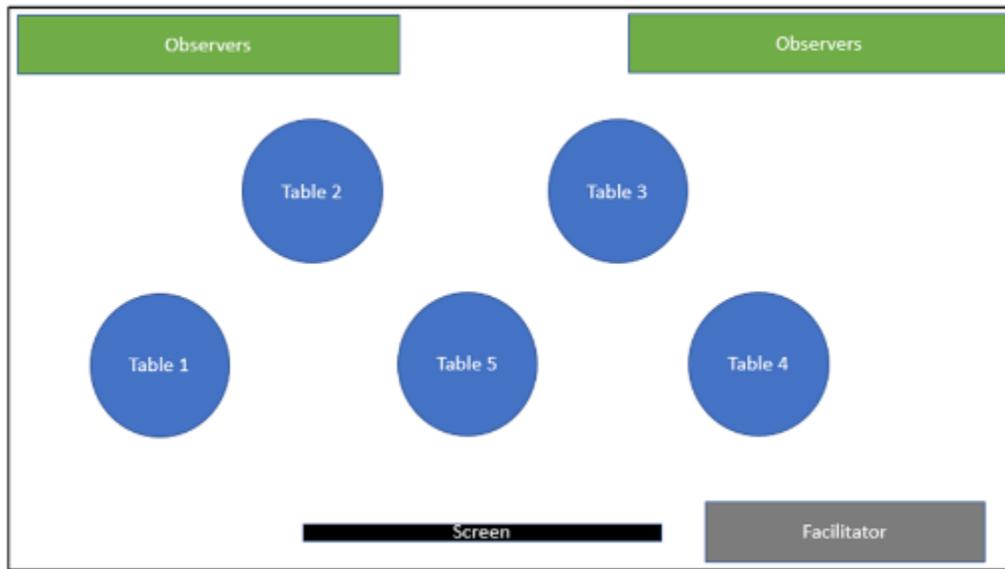
TEA selected the panelists to represent statewide educators with relevant content knowledge and experience with a variety of student groups. When setting cut scores, it is important to obtain the best judgments from people in the best possible position to make those judgments. To meet this goal, panelists should have the following qualifications:

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

The English I panel consisted of 14 panelists, and the English II panel had nine panelists. To ensure that the panelists were subject-matter experts, educators recruited for the meeting possessed experience in the course for which performance standards were being established. Appendix C presents the composition of the committees.

The panelists in each committee were assigned to table groups. Panelists assigned to each table were 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. 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 good candidate for this role. The table leader assisted the facilitator in maintaining appropriate discussions among the panelists, distributed and collected materials, maintained security measures, and performed other duties as deemed appropriate by the facilitator.

Facilitators and Staff

Staff members from TEA and Pearson collaborated to conduct the STAAR Alternate 2 English I and English II standard setting meeting. These staff members worked in facilitative and observational roles and did not contribute to the cut score recommendations during the meeting.

Facilitators

The lead facilitator of the standard setting meeting was Eric L. Moyer, Ph.D., from Pearson. Each breakout committee meeting was led by a process facilitator, who was a Pearson psychometrician 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 available also, 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 5 presents the process and content facilitators for each standard setting committee.

Table 5. Process and Content Facilitators

Committee	Process Facilitator	Content Facilitator
English I	Jie Lin	Megan Carlin
English II	Jiajing Huang	Kathleen McDonnold

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, other than vendor staff, were invited to attend the meeting by TEA. The number of observers in a committee meeting was kept to a maximum of one to two individuals, so the panelists did not feel overwhelmed.

Data Analysts

Two data analysts performed all analyses for the standard setting committees: Paige Rainforth and Morgen Hickey. During the meeting, the analysts collected panelist judgment data, performed independent analysis to verify analysis results, and prepared panelists' feedback.

TEA Staff

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

Materials

Extensive materials are needed for the successful implementation of the standard setting meeting. The following section describes the development of meeting materials.

Pearson Standard Setting Website

The Pearson standard setting website is the online platform for 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
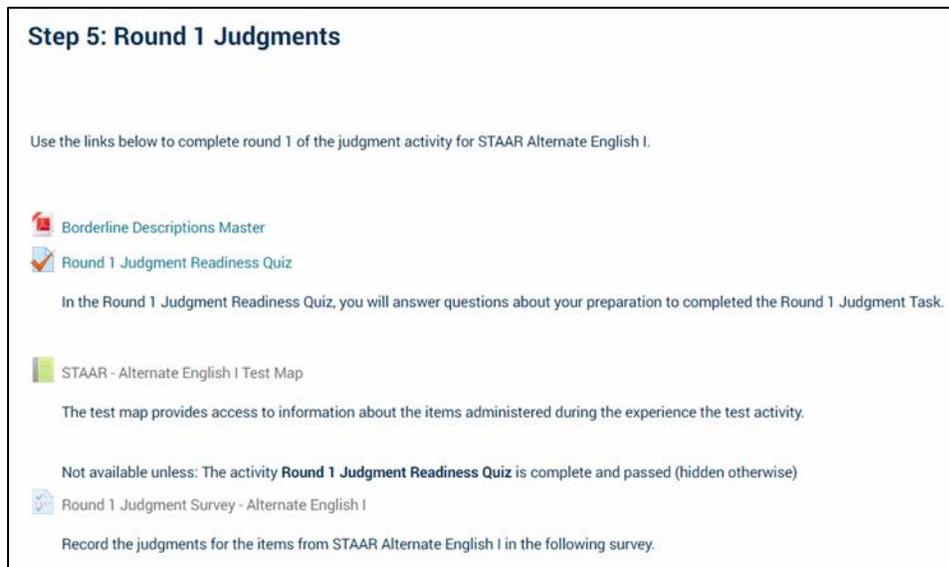
Even though the assessment was paper based, the panelists provided their judgments on the Pearson standard setting website. During the meeting, panelists accessed the website using a computer provided by Pearson and set up specifically for this meeting. The facilitator provided training to all panelists 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 a variety of hard copy materials they used throughout the meeting. The materials provided to panelists in their folders included the following:

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

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 of their meetings. 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 the meeting room to access the online resources through the Pearson standard setting website. Panelists were also provided an external monitor so they could access the online materials with limited switching between online materials. 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 list of permitted websites to restrict panelists' use of the computers to work associated with the standard setting meeting.

Procedure

The Modified Angoff method (Angoff, 1971) was used during the meeting to assist panelists in recommending performance level cut scores for each assessment. This standard setting procedure is both 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 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 is typical of this method. The set of judgments made by panelists are used to determine both individual and committee cut score recommendations for each performance level. Between the item judgment rounds, the panelists were provided feedback information, including data relative to panelist agreement, student performance on the items, and student performance on the test, as a whole. During Round 3, the panelists were asked the following for the Level 2 (*Satisfactory*) cut:

> *"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, they were informed additionally:

> *"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 two days, February 13–14, 2023. Appendix D includes the complete agenda for the standard setting meetings. Table 6 presents an overview of the agenda.

Table 6. 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 EOC 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 individuals recruited as panelists were registered into the standard setting website one to two weeks prior to the standard setting meeting. In an email from the website, panelists were provided with their unique user ID, a temporary password, and a link to the Pearson standard setting website. When panelists first logged in to the website, 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 in to 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. The access to the website prior to the standard setting ensured that panelists were oriented and trained to perform each step of the process during the meetings.

Panelists were asked to complete a set of tasks as pre-work at a convenient time prior to attending the standard setting meeting. Completion of meeting pre-work 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

To set the stage for the standard setting activity, a training video was included as part of the pre-work materials 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 site. 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 EOC 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, was presented by the lead psychometrician from Pearson. A clear description of the review process after the meetings was included to emphasize that committees are 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 course-specific breakout sessions for the remainder of the standard setting meeting. Each committee was responsible for providing recommendations for cut scores for each of the performance levels for the assessment associated with the committee. The committee provided recommendations using each of the activities described below.

Introductions and Overview

To begin the breakout session, the individuals in the room—facilitator, panelists, and observers—introduced themselves. The facilitator then distributed the meeting folders with panelist materials 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 a brief overview of the overall test design and item types that appeared on the spring 2023 STAAR Alternate 2 EOC assessment for their subject-specific committee. Panelists then reviewed a sample of assessment items students took during the spring 2023 administration. As panelists reviewed the items, they were encouraged to think from a student's perspective and take notes of the specific knowledge and skills a student would need to correctly respond to the item.

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 any 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 to complete the test because it was expected that content experts would need less time to complete the test than students. 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 of the Modified Angoff standard setting process is the development of borderline descriptions to provide all panelists 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 EOC 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 with performance at the borderline (i.e., a student that is "just-barely" past the point-of-entry for that performance level). Because the STAAR Alternate 2 EOC assessments have three performance levels, panelists were asked to develop borderline descriptions for two performance levels:

- Level 2: *Satisfactory*
- Level 3: *Accomplished*

Panelists were led through a 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*—After the modeling example, panelists worked within their table groups to review the draft 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. Questions panelists were asked to consider included:
  - 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 a 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.

A final whole-group review of the entire set of borderline descriptions was used 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 throughout 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 is "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 EOC assessments, the Modified Angoff method was extended to support judgments with polytomously scored items, where multiple score points are 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.

- "*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" is an aspirational expectation, "would" is 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?*" — In this context, likely was defined as two out of three times, or 67%. To make this concrete for panelists, facilitators asked panelists to think about three students performing at the borderline of a performance level for a specific point value, starting with one point. If panelists believed two out of three students performing at the borderline would earn a

specific number of points, the panelists were instructed to enter that number of points for that question. If they did not, the panelists were instructed to consider 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 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. The items were shown in the order they were administered during the operational test. Panelists reviewed the operational test items from the test booklets.
- *Test map*—A summary of the items in the test form. The test map included the following information:
    - Item position from the order of presentation
    - Item scoring key, item type, and item position in the cluster
    - 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*—Panelists recorded their judgments in the standard setting website and on the judgment record sheet for each judgment round.

Panelists reviewed each item and made a judgment for each borderline performance level, starting with Level 2, and then for Level 3. Because student performance on an item is 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, which 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 meeting facilitator.

Practice Judgment Activity

At the end of the training session, panelists were provided the opportunity to practice making judgments prior to beginning the actual judgment rounds. The goals of this activity were to

- give panelists experience reviewing and making judgments for different types of items,
- familiarize panelists with the paper judgment record sheet and judgment survey in the standard setting website, and
- 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 showed item-level results interactively 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. Additionally, panelists were reminded to refer to the borderline descriptions as well as other key considerations when making judgments. Finally, the facilitator demonstrated how the judgments were used to calculate individual and committee cut score recommendations.

Judgment Rounds

After receiving training on the standard setting process, the panelists participated in three rounds of judgments, with feedback discussion after each round. Prior to starting each judgment round, panelists were asked the following readiness questions to verify that they understood their task and were ready to begin. Panelists were unable to start the judgment survey until they answered "yes" to each readiness question.

- 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 and skills 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 record sheet also, 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, 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 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:

- *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 of their item judgments for each level.
- *Committee cut score recommendations and statistics*—Committee-level recommendations for each performance level were the median cut score 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 showing 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.
- *Cut scores*—The estimated cut score for each performance level.

- *Benchmark reasonable ranges*—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 7 presents 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. That is, they should listen to other panelists and consider the rationales given for their judgments, but they should not feel pressured to change their judgments to reach consensus.

Table 7. Feedback Data by Round

Feedback Data	Round 1	Round 2	Round 3
Panelist Item-Level 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. Panelists completed process evaluation surveys at specific points throughout the process, including after the practice judgment activity and after Round 3.

The purpose of the evaluation surveys is to determine the perceived effectiveness of the standard setting meeting, including panelists' understanding of the process, their comfort with the overall process, and their level of agreement with the results. The evaluation surveys were delivered through the standard setting website. Results from the evaluations were aggregated and included in this report.

Closing

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

Benchmark Reasonable Ranges

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 placement outside the range.

The benchmark reasonable ranges were created by mapping the performance level cuts 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 measures (CSEM) 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 of one performance level range 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, so ranges indicated subsequently greater expectations.

Table 8 presents the benchmark reasonable ranges presented to the panelists.

Table 8. Benchmark Reasonable Ranges (Raw Score Points)

Course	Satisfactory	Accomplished
English I	18–24	30–36
English II	18–24	29–35

Recommended Performance Level Cut Scores

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 the Round 3 judgments was considered the committee's recommendation for each performance level. Table 9 presents the recommended cut scores for each performance level based on the Round 3 recommendations for each EOC assessment.

Table 9. Cut Score Recommendations from Standard Setting Committees

Course	Max. Score	*Satisfactory Raw Score Cut*	*Accomplished Raw Score Cut*
English I	40	19	31
English II	40	21	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.

Chapter 4 – Post-Standard Setting

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

- Linear Scaling Process
- TEA Reasonableness Review
- Final Approval

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 would not be comparable across years. To address this, student results on the STAAR Alternate 2 EOC assessments are reported using scale scores that are comparable across administration years.

To be consistent with other STAAR Alternate 2 titles, it was determined that the STAAR Alternate 2 EOC reporting scale would have the following properties:

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

The reporting scale was set using the cut scores for Level 2: *Satisfactory* and using a fixed target standard deviation of 50. This was consistent with the previous practice used in scaling STAAR Alternate 2 tests. While the cut scores were defined with the same scale score cuts for Level 2: *Satisfactory* and the same standard deviation across the courses, they are not identical, and direct comparisons through averaging and aggregation across grades should not be made without study and/or statistical adjustments. The scale scores and distributions of students resulting from the cuts were not designed for direct comparison.

TEA Reasonableness Review

To support the rapid reporting of results, the standard setting was conducted prior to the administration of the STAAR Alternate 2 EOC assessments so the TEA commissioner could review and approve the cut score recommendations in time to support the reporting of student performance on these assessments shortly after the administration window ends. The standard setting process did not include the presentation of student performance on the assessment due to when the meeting occurred. As part of the process for validating the reporting scale for the STAAR Alternate 2 EOC assessments, TEA reviewed the distribution of student classification across the performance levels based on the approved cut score recommendations from the standard setting meeting.

TEA reviewed the recommendations from the standard setting committees in a reasonableness review to examine the performance level cut score recommendations 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 2022 administration, reasonable ranges for the cut scores, and the committee-recommended cut score ranges. The focus was on honoring the work of the standard setting committees while establishing performance levels that would work for the assessment program. Table 10 presents the final performance level cut scores on the IRT scale following the TEA reasonableness review.

Table 10. Final Recommended Cut Scores on the IRT Scale

Course	*Satisfactory*	*Accomplished*	*Satisfactory*	*Accomplished*	A (Slope)	B (Intercept)
English I	21	32	300	365	49.3225	294.3526
English II	21	33	300	370	49.5385	294.8480

Note. The first set of cuts is the raw score cut scores, and the second set is the IRT cuts.

Final Approval

Mike Morath, the Commissioner of Education at TEA, reviewed and approved the final performance level cut scores for the STAAR Alternate 2 EOC assessments on February 27, 2023.

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

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

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

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

All panelists were currently working in school districts, as presented in Table C.9 and represented the various types of districts across the state, including size, type, and socioeconomic status. Teachers representing schools from a rural area were a slight majority of the panelists for English I, whereas the English II committee had most panelists from suburban schools. Nevertheless, in both committees, the panelists all taught in districts that were considered either 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 methodology. The training on the standard setting methodology was provided during the general session and in the individual standard setting committees. The training on the implementation of the standard setting process was standardized across committees through the PowerPoint training slides.

Panelists completed a practice judgment round as an opportunity to implement the standard setting methodology without consequence, including making judgments within the standard setting website. During the practice judgment round, the panelists reviewed a reduced set of items and provided judgments for two performance levels (*Satisfactory* and *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 panelists were prepared for making their judgments. Panelists were not able to continue to the judgment survey unless they answered "yes" to both questions on the readiness survey. They were encouraged to ask the facilitator questions if they responded "no" to either question.

At various points within 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 committees for several questions related to the training on the standard setting process. Appendix I presents the results for all evaluation survey questions.

As part of the evaluation survey, panelists were specifically asked about the effectiveness of the training they received on the standard setting process. For one question, panelists rated the level of success of the introduction to the standard setting process during the general session. Overall, the initial introduction to the standard setting process was perceived as successful, with all but one panelist responding that it was either *Successful* or *Very Successful*. The same favorable balance of responses held for the practice judgment activity for the standard setting process. Even though one panelist for English II did not think the training on the standard setting process in the breakout groups was adequate, most panelists on both committees perceived that the training provided on the standard setting process was either Adequate or More than Adequate.

Figure 3. Evaluation Results of the Standard Setting Process Training Activities

Introduction to the standard setting process	Rating	English I	English II
	Not successful	–	–
	Partially successful	1	1
	Successful	6	6
	Very Successful	7	3

Practice exercise for the standard setting procedures	Rating	English I	English II
	Not successful	–	–
	Partially successful	1	1
	Successful	7	6
	Very Successful	6	3

Training provided on the standard setting process	Rating	English I	English II
	Not Adequate	–	1
	Somewhat Adequate	2	2
	Adequate	5	4
	More Than Adequate	7	3

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

Generally, 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 had the opportunity to indicate their confidence that the PLDs were reasonable for each of the performance levels. 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 by Performance Level

Satisfactory PLDs	Rating	English I	English II
	Not Confident	–	–
	Somewhat Confident	1	3
	Confident	6	3
	Very Confident	7	3

Accomplished PLDs	Rating	English I	English II
	Not Confident	–	–
	Somewhat Confident	1	3
	Confident	5	3
	Very Confident	8	3

All but one panelist on the English I committee were either Confident or Very Confident that the PLDs were reasonable for both performance levels. The English II committee panelists were evenly divided between Somewhat Confident, Confident, and Very Confident. Given that most panelists were either Confident or Very Confident in the PLDs for both courses and performance levels, these results provide evidence that, overall, the PLDs were perceived by the panelists as providing reasonable expectations for each performance level.

The panelists were also provided the opportunity to indicate 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 by Performance Level

Satisfactory Cut Scores	Rating	English I	English II
	Not Confident	–	1
	Somewhat Confident	1	3
	Confident	7	2
	Very Confident	6	3

Accomplished Cut Scores	Rating	English I	English II
	Not Confident	–	1
	Somewhat Confident	–	3
	Confident	7	2
	Very Confident	7	3

All but one panelist had some level of confidence in both performance level cuts for the respective courses, although this was stronger for English I. More than half of the panelists indicated that they were Confident or Very Confident in their perception of the cut scores for both performance levels.

Overall, this feedback from the cut score setting panelists provides evidence for the validity of the cut score recommendations for each of the performance levels from the standard setting committee, though it was much stronger for English I than for English II.

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

STAAR Alternate 2 English I Performance Level Descriptors

State of Texas Assessments of Academic Readiness (STAAR®) Alternate 2
Performance Level Descriptors
English I

Performance Level Descriptors
When reading texts of increasing complexity,* students achieving Accomplished Performance can   - Use context such as definition and examples to clarify the meaning of unfamiliar and multiple-meaning words   - Make inferences and use text evidence to support understanding of informational and narrative text, with adult assistance   - Discuss plot elements including conflict and resolution   - Identify the author's purpose for writing text   - Use coordinating conjunctions to form compound subjects, predicates, and sentences with subject-verb agreement   - Revise drafts by adding, deleting, and combining ideas to improve sentence structure and word choice
When reading texts of increasing complexity,* students achieving Satisfactory Performance can   - Use context to determine the meaning of unfamiliar words and multiple-meaning words   - Evaluate details to determine what is most important   - Describe characters, setting, and plot and recognize chronological order   - Use coordinating conjunctions in sentences with subject-verb agreement   - Revise drafts by adding or deleting sentences for improved clarity
When reading texts of increasing complexity,* students achieving Developing Performance can   - Use context and illustrations to determine the meaning of unknown words, with adult assistance   - Synthesize information to create new understanding, with adult assistance   - Identify characters, events, or setting in literary text, with adult assistance   - Identify prepositions and prepositional phrases, with adult assistance   - Revise drafts by adding details using words, with adult assistance

*Text complexity increases from grade to grade. Texts can become increasingly complex for a variety of reasons: (1) vocabulary/use of language may be more varied and challenging; (2) sentence structures may be more varied and sophisticated; (3) the topic/content may be less familiar or more challenging; and (4) 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.

STAAR Alternate 2 English II Performance Level Descriptors

State of Texas Assessments of Academic Readiness (STAAR®) Alternate 2
Performance Level Descriptors
English II

Performance Level Descriptors
When reading texts of increasing complexity,* students achieving Accomplished Performance can • Use context such as definition and examples to clarify the meaning of unfamiliar and multiple-meaning words • Make inferences and use text evidence to support understanding of informational and narrative text, with adult assistance • Discuss plot elements including conflict and resolution • Identify the author's purpose for writing text • Use coordinating conjunctions to form compound subjects, predicates, and sentences with subject-verb agreement • Revise drafts by adding, deleting, and combining ideas to improve sentence structure and word choice
When reading texts of increasing complexity,* students achieving Satisfactory Performance can • Use context to determine the meaning of unfamiliar words and multiple-meaning words • Evaluate details to determine what is most important, with adult assistance • Describe characters, setting, and plot and recognize chronological order • Use coordinating conjunctions in sentences with subject-verb agreement • Revise drafts by adding or deleting sentences for improved clarity
When reading texts of increasing complexity,* students achieving Developing Performance can • Use context and illustrations to determine the meaning of unknown words, with adult assistance • Synthesize information to create new understanding, with adult assistance • Identify characters, events, or setting in literary text, with adult assistance • Identify prepositions and prepositional phrases, with adult assistance • Revise drafts by adding details using words, with adult assistance

*Text complexity increases from grade to grade. Texts can become increasingly complex for a variety of reasons: (1) vocabulary/use of language may be more varied and challenging; (2) sentence structures may be more varied and sophisticated; (3) the topic/content may be less familiar or more challenging; and (4) 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 Agenda

The following is an example of the agenda that was handed to the panelists at the standard setting meeting.

STAAR Alternate 2 English EOC Assessments
Standard Setting Meeting
February 2023

Agenda

Day 1 – February 13

7:30 am	*Breakfast*
8:30 am	*General Session*

> Welcome
>
> Overview of STAAR Alternate 2 English Assessments
>
> Standard Setting Overview

9:50 am	*Break*
10:00 am	*Breakout Sessions (STAAR English I and II,* STAAR Alternate 2 *English I and II)*

> Welcome and Introductions
>
> Assessment Overview
>
> Experience the Assessment Activity

11:30 am	*Lunch*

> Performance Level Descriptors Discussion
>
> Borderline Descriptor Development

2:15 pm	*Break*

> Borderline Descriptor Development (Cont.)
>
> Standard Setting Training
>
> Practice Judgment Activity and Discussion

5:00 pm	*End-of-Day*

Day 2 – February 14

7:30 am   *Breakfast*

8:30 am   *Breakout Session (STAAR English I and II,* STAAR Alternate 2 *English I and II)*
            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 *(cont.)*
            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*

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 Response Record Form

Only the first page of this document is presented as an example.

<div align="center">

Experience the Assessment
Notes Sheet
STAAR Alternate 2 English I EOC

</div>

Sequence	
1	
2	
3	
4	
5	
6	
7	

Item Judgment Round Record Form

Only the first page of this document is presented as an example.

<div align="center">

Judgment Round Record Sheet
STAAR Alternate 2 English I EOC

</div>

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

Sequence	Item Code	Judgment Round			
		Round 1		Round 2	
		L2	L3	L2	L3
1					
2					
3					
4					
5					
6					
7					
8					

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 of the questions?

Level	Total Points
Satisfactory Academic Performance	
Accomplished Academic Performance	

Item Judgment Survey

The survey for only the first two items is shown.

Rounds 1 and 2

For each of the items, answer the following question:

"How many points would a borderline student of the performance level receive?"

To answer the question, you will select the option between 0 and the maximum points that would best answer the question.

1 **Item**

Key:

		0	1	2
Satisfactory	●	○	○	○
Accomplished	●	○	○	○

2 **Item:**

		0	1	2
Satisfactory	●	○	○	○
Accomplished	●	○	○	○

Round 3

Directions: Review the items and your item judgments for each performance level to answer the following question.

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

Write your recommended cut score for the **Level 2: Satisfactory** Performance Level in the box.

This would be a number between 1 and 40.

Write your recommended cut score for the **Level 3: Accomplished** Performance Level in the box.

This would be a number between 1 and 40 <u>and</u> greater than your recommended cut score for Satisfactory.

Process Evaluation #1

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

Process Evaluation Survey #1

English II

The purpose of this evaluation is to collect information about your experience in recommending cut scores associated with the achievement levels for the STAAR 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 EOC 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

English II

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 - table-level statistics	○	○	○	○
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
Table-level statistics after Rounds 1 and 2	○	○	○	○
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 four performance levels) for student performance on STAAR Alternate 2 EOC assessments.

How confident do you feel that the Performance Level Descriptors (PLDs) for English II 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 four performance levels) for student performance on STAAR Alternate 2 EOC assessments.

How confident do you feel that the recommended cut scores for English II 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.

Paragraph ▼ | B | I | ☰ | ☰ | 🔗 | ✂ | 🖼 | 🔲

Path: p

50

Appendix C – Committee Panelist Composition

Table C.1. Panelist Position

Response Option	English I	English II
Teacher (K–12)	10	9
Teacher (Higher Ed.)	1	1
Administrator (School)	1	–
Administrator (District)	1	–
Instructional Specialist/Coach	1	–
Other Position (e.g., Coordinator/Coach)	–	–
Total	14	10

Table C.2. Years of Total Teaching Experience

Response Option	English I	English II
None	–	–
1 to 5 years	1	–
6 to 10 years	3	4
11 to 15 years	4	2
16 to 20 years	1	2
More than 20 years	5	2
Total	14	10

Table C.3. Years of Experience Teaching STAAR Alternate 2 English at This Level

Response Option	English I	English II
None	5	5
1 to 5 years	1	2
6 to 10 years	5	1
11 to 15 years	2	1
16 to 20 years	–	–
More than 20 years	1	1
Total	14	10

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

Response Option	English I	English II
Mainstream special education	9	8
Self-contained special education	9	5
English learners (ELs)	9	7
General education instruction	8	7
Vocational technical instruction	2	3

Table C.5. Highest Degree Completed

Response Option	English I	English II
High School Diploma	–	–
Associate degree (A.A., A.S.)	–	–
Bachelor's degree (B.A., B.S.)	5	3
Master's degree (M.A., M.S.)	8	7
Doctoral degree (Ph.D., Ed.D.)	1	–
Total	14	10

Table C.6. Demographic: Gender

Response Option	English I	English II
Female	12	9
Male	2	1
Other/No answer	12	9

Table C.7. Demographic: Ethnicity

Response Option	English I	English II
Hispanic or Latino	3	3
Not Hispanic or Latino	11	7
No answer	–	–

Table C.8. Demographic: Race

Response Option	English I	English II
American Indian or Alaskan Native	–	–
Asian	–	1
Black or African American	1	–
Native Hawaiian or Pacific Islander	–	–
White	11	9
No answer	2	–

Table C.9. Currently Working in a School District

Response Option	English I	English II
Yes	14	10
No (Higher Ed)	–	–

Table C.10. Size of School District

Response Option	English I	English II
Small	6	1
Medium	6	4
Large	2	5

Table C.11. Type of School District

Response Option	English I	English II
Rural	8	2
Metropolitan/Urban	–	1
Suburban	6	7

Table C.12. Socioeconomic Status of School District

Response Option	English I	English II
Low	8	7
Moderate	6	3
High	–	–

Appendix D – Standard Setting Meeting Agenda

Standard Setting Meeting Facilitator Agenda

Day 1 – February 13

Breakfast available in the hotel from 7:30–8:30 a.m. (CT)

Start Time	End Time	
General Session		
8:30 am	8:45 am	Welcome, 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		
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 3 Group Work
1:45 pm	2:15 pm	Borderline Description Development – Level 3 Whole-Group Discussion
2:15 pm	2:30 pm	*Break*
2:30 pm	3:00 pm	Borderline Description Development – Level 2 Group Work
3:00 pm	3:30 pm	Borderline Description Development – Level 2 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</u> February 14

Breakfast available in the hotel from 7:30-8:30am (CT)

Start Time	End Time	
Breakout Session		
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

Appendix E – Examples of Feedback Data

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

Individual Item-Level Judgments

This provided the panelist with the actual item-level judgments that were recorded in Moodle for the panelist. This was provided so that the panelist could check that the system recorded the judgments correctly.

Alternate English I – Individual Rating – Round 1

Table=1 Name=-

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

Individual Test-Level Recommendation

This provided the panelist with the recommendations for test-level cut scores based on his or her item judgments for the *Satisfactory* and *Accomplished* performance levels.

Alternate English I – Individual Cut Scores – Round 2

Table=1 Name=

L2 Raw Score	L3 Raw Score
20	32

Overall Test-Level Recommendations

This provided the panelist 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 English I Round 1 Summary Statistics – Overall

	N	Mean	Median	Min	Max	Q1	Q3
L2 Raw Score	14	20.71	19.00	16.00	28.00	17.00	23.00
L3 Raw Score	14	31.29	31.00	27.00	38.00	29.00	32.00

Item-Level Judgment Agreement

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

Alternate English I Round 1 Level L2

SeqNo	UIN	Max Points	0	1	2
1ALT		2			100%
2ALT		2		57%	43%
3ALT		2	7%	86%	7%
4ALT		2	100%		
5ALT		2			100%
6ALT		2	7%	71%	21%
7ALT		2	36%	64%	
8ALT		2	57%	43%	
9ALT		2		36%	64%
10ALT		2	7%	71%	21%

Test-Level Panelist Recommendation Agreement

This feedback was presented to panelists by the facilitator. It presented as bar graphs displaying the distribution of panelist recommendations for the cut score, by raw score, for each performance level: *Satisfactory* and *Accomplished*.

Alternate English II Panelist Agreement at Level L2 and L3 - Round 1
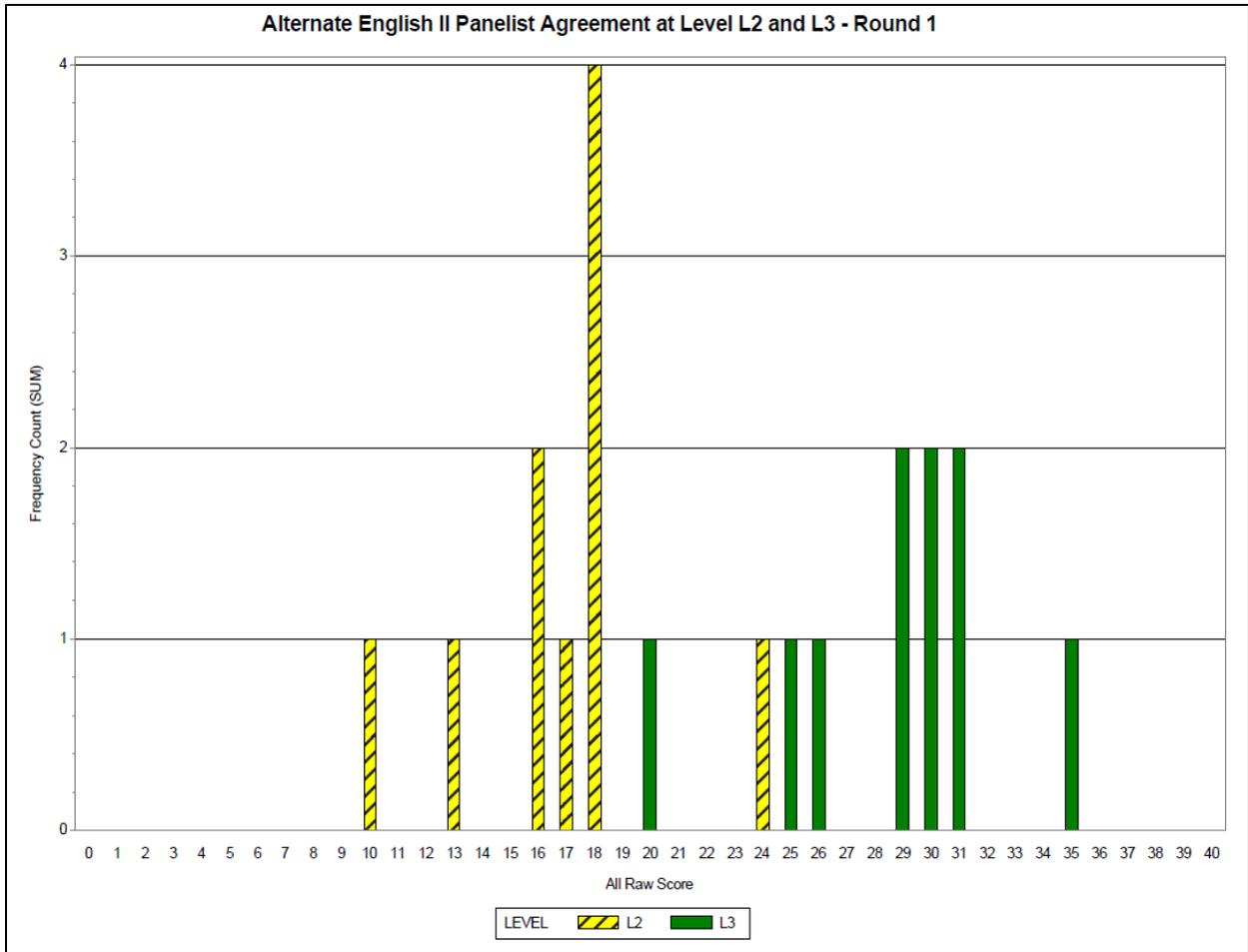
Appendix F – Committee Recommended Cut Scores by Round

Table F.1. Committee Recommended Cut Scores by Round, English I

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

Table F.2. Committee Recommended Cut Scores by Round, English II

Performance Level	Max. Score	Round 1	Round 2	Round 3
Level 2: *Satisfactory*	40	18	16	21
Level 3: *Accomplished*	40	30	30	31

Appendix G – Recommended Cut Score Summary Statistics

Table G.1. Recommended Cut Score Summary Statistics, English I

Round	Statistic	*Satisfactory*	*Accomplished*
1	Mean	20.71	31.29
	Minimum	16	27
	Q1	17	29
	Median	**19**	**31**
	Q3	23	32
	Maximum	28	38
2	Mean	19.79	30.86
	Minimum	18	29
	Q1	18	30
	Median	**19**	**31**
	Q3	20	32
	Maximum	25	32
3	Mean	19.50	30.90
	Minimum	18	30
	Q1	19	30
	Median	**19**	**31**
	Q3	20	31
	Maximum	23	32

Table G.2. Recommended Cut Score Summary Statistics, English II

Round	Statistic	*Satisfactory*	*Accomplished*
1	Mean	16.80	28.60
	Minimum	10	20
	Q1	16	26
	Median	**18**	**30**
	Q3	18	31
	Maximum	24	35
2	Mean	17.00	29.22
	Minimum	15	25
	Q1	16	29
	Median	**16**	**30**
	Q3	18	31
	Maximum	20	31
3	Mean	21.30	31.10
	Minimum	18	29
	Q1	20	30
	Median	**21**	**31**
	Q3	23	32
	Maximum	23	33

Appendix H – Test-Level Panelist Judgement Agreement

This appendix presents the raw score cuts selected by panelists for each performance 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: STAAR Alternate 2 English I, Round 1

Raw Score	Satisfactory	Accomplished
16	1	
17	3	
18	3	
20	1	
22	1	
23	2	
25	1	
27		1
28	2	1
29		2
30		2
31		2
32		3
34		1
35		1
38		1

Table H.2. Panelist Agreement Data: STAAR Alternate 2 English I, Round 2

Raw Score	Satisfactory	Accomplished
18	4	
19	5	
20	2	
22	1	
23	1	
25	1	
29		3
30		1
31		5
32		5

Table H.3. Panelist Agreement Data: STAAR Alternate 2 English I, Round 3

Raw Score	Satisfactory	Accomplished
18	3	
19	6	
20	2	
21	2	
23	1	
30		4
31		7
32		3

Table H.4. Panelist Agreement Data: STAAR Alternate 2 English II, Round 1

Raw Score	Satisfactory	Accomplished
10	1	
13	1	
16	2	
17	1	
18	4	
20		1
24	1	
25		1
26		1
29		2
30		2
31		2
35		1

Table H.5. Panelist Agreement Data: STAAR Alternate 2 English II, Round 2

Raw Score	Satisfactory	Accomplished
15	2	
16	3	
17	1	
18	1	
20	2	
25		1
26		1
29		1
30		3
31		3

Table H.6. Panelist Agreement Data: STAAR Alternate 2 English II, Round 3

Raw Score	Satisfactory	Accomplished
18	1	
20	2	
21	2	
23	4	
29		1
30		2
31		2
32		3
33		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 were designed to help you both understand the process and be supportive of the recommendations made by the committee.

Overview of the STAAR Alternate 2 EOC Assessment

Rating	English I	English II
Not successful	–	–
Partially successful	1	1
Successful	5	6
Very Successful	–	3

Introduction to the Standard Setting Process

Rating	English I	English II
Not successful	–	–
Partially successful	1	1
Successful	6	6
Very Successful	7	3

Experiencing the Actual Assessment

Rating	English I	English II
Not successful	–	–
Partially successful	–	1
Successful	4	4
Very Successful	10	6

Discussion of Scoring Items on the Assessment

Rating	English I	English II
Not successful	–	–
Partially successful	1	1
Successful	6	5
Very Successful	7	5

Discussion of Performance Level Descriptors (PLDs)

Rating	English I	English II
Not successful	2	–
Partially successful	1	2
Successful	5	4
Very Successful	6	4

Overview of the Standard Setting Procedure

Rating	English I	English II
Not successful	–	1
Partially successful	1	–
Successful	6	6
Very Successful	7	3

Practice Exercise for the Standard Setting Procedure

Rating	English I	English II
Not successful	–	–
Partially successful	1	1
Successful	7	6
Very Successful	6	3

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

Performance Level Descriptors (PLDs)

Rating	English I	English II
Very Useful	8	4
Useful	4	3
Somewhat Useful	1	2
Not Useful	1	1

Borderline Description Development

Rating	English I	English II
Very Useful	8	5
Useful	3	2
Somewhat Useful	3	3
Not Useful	–	–

Standard Setting Training

Rating	English I	English II
Very Useful	8	4
Useful	4	4
Somewhat Useful	2	2
Not Useful	–	–

How adequate were the following elements of the session?

Training Provided on the Standard Setting Process

Rating	English I	English II
Not Adequate	–	1
Somewhat Adequate	2	2
Adequate	5	4
More Than Adequate	7	3

Amount of Time Spent Training

Rating	English I	English II
Not Adequate	–	–
Somewhat Adequate	1	2
Adequate	5	4
More Than Adequate	8	4

Total Amount of Time to Discuss the PLDs

Rating	English I	English II
Not Adequate	–	–
Somewhat Adequate	–	–
Adequate	4	5
More Than Adequate	10	5

Total Amount of Time to Create and Discuss Borderline Descriptions

Rating	English I	English II
Not Adequate	–	–
Somewhat Adequate	–	1
Adequate	4	4
More Than Adequate	10	5

Total amount of Time to Discuss the Practice Judgments

Rating	English I	English II
Not Adequate	–	–
Somewhat Adequate	–	2
Adequate	4	4
More Than Adequate	10	4

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 were designed to help you both understand the process and be supportive of the recommendations made by the committee.

Judgment Rounds

Rating	English I	English II
Not successful	–	–
Partially successful	–	2
Successful	6	6
Very Successful	8	1

Judgment Round Feedback – Table-level Statistics

Rating	English I	English II
Not successful	–	2
Partially successful	–	2
Successful	5	3
Very Successful	9	2

Judgment Round Feedback – Committee-Level Statistics

Rating	English I	English II
Not successful	–	–
Partially successful	–	2
Successful	5	5
Very Successful	9	2

Judgment Round Feedback – Panelist Agreement Data

Rating	English I	English II
Not successful	–	–
Partially successful	–	1
Successful	4	5
Very Successful	10	3

Discussions After Each Round

Rating	English I	English II
Not successful	–	–
Partially successful	1	1
Successful	3	6
Very Successful	10	2

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

Table-Level Statistics After Rounds 1 and 2

Rating	English I	English II
Very Useful	10	3
Useful	4	4
Somewhat Useful	–	2
Not Useful	–	–

Committee-level Statistics After Round 2

Rating	English I	English II
Very Useful	10	3
Useful	4	4
Somewhat Useful	–	2
Not Useful	–	–

Panelist Agreement Data Provided After Round 1

Rating	English I	English II
Very Useful	10	3
Useful	4	3
Somewhat Useful	–	2
Not Useful	–	1

Panelist Agreement Data Provided After Round 2

Rating	English I	English II
Very Useful	9	3
Useful	5	5
Somewhat Useful	–	1
Not Useful	–	–

Discussion After Each Judgment Round

Rating	English I	English II
Very Useful	12	4
Useful	2	4
Somewhat Useful	–	1
Not Useful	–	–

How adequate were the following elements of the session?

Amount of Time to Make Judgments

Rating	English I	English II
Not Adequate	–	1
Somewhat Adequate	–	1
Adequate	4	2
More Than Adequate	10	5

Visual Presentation of the Feedback Provided

Rating	English I	English II
Not Adequate	–	–
Somewhat Adequate	–	2
Adequate	5	3
More Than Adequate	9	4

Number of Judgment Rounds

Rating	English I	English II
Not Adequate	–	–
Somewhat Adequate	–	1
Adequate	6	5
More Than Adequate	8	3

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

How confident do you feel that the Performance Level Descriptors (PLDs) for your committee are reasonable for each student performance level?

Level 2 – *Satisfactory*

Rating	English I	English II
Not Confident	–	–
Somewhat Confident	1	3
Confident	6	3
Very Confident	7	3

Level 3 – *Accomplished*

Rating	English I	English II
Not Confident	–	–
Somewhat Confident	1	3
Confident	5	3
Very Confident	8	3

How confident do you feel that the recommended cut scores for your course represent appropriate levels of student performance level?

Level 2 – *Satisfactory*

Rating	English I	English II
Not Confident	–	1
Somewhat Confident	1	3
Confident	7	2
Very Confident	6	3

Level 3 – *Accomplished*

Rating	English I	English II
Not Confident	–	1
Somewhat Confident	–	3
Confident	7	2
Very Confident	7	3

How adequate were the following elements of the session?

Facilities Used for the Meeting

Rating	English I	English II
Not Adequate	1	–
Somewhat Adequate	1	–
Adequate	2	2
More Than Adequate	10	7

Computers Used During the Meeting

Rating	English I	English II
Not Adequate	–	–
Somewhat Adequate	–	–
Adequate	3	3
More Than Adequate	11	6

Pearson Standard Setting Website for Accessing Materials and Making Judgments

Rating	English I	English II
Not Adequate	–	–
Somewhat Adequate	–	1
Adequate	5	5
More Than Adequate	9	3

Content Review System for Viewing Items

Rating	English I	English II
Not Adequate	–	–
Somewhat Adequate	–	1
Adequate	4	5
More Than Adequate	10	3

Materials Provided in the Folder

Rating	English I	English II
Not Adequate	–	–
Somewhat Adequate	1	1
Adequate	3	5
More Than Adequate	10	3

Workspace in Table Groups During the Meeting

Rating	English I	English II
Not Adequate	–	–
Somewhat Adequate	–	2
Adequate	11	2
More Than Adequate	3	5

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

Express Your Opinions About Student Performance Levels

Rating	English I	English II
Not Adequate	–	–
Somewhat Adequate	–	3
Adequate	5	4
More Than Adequate	9	2

Ask Questions About the Cut Scores and How They Will be Used

Rating	English I	English II
Not Adequate	–	1
Somewhat Adequate	–	2
Adequate	4	4
More Than Adequate	10	2

Ask Questions About the Process of Making Cut Score Recommendations

Rating	English I	English II
Not Adequate	–	1
Somewhat Adequate	–	2
Adequate	4	4
More Than Adequate	10	2

Interact with Your Fellow Panelists

Rating	English I	English II
Not Adequate	–	–
Somewhat Adequate	–	1
Adequate	4	5
More Than Adequate	10	3

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

Fellow Panelists

Rating	English I	English II
Yes	12	9
Sometimes	2	–
No	–	–

Facilitators

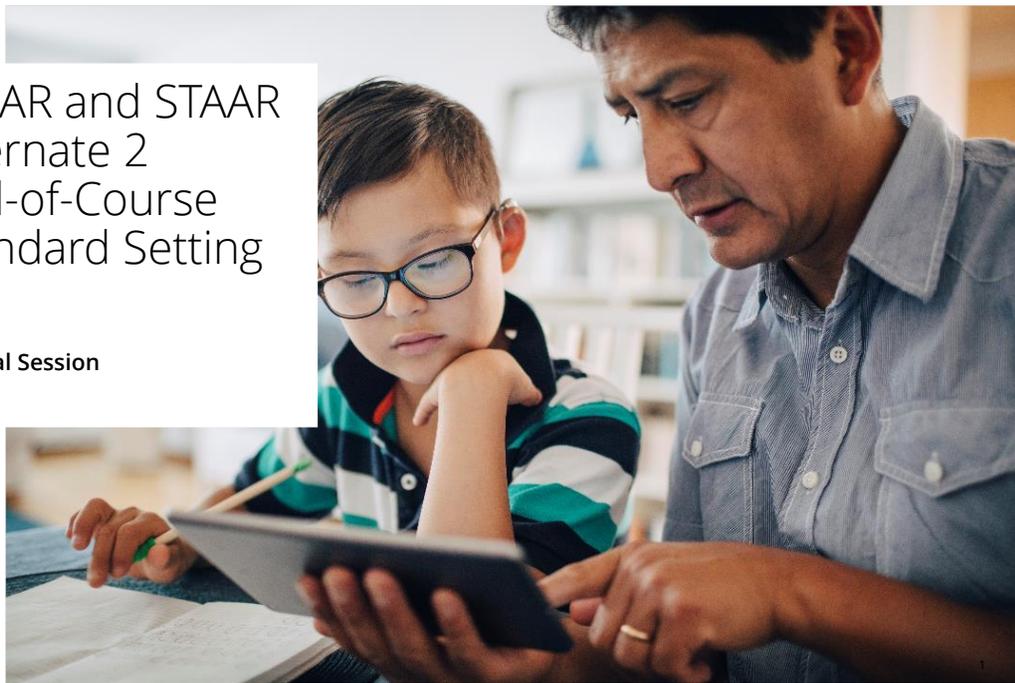| Rating	English I	English II
Yes	13	5
Sometimes	1	4
No	–	–

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 – English I