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

STAAR EOC Assessments – English I, English II, Biology, Algebra I, and U.S. History

February 2023

Pearson

Version 4.1

Table of Contents

Executive Summary 4
Chapter 1 – Overview of the Standard Setting Process 9
Goals of the Standard Setting Meeting 9
Performance Levels 9
Standard Setting Process 10
Chapter 2 – Pre-meeting Development 11
Performance Level Descriptors 11
Pearson Standard Setting Website 12
Development of Panelist Materials 12
Development of Presentation Materials 13
Facilitator Training 13
Preparation for Data Analysis During the Meetings 14
Chapter 3 – Standard Setting Meetings 15
Purpose of the Standard Setting Meetings 15
Committee Composition 15
Facilitators and Staff 17
Materials 18
Procedure 19
Standard Setting Meeting Proceedings 20
Recommended Performance Level Cut Scores 30
Chapter 4 – Post-Standard Setting 31
Linear Scaling Process 31
TEA Reasonableness Review 31
Final Approval 32
Chapter 5 – Evidence of Procedural Validity of the Standard Setting Process 33
Committee Representation 33
Committee Training 33
Perceived Validity of the Workshop 35
References 37
Appendix A – Performance Level Descriptors 38
Appendix B – Panelist Meeting Materials 45

Appendix C – Committee Panelist Composition 55
Appendix D – Standard Setting Meeting Agenda 57
Appendix E – Examples of Feedback Data 59
Appendix F – Committee Recommended Cut Scores by Round 62
Appendix G – Recommended Cut Score Summary Statistics 63
Appendix H – Test-Level Panelist Judgment Agreement 66
Appendix I – Panelist Evaluation Results 76
    Process Evaluation Survey #1 76
    Process Evaluation Survey #2 79
Appendix J – PowerPoint Presentations 89

Executive Summary

This report describes the standard setting process for the State of Texas Assessments (STAAR®) End-of-Course (EOC) assessments in English I, English II, Algebra I, Biology, and U.S. History. A summary of the results is also provided.

Standard Setting Process and Results

Performance levels are used to classify and describe student performance on an assessment. To classify student performance into the different performance levels, performance level descriptors (PLDs) and cut scores are generally required. The PLDs 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 STAAR 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 STAAR EOC assessments for English I and II. The draft PLDs were reviewed by a panel of teachers from across the state who provided feedback and recommended revisions for TEA to consider when finalizing the PLDs. New PLDs were created and reviewed for English I and II from the test redesign as these had new standards, whereas the PLDs for the other EOC assessments (Algebra I, Biology and U.S. History) were not changed for this standard setting meeting. The performance levels for the STAAR EOC assessments are as follows:

- Level 4: *Masters Grade Level*
- Level 3: *Meets Grade Level*
- Level 2: *Approaches Grade Level*
- Level 1: *Did Not Meet Grade Level*

Standard Setting Meeting

A standard setting meeting was convened from February 13–16, 2023, to recommend cut scores for the STAAR EOC assessments. The panels for English I and II met on February 13–14, while the panels for Algebra I, Biology, and U.S. History met on February 15–16. The committees were composed of 14–17 panelists who were educators with relevant content area teaching experience. The panelists were selected to provide content expertise during the meeting and represent diverse state geographic regions, gender, ethnicity, educational experience, community size, and community socioeconomic status.

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

To begin Day 1, the panelists participated in a general session where 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, followed by the "Experience the Assessment" activity that allowed them 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 one of two questions for each performance level. If the item was a 1-point (dichotomously scored) item, the panelists were to consider the question,

> "*What is the probability that a student with performance at the borderline of the [given] level would likely answer the question correctly?*"

If the item was a multi-point (polytomously scored) item, the panelists were to consider the question,

> "*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 beginning the 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 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 be conducted to establish the performance standards for the new STAAR EOC assessments before the spring 2023 administration of the assessments. 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 the test design and administration mode, there are policy expectations that trends in student distribution across the performance levels do not vary significantly from previous administrations. To assist in aligning the resulting standard setting cut score recommendations with the academic expectations defined in the PLDs while also maintaining similar impact data to previous administrations, benchmark values were established as reasonable ranges.

The benchmark values represented a reasonable range for each performance level of every STAAR 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 measurements (CSEMs) were determined for each performance level of each assessment. The CSEM was used to create a reasonable range around the cut score for each performance level. The reasonable range values on the spring 2022 theta scale were matched with the nearest theta values on the raw score look-up table from the 2023 pre-equated raw score look-up table for each assessment. The raw scores associated with the reasonable range of theta values from the spring 2023 pre-

equated raw score look-up table were used to establish the benchmark reasonable ranges. If the raw score values associated with the maximum of one performance level range were 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	*Approaches*	*Meets*	*Masters*
English I	24–31	32–40	51–57
English II	23–31	32–41	54–59
Algebra I	17–23	29–36	37–44
Biology	12–17	22–28	35–42
U.S. History	19–26	32–40	45–54

Results

During Round 3, panelists made their judgments at the test level instead of 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 EOC assessments.

Table 2. Standard Setting Recommendations

Course	Max. Score	*Approaches* Cut	*Approaches* Range	*Meets* Cut	*Meets* Range	*Masters* Cut	*Masters* Range
English I	64	27	24–31	37	32–40	53	51–57
English II	64	23	22–27	35	32–37	54	52–56
Algebra I	59	18	18–18	31	31–31	41	40–43
Biology	53	14	13–17	25	23–27	37	34–44
U.S. History	78	22	19–24	35	33–38	50	45–54

After Round 3, the panelists completed an evaluation of the standard setting process and their confidence in the 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 before the administration of the spring 2023 STAAR 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 after the administration window ends. The standard setting process did not include the presentation of student performance on the assessment due to meeting scheduling. As part of the process for validating the reporting scale for the STAAR 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 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 EOC assessments.

Table 3. Final Performance Level Cut Scores

Course	Max. Score	*Approaches*	*Meets*	*Masters*
English I	64	27	36	54
English II	64	27	36	56
Algebra I	59	20	32	41
Biology	53	14	25	38
U.S. History	78	22	36	50

Final Approval

Mike Morath, the Commissioner of Education at TEA, reviewed and approved the final performance level cut scores for the STAAR 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 the State of Texas Assessments (STAAR®) End-of-Course (EOC) assessments 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 *Texas Essential Knowledge and Skills (TEKS) Standards,* 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 EOC assessments directly to what students are expected to learn based on the *TEKS Standards*. Student achievement on all STAAR EOC assessments is classified into four performance levels that delineate the knowledge, skills, and abilities for which students are able to demonstrate mastery.

The policy-level PLDs for the performance levels provide general expectations for student achievement on the STAAR assessments to be classified into each performance level. These do not differentiate student performance between courses. Table 4 presents the four performance levels with their respective policy descriptions.

[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

Table 4. Policy Performance Level Descriptors

Label	Description
Masters Grade Level	Performance in this category indicates that students are expected to succeed in the next grade or course with little or no academic intervention. Students in this category demonstrate the ability to think critically and apply the assessed knowledge and skills in varied contexts, both familiar and unfamiliar.
Meets Grade Level	Performance in this category indicates that students have a high likelihood of success in the next grade or course but may still need some short-term, targeted academic intervention. Students in this category generally demonstrate the ability to think critically and apply the assessed knowledge and skills in familiar contexts.
Approaches Grade Level	Performance in this category indicates that students are likely to succeed in the next grade or course with targeted academic intervention. Students in this category generally demonstrate the ability to apply the assessed knowledge and skills in familiar contexts.
Did Not Meet Grade Level	Performance in this category indicates that students are unlikely to succeed in the next grade or course without significant, ongoing academic intervention. Students in this category do not demonstrate a sufficient understanding of the assessed knowledge and skills.

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. This includes content standards and educator judgments regarding what students should know based on the *TEKS Standards* 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 possess 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 EOC assessments have four 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 in the standards defined in the *TEKS Standards* through classification of their achievement on a STAAR 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 Standards* provide 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 three of the four STAAR 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. No descriptors were developed for the lowest performance level because the most accurate way to describe the performance of a student classified as *Did Not Meet Grade Level* is as that of a student who has not demonstrated the knowledge and skills necessary to achieve *Approaches Grade Level*. Appendix A presents the full version of the PLDs.

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.

Because the STAAR EOC assessments are computer-delivered and the online test form was used for the standard setting process, the standard setting website allows panelists to access the assessment items in Pearson's secure online environment. During the meeting, panelists accessed the website using a computer provided by Pearson and set up specifically for the 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 website. Table 5 presents a list of the materials developed for panelists and their mode of presentation.

Table 5. Materials Prepared for Panelists

Panelist Material	Paper	Online
Meeting agenda	✓	✓
Panelist information survey		✓
Non-disclosure agreement		✓
TEKS Standards		✓
Subject-area test forms/items		✓*
"Experience the Assessment" items		✓*
"Experience the Assessment" response form	✓	
Test form item map/answer key		✓
Practice judgment items		✓*
Practice judgment record sheet	✓	
Practice judgment survey		✓

Panelist Material	Paper	Online
Practice judgment form test map/answer key		✓
Judgment items*		✓
Judgment round record sheet	✓	
Judgment round surveys		✓
Performance level descriptors	✓	✓
Borderline descriptions	✓	
Process evaluations		✓

*Items were accessed through Cambium Assessment software.

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

Development of Presentation Materials

Customized 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 EOC standard setting meetings. Facilitator training included the following:

- *STAAR EOC assessments*—The facilitators were provided an overview of the STAAR 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 the 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 that were considered the committee cut score recommendations. Between judgment rounds, the estimated performance level cut score and ranges from the judgment process were presented so panelists could compare their content judgments to those from the process.

Chapter 3 – Standard Setting Meetings

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

- Purpose of the Standard Setting Meetings
- Committee Composition
- Facilitators and Staff
- Materials
- 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 them into cut scores for the performance levels. The purpose of the STAAR 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 EOC assessment in each content area.

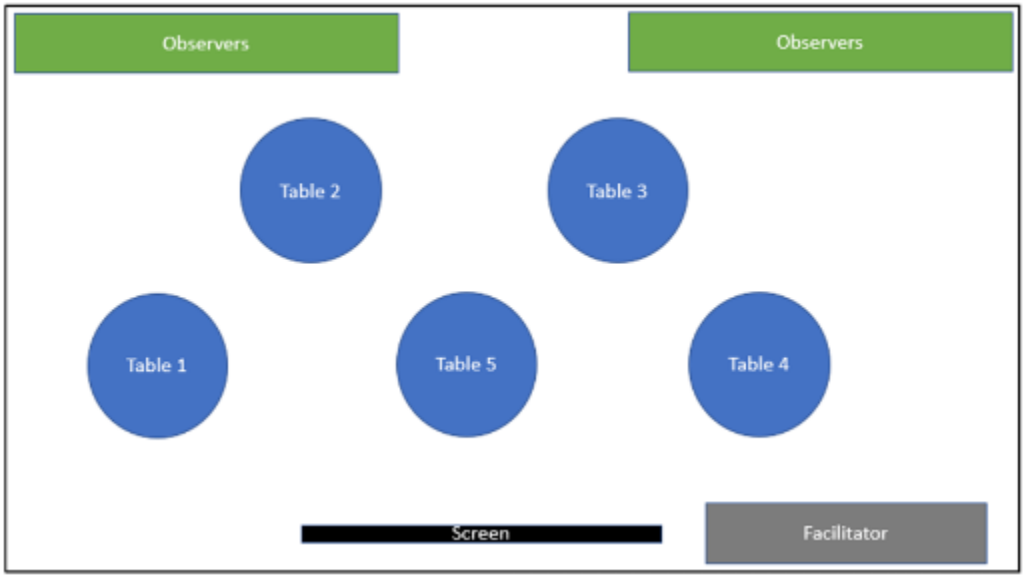
Student performance on each STAAR EOC assessment is classified into one of four performance levels. Each standard setting committee was asked to recommend three cut scores that would define the boundaries between the different performance levels for STAAR EOC. These recommended cut scores represent the 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 EOC assessment. Individuals in each meeting included three distinct groups, as illustrated in Figure 1:

- Meeting facilitators
- Committee panelists
- Observers and staff

Figure 1. General Room Setup for the Meeting



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 Standards*
- Understand the student population
- Be able to estimate item difficulty
- Understand the instructional environment
- Appreciate the consequences of the standards
- Be representative of key stakeholder groups

Each committee panel consisted of 14–17 members, resulting in a total of 74 panelists for the five courses covered at the STAAR EOC standard setting. To ensure that the panelists were subject-matter experts with grade-level expertise, educators recruited for the meeting possessed experience in the course for which the cut scores 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 EOC 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 with knowledge of and experience in facilitating standard setting meetings. The process facilitator was responsible for ensuring that appropriate processes were followed throughout all phases of the meeting and verifying that panelists had a solid understanding of the tasks they were being asked to complete. Content experts from Pearson and TEA were also available as observers to help answer content and policy questions that arose during the meeting.

Prior to the meeting, a staffing plan was provided to TEA that communicated the psychometric, content, and support staff required to attend each committee meeting. Table 6 presents the process and content facilitators for each standard setting committee.

Table 6. Process and Content Facilitators

Committee	Process Facilitator	Content Facilitator
English I	Jiawei Xiong/Winnie Reid	Cyndi Johnson
English II	Phyllis Echols	Sarah Cattan
Algebra I	Kuo-Feng Chang	Kathy Pieper
Biology	Phyllis Echols	Karen Kreder
U.S. History	Jiawei Xiong	Brian Vogel

Observers

Observers did not participate in the standard setting process. The individuals who attended as observers consisted of TEA staff, vendor (Pearson and Cambium Assessment) staff, 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

Four data analysts performed all analyses for the standard setting committees. Bailey Trip and Andrea Olson were the data analysts for English I and II, Algebra I, and Biology. Paige Rainforth and Morgen Hickey were the data analysts for U.S. History. During the meeting, the analysts collected panelist judgment data, performed independent analysis to verify the results, and prepared panelists' feedback. Bailey Trip was the lead analyst and performed the analysis onsite, while Jack Kissell was the replicator and completed the analysis offsite.

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 meetings. TEA was represented by Chris Rozunick, Director, Assessment Development Division; Chelaine Marion, Director of Foundation Education; Jo Ann Bilderback, Math, Science and Social Studies Director; and Mi-Suk Shim, Director of Psychometric Services. Additional TEA staff, including content and assessment specialists, assisted these individuals in monitoring the standard setting meeting.

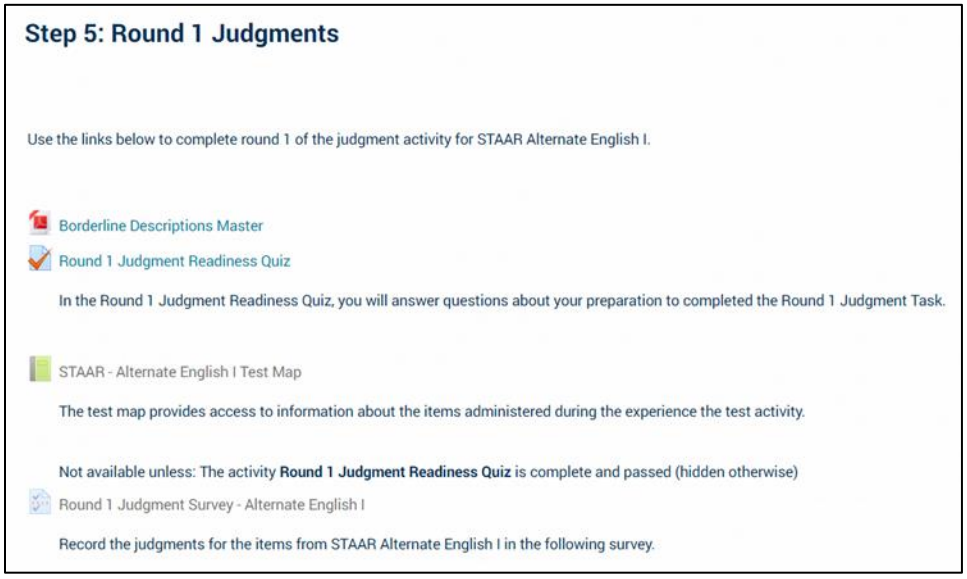
Materials

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



Because the STAAR EOC assessments are computer-delivered and the online test form was used for the standard setting process, Cambium Assessment's Content Rater system allowed the panelists to access the items. The standard setting website allowed panelists to access other materials in Pearson's secure online environment. During the meeting, panelists accessed the Content Rater system and 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 Content Rater system, 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 hard copy materials used throughout the meeting, including the following:

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

Panelists were required to check in at the start of each day and to return their folders and check out at the end of each day. Panelists were provided with additional materials throughout the meeting and instructed to insert them 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 with feedback information, including data relative to panelist agreement, student performance on the items, and student performance on the overall test. During Round 3, the panelists were asked the following for the Level 2 (*Approaches*) 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 [maximum points for that form]."*

For the Level 3 (*Meets*) cut, they were informed additionally that:

> *"This would be a number between 1 and [maximum points for that form] and greater than your recommended cut score for Approaches."*

For Level 4 (*Masters*), they were informed that:

> *"This would be a number between 1 and [maximum points for that form] and greater than your recommended cut score for Meets."*

The maximum number of points for each form was 64 for English I and II, 53 for Biology, 78 for U.S. History, and 59 for Algebra I.

Standard Setting Meeting Proceedings

Standard setting was conducted in two 2-day meetings, February 13–14 and 15–16, 2023. Appendix D presents a complete agenda for the meetings. Table 7 presents an overview of the agenda.

Table 7. Overview of Agenda

Session	Activity
Pre-Work	Complete Pre-Meeting Activities, Including a Review of PLDs
General Session	Welcome and Overview of STAAR 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 Results Round 3 Recommendations Closing Remarks and Final Evaluation

Pre-Work

The individuals recruited as panelists were registered into the Pearson standard setting website one to two weeks before 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 website. When panelists first logged in, they were required to create a unique, strong password consisting of at least eight characters, including at least one lowercase letter, one uppercase letter, one number, and one symbol.

Once panelists logged into the website, they only had limited access to certain materials for their assigned committee, as this occurred before the first day of the standard setting meeting. 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 meeting. Completion of the pre-work maximized the efficiency of time usage during the meetings. Pre-work activities included the following:

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

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 before 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 with an overview of the STAAR 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 cut scores. 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 performance level for the test associated with the committee. The committee provided recommendations using each activity 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 the use of those resources during the standard setting process. The panelists had an opportunity to ask questions before proceeding.

"Experience the Assessment" Activity

The panelists were given an overview of the test design and item types on the STAAR EOC assessment for their committee. Panelists then reviewed the test items students will take. As panelists reviewed the items, they were encouraged to think from a student's perspective and take notes on 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. A good reference point was thereby provided 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 assessment because it was expected that content experts would need less time to complete the test than students. If panelists did not complete the assessment 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 considered the knowledge and skills of students with performance at the borderline (i.e., a student who is just barely past the point of entry for that performance level). The STAAR EOC assessments have four performance levels, and panelists were asked to develop borderline descriptions for three of them:

- Level 2: *Approaches Grade Level*
- Level 3: *Meets Grade Level*
- Level 4: *Masters Grade Level*

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 3 performance level with the entire committee to create a framework for the activity.
- *Step 2*—After the modeling example, panelists worked in their table groups to review the draft PLDs for Level 3. Each table group created a set of descriptions that identified the key characteristics of student performance at the borderline of Level 3. Questions panelists were asked to consider included the following:
    - What would a student with performance just barely at Level 3 be able to do with respect to the PLDs?
    - What differentiates student performance at the borderline of Level 3 from a student in the middle or upper end of this level?
    - What differentiates a student performance at the borderline of Level 3 from the upper end of Level 2?
- *Step 3*—The facilitator collected the Level 3 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 2 and Level 4 performance levels, 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, skills, and abilities 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 outside of 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 with 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 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. For example:

- *"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 them 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 the panelists did not, they 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 was found that two out of three students performing at the borderline would earn. 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 through the standard setting website.
- Test map—A summary of the items on the test form that includes the following:
    - Item position from the order of presentation
    - Item scoring key and scoring rubrics, notes, and exemplars

    - Maximum number of possible points for each item
    - TEKS Standard(s) aligned to each item
- Judgment record form—Used by panelists to record 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 and Level 4. 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 and 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 process review by the meeting facilitator.

Practice Judgment Activity

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

- give panelists experience reviewing and making judgments for different item types,
- familiarize panelists with the paper judgment record sheet and judgment survey in the standard setting website, 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 was available that panelists would review during the actual judgment rounds. The practice activity included a range of item types, item difficulties, and scoring types.

Following the practice judgments, the facilitators showed item-level results interactively through the standard setting website, including the percentage of panelists who selected each point value for each performance level. The facilitator walked through the judgment materials for the first few items to ensure that panelists knew where to locate key information when making their judgments. The group also discussed a few practice items to better understand that various judgments were possible. Panelists were reminded to refer to the borderline descriptions along with 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 independent 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)
- Do you understand the feedback data provided? (Rounds 2 and 3)
- Are you ready to begin the judgment activity? (Rounds 1, 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 Level 2 based on the borderline descriptions and the knowledge and skills the item required. The panelists then made judgments about the same item for Level 3 and Level 4 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 also provided with a paper record sheet so they could keep a record of their judgments. 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 the sum of their individual item judgments as a reference point for the Round 3 judgments. Panelists made a judgment regarding the number of points a student with performance at the borderline of the level would likely earn across all items on the test form.

After all panelists completed the judgment activity for the round, the data analysts from Pearson analyzed 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 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 members 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 show the frequency of individual cut score recommendations for each performance level and across adjacent performance levels.
- *Item-level judgment agreement across panelists*—This is the distribution of individual judgments for each item and performance level.
- *Cut scores*—The estimated cut score is provided 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 8 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. They were told 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 8. 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 cut scores from the previous administration onto the spring 2023 standard setting form and determining ranges around each performance level. Specifically, the raw cut scores from the spring 2022 raw score look-up table along with the associated theta values and conditional standard error of measurements (CSEMs) were determined for each performance level of each assessment. The CSEM was used to create a reasonable range around the cut score for each performance level. The reasonable range values on the spring 2022 theta scale were matched with the nearest theta values on the raw score look-up table from the 2023 pre-equated raw score look-up table for each assessment. The raw scores associated with the reasonable range of theta values from the spring 2023 pre-equated raw score look-up table were used to establish the benchmark reasonable ranges. If the raw score values associated with the maximum of one performance level range were 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 9 presents the benchmark reasonable ranges that were presented to the panelists.

Table 9. Benchmark Reasonable Ranges (Raw Score Points)

Course	*Approaches*	*Meets*	*Masters*
English I	24–31	32–40	51–57
English II	23–31	32–41	54–59
Algebra I	17–23	29–36	37–44
Biology	12–17	22–28	35–42
U.S. History	19–26	32–40	45–54

Recommended Performance Level Cut Scores

During the standard setting meeting, variation was expected 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 10 presents the recommended cut scores for each performance level based on the Round 3 recommendations for each EOC assessment.

Table 10. Cut Score Recommendations from Standard Setting Committees

Course	Max. Score	*Approaches*	*Meets*	*Masters*
English I	64	27	37	53
English II	64	23	35	54
Algebra I	59	18	31	41
Biology	53	14	25	37
U.S. History	78	22	35	50

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 EOC assessments are reported using scale scores that are comparable across administration years. Table 11 presents the lowest and highest obtainable scores for each course.

Table 11. Obtainable Score Range

Course	LOSS	HOSS
English I	1750	6000
English II	1650	6050
Algebra I	1500	6430
Biology	1900	6260
U.S. History	1420	6750

The reporting scale was set using the two cut scores for the *Approaches* and *Meets* performance levels. The scale score for the *Masters* cut was found empirically Direct comparisons through averaging and aggregation across courses should not be made without study and/or statistical adjustments. The scaled 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 before the administration of the spring 2023 STAAR EOC assessments so the TEA commissioner could review and approve the performance level cut score recommendations in time to support the reporting of student performance on these assessments after the administration window ends. The standard setting process did not include the presentation of student performance on the assessment due to meeting scheduling. As part of the process for validating the reporting scale for the STAAR EOC assessments, TEA reviewed the distribution of student classification across the performance levels based on the approved performance level 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 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 12 presents the final performance level cut scores on the IRT scale following the TEA reasonableness review.

Table 12. Final Recommended Cut Scores on the IRT Scale

Course	*Approaches*	*Meets*	*Masters*	*Approaches*	*Meets*	*Masters*	A (Slope)	B (Intercept)
English I	27	36	54	3775	4000	4606	429.3074	3845.4064
English II	27	36	56	3775	4000	4734	444.4006	3852.8590
Algebra I	20	32	41	3550	4000	4345	460.7351	3919.0028
Biology	14	25	38	3550	4000	4531	435.9620	4042.0267
U.S. History	22	36	50	3550	4000	4424	487.6991	4073.2524

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 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
- Technical Advisors' 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, mainstream special education, and English language learners.

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 the most represented, although each committee also included a significant number of teachers from urban and suburban districts. The set of panelists for this standard setting was well-selected to represent the teachers across the state, and the facilitators of the meeting noted this. Most teachers for four of the EOC committees (excluding U.S. History) were currently teaching in districts with low socioeconomic status (Table C.13).

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 in the standard setting website. During the practice judgment round, the panelists reviewed a reduced set of items and provided judgments for three performance levels. 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 judgement round, panelists responded to a readiness survey that asked whether panelists were prepared to make their judgments. Panelists were unable 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 in 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 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 introduction to the standard setting process was perceived as successful, with 100% of panelists responding that it was either Successful or Very Successful. The perception of the training on the standard setting process in the breakout groups was also very good; 100% of panelists in the committees responded that it was either Adequate or More than Adequate. More than 88% of panelists in the committees indicated that the practice judgment activity for the standard setting process was either Successful or Very Successful. These responses indicate that, overall, most panelists believed that the training prepared them to implement the standard setting procedure.

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

Introduction to the standard setting process

Rating	English I	English II	Algebra I	Biology	U.S. History
Not successful	1	–	–	–	–
Partially successful	2	–	–	–	1
Successful	6	6	5	7	8
Very Successful	5	11	9	8	5

Practice exercise for the standard setting procedures

Rating	English I	English II	Algebra I	Biology	U.S. History
Not successful	1	–	–	–	–
Partially successful	5	1	–	1	1
Successful	6	8	9	4	4
Very Successful	2	8	5	10	9

Training provided on the standard setting process

Rating	English I	English II	Algebra I	Biology	U.S. History
Not Adequate	2	–	–	–	1
Somewhat Adequate	1	1	1	1	–
Adequate	9	9	8	4	10
More Than Adequate	2	7	5	10	3

Perceived Validity of the Workshop

Panelists and reviewers communicated their perceived validity of the standard setting meeting and the recommended cut scores as part of the workshop evaluation. Evaluations are important as evidence to establish the validity of recommended cut scores for the performance levels.

Panelist Evaluations

Generally, the panelists were satisfied with their recommendations and with the overall workshop. 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 performance level. Figure 4 presents the results of the evaluation survey across committees and indicates that the PLDs were reasonable for each performance level. Appendix I presents the results for all evaluation survey questions.

Figure 4. Evaluation Results on Reasonableness of the PLDs by Performance Level

Approaches Grade Level PLDs

Rating	English I	English II	Algebra I	Biology	U.S. History
Not Confident	–	1	1	–	1
Somewhat Confident	3	1	–	1	1
Confident	4	7	5	5	3
Very Confident	7	7	7	9	9

Meets Grade Level PLDs

Rating	English I	English II	Algebra I	Biology	U.S. History
Not Confident	–	–	–	–	–
Somewhat Confident	1	1	1	–	–
Confident	7	9	5	5	6
Very Confident	6	6	7	10	8

Masters Grade Level PLDs

Rating	English I	English II	Algebra I	Biology	U.S. History
Not Confident	–	–	–	–	–
Somewhat Confident	1	2	3	–	1
Confident	6	8	3	8	7
Very Confident	7	6	7	7	6

Most panelists had confidence that the PLDs were reasonable for each performance level. In all committees, at least 80% of the panelists were Confident or Very Confident that the PLDs were reasonable for the performance levels. These responses provide evidence that, overall, the panelists perceived the PLDs 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

Approaches Grade Level Cut Scores

Rating	English I	English II	Algebra I	Biology	U.S. History
Not Confident	–	1	–	–	–
Somewhat Confident	3	1	2	–	1
Confident	4	4	3	1	3
Very Confident	7	10	7	14	10

Meets Grade Level Cut Scores

Rating	English I	English II	Algebra I	Biology	U.S. History
Not Confident	–	–	–	–	–
Somewhat Confident	1	2	1	–	1
Confident	5	4	4	2	3
Very Confident	8	10	8	13	10

Masters Grade Level Cut Scores

Rating	English I	English II	Algebra I	Biology	U.S. History
Not Confident	1	1	–	–	–
Somewhat Confident	1	2	2	–	2
Confident	5	5	4	5	4
Very Confident	7	8	7	10	8

As with the PLDs, all but one panelist (in English II, Algebra I, and U.S. History, each for *Approaches*) indicated that they had at least some confidence that the recommended cut scores represented appropriate levels of student performance for each performance level. The panelists demonstrated greatest confidence in the *Meets* cut score recommendation, with nearly all panelists selecting Confident or Very Confident.

Overall, this feedback from the standard setting panelists provides evidence for the validity of the cut score recommendations for each performance level from the standard setting committee.

References

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

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

Appendix A – Performance Level Descriptors

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

Performance Level Descriptors
When reading texts of increasing complexity,* students achieving Masters Grade Level Performance can • Synthesize information across multiple texts from a variety of genres to create new understanding and develop thoughtful interpretations of the text • Make complex inferences about texts based on explicit and implicit text evidence • Use critical inquiry to analyze the authors' choices and how they influence and communicate meaning within a variety of texts • Write skillfully developed essays that demonstrate grade-level mastery of writer's craft with advanced command of language conventions
When reading texts of increasing complexity,* students achieving Meets Grade Level Performance can • Analyze how the author's use of language informs and shapes the perception of readers • Analyze literary texts by examining themes and the ways in which literary devices contribute to the development of complex yet believable characters and linear and/or nonlinear plots • Demonstrate an understanding of informational and argumentative texts by analyzing the thesis or claim and evaluating structural elements and characteristics such as organizational patterns, pertinent examples, and counterarguments • Synthesize information across multiple texts to create new understanding • Make logical inferences and predictions based on explicit and implicit text evidence • Write well-developed essays that are suited to the writing task, with consistent command of grade-level appropriate conventions • Demonstrate proficient skills in revising and editing

When reading texts of increasing complexity,* students achieving Approaches Grade Level Performance can • Distinguish between the denotative and connotative meanings of words using context and reference materials • Recognize how literal and figurative language conveys meaning in texts • Demonstrate a basic understanding of literary texts by recognizing elements such as theme and plot development • Demonstrate a basic understanding of informational and argumentative texts by recognizing characteristics and structural elements such as key ideas, supporting evidence, and print and graphic features • Recognize the use of literary devices and their meaning in a text • Make logical connections and comparisons between texts representing similar or different genres • Make simple inferences and predictions based on explicit and implicit text evidence • Write basic essays that are generally suited to the writing task, with a partial command of grade-level appropriate conventions • Demonstrate developing skills in revising and editing
When reading texts of increasing complexity,* students achieving Did Not Meet Grade Level Performance can • Determine the denotative meaning of words using context and reference materials • Demonstrate limited understanding of the fundamental elements of literary texts such as character, setting, and themes and the characteristics of informational and argumentative texts such as thesis, claim, and organizational patterns • Make simple inferences about texts based on explicit text evidence • Write limited essays that are minimally developed and only marginally suited to the writing task, with little to no command of grade-level appropriate conventions • Demonstrate limited skills in revising and editing

*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 because it is nonliteral/figurative, abstract, or academic/technical; (2) sentence structure may be more varied, dense, and sophisticated; (3) the author's use of literary elements/devices, rhetorical strategies, organizational patterns, and text features may be more nuanced or sophisticated; (4) the topic/content may be less familiar or more cognitively demanding; and (5) relationships among ideas may be less explicit and require more interpretation, reasoning, and inferential thinking to understand the subtlety, nuances, and depth of ideas. The rigor of the writing task also increases from grade to grade due to the text complexity of the source text(s) students use in developing the essay and the sophistication of the topic.

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

Performance Level Descriptors
When reading texts of increasing complexity,* students achieving Masters Grade Level Performance can • Synthesize information across multiple texts from a variety of genres to create new understanding and develop thoughtful interpretations of the text • Make complex inferences about texts based on explicit and implicit text evidence • Use critical inquiry to analyze the authors' choices and how they influence and communicate meaning within a variety of texts • Write skillfully developed essays that demonstrate grade-level mastery of writer's craft with advanced command of language conventions
When reading texts of increasing complexity,* students achieving Meets Grade Level Performance can • Analyze how the author's use of language informs and shapes the perception of readers • Analyze literary texts by examining themes and the ways in which literary devices contribute to the development of complex yet believable characters through historical and cultural settings and events • Demonstrate an understanding of informational and argumentative texts by analyzing the thesis or claim and evaluating structural elements and characteristics such as organizational patterns, pertinent examples, and counterarguments • Synthesize information across multiple texts to create new understanding • Make logical inferences and predictions based on explicit and implicit text evidence • Write well-developed essays that are suited to the writing task, with consistent command of grade-level appropriate conventions • Demonstrate proficient skills in revising and editing
When reading texts of increasing complexity,* students achieving Approaches Grade Level Performance can • Distinguish between the denotative, connotative, and figurative meanings of words using context and reference materials • Recognize how literal and figurative language convey meaning in texts • Demonstrate a basic understanding of literary texts by recognizing elements such as theme and plot development • Demonstrate a basic understanding of informational and argumentative texts by recognizing characteristics and structural elements such as key ideas, supporting evidence, and print and graphic features • Recognize the use of literary devices and their meaning in a text • Make logical connections and comparisons between texts representing similar or different genres

- Make simple inferences and predictions based on explicit and implicit text evidence
- Write basic essays that are generally suited to the writing task, with a partial command of grade-level appropriate conventions
- Demonstrate developing skills in revising and editing

When reading texts of increasing complexity,* students achieving Did Not Meet Grade Level Performance can

- Determine the denotative meaning of words using context and reference materials
- Demonstrate limited understanding of the fundamental elements of literary texts such as character, setting, and themes and the characteristics of informational and argumentative texts such as thesis, claim, and organizational patterns
- Make simple inferences about texts based on explicit text evidence
- Write limited essays that are minimally developed and only marginally suited to the writing task, with little to no command of grade-level appropriate conventions
- Demonstrate limited skills in revising and editing

*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 because it is nonliteral/figurative, abstract, or academic/technical; (2) sentence structure may be more varied, dense, and sophisticated; (3) the author's use of literary elements/devices, rhetorical strategies, organizational patterns, and text features may be more nuanced or sophisticated; (4) the topic/content may be less familiar or more cognitively demanding; and (5) relationships among ideas may be less explicit and require more interpretation, reasoning, and inferential thinking to understand the subtlety, nuances, and depth of ideas. The rigor of the writing task also increases from grade to grade due to the text complexity of the source text(s) students use in developing the essay and the sophistication of the topic.

State of Texas Assessments of Academic Readiness (STAAR®)
Performance Level Descriptors
Algebra I

Performance Level Descriptors
The mathematical process skills describe ways in which students are expected to engage in the content. They are not assessed in isolation but are incorporated into questions that assess Algebra I content. The process skills focus on applying mathematics to solve problems, analyze mathematical relationships, and communicate mathematical ideas.
Students achieving Masters Grade Level Performance can • Evaluate the reasonableness of the domain and range of linear functions • Generate representations of exponential functions • Make predictions from exponential functions that provide a reasonable fit to data for real-world problems • Divide polynomial expressions
Students achieving Meets Grade Level Performance can • Factor and multiply polynomial expressions • Determine the domain and range of linear, quadratic, and exponential functions • Calculate the rate of change of linear functions in mathematical and real-world problems • Determine solutions to quadratic equations, linear inequalities, and systems of linear equations in mathematical and real-world problems • Formulate linear and quadratic equations, linear inequalities, and systems of linear equations to solve problems • Estimate solutions and make predictions from linear and quadratic functions that provide a reasonable fit to data for real-world problems • Identify attributes of an exponential function from its graph • Use the properties of exponents
Students achieving Approaches Grade Level Performance can • Identify solutions to systems of equations and inequalities from a graph • Factor quadratic expressions • Determine the domain and range of linear, quadratic, and exponential functions using a graph • Add and subtract polynomial expressions • Formulate linear and quadratic equations, linear inequalities, and systems of linear equations • Generate representations of linear and quadratic functions and linear inequalities • Analyze the effects of parameter changes on the graph of linear and quadratic parent functions • Solve a linear equation
Students achieving Did Not Meet Grade Level Performance can • Identify slopes and y-intercepts of linear functions from tables, graphs, and equations in slope- intercept form • Identify attributes of a linear or quadratic function from its graph • Write a linear equation, function, inequality or system of equation given a verbal description • Simplify a square root expression • Calculate the rate of change of linear functions from a table or graph

State of Texas Assessments of Academic Readiness (STAAR®)
Performance Level Descriptors
Biology

Performance Level Descriptors
Scientific process skills are not assessed in isolation but are incorporated into questions that assess the biology content. These process skills focus on safe, environmentally appropriate, and ethical laboratory and field investigations; using scientific methods and equipment in investigations; and using critical thinking, scientific reasoning, and problem solving to make informed decisions.
Students achieving Masters Grade Level Performance can • Explain the effects of a variety of evolutionary mechanisms • Apply the regulation of gene expression to its role in protein synthesis • Evaluate how genes affect both Mendelian and non-Mendelian inheritance patterns • Analyze the impact of environmental change on ecosystem stability
Students achieving Meets Grade Level Performance can • Summarize the role of biomolecules in the metabolic, homeostatic, and reproductive processes that occur in cells • Analyze how viruses are different from cells and how viruses can affect cells • Describe the roles of DNA and RNA in gene expression • Describe how genes affect inheritance patterns and use this information to predict outcomes of monohybrid and dihybrid crosses • Analyze and evaluate the evidence, processes, and effects of evolutionary theory • Classify organisms based upon similarities and differences • Interpret interactions between organisms and their environment • Describe how changes in the environment alter ecosystems
Students achieving Approaches Grade Level Performance can • Distinguish between prokaryotic and eukaryotic organisms based on their cellular structures • Recognize that the genetic code in DNA is universal • Predict the outcomes of a simple Mendelian genetic cross • Identify the body systems that interact to carry out biological processes in animals • Relate the stages of ecological succession to the diversity of species in an ecosystem
Students achieving Did Not Meet Grade Level Performance can • Recognize the components of DNA • Identify the structures and functions of cells and viruses • Recognize that genes affect inheritance • Recognize the effects of evolution • Identify relationships among organisms

Texas Education Agency
Student Assessment Division
April 2019

State of Texas Assessments of Academic Readiness (STAAR®)
Performance Level Descriptors
U.S. History Studies Since 1877

Performance Level Descriptors
Students' social studies skills are not assessed in isolation but are incorporated into questions that assess understanding of U.S. history content. These skills focus on applying critical- thinking skills to interpret, organize, and analyze social studies information from a variety of sources.
Students achieving Masters Grade Level Performance can • Evaluate historical perspectives on major events and issues in U.S. history • Apply content knowledge in multiple contexts to make historical connections and evaluate change over time • Evaluate historical justifications and interpretations through the examination of multiple and varied sources
Students achieving Meets Grade Level Performance can • Apply understanding of U.S. constitutional principles to major events in U.S. history • Analyze the domestic and international impact of U.S. participation in wars and international relations • Analyze issues related to the development of the U.S. economic system • Analyze geographic and cultural influences on the United States • Evaluate the impact of reform movements, court cases, and legislation on the civil and political rights of citizens.
Students achieving Approaches Grade Level Performance can • Describe the effects of U.S. government and economic policies and actions • Describe the role and influence of the United States in the international community • Describe geographic and cultural influences on the United States • Describe the impact of significant individuals, groups, organizations, and policies on • U.S. history • Explain the impact of science and technology on the United States
Students achieving Did Not Meet Grade Level Performance can • Identify significant individuals, events, and issues in U.S. history • Define major social studies terminology • Identify and use social studies sources • Recognize major historical points of reference

Texas Education Agency
Revised July 2019

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 to panelists through TestNav8, the online testing platform used for administering the EOC assessments.
- *Open-ended item rubrics*—These documents presented the scoring rubrics and notes and student-produced response examples for each open-ended item presented to panelists.
- *Practice item judgment set*—This was presented to panelists through TestNav8, the online testing platform used for test administration.

Panelist Agendas

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

English I and English II
Agenda

Day 1 – February 13

7:30 am *Breakfast*

8:30 am *General Session*
- Welcome
- Overview of STAAR EOC 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
County of _____

Texas Education Agency
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.

Texas STAAR EOC Assessments
Standard Setting Meeting
February 2023

Experience the Assessment
Notes Sheet
English I

Sequence	
1	
2	
3	
4	
5	
6	
7	

Item Judgment Round Record Form

Item Code removed to protect item security. Only the first pages of this document are presented as an example.

Panelist Name: _____

Texas STAAR EOC Assessments
Standard Setting Meeting
February 2023

Judgment Rounds Record Sheet

English I

"What is the probability that a student with performance at the borderline of the level would answer the question correctly?"

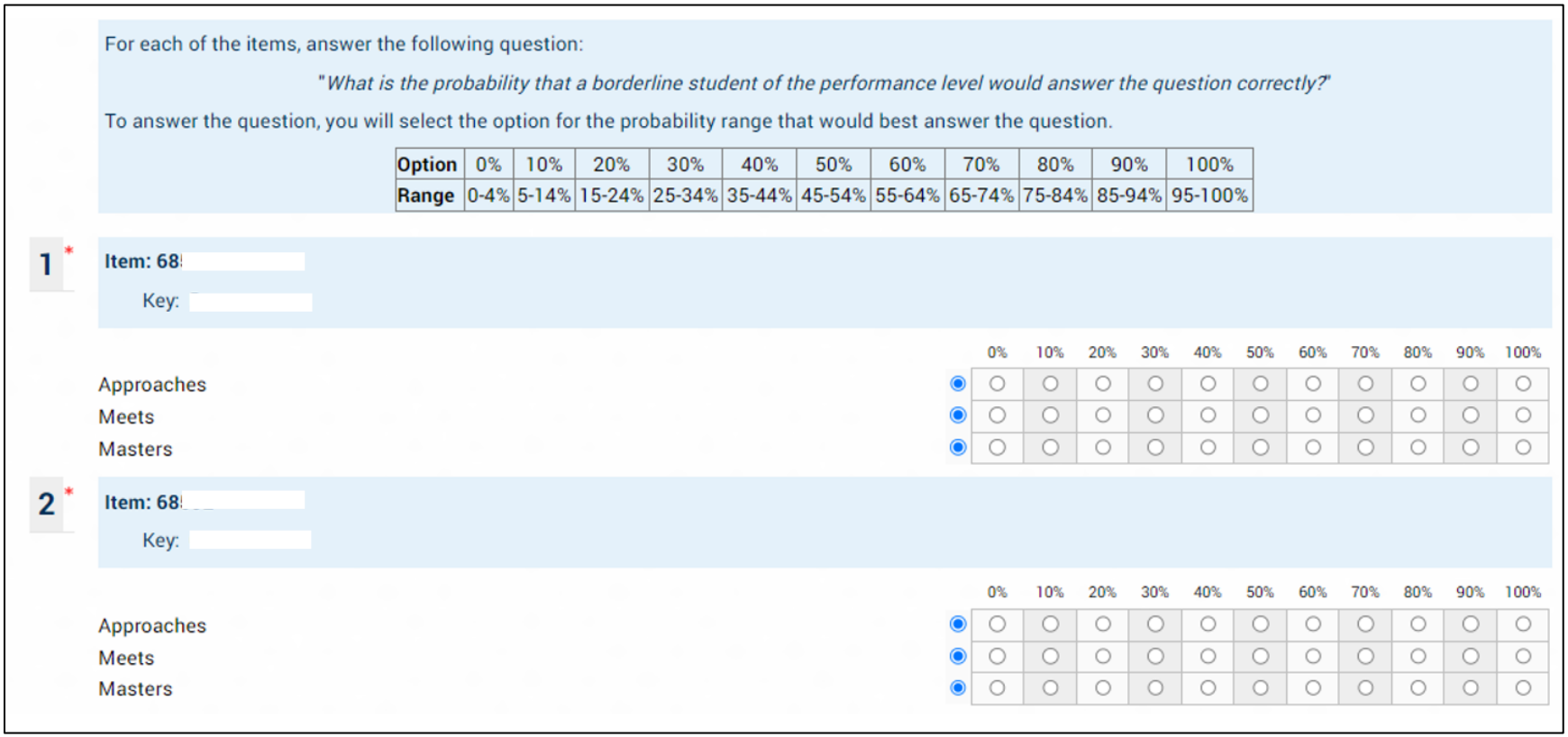
Seq.	Item Code	Judgment Round					
		Round 1			Round 2		
		L2	L3	L4	L2	L3	L4
1							
2							
3							
4							
5							
6							
7							
8							

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

Seq.	Item Code	Judgment Round					
		Round 1			Round 2		
		L2	L3	L4	L2	L3	L4
10							

Item Judgment Survey

The survey for only the first two items is shown.



For each of the items, answer the following question:

"What is the probability that a borderline student of the performance level would answer the question correctly?"

To answer the question, you will select the option for the probability range that would best answer the question.

Option	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
Range	0-4%	5-14%	15-24%	25-34%	35-44%	45-54%	55-64%	65-74%	75-84%	85-94%	95-100%

1 *

Item: 68

Key:

	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
Approaches	○	○	○	○	○	○	○	○	○	○	○
Meets	○	○	○	○	○	○	○	○	○	○	○
Masters	○	○	○	○	○	○	○	○	○	○	○

2 *

Item: 68

Key:

	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
Approaches	○	○	○	○	○	○	○	○	○	○	○
Meets	○	○	○	○	○	○	○	○	○	○	○
Masters	○	○	○	○	○	○	○	○	○	○	○

Process Evaluation #1

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

Process Evaluation Survey #1

English I

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)
Standard Setting Meeting

Process Evaluation Survey #2

English I

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 English I 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	○	○	○	○

Appendix C – Committee Panelist Composition

Table C.1. Panelist Position

Response Option	English I	English II	Algebra I	Biology	U.S. History
Teacher (K–12)	12	16	12	14	11
Teacher (Higher Ed.)	–	–	–	1	1
Administrator (School)	–	–	–	–	1
Administrator (District)	–	–	–	–	–
Other Position (e.g., Coordinator/Coach)	2	1	2	–	1
Total	14	17	14	15	14

Table C.2. Years of Total Teaching Experience

Response Option	English I	English II	Algebra I	Biology	U.S. History
1 to 5 years	–	1	–	1	–
6 to 10 years	5	4	5	2	–
11 to 15 years	3	4	1	2	5
16 to 20 years	2	4	4	5	2
More than 20 years	4	4	4	5	7
Total	14	17	14	15	14

Table C.3. Years of Experience Teaching this Subject

Response Option	English I	English II	Algebra I	Biology	U.S. History
None	–	1	–	–	–
1 to 5 years	3	4	3	3	–
6 to 10 years	5	7	6	3	4
11 to 15 years	3	4	1	2	3
16 to 20 years	2	1	4	4	3
More than 20 years	1	–	–	3	4
Total	14	17	14	15	14

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

Response Option	English I	English II	Algebra I	Biology	U.S. History
Mainstream special education	12	16	14	15	13
Self-contained special education	1	–	3	2	2
English language learners (ELL)	13	16	13	14	13
General education	14	17	14	15	14
Vocational technical instruction	2	4	2	7	3

Table C.5. Highest Degree Completed

Response Option	English I	English II	Algebra I	Biology	U.S. History
Bachelor's degree	3	5	6	3	6
Master's degree	10	11	8	11	8
Doctoral degree	1	1	–	1	–
Total	14	17	14	15	14

Table C.6. Demographic: Gender

Response Option	English I	English II	Algebra I	Biology	U.S. History
Female	10	14	12	11	10
Male	3	3	1	2	4
Other/No answer	1	–	–	2	–

Table C.7. Demographic: Ethnicity

Response Option	English I	English II	Algebra I	Biology	U.S. History
Hispanic or Latino	6	2	3	1	4
Not Hispanic or Latino	8	15	10	11	9
No answer	–	–	1	3	1

Table C.8. Demographic: Race

Response Option	English I	English II	Algebra I	Biology	U.S. History
American Indian or Alaskan Native	–	–	1	–	–
Asian	–	1	–	–	–
Black or African American	2	2	2	1	2
Middle Eastern	–	–	–	–	–
Native Hawaiian or Pacific Islander	–	–	–	–	1
White	11	12	10	11	10
No answer	1	2	1	3	1

Table C.9. Currently Work in a School District

Response Option	English I	English II	Algebra I	Biology	U.S. History
Yes	14	17	14	15	14
No (Higher Ed)	–	–	–	–	–

Table C.10. Size of School District

Response Option	English I	English II	Algebra I	Biology	U.S. History
Small	3	4	5	4	4
Medium	5	5	6	6	5
Large	6	8	3	5	5

Table C.11. Type of School District

Response Option	English I	English II	Algebra I	Biology	U.S. History
Rural	5	6	9	9	4
Metropolitan/Urban	5	6	1	1	4
Suburban	4	5	4	5	6

Table C.12. Socioeconomic Status of School District

Response Option	English I	English II	Algebra I	Biology	U.S. History
Low	11	10	12	9	5
Moderate	1	6	1	6	7
High	2	1	1	–	2

Appendix D – Standard Setting Meeting Agenda

TX Standard Setting

STAAR End-of-Course Assessments

Standard Setting Meeting Facilitator Agenda

Day 1 - February 13 and February 15

Breakfast available in the hotel from 7:30-8:30am 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 – Meets Group Work
1:45 pm	2:15 pm	Borderline Description Development – Meets Whole-Group Discussion
2:15 pm	2:30 pm	*Break*
2:30 pm	3:00 pm	Borderline Description Development – Approaching and Masters Group Work
3:00 pm	3:30 pm	Borderline Description Development – Approaching and Masters 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

Day 2 February 14 and February 16

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 the Pearson standard setting website. This was provided so that the panelist could check that the system recorded the judgments correctly.

English I – Individual Rating – Round 1

Table=1 Name=

SeqNo	UIN	L2	L3	L4
1MC		0.1	0.3	0.6
2MC		0.4	0.7	0.9
3MC		0.2	0.4	0.8
4MC		0.1	0.3	0.8
5MC		0.3	0.6	0.9
6MC		0.1	0.3	0.7
7MC		0.4	0.7	0.9
8MC		0.3	0.5	0.9
9MC		0.2	0.3	0.8
10XI		0.0	1.0	1.0

Individual Test-Level Recommendation

This provided the panelist with the recommendations for test-level cut scores based on their item judgments for the *Approaches Grade Level*, *Meets Grade Level*, and *Masters Grade Level* performance levels.

English I – Individual Cut Scores – Round 2

Table=1 Name=

Raw Score	A Roundup Raw Score	ME Raw Score	ME Roundup Raw Score	MA Round Score	MA Roundup Raw Score
20.5	21	37.2	38	55.1	56

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, roundup median, the minimum and maximum recommendation, and the first and third quartiles for each performance level.

English I Round 2 Summary Statistics - Overall

	N	Mean	Median	Roundup Median	Min.	Max.	Q1	Q3
A Raw Score	14	25.56	26.90	27	17.60	32.20	21.80	28.20
ME Raw Score	14	38.39	38.65	39	29.70	47.10	36.40	40.60
MA Raw Score	14	49.67	50.20	51	38.60	55.90	47.20	53.30

Item-Level Judgment Agreement

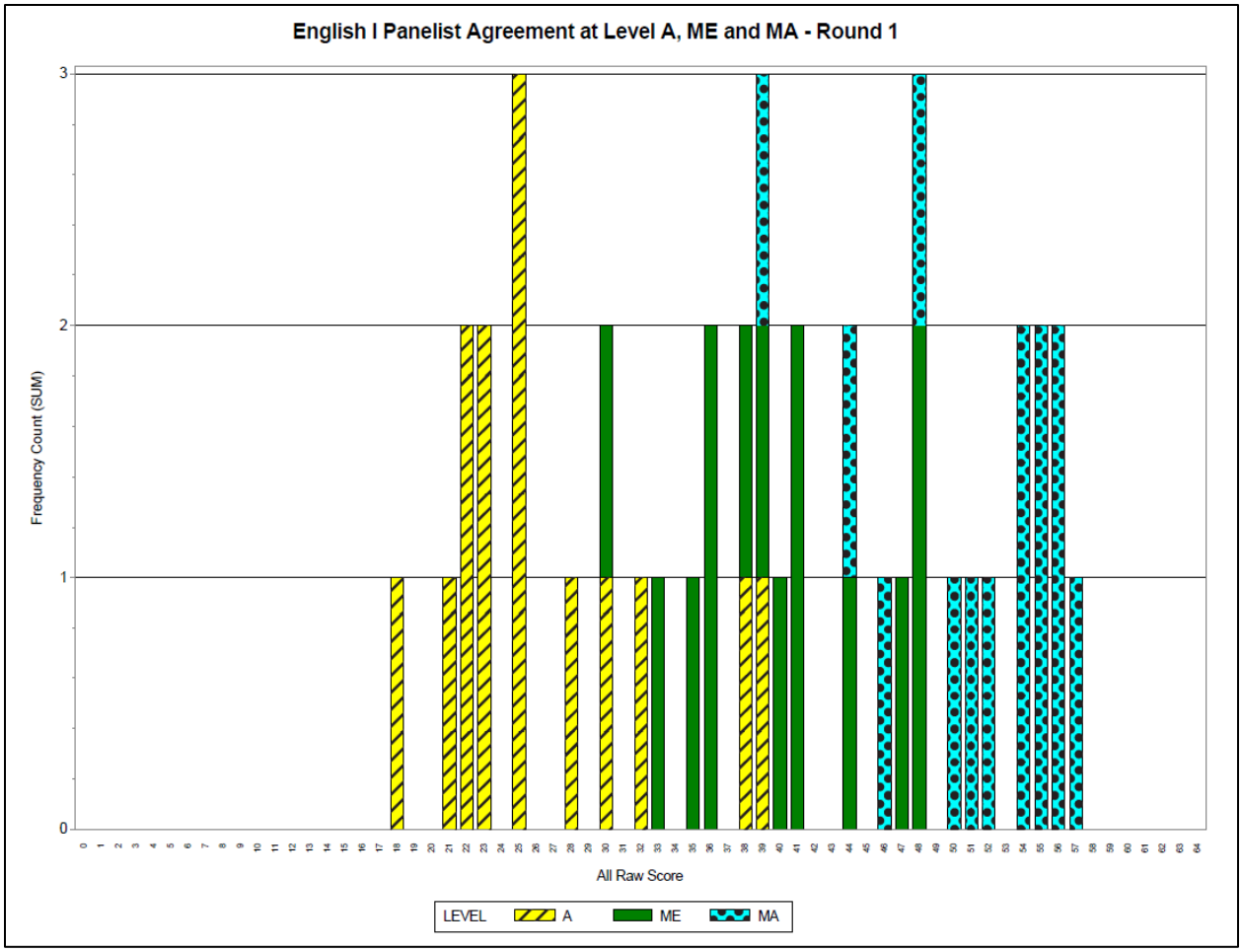
This provided the panelists with item-level judgment distributions for the committee for each item. Additionally, for each performance level, the items with the greatest level of judgment disagreement were identified.

English I Round 1 Round 1 Level A Flagged Items

SeqNo	UIN	Max. Points	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
24MC		1	.	7%	.	21%	21%	21%	14%	7%	7%	.	.
21MC		1	.	.	7%	7%	21%	21%	14%	21%	7%	.	.
26MC		1	.	7%	.	21%	21%	7%	21%	21%	.	.	.
14MC		1	.	7%	7%	29%	14%	7%	21%	7%	7%	.	.
5MC		1	.	.	7%	29%	14%	7%	21%	14%	7%	.	.
28MC		1	.	.	.	21%	29%	7%	14%	21%	7%	.	.
19MC		1	.	.	.	29%	14%	21%	21%	7%	7%	.	.
52MC		1	.	.	.	14%	21%	21%	14%	29%	.	.	.

Test-Level Panelist Recommendation Agreement

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


English I Panelist Agreement at Level A, ME and MA - Round 1
Frequency Count (SUM)
All Raw Score
LEVEL
A
ME
MA

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: *Approaches*	64	25	27	27
Level 3: *Meets*	64	40	39	37
Level 4: *Masters*	64	53	51	53

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

Performance Level	Max. Score	Round 1	Round 2	Round 3
Level 2: *Approaches*	64	21	22	23
Level 3: *Meets*	64	36	36	35
Level 4: *Masters*	64	50	51	54

Table F.3. Committee Recommended Cut Scores by Round, Algebra I

Performance Level	Max. Score	Round 1	Round 2	Round 3
Level 2: *Approaches*	59	20	18	18
Level 3: *Meets*	59	33	31	31
Level 4: *Masters*	59	48	45	41

Table F.4. Committee Recommended Cut Scores by Round, Biology

Performance Level	Max. Score	Round 1	Round 2	Round 3
Level 2: *Approaches*	53	14	14	14
Level 3: *Meets*	53	25	25	25
Level 4: *Masters*	53	37	37	37

Table F.5. Committee Recommended Cut Scores by Round, History

Performance Level	Max. Score	Round 1	Round 2	Round 3
Level 2: *Approaches*	78	24	20	22
Level 3: *Meets*	78	45	38	35
Level 4: *Masters*	78	64	58	50

Appendix G – Recommended Cut Score Summary Statistics

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

Round	Statistic	*Approaches*	*Meets*	*Masters*
1	Mean	26.12	39.19	50.71
	Minimum	17.8	30.0	38.8
	Q1	21.9	35.6	47.8
	Median	**24.2**	**39.2**	**52.2**
	Roundup Median	**25.0**	**40.0**	**53.0**
	Q3	29.5	43.7	54.8
	Maximum	38.9	47.5	56.4
2	Mean	25.56	38.39	49.67
	Minimum	17.6	29.7	38.6
	Q1	21.8	36.4	47.2
	Median	**26.9**	**38.7**	**50.2**
	Roundup Median	**27.0**	**39.0**	**51.0**
	Q3	28.2	40.6	53.3
	Maximum	32.2	47.1	55.9
3	Mean	27	37	53
	Minimum	27.1	36.5	53
	Q1	24	32	51
	Median	**31**	**40**	**57**
	Q3	26	35	51
	Maximum	28	38	54

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

Round	Statistic	*Approaches*	*Meets*	*Masters*
1	Mean	21.22	35.18	49.48
	Minimum	12.2	25.3	39.6
	Q1	16.4	32.4	47.9
	Median	**20.3**	**35.3**	**50.0**
	Roundup Median	**21.0**	**36.0**	**50.0**
	Q3	24.9	37.2	52.5
	Maximum	36.4	45.6	57.7
2	Mean	22.56	36.64	51.14
	Minimum	17.7	32.2	46.3
	Q1	19.6	34.2	49.8
	Median	**21.5**	**35.9**	**50.9**
	Roundup Median	**22.0**	**36.0**	**51.0**
	Q3	24.3	36.8	52.0
	Maximum	36.1	45.9	58.1
3	Mean	23.10	34.60	53.50
	Minimum	22	32	52
	Q1	23	33	52
	Median	**23**	**35**	**54**
	Q3	23	36	54
	Maximum	27	37	56

Table G.3. Recommended Cut Score Summary Statistics, Algebra I

Round	Statistic	*Approaches*	*Meets*	*Masters*
1	Mean	19.41	32.37	46.16
	Minimum	13.0	24.1	33.0
	Q1	17.2	27.7	41.7
	Median	**19.6**	**32.8**	**47.1**
	Roundup Median	**20.0**	**33.0**	**48.0**
	Q3	20.9	36.6	50.8
	Maximum	25.4	38.4	54.3
2	Mean	17.98	30.91	45.10
	Minimum	13.0	24.2	33.1
	Q1	17.0	28.9	43.9
	Median	**18.0**	**31.0**	**44.7**
	Roundup Median	**18.0**	**31.0**	**45.0**
	Q3	18.5	33.7	49.0
	Maximum	25.3	37.0	50.4
3	Mean	18.00	31.00	41.10
	Minimum	18	31	40
	Q1	18	31	41
	Median	**18**	**31**	**41**
	Q3	18	31	41
	Maximum	18	31	43

Table G.4. Recommended Cut Score Summary Statistics, Biology

Round	Statistic	*Approaches*	*Meets*	*Masters*
1	Mean	14.54	24.17	36.46
	Minimum	10.4	19.4	26.8
	Q1	12.6	21.3	31.6
	Median	**14.0**	**24.1**	**36.1**
	Roundup Median	**14.0**	**25.0**	**37.0**
	Q3	15.4	26.6	40.2
	Maximum	20.4	30.1	46.7
2	Mean	14.22	24.99	37.61
	Minimum	10.8	22.1	30.9
	Q1	11.9	23.2	35.7
	Median	**13.7**	**24.2**	**36.9**
	Roundup Median	**14.0**	**25.0**	**37.0**
	Q3	16.4	27.0	40.1
	Maximum	19.2	29.5	45.1
3	Mean	14.40	24.90	37.30
	Minimum	13	23	34
	Q1	14	24	37
	Median	**14**	**25**	**37**
	Q3	15	25	38
	Maximum	17	27	44

Table G.5. Recommended Cut Score Summary Statistics, U.S. History

Round	Statistic	*Approaches*	*Meets*	*Masters*
1	Mean	23.46	44.71	62.83
	Minimum	15.1	37.3	54.0
	Q1	18.5	41.6	57.8
	Median	**23.5**	**44.8**	**63.1**
	Roundup Median	**24.0**	**45.0**	**64.0**
	Q3	28.4	48.4	67.5
	Maximum	32.1	52.1	72.7
2	Mean	19.74	38.51	55.31
	Minimum	14.0	33.9	41.6
	Q1	15.6	36.4	53.3
	Median	**19.8**	**37.2**	**57.2**
	Roundup Median	**20.0**	**38.0**	**58.0**
	Q3	21.6	42.0	58.2
	Maximum	27.8	42.9	59.5
3	Mean	21.00	35.00	49.40
	Minimum	19	33	45
	Q1	19	34	47
	Median	**22**	**35**	**50**
	Q3	22	36	50
	Maximum	24	38	54

Appendix H – Test-Level Panelist Judgment 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: English I, Round 1

Raw Score	*Approaches*	*Meets*	*Masters*
18	1		
21	1		
22	2		
23	2		
25	3		
28	1		
30	1	1	
32	1		
33		1	
35		1	
36		2	
38	1	1	
39	1	1	1
40		1	
41		2	
44		1	1
46			1
47		1	
48		2	1
50			1
51			1
52			1
54			2
55			2
56			2
57			1

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

Raw Score	*Approaches*	*Meets*	*Masters*
18	1		
21	2		
22	1		
24	2		
27	1		

Raw Score	*Approaches*	*Meets*	*Masters*
28	2		
29	3		
30		1	
31	1		
32		1	
33	1		
35		1	
37		1	
38		2	
39		2	1
40		1	
41		2	
43		2	
44			1
47			1
48		1	1
49			2
51			2
52			2
54			1
55			1
56			2

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

Raw Score	*Approaches*	*Meets*	*Masters*
24	2		
25	1		
26	1		
27	4		
28	3		
30	1		
31	1		
32		2	
34		1	
35		2	
36		1	
37		1	
38		3	
39		1	
40		2	
51			5
52			1

Raw Score	*Approaches*	*Meets*	*Masters*
53			2
54			2
55			1
56			1
57			1

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

Raw Score	*Approaches*	*Meets*	*Masters*
13	1		
14	1		
15	1		
16	1		
17	1		
19	1		
20	1		
21	3		
22	2		
25	2		
26		1	
27		1	
29	1		
31	1	1	
33		3	
34		2	
36		3	
37	1	1	
38		1	
40			1
42		2	1
44		1	
46		1	1
48			2
49			3
50			1
51			2
52			1
53			2
54			1
55			1
58			1

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

Raw Score	*Approaches*	*Meets*	*Masters*
18	3		
19	1		
20	2		
21	1		
22	2		
23	2		
24	1		
25	2		
27	1		
29	1		
33		2	
34		1	
35		3	
36		3	
37	1	4	
38		1	
41		1	
44		1	
46		1	
47			1
49			2
50			3
51			3
52			4
53			1
54			1
56			1
59			1

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

Raw Score	*Approaches*	*Meets*	*Masters*
22	4		
23	11		
24	1		
27	1		
32		4	
33		1	
34		3	
35		1	
36		6	
37		2	

Raw Score	*Approaches*	*Meets*	*Masters*
52			6
54			9
55			1
56			1

Table H.7. Panelist Agreement Data: Algebra I, Round 1

Raw Score	*Approaches*	*Meets*	*Masters*
13	1		
17	2		
18	1		
19	2		
20	2		
21	3		
23	2		
25		1	
26	1		
28		3	
30		1	
31		1	
32		1	
33			1
34		1	
36		1	
37		3	
38		1	
39		1	
41			1
42			2
44			2
45			1
50			2
51			3
53			1
55			1

Table H.8. Panelist Agreement Data: Algebra I, Round 2

Raw Score	*Approaches*	*Meets*	*Masters*
13	1		
14	2		
17	1		
18	3		
19	4		

Raw Score	*Approaches*	*Meets*	*Masters*
21	1		
23	1		
25		1	
26	1		
27		1	
29		2	
30		2	
31		1	
32		2	
33		1	
34		1	1
35		2	
37		1	
42			1
43			1
44			1
45			4
46			1
49			2
50			2
51			1

Table H.9. Panelist Agreement Data: Algebra I, Round 3

Raw Score	*Approaches*	*Meets*	*Masters*
18	14		
31		14	
40			1
41			12
43			1

Table H.10. Panelist Agreement Data: Biology, Round 1

Raw Score	*Approaches*	*Meets*	*Masters*
11	1		
12	2		
13	2		
14	3		
15	1		
16	3		
19	1		
20	1	3	
21	1		
22		1	

Raw Score	*Approaches*	*Meets*	*Masters*
23		2	
24		1	
25		2	
26		2	
27		1	1
28		1	
29		1	
31		1	
32			4
36			2
37			2
38			1
40			1
41			1
43			1
45			1
47			1

Table H.11. Panelist Agreement Data: Biology, Round 2

Raw Score	*Approaches*	*Meets*	*Masters*
11	1		
12	3		
13	1		
14	4		
15	2		
17	1		
18	1		
19	1		
20	1		
23		2	
24		5	
25		2	
27		3	
28		2	
30		1	
31			1
33			1
35			1
36			1
37			4
38			2
40			1

Raw Score	*Approaches*	*Meets*	*Masters*
41			2
44			1
46			1

Table H.12. Panelist Agreement Data: Biology, Round 3

Raw Score	*Approaches*	*Meets*	*Masters*
13	1		
14	9		
15	4		
17	1		
23		1	
24		3	
25		8	
26		2	
27		1	
34			1
35			1
36			1
37			8
38			3
44			1

Table H.13. Panelist Agreement Data: U.S. History, Round 1

Raw Score	*Approaches*	*Meets*	*Masters*
16	2		
17	1		
19	1		
22	2		
23	1		
25	1		
26	1		
28	1		
29	1		
30	2		
33	1		
38		2	
40		1	
42		1	
43		1	
44		1	
45		2	
47		1	

Raw Score	*Approaches*	*Meets*	*Masters*
48		1	
49		2	
51		1	
53		1	
54			1
56			1
58			2
60			1
62			1
63			1
64			1
65			1
68			3
69			1
73			1

Table H.14. Panelist Agreement Data: U.S. History, Round 2

Raw Score	*Approaches*	*Meets*	*Masters*
14	1		
15	1		
16	2		
19	2		
20	1		
21	1		
22	3		
24	1		
25	1		
28	1		
34		1	
35		1	
36		1	
37		4	
38		1	
40		1	
41		1	
42		1	1
43		3	
51			1
54			3
57			2
58			3
59			1

Raw Score	*Approaches*	*Meets*	*Masters*
60			3

Table H.15. Panelist Agreement Data: U.S. History, Round 3

Raw Score	*Approaches*	*Meets*	*Masters*
19	4		
20	1		
21	2		
22	6		
24	1		
33		2	
34		3	
35		5	
36		2	
37		1	
38		1	
45			1
47			3
48			2
50			5
52			1
53			1
54			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 EOC Assessment

Rating	English I	English II	Algebra I	Biology	U.S. History
Not successful	–	–	–	–	–
Partially successful	–	1	–	–	–
Successful	7	7	4	6	6
Very Successful	6	10	10	9	8

Introduction to the Standard Setting Process

Rating	English I	English II	Algebra I	Biology	U.S. History
Not successful	1	–	–	–	–
Partially successful	2	–	–	–	1
Successful	6	6	5	7	8
Very Successful	5	11	9	8	5

Experiencing the Actual Assessment

Rating	English I	English II	Algebra I	Biology	U.S. History
Not successful	–	–	–	–	–
Partially successful	4	1	–	–	–
Successful	5	3	6	3	8
Very Successful	5	13	8	12	6

Discussion of Scoring Items on the Assessment

Rating	English I	English II	Algebra I	Biology	U.S. History
Not successful	2	–	–	–	1
Partially successful	3	–	–	–	1
Successful	7	4	7	7	9
Very Successful	2	13	7	8	3

Discussion of Performance Level Descriptors (PLDs)

Rating	English I	English II	Algebra I	Biology	U.S. History
Not successful	1	–	–	–	–
Partially successful	2	–	–	2	1
Successful	7	6	8	5	8
Very Successful	4	12	6	8	5

Overview of the Standard Setting Procedure

Rating	English I	English II	Algebra I	Biology	U.S. History
Not successful	1	–	–	–	1
Partially successful	3	1	–	–	–
Successful	7	6	10	7	9
Very Successful	3	10	4	8	4

Practice Exercise for the Standard Setting Procedure

Rating	English I	English II	Algebra I	Biology	U.S. History
Not successful	1	–	–	–	–
Partially successful	5	1	–	1	1
Successful	6	8	9	4	4
Very Successful	2	8	5	10	9

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	Algebra I	Biology	U.S. History
Very Useful	5	13	7	10	7
Useful	6	2	5	3	4
Somewhat Useful	2	2	2	1	2
Not Useful	1	–	–	1	1

Borderline Description Development

Rating	English I	English II	Algebra I	Biology	U.S. History
Very Useful	–	12	8	10	6
Useful	6	3	4	4	5
Somewhat Useful	5	1	2	–	2
Not Useful	4	1	–	1	1

Standard Setting Training

Rating	English I	English II	Algebra I	Biology	U.S. History
Very Useful	2	13	5	12	5
Useful	9	2	8	2	7
Somewhat Useful	1	2	1	–	–
Not Useful	2	–	–	1	2

How adequate were the following elements of the session?

Training Provided on the Standard Setting Process

Rating	English I	English II	Algebra I	Biology	U.S. History
Not Adequate	2	–	–	–	1
Somewhat Adequate	1	1	1	1	–
Adequate	9	9	8	4	10
More Than Adequate	2	7	5	10	3

Amount of Time Spent Training

Rating	English I	English II	Algebra I	Biology	U.S. History
Not Adequate	2	–	–	–	–
Somewhat Adequate	3	1	1	2	–
Adequate	9	9	10	4	10
More Than Adequate	–	7	3	9	4

Total Amount of Time to Discuss the PLDs

Rating	English I	English II	Algebra I	Biology	U.S. History
Not Adequate	1	–	–	–	–
Somewhat Adequate	2	–	–	1	–
Adequate	6	9	7	6	7
More Than Adequate	5	8	7	8	7

Total Amount of Time to Create and Discuss Borderline Descriptions

Rating	English I	English II	Algebra I	Biology	U.S. History
Not Adequate	3	–	–	–	–
Somewhat Adequate	2	–	–	2	–
Adequate	5	5	4	6	9
More Than Adequate	4	12	10	7	5

Total Amount of Time to Discuss the Practice Judgments

Rating	English I	English II	Algebra I	Biology	U.S. History
Not Adequate	2	3	–	1	–
Somewhat Adequate	3	4	4	3	3
Adequate	8	7	10	3	10
More Than Adequate	1	5	2	8	1

Process Evaluation Survey #2

Judgment Rounds

Rating	English I	English II	Algebra I	Biology	U.S. History
Not successful	–	–	–	–	–
Partially successful	1	2	–	–	–
Successful	7	6	6	3	6
Very Successful	6	8	7	12	8

Judgment Round Feedback – Table-level Statistics

Rating	English I	English II	Algebra I	Biology	U.S. History
Not successful	–	1	–	–	–
Partially successful	–	1	–	–	–
Successful	7	6	8	5	7
Very Successful	7	8	5	10	7

Judgment Round Feedback – Committee-Level Statistics

Rating	English I	English II	Algebra I	Biology	U.S. History
Not successful	–	1	–	–	–
Partially successful	1	3	–	–	–
Successful	8	4	7	4	6
Very Successful	–	8	6	11	8

Judgment Round Feedback – Panelist Agreement Data

Rating	English I	English II	Algebra I	Biology	U.S. History
Not successful	–	–	–	–	–
Partially successful	–	3	1	–	–
Successful	8	6	4	4	6
Very Successful	6	7	8	10	6

Discussions After Each Round

Rating	English I	English II	Algebra I	Biology	U.S. History
Not successful	–	–	–	–	–
Partially successful	1	3	–	–	–
Successful	7	6	3	3	4
Very Successful	6	7	10	12	10

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	Algebra I	Biology	U.S. History
Very Useful	5	9	7	11	10
Useful	7	6	5	2	4
Somewhat Useful	1	–	1	–	–
Not Useful	1	1	–	2	–

Committee-level Statistics After Round 2

Rating	English I	English II	Algebra I	Biology	U.S. History
Very Useful	4	8	8	13	11
Useful	8	8	3	–	3
Somewhat Useful	1	–	2	–	–
Not Useful	1	–	–	2	–

Panelist Agreement Data Provided After Round 1

Rating	English I	English II	Algebra I	Biology	U.S. History
Very Useful	4	10	7	12	11
Useful	8	5	5	1	3
Somewhat Useful	1	1	1	–	–
Not Useful	1	–	–	2	–

Panelist Agreement Data Provided After Round 2

Rating	English I	English II	Algebra I	Biology	U.S. History
Very Useful	4	10	7	11	11
Useful	8	4	5	2	3
Somewhat Useful	–	2	1	–	–
Not Useful	2	–	–	2	–

Discussion After Each Judgment Round

Rating	English I	English II	Algebra I	Biology	U.S. History
Very Useful	4	7	9	11	12
Useful	7	5	3	1	2
Somewhat Useful	2	4	1	1	–
Not Useful	1	–	–	2	–

How adequate were the following elements of the session?

Amount of Time to Make Judgments

Rating	English I	English II	Algebra I	Biology	U.S. History
Not Adequate	1	2	–	–	–
Somewhat Adequate	3	–	1	–	3
Adequate	7	8	6	6	5
More Than Adequate	3	6	6	9	5

Visual Presentation of the Feedback Provided

Rating	English I	English II	Algebra I	Biology	U.S. History
Not Adequate	–	1	–	–	–
Somewhat Adequate	2	1	–	–	–
Adequate	6	9	9	8	8
More Than Adequate	6	5	4	7	6

Number of Judgment Rounds

Rating	English I	English II	Algebra I	Biology	U.S. History
Not Adequate	–	1	1	–	–
Somewhat Adequate	1	–	–	–	–
Adequate	11	11	6	7	7
More Than Adequate	2	4	6	8	7

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

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

Level 2 – Approaches Grade Level

Rating	English I	English II	Algebra I	Biology	U.S. History
Not Confident	–	1	1	–	1
Somewhat Confident	3	1	–	1	1
Confident	4	7	5	5	3
Very Confident	7	7	7	9	9

Level 3 – Meets Grade Level

Rating	English I	English II	Algebra I	Biology	U.S. History
Not Confident	–	–	–	–	–
Somewhat Confident	1	1	1	–	–
Confident	7	9	5	5	6
Very Confident	6	6	7	10	8

Level 4 – Masters Grade Level

Rating	English I	English II	Algebra I	Biology	U.S. History
Not Confident	–	–	–	–	–
Somewhat Confident	1	2	3	–	1
Confident	6	8	3	8	7
Very Confident	7	6	7	7	6

How confident do you feel that the recommended cut scores for your course represent are reasonable for each student performance level?

Level 2 – Approaches Grade Level

Rating	English I	English II	Algebra I	Biology	U.S. History
Not Confident	–	1	–	–	–
Somewhat Confident	3	1	2	–	1
Confident	4	4	3	1	3
Very Confident	7	10	7	14	10

Level 3 – Meets Grade Level

Rating	English I	English II	Algebra I	Biology	U.S. History
Not Confident	–	–	–	–	–
Somewhat Confident	1	2	1	–	1
Confident	5	4	4	2	3
Very Confident	8	10	8	13	10

Level 4 – Masters Grade Level

Rating	English I	English II	Algebra I	Biology	U.S. History
Not Confident	1	1	–	–	–
Somewhat Confident	1	2	2	–	2
Confident	5	5	4	5	4
Very Confident	7	8	7	10	8

How adequate were the following elements of the session?

Facilities Used for the Meeting

Rating	English I	English II	Algebra I	Biology	U.S. History
Not Adequate	–	–	–	–	–
Somewhat Adequate	–	–	–	–	–
Adequate	4	4	4	2	3
More Than Adequate	10	12	9	13	11

Computers Used During the Meetings

Rating	English I	English II	Algebra I	Biology	U.S. History
Not Adequate	–	–	–	–	–
Somewhat Adequate	1	–	–	1	–
Adequate	4	4	6	4	7
More Than Adequate	9	12	7	10	7

Pearson Website for Accessing Materials and Making Judgments

Rating	English I	English II	Algebra I	Biology	U.S. History
Not Adequate	–	–	–	–	–
Somewhat Adequate	–	–	–	–	–
Adequate	5	8	6	4	6
More Than Adequate	9	8	7	11	8

Content Review System for Viewing Items

Rating	English I	English II	Algebra I	Biology	U.S. History
Not Adequate	–	–	–	–	–
Somewhat Adequate	–	–	1	–	1
Adequate	5	7	3	5	6
More Than Adequate	9	9	9	10	7

Materials Provided in the Folder

Rating	English I	English II	Algebra I	Biology	U.S. History
Not Adequate	–	–	–	–	–
Somewhat Adequate	–	–	–	1	1
Adequate	6	8	6	2	4
More Than Adequate	8	8	7	12	9

Workspace in Table Groups During the Meeting

Rating	English I	English II	Algebra I	Biology	U.S. History
Not Adequate	–	–	1	2	–
Somewhat Adequate	–	–	2	1	2
Adequate	7	8	6	5	7
More Than Adequate	7	8	4	7	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	Algebra I	Biology	U.S. History
Not Adequate	–	–	–	–	–
Somewhat Adequate	1	1	–	–	–
Adequate	6	7	4	4	5
More Than Adequate	7	10	9	11	9

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

Rating	English I	English II	Algebra I	Biology	U.S. History
Not Adequate	–	–	–	–	–
Somewhat Adequate	2	2	–	–	–
Adequate	5	7	6	4	5
More Than Adequate	7	7	7	11	9

Ask Questions About the Process of Making Cut Score Recommendations

Rating	English I	English II	Algebra I	Biology	U.S. History
Not Adequate	–	–	–	–	–
Somewhat Adequate	2	2	–	–	–
Adequate	4	7	5	6	5
More Than Adequate	8	7	8	9	9

Interact with Your Fellow Panelists

Rating	English I	English II	Algebra I	Biology	U.S. History
Not Adequate	–	–	–	–	–
Somewhat Adequate	–	1	–	–	–
Adequate	7	5	4	2	4
More Than Adequate	7	10	9	13	10

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

Fellow Panelists

Rating	English I	English II	Algebra I	Biology	U.S. History
Yes	13	15	13	15	13
No	–	–	–	–	–
Sometimes	1	1	–	–	1

Facilitators

Rating	English I	English II	Algebra I	Biology	U.S. History
Yes	11	14	13	15	14
No	–	–	–	–	–
Sometimes	3	2	–	–	–

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

English I:

- not applicable

- Developing the borderline PLD could have been done in one day and further discussion could have been done and not felt so rushed. Facilitators could have done a better job of facilitating group discussion to move the conversation along. Overall great group discussion and hearing others opinions to help determine cut score. Enjoyed when the Dr. explained in more detail how each discussion needed to take place
- It would be helpful for the presenter to have a mic to present as his soft-spoken voice could get lost within the conversational pieces. The presenter showed great growth as he got settled into the position.
- I appreciate the facilitator's patience and professionalism. He did a good job.
- I was a little confused initially, however, as the process was explained it all made since. Enjoyed the collaborations and conversations. Excellent job.
- I believe Day 1 could have been more efficient if there were examples provided of what the setting process looked like or what was wanted from the panelists. It would have been critical for me to see how well I did on the test but we did not have time for that. I also would have liked to know which TEKs corresponded to each question.
- None at this time.
- Our facilitator was a delightful young man with insufficient experience in presenting to groups. I think he would have benefitted from having a lead presenter, but the second day was much better. I also think we really needed 3 days for the process. Thank you for inviting me to serve -- I learned a great deal and enjoyed the experience.
- The session needed to be longer. There was not enough time to provide judgments for 52 items with considerable though for each one. We also needed more time to have discussions on the results. I think it should have been 3 days, vs. 2 days, with one day devoted to Borderline Descriptors. We needed more time to hash that out so it would be more useful during judgement. All of it was rushed. Facilitator needs to be much more informed when teaching the material as well. First day was rough. Second day was better
- The meeting was very successful; however, our presenter should be a master in the content being discussed in order for him/her to be able to address any subject related concerns brought forth by participants. A presenter in training should not be the person to direct a meeting where such high stakes come into play.
- This was an incredibly informative experience! I look forward to future sessions. The facilitator was fantastic!
- Thank you for the invitation! I look forward to working on more committees in the future.
- At the beginning of this meeting there was some confusion as to what the expectations were, however, by day 2 things became clear. Great experience.
- The standard setting process, facilitators, materials was very informative and allowed me the opportunity to adjust my approach to standards and assessment performance

English II:

- Thank you for this opportunity. Please consider me for future opportunities.
- I understand the range, however, I think telling us that we had to make our numbers fit in the range regardless of the work we had done disregarded our efforts & our thoughts on what students could do, especially when we were working diligently to use the borderline descriptors in making our judgements.

- Quite informative. Cannot wait to see how our input factors into the Spring administration of this test.
- In all honesty, once TEA's recommended scores came out, this whole thing felt like a formality. I think it's impossible to ask teachers to separate their judgments from their students' needs and abilities after Covid.
- Thank You for this opportunity!
- The item types have increased in rigor from the last iteration of STAAR testing. It is reasonable to expect that there will be a decrease in student performance based on this variable alone. I hope that in the next round of STAAR redesign we will explore ways for teachers to score performance tasks to evaluate components like Inquiry.
- I absolutely loved this experience and hope to be back soon!
- The second round judgment discussion was quite repetitive, after having discussed thoroughly in the first round.
- Great experience!
- Thank you!
- Great opportunity and great presenter
- The process was well guided and positive. I felt that the entire process ran very smoothly.
- Given the depth of thinking, it would be nice to have 2 1/2 days for this process. By the end of day 2, it was difficult to give input.
- na
- It is a great process to be included in.
- The standard setting process is so intricate and important, so glad that I was able to experience this. The use of equipment and layout of the room/tables was so incredibly helpful. Thank you!

Algebra I:

- If this information is taken and used accurately, it will have a positive impact on our students. A touch screen for judgement would be incredibly helpful. This was valuable to be able to have participated in this process. These cut scores were chosen with the rigor of this test in mind and with our students across the state in mind. The use of multiple TEKS combined into one question and the use of prior knowledge we cannot guarantee students have had makes the rigor even higher. Please take this into consideration.
- The standard setting process was very organized and I truly feel like the group of panelists represented a variety of students from various demographics and socioeconomic statuses. I feel like these cut scores truly represent the students abilities based on the learning gaps that have progressed during Covid and that this truly reflects the current level of students.
- Can I please be invited again for sessions like this?
- I really appreciate the ability to participate in the meeting. I love being able to see and work with the standards to see how the test is measured and put in towards the TEKS, and the ability level of each question.
- No comment
- I enjoyed it, and I really want to participate in future committees. The facilitators were great, the materials were clear and concise, and I loved the facilities that we used this time. Thank you for having me!

- Thank you for inviting me to participate in this committee. I feel like the work we did will have a positive impact o student performance. I do have concerns about how this year's scores will impact accountability ratings for our schools and districts.
- Great opportunity.
- Thank you for selecting me for this session.
- We justified the lower cut scores based on the higher rigor of the test.
- Thank for this opportunity. I truly believe the scores our committee came up with are justifiable and defendable.
- I would have liked the opportunity to adjust data between the Round 2 and the final determination for Round 3. I think that extra step would have been very helpful. The discussion after Round 2 was even more helpful than the discussion after Round 1.
- Thank you

Biology:

- Our Biology Facilitator did an excellent job facilitating our session. The standard setting process was brilliant process and well structured.
- Everything went very smoothly and I really enjoyed the opportunity to be on this committee and participate in this process.
- This was very informative and helpful. I do feel like we gave a very inclusive and accurate depiction of Texas students, and I hope that the information and standards remain where we set them. The work we put into these numbers is more than you could ask for, and the best you could hope for.
- na
- Great hotel and meals
- Thank you for giving us the opportunity to help with the process of testing out kids
- I feel that this was a great experience and I have a lot of useful information to take back to my district.
- This is a great process and will definitely continue to ensure that my instructional practices are aligned with the TEK(s) for Biology. In addition, I will continue to encourage other members of the profession to apply for and participate in these committee opportunities for the overall success of all student achievement. Thank you so much for this opportunity. It is very beneficial to educators.
- I loved this process. Even though I have taught Biology for many years, this process will help me see the test in a different light. Thank you for including me in this session!
- I would have liked a mouse for computer.. I loved that we had multiple presentation from every organization.
- They did a great job. I would recommend having students data to help us gage how well we think students will perform on different TEKS. The food was great and the facilitators, and TEA representative were kind and helpful.
- There should more guidance when setting the PLD s for Borderline descriptors so that they are complete for all levels. We had only some of them and if they were not created initially then we should have been guided to complete the chart so that we had those to use as we assessed the questions.

- I would have liked to have had a touch screen computer or a mouse or a track pad that more easily right clicks. There was too much wait time for the data to be processed. The Judgement Round Record Sheet was hard to write on/see your writing on the dark grey stripes. It was also frustrating that the paper pages that don't line up with the screen pages. (ex paper had numbers 1-3 but the online had more questions) The table space was too crowded.
- I thought it was handled and organized quite well.
- I appreciate this opportunity. I feel like the process was planned and executed very well. The facilitators were knowledgeable and helpful at all times and the materials were useful.

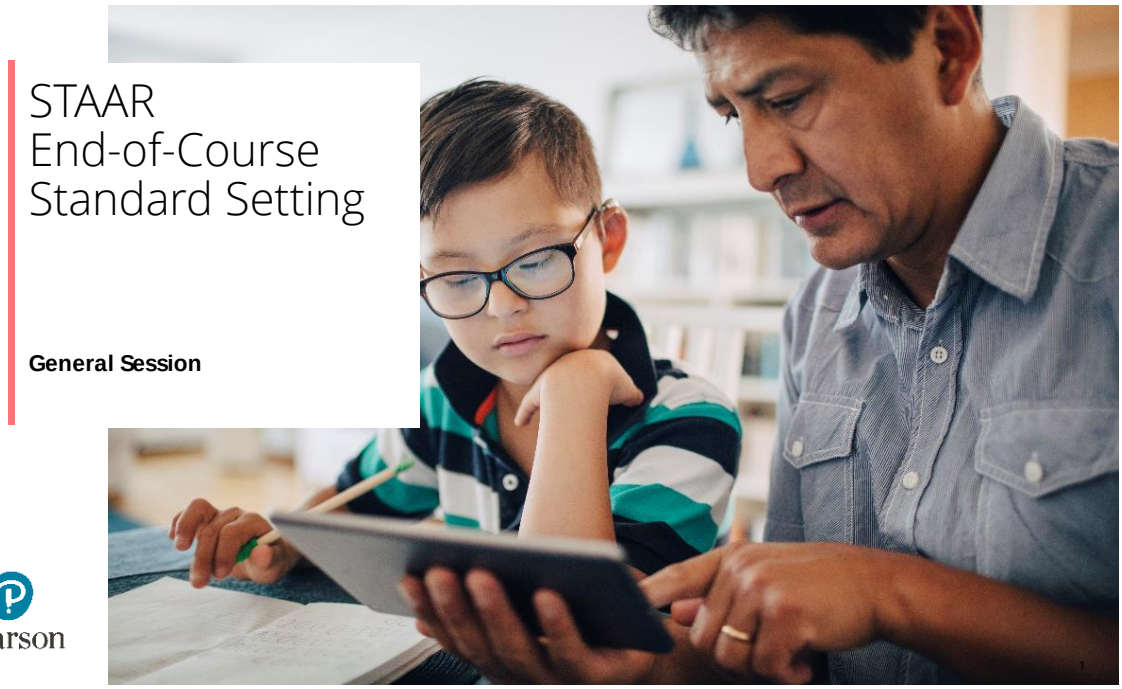
U.S. History:

- Great meeting.
- Overall everything was well, we could have spend more time with the descriptors and looking at student data to help set standards.
- Our facilitators for the US History session were excellent. They kept on task and explained our mission very well. On top of this--they were delightful. Thank you for this opportunity.
- A mouse for the computer would have been helpful.
- Computer mouses would have been useful for navigation.
- Great process! Great facility!
- I think more time could have been allotted for discussion if we had 2 and half days to complete the meeting.
- This was a wonderful experience, very interesting how this process works. Thank you so much for allowing me to participate on his panel, and providing such a wonderful place to stay :)
- Although the equipment and materials were adequate, it would have been nice to have access to a mouse for the computer to make the process easier. Additionally, there were times that I felt that the meeting might have had a little bit more time to accomplish the tasks without being rushed (Granted that there were some technical issues that may have lead to the feeling of being rushed on Day 1) Other than the items mentioned above, all of the facilitators were extremely helpful, supportive, and appreciative of our efforts. That goes along way! Thank you for this opportunity.
- I very much appreciate being chosen for this committee and I have learned so much about the testing process. The facilitator was great!
- great meeting!
- This is a very valuable process and I was pleased to be a part of it.
- Thank you.
- none
- It would have been nice to have a copy of Blooms Taxonomy to work with during the discussions about the PDL's and creating borderline expectations. Thank you for a great experience.

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