

Chapter 12: Redesigned STAAR English EOC Standard Setting

This chapter summarizes the procedures and results of the standard setting conducted in January 2014 for the redesigned STAAR English assessments. It includes the following sections:

- Background and Legislation
- Validity and Linking Studies
- Performance Level Descriptors
- Standard-Setting Committee
- Post-Standard-Setting Activities

Background and Legislation

The 83rd Texas Legislature’s passage of House Bill 5 (HB 5) mandates that, beginning in spring 2014, the STAAR English I and English II¹ should measure reading and writing achievement in a combined assessment that provides a single score. (HB 5, Section 39.023c).

To reflect the intent of this legislation, the Texas Education Agency (TEA) has redesigned the STAAR English I and II assessments to combine reading and writing into a single assessment with a single test score. Each redesigned English assessment must also be administered in a single day. The blueprint for the redesigned STAAR English assessments is provided in Table 12.1. For reference, the base test for each redesigned STAAR English assessment includes a total of 50 multiple-choice questions, two short-answer questions, and one written composition.

Prior to the initial administration of STAAR in spring 2012, separate performance standards were established for the STAAR English reading and writing assessments by separate reading and writing standard-setting committees. Because reading and writing have now been combined into a single assessment and only one score should be reported, new performance standards for STAAR English I, II, and III had to be established.²

The evidence-based standard-setting approach (O’Malley, Keng, & Miles, 2012) was used to establish the performance standards for the redesigned STAAR English assessments. This approach combines considerations regarding policy, the TEKS content standards, knowledge and experience of Texas educators, and information about how student performance on STAAR aligns with performance on related tests and measures. This was the same approach used to set standards for all STAAR assessments, including the original STAAR English reading and

¹ Note that per HB 5, STAAR English III was not to be administered after spring 2013. However, the Texas Education Agency (TEA) is still required to develop an appropriate postsecondary readiness assessment for English III that may be administered at each district’s option beginning in May of the 2015–2016 school year.

² Even though STAAR English III was not administered after spring 2013, performance standards for a redesigned English III assessment were considered in the standard-setting process. This was done to help facilitate alignment of the recommended standards across the STAAR English I, II, and III assessments. It is intended that the preliminary English III standards be reviewed following the first optional administration in spring 2016.

writing assessments in 2012. Please refer to Chapter 2: Overview of the STAAR Standard-Setting Process for a detailed description of the evidence-based standard-setting approach.

Table 12.1: Redesigned STAAR English Blueprint

Test Characteristics	Redesigned STAAR English (Embedded Field Testing Except for Field-test Prompt)
Reporting Category 1: Understanding/Analysis Across Genres	6 multiple choice (MC) 2 short answer questions (SA)
Reporting Category 2: Understanding/Analysis of Literary Texts	11 MC (over 2 literary selections)
Reporting Category 3: Understanding/Analysis of Informational Texts	11 MC (over 2 informational selections)
Selection Breakdown	4 selections (1 pair/2 single selections)
Reading Base-test Section	28 MC (reduction of 1 single selection) 2 SA
Reporting Category 4: Composition	1 essay <i>English I (1 expository)</i> <i>English II (1 persuasive)</i>
Reporting Category 5: Revision	11 MC
Reporting Category 6: Editing	11 MC
Passage Breakdown	4 passages (2 revision/2 editing)
Writing Base-Test Section	22 MC (reduction of 1 writing passage) 1 essay (reduction of 1 prompt)
Total Number of Base-Test Items	50 MC, 2 SA, 1 essay
Time Limit	1 day, 5-hour testing window
Weighting of Each Component	Reading and writing contribute equally to the total score: 30% – reading multiple choice 20% – short answer questions 24% – writing multiple choice 26% – essay

Validity and Linking Studies

This section provides a summary of results of the validity and linking studies conducted for the redesigned STAAR English standard-setting process. The following studies were conducted:

- Bridge Studies
- Linking Studies
- External Validity Studies
- Grade Correlation Studies

Please refer to Chapter 3: Validity and Linking Studies for a complete description of the methods used in the validity and linking studies. Prior to conducting the validity and linking studies, the impact (i.e., the percentage of students expected to meet the Level II performance standard) was estimated for each possible cut score.

ESTIMATING IMPACT DATA

Because standards were set several months before the first operational administration of the redesigned English assessments in spring 2014, data from the spring 2013 administrations of the STAAR English reading and writing assessments were combined to estimate how students might have performed had they taken the redesigned English assessment. This was accomplished by creating “prototype builds” of the redesigned English assessments using select items from the separate reading and writing assessments administered in spring 2013. New combined scores were computed for all students who took both reading and writing in spring 2013. These scores were used throughout the validity and linking studies, and they were eventually used to estimate impact data, which are the percentage of students who would meet a particular performance standard (i.e., Level II or Level III) on the redesigned English assessments.

BRIDGE STUDIES

The bridge studies were designed to empirically link student performance on the original STAAR English assessments (reading and writing) and the redesigned STAAR English assessments. The end results of this study were estimated locations of the original Level II and Level III cut scores on the new STAAR English scales. These results provided information about how the new standards compare to the original standards.

Because separate cut scores were established for the reading and writing components of the STAAR English assessments during the original STAAR standard-setting process, there were two “bridge” cut scores—one for reading and one for writing—estimated for every performance standard (Level II and Level III) on the new STAAR English scale. Bridge cut scores were established by (1) using item parameters on the new and original STAAR English score scales to estimate the linear relationship between those scales and (2) using that relationship to estimate the location of the original performance standard on the new score scale. Table 12.2 shows results of the bridge studies expressed on the new STAAR English I and II score scale.

Table 12.2: Estimated STAAR Reading and Writing Cut Scores on the Redesigned STAAR English Scale

Assessment	Level II: Satisfactory Academic Performance		Level III: Advanced Academic Performance	
	Reading	Writing	Reading	Writing
English I	4072	4196	4692	5013
English II	3983	4199	4608	5001

LINKING STUDIES

The linking studies establish empirical links between STAAR assessments in adjacent grades. In this case, links were estimated between grade 8 reading and English I, between English I and English II, and between English II and English III. In these studies, regression-based linking using logistic regression was employed to address two questions: “For students with a certain score on the preceding assessment, what score do they have a 50% chance of attaining on the

current assessment?” and “What level of performance on the current assessment is associated with a 50% chance of meeting a certain score on the following assessment?” The results of such studies can be used to inform the alignment of performance standards across assessments. For example, it was found that students who met the grade 8 reading Level II standard had a 50% chance of scoring 4110 on English I, and students who met the grade 8 reading Level III standard had a 50% chance of scoring 4432 on English I. Other studies were used to link the lower and upper boundaries of the performance-standard “neighborhoods” (see Chapter 6: Policy Committee) to the score scale of assessments in adjacent grades.

EXTERNAL VALIDITY STUDIES

External validity studies establish empirical links between student performance on the STAAR English assessments and other assessments measuring similar constructs that are administered nationally. Two such studies were conducted. STAAR English II performance was linked to PSAT performance, which provides concurrent validity evidence that the STAAR English II cut scores relate to the College Board’s [PSAT College and Career Readiness Benchmarks](#) for Critical

Table 12.3: Estimated STAAR English Scores Associated with External Assessment Benchmarks

Assessment	Score	Probability	Benchmark
English II	4218	.50	PSAT Critical Reading College and Career Readiness Benchmark for 10 th graders (42)
English II	4458	.75	PSAT Critical Reading College and Career Readiness Benchmark for 10 th graders (42)
English II	4293	.50	PSAT Writing College and Career Readiness Benchmark for 10 th graders (42)
English II	4549	.75	PSAT Writing College and Career Readiness Benchmark for 10 th graders (42)
English II	4700	.75	PSAT Critical Reading College and Career Readiness Benchmark for 11 th graders (45)
English II	4755	.75	PSAT Writing College and Career Readiness Benchmark for 11 th graders (45)
English III	4056	.50	SAT Critical Reading College Readiness Benchmark (500)
English III	4791	.75	SAT Critical Reading College Readiness Benchmark (500)
English III	4284	.50	SAT Writing College Readiness Benchmark (500)
English III	4954	.75	SAT Writing College Readiness Benchmark (500)

Reading and Writing. Similarly, STAAR English III performance was linked to SAT performance, which provides concurrent validity evidence that the STAAR English III cut scores relate to the College Board’s SAT [College Readiness Benchmarks](#) for Critical Reading and Writing. Using the first row of Table 12.3 as an example, external validity study results can be interpreted as follows: “Students who score 4218 on STAAR English II have a 50% chance of meeting the PSAT Critical Reading College and Career Readiness Benchmark for 10th graders.

Performance Level Descriptors

Performance level descriptors (PLDs) are statements that articulate the specific knowledge and skills students typically demonstrate at each performance level of a test given for a specific grade or course. The PLDs developed for STAAR provide a snapshot of students’ academic characteristics based on performance on a given STAAR assessment and reflect the breadth and depth of the content, skills, cognitive demand, and performance requirements evident in the TEKS.

PLDs for STAAR English reading and writing were developed previously, as described in Chapter 5: Performance Level Descriptors. These PLDs were combined by TEA content specialists in advance of standard setting. The PLDs for [English I](#), [English II](#), and [English III](#) are shown in Tables 12.4, 12.5, and 12.6, respectively.

Table 12.4: STAAR English I Performance Level Descriptors

Performance Level Descriptors
When writing essays in response to progressively demanding writing tasks and reading texts of increasing complexity,* students achieving Level III: Advanced Academic Performance can • Write skillfully crafted expository essays with sustained focus, a logical organizing structure, and development that lends substance to the essay • Choose sentences that are purposeful and well controlled • Evaluate how the author’s use of diction and figurative language creates meaning • Make subtle inferences about literary and informational texts, supporting those inferences with specific and well-chosen textual evidence
When writing essays in response to progressively demanding writing tasks and reading texts of increasing complexity,* students achieving Level II: Satisfactory Academic Performance can • Write expository essays that contain a clear thesis statement, use an appropriate organizing structure, sufficiently develop ideas with specific details and examples, choose words that reflect an understanding of the explanatory purpose and demonstrate an adequate command of written conventions • Use a variety of sentence structures • Revise drafts to strengthen the introductory and concluding paragraphs, add information that enhances the supporting details, strengthen transitions within and between paragraphs, improve the effectiveness of sentences, and recognize appropriate style and word choice • Edit drafts to correct errors in grammar, sentence structure, capitalization, punctuation,

- and spelling - Distinguish between the denotative and connotative meanings of words using context, structural analyses, and reference materials - Analyze how the author's use of diction and figurative language supports meaning - Analyze literary texts by recognizing universal themes and the ways in which literary devices contribute to the development of linear and non-linear plots and complex, believable characters - Demonstrate an understanding of informational texts by recognizing the controlling idea or argument, identifying the author's purpose, and summarizing the text by determining which ideas are most important - Recognize the logical connections and thematic links between texts representing similar or different genres - Make reasonable inferences about literary and informational texts, supporting those inferences with accurate, relevant textual evidence
When writing essays in response to progressively demanding writing tasks and reading texts of increasing complexity,* students achieving Level I: Unsatisfactory Academic Performance can - Write basic or limited essays that are only marginally suited to the expository task and minimally developed, with a partial command of conventions - Demonstrate a minimal control of sentence structures - Demonstrate basic skills in revision and editing - Determine the denotative meaning of words using context and reference materials - Demonstrate a basic understanding of literary and informational texts and recognize a summary - Make plausible inferences about literary and informational texts

*The rigor of the expository writing task increases from grade 7 to English I in that the prompt is more demanding in English I, specifically with regard to the cognitive complexity of the stimulus (the synopsis or quotation students use in developing the essay) and the sophistication of the topic. In addition, the text complexity of the reading selections increases from grade 8 to English I. 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 structures 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.

Table 12.5: STAAR English II Performance Level Descriptors

Performance Level Descriptors
When writing essays in response to progressively demanding writing tasks and reading texts of increasing complexity,* students achieving Level III: Advanced Academic Performance can - Write persuasive essays that maintain a convincing position and sustain focus with a skillful organizing structure, compelling evidence and support, purposeful and precise word choice, and an understanding and control of rhetorical techniques that enhance effectiveness - Choose sentences that are purposeful and well controlled - Evaluate how the author's use of syntax, diction, and sensory language creates voice, tone, and meaning - Make discerning inferences about literary and informational texts, supporting those inferences with specific and well-chosen textual evidence

When writing essays in response to progressively demanding writing tasks and reading texts of increasing complexity,* students achieving Level II: Satisfactory Academic Performance can

- Write persuasive essays that contain a clear position, use a logical organizing structure, sufficiently develop relevant reasons and evidence, create an appropriate tone through clear and specific word choice, and demonstrate an adequate command of written conventions
- Use a variety of sentence structures
- Revise drafts to strengthen the introductory and concluding paragraphs, add information that enhances the supporting details, strengthen transitions within and between paragraphs, improve the effectiveness of sentences, and recognize appropriate style and word choice
- Edit drafts to correct grammar, sentence structure, capitalization, punctuation, and spelling
- Distinguish between the denotative and connotative meanings of words using context, structural analyses, and reference materials
- Analyze how the author's use of syntax, diction, and sensory language supports meaning
- Analyze literary texts by recognizing universal themes and the ways in which literary devices contribute to the development of linear and non-linear plots and complex, believable characters
- Demonstrate an understanding of informational texts by analyzing the controlling idea or argument, determining the author's purpose, identifying organizational patterns, and distinguishing between a summary and a critique of the text
- Identify the implicit connections and thematic links between texts representing similar or different genres
- Make logical inferences about literary and informational texts, supporting those inferences with accurate, relevant textual evidence

When writing essays in response to progressively demanding writing tasks and reading texts of increasing complexity,* students achieving Level I: Unsatisfactory Academic Performance can

- Write basic or limited essays that are only marginally suited to the persuasive task and minimally developed, with a partial command of conventions
- Demonstrate a minimal control of sentence structure
- Demonstrate basic skills in revision and editing
- Determine the denotative meaning of words using context, structural analyses, and reference materials
- Demonstrate a basic understanding of literary and informational texts and identify universal themes and controlling ideas
- Make plausible inferences about literary and informational texts

* The rigor of the writing task increases from English I to English II in that the prompt is more demanding in English II, specifically with regard to the cognitive complexity of the stimulus (the synopsis or quotation students use in developing the essay) and the sophistication of the topic. Persuasive writing, which is assessed on STAAR for the first time, also increases the rigor of English II writing. The persuasive task requires students to take a position on a specific issue and to develop an argument that not only supports this position but also convinces the reader of its merit. In addition, the text complexity of the reading selections increases from English I to English II. 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 structures 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.

Table 12.6: STAAR English III Performance Level Descriptors

Performance Level Descriptors
When writing essays in response to progressively demanding writing tasks and reading texts of increasing complexity,* students achieving Level III: Advanced Academic Performance can • Write analytical essays that demonstrate a thorough understanding of both the text and the analytical writing task by establishing a cogent thesis statement, providing an insightful interpretation of text, and smoothly integrating well-chosen textual evidence • Choose sentences that are purposeful and well controlled • Evaluate how an author’s use of language advances purpose, creates tone, evokes emotion, and shapes meaning • Make perceptive inferences about literary and informational texts, supporting those inferences with specific and well-chosen textual evidence
When writing essays in response to progressively demanding writing tasks and reading texts of increasing complexity,* students achieving Level II: Satisfactory Academic Performance can • Write analytical essays that contain a clear thesis statement, provide a reasonable interpretation of a text through the use of appropriate language, sufficiently support this interpretation with relevant textual evidence, and demonstrate an adequate command of written conventions • Use a variety of sentence structures • Revise drafts to clarify the thesis, strengthen supporting ideas, use appropriate transitions within and between paragraphs, improve the effectiveness of sentences, and recognize that word choice clarifies meaning and creates an appropriate tone • Edit drafts to correct grammar, sentence structure, capitalization, punctuation, and spelling • Determine the meaning of technical or nuanced words using context, structural analyses, and reference materials • Analyze how the author’s use of language, including common rhetorical techniques and literary elements and devices, supports meaning • Analyze literary texts by examining universal themes and the ways in which literary elements and devices support meaning • Demonstrate an understanding of informational texts by determining the implicit main idea or argument, summarizing accurately, recognizing organizational patterns, and making inferences about the author’s purpose and viewpoint • Analyze the underlying connections and thematic links between texts representing similar or different genres • Make complex inferences about literary and informational texts, supporting those inferences with accurate, relevant textual evidence
When writing essays in response to progressively demanding writing tasks and reading texts of increasing complexity,* students achieving Level I: Unsatisfactory Academic Performance can • Write basic or limited analytical essays that are based on a literal or obvious interpretation of the text and a minimal understanding of the analytical writing task • Demonstrate a minimal control of sentence structures • Demonstrate basic skills in revision and editing • Determine the meaning of unfamiliar words using context, structural analyses, and reference materials • Demonstrate a basic understanding of literary and informational texts and identify

- | |
|--|
| <ul style="list-style-type: none">• common rhetorical techniques, literary elements and devices, and universal themes• Make plausible inferences about literary and informational texts |
|--|

* The rigor of the writing task increases from English II to English III in that the prompt is more demanding in English III, specifically with regard to the cognitive complexity of the stimulus and the sophistication needed to write an analysis. The analytical task increases the rigor of English III writing because students must be able to interpret one aspect of a literary or informational text approximately 350–450 words in length, support this interpretation with relevant textual evidence, and write a rhetorically effective expository essay. In addition, the text complexity of the reading selections increases from English II to English III. 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 structures 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.

Standard-Setting Committee

COMMITTEE COMPOSITION

On January 15–17, 2014, a standard-setting committee consisting of Texas educators was convened in Austin to recommend performance standards for the redesigned STAAR English I, II and III assessments. The committee was comprised of a subset of the panelists that established the original standards in 2012 and included both K–12 and higher-education representatives. The demographics and educational experiences of the committee are summarized in Table 12.7.

COMMITTEE MEETING PROCEEDINGS

The three-day standard-setting meeting included sessions in which panelists participated in the following activities to set recommended cut points:

- 1) reviewed the test questions, reading selections, writing passages, and writing prompts;
- 2) became familiar with the proficiency level descriptors (PLDs) and borderline student descriptors for the assessments; and
- 3) applied an item-mapping procedure to make cut-score recommendations (Lewis, Mitzel, Green, & Patz, 1999³).

Prior to the standard-setting meeting, reasonable ranges (“neighborhoods”) for the performance standards were determined by considering student performance data, results from empirical validity and linking studies conducted on the STAAR English assessments, the original reasonable ranges recommended by the policy committee in 2012 (see Chapter 6: Policy Committee), the content of the questions falling within each range, and the estimated percentage of total points students would need to reach each performance category. The reasonable ranges provided panelists with a target area in which to make their judgments; however, panelists were not required to keep their cut-score recommendations within these ranges.

³ Lewis, D. M., Mitzel, H. C., Green, D. R., & Patz, R. J. (1999). *The bookmark standard setting procedure*. Monterey, CA: McGraw-Hill.

Table 12.7: Demographics and Educational Experiences of the STAAR English Standard-Setting Committee.

Current Position and Years of Experience in Education

		Years of Professional Experience in Education					
		1–5 years	6–10 years	11–15 years	16–20 years	More Than 20 years	Total
Current Position	Administrator	0	0	1	2	2	5
	Higher Education	0	0	0	1	1	2
	Teacher	0	2	0	2	4	8
	Other	0	1	0	0	0	1
	Total	0	3	1	5	7	16

Gender Distribution

Gender	N-Count
Female	13
Male	3

Ethnicity Distribution

Ethnicity	N-Count
African American	2
Asian or Pacific Islander	0
Hispanic	3
Multi-racial	2
Native American	1
White	8

Experience with Student Populations

Student Population	N-Count
General Education	16
Special Education	14
English Language Learners	14
Low Socioeconomic Status	14

District Type

Type	N-Count
Metro	2
Suburban	7
Rural	5
Did Not Respond	2

District Size

Type	N-Count
Large	7
Medium	5
Small	2
Did Not Respond	2

District Socioeconomic Status

Type	N-Count
High	1
Moderate	8
Low	5
Did Not Respond	2

The standard-setting meeting was conducted using the process described in Chapter 7: Standard-Setting Committees, except that descriptions of “borderline students” were provided rather than developed by the standard-setting committee. Like the PLDs, TEA content experts created the borderline student descriptors by combining the borderline student descriptors generated by panelists during the separate reading and writing standard-setting meetings conducted in 2012.

During three rounds of the item-mapping procedure, panelists reviewed the content assessed by the test questions, engaged in table and whole-group discussions, and considered the potential impact on the distribution of students’ performance categories. With these things in mind, the committee members recommended performance standards (or cut scores) for each STAAR English assessment to establish the following performance categories:

- Level III: Advanced Academic Performance
- Level II: Satisfactory Academic Performance
- Level I: Unsatisfactory Academic Performance

The policy definitions for each of the performance categories are provided in Chapter 4: Performance Labels and Policy Definitions.

The final activity in which the standard-setting panelists participated was cross-course articulation. The purpose of the cross-course articulation was to look at the cut-score recommendations (presented as page numbers in the ordered item booklet; OIB) that were made across all STAAR English assessments to evaluate the reasonableness of these cut scores (see Chapter 7: Standard-Setting Committees for a description of the process for developing OIBs). Panelists were shown the impact data resulting from their Round 3 cut-score recommendations across STAAR English I, II and III. Recommendations for cut-score adjustments could be made by the committee as a group after reviewing the Round 3 feedback and group discussion. Any recommended changes made during the cross-course articulation had to be supported by a review of the OIB and the PLDs for that assessment.

COMMITTEE MEETING RESULTS

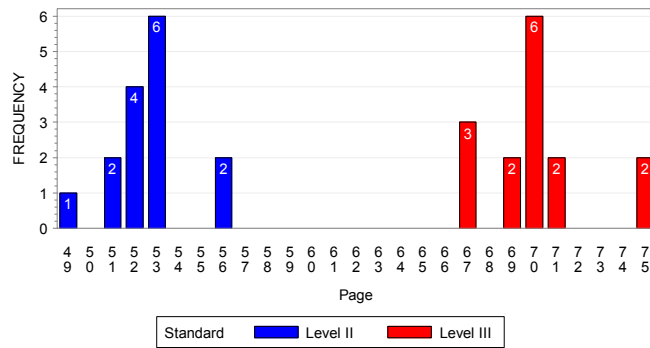
This section includes several sets of tables and figures that describe the results of the STAAR English standard-setting committee meetings. Similar tables and figures for other assessments are located in Appendices 11, 12, 13, 14, and 15.

Figures 12.1, 12.2, and 12.3 show the frequency distributions of the recommended cuts (i.e., bookmarked page numbers) after each round of judgment. The spread of the judgments provides a sense of overall agreement among the committee members. Table 12.8 summarizes the committee members’ judgments by providing the median page numbers at each stage of standard setting. The median page numbers represent the committee’s cut-score recommendations after each round of judgment. Additional descriptive statistics for the distributions of judgments are shown in Table 12.9. This includes the minimum, maximum, mean, standard deviation, and median of the standard-setting panelists’ cut score

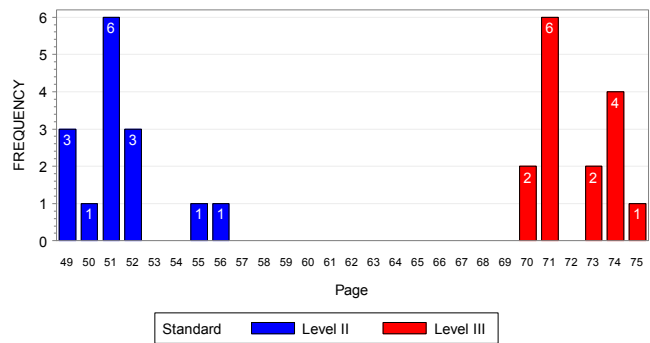
recommendations (based on the OIB page number) during each judgment round of the committee meetings.

Figures 12.4, 12.5, and 12.6 show the estimated impact data (i.e., the percentage of students at each performance level) based on the cut-score recommendations after Round 3 of the standard-setting committee meetings, cross-course articulation, and the reasonableness review. As described earlier, the impact data were computed using student performance on the prototype builds of the redesigned STAAR English assessments (i.e., using spring 2013 performance on a selection of STAAR reading and writing items). Note that the available impact data for English III were based on “accelerated” students who took STAAR English III before their junior year of high school. Thus, the estimated impact for English III is likely different than it would be if the assessment was administered to the entire student population taking English III.

Round 1 Panelist Agreement Data STAAR English I



Round 2 Panelist Agreement Data STAAR English I



Round 3 Panelist Agreement Data STAAR English I

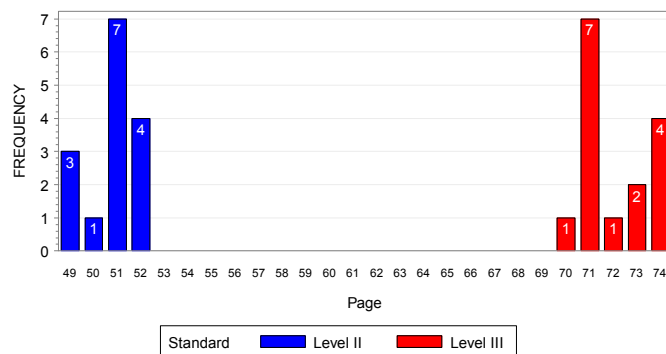
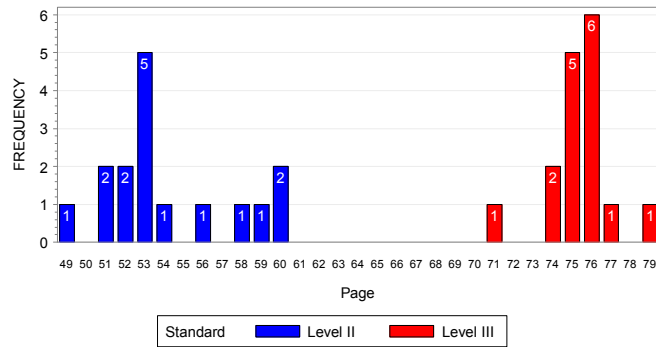
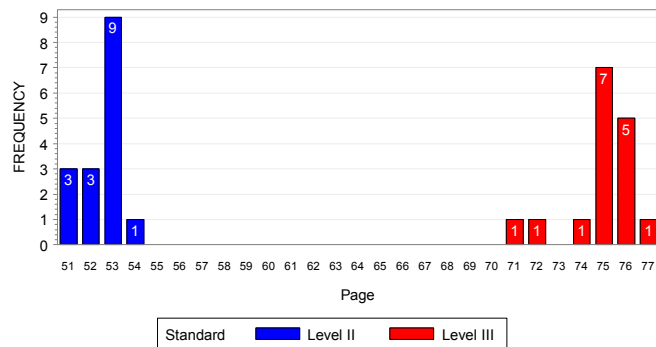


Figure 12.1. Standard-Setting Panelists' Agreement Data for English I.

Round 1 Panelist Agreement Data STAAR English II



Round 2 Panelist Agreement Data STAAR English II



Round 3 Panelist Agreement Data STAAR English II

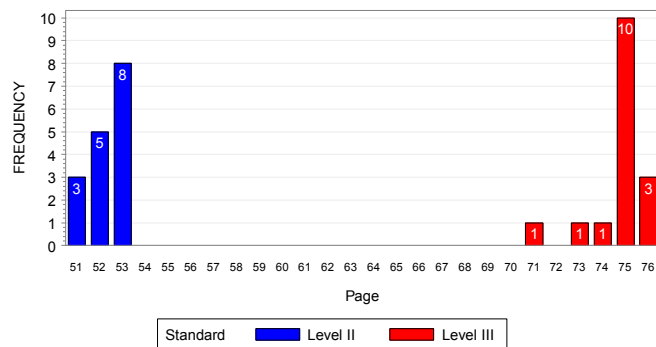
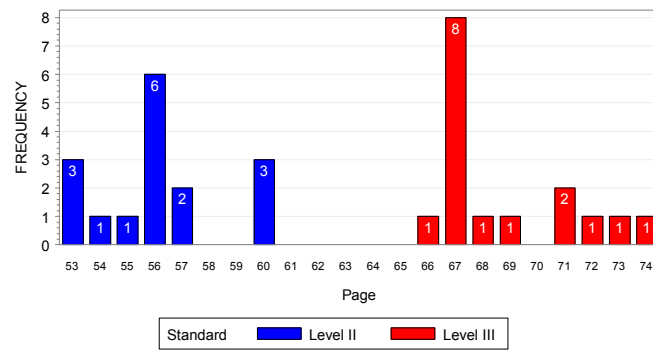
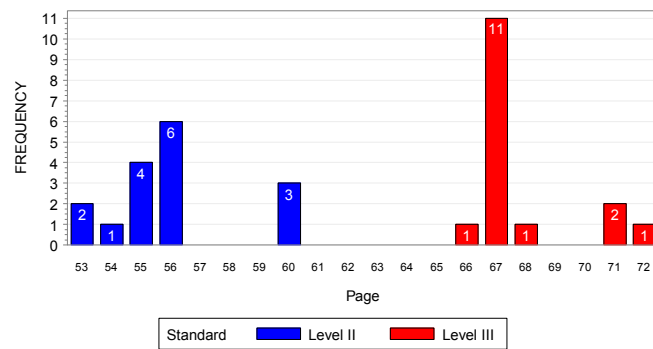


Figure 12.2. Standard-Setting Panelists' Agreement Data for English II.

Round 1 Panelist Agreement Data STAAR English III



Round 2 Panelist Agreement Data STAAR English III



Round 3 Panelist Agreement Data STAAR English III

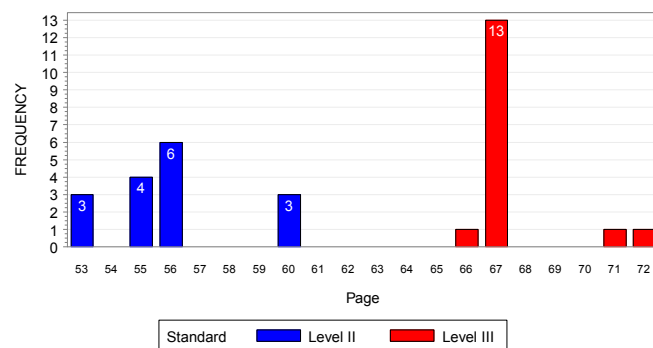


Figure 12.3. Standard-Setting Panelists' Agreement Data for English III.

Table 12.8: Summary of Cut-Score Recommendations for STAAR English I, II, and III.

	English I		English II		English III	
	Level II	Level III	Level II	Level III	Level II	Level III
Round 1	53	70	53	76	56	67
Round 2	51	71	53	75	56	67
Round 3	51	71	53	75	56	67
Articulation	50	71	53	75	56	67
Reasonableness Review	50	71	53	75	56	67

Table 12.9: Summary of Standard-Setting Panelists' Judgments for STAAR English I, II, and III.

		Round	Minimum	Maximum	Mean	Standard Deviation	Median
English I	Level II	1	49	56	52.6	1.8	53
		2	49	56	51.3	2.0	51
		3	49	52	50.8	1.1	51
	Level III	1	67	75	70.1	2.4	70
		2	70	75	72.2	1.7	71
		3	70	74	72.1	1.4	71
English II	Level II	1	49	60	54.2	3.4	53
		2	51	54	52.5	0.9	53
		3	51	53	52.3	0.8	53
	Level III	1	71	79	75.4	1.7	76
		2	71	77	74.9	1.5	75
		3	71	76	74.8	1.2	75
English III	Level II	1	53	60	56.1	2.3	56
		2	53	60	56.0	2.2	56
		3	53	60	55.9	2.3	56
	Level III	1	66	74	68.8	2.6	67
		2	66	72	67.8	1.8	67
		3	66	72	67.5	1.6	67

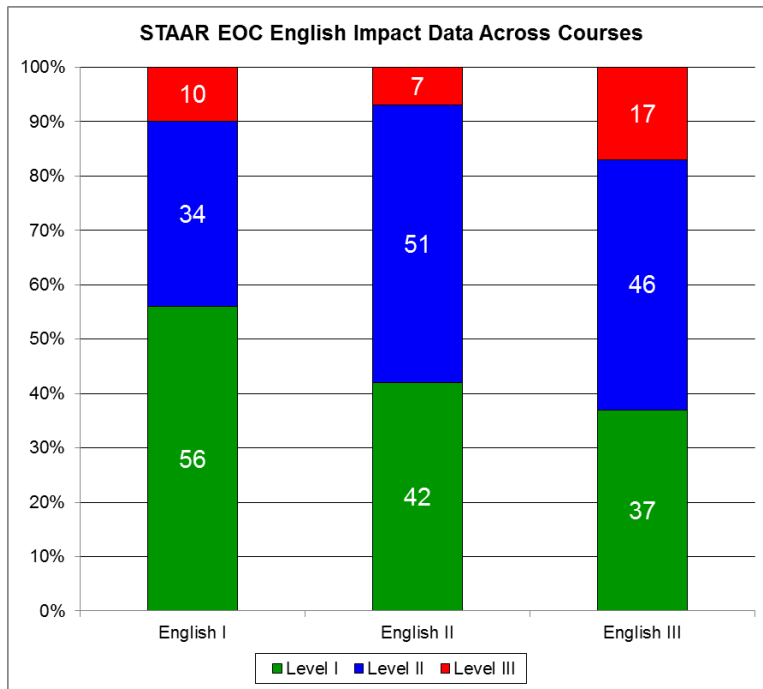


Figure 12.4. Estimated Impact Across Courses after Round 3.

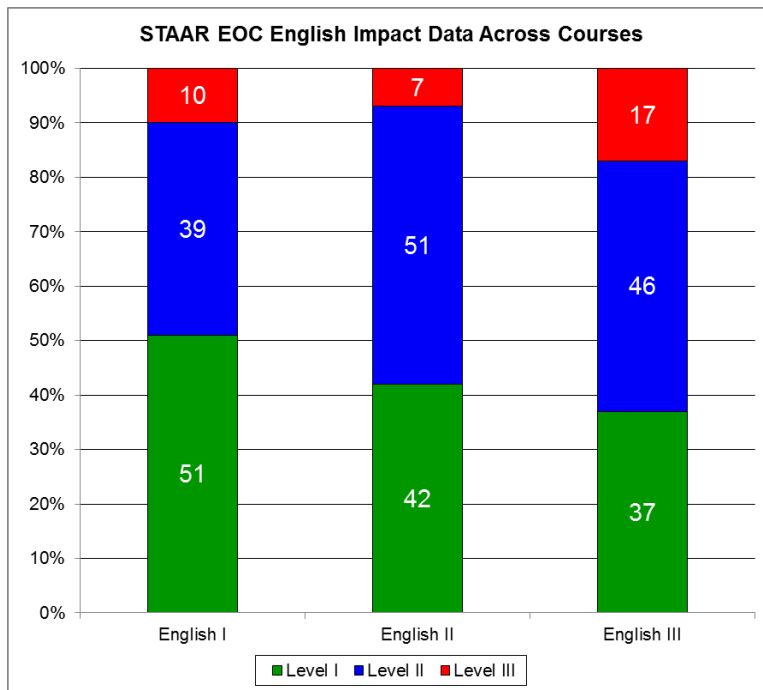


Figure 12.5. Estimated Impact Across Courses after Articulation and Reasonableness Review.

STANDARD-SETTING PROCESS EVALUATION

At the end of the standard-setting meeting, panelists were asked to complete a process-evaluation survey. The purpose of the survey was to collect information about each panelist's experience in recommending cut scores for the STAAR English assessments. Panelists' responses to the evaluation form are summarized in Table 12.10.

Table 12.10: Standard-Setting Process Evaluation Summary Results for STAAR English I, II, and III (a total of 15 panelists responded to the process evaluation survey).

Section 1: Success of the Meeting Components

Meeting Component	Not Successful	Partially Successful	Successful	Very Successful	Omit
Introduction to the process of setting performance standards	0%	0%	13%	87%	0%
Discussion of the performance labels and the definitions	0%	0%	20%	80%	0%
Taking the actual assessment(s)	0%	0%	13%	87%	0%
Overview of the item mapping procedure	0%	0%	27%	73%	0%
Practice exercise for the item-mapping procedure	0%	13%	13%	73%	0%
Feedback data provided in each round	0%	0%	7%	93%	0%
Discussion after each round	0%	7%	20%	73%	0%
Articulation	0%	0%	33%	67%	0%

Section 2: Usefulness of Activities and Information

Activity or Information	Not Useful	Somewhat Useful	Useful	Very Useful	Omit
Specific Performance Level Descriptors (PLDs)	0%	0%	27%	73%	0%
Training in the bookmark standard setting method	0%	7%	27%	67%	0%
Feedback data provided after Round 1	0%	0%	7%	93%	0%
Feedback data provided after Round 2	0%	0%	7%	93%	0%
Presentation of data across courses	0%	0%	13%	87%	0%

Section 3: Adequacy of Meeting Elements

Meeting Element	Not Adequate	Somewhat Adequate	Adequate	More Than Adequate	Omit
Training provided	0%	0%	7%	93%	0%
Amount of time spent training	0%	0%	2%	80%	0%
Feedback provided between rounds	0%	0%	2%	80%	0%
Facilities used for the session	0%	0%	7%	93%	0%
Total amount of time in breakout groups to make judgments	0%	0%	7%	93%	0%
Number of rounds for the judgments	0%	0%	13%	87%	0%

Section 4: Specific PLDs

Performance Category	Not Confident	Somewhat Confident	Confident	Very Confident	Omit
Level I: Unsatisfactory Academic Performance	0%	7%	40%	47%	7%
Level II: Satisfactory Academic Performance	0%	7%	53%	40%	0%
Level III: Advanced Academic Performance	0%	13%	40%	47%	0%

Section 5: Cut-Score Recommendations

Cut Score	Not Confident	Somewhat Confident	Confident	Very Confident	Omit
Level II: Satisfactory Academic Performance	0%	7%	73%	20%	0%
Level III: Advanced Academic Performance	0%	13%	60%	27%	0%

Section 6: Opportunities to Express Opinions

Category	Not Adequate	Somewhat Adequate	Adequate	More Than Adequate	Omit
Express your opinions about student performance levels	0%	0%	7%	93%	0%
Ask questions about the standards and how they will be used	0%	0%	7%	93%	0%
Ask questions about the process of making cut score recommendations	0%	0%	7%	93%	0%
Interact with your fellow panelists	0%	0%	7%	93%	0%

Section 7: Respect

Party	No	Sometimes	Yes	Omit
Fellow panelists	0%	0%	100%	0%
Facilitators	0%	0%	100%	0%

In summary, most committee members thought that the various components of the meeting were “successful” or “very successful.” The majority of panelists thought that the activities conducted during the meeting were either “useful” or “very useful.” In general, they reported that the time spent on training, table discussions, and judgment tasks was “more than adequate.” Regarding the specific PLDs provided by TEA, the panelists reported being “confident” or “very confident.” When asked about their confidence in the cut scores, most panelists felt “confident.” Virtually all committee members thought that they were given adequate opportunity to express their opinions, ask questions, and interact with other committee members. Additionally, the majority of panelists indicated that they believed that their opinions and judgments were respected by others.

Post-Standard-Setting Activities

REASONABLENESS REVIEW

Following the standard-setting meeting, TEA conducted a reasonableness review of the STAAR English I and II cut-score recommendations. The cut scores were reviewed in comparison to other STAAR EOC content areas and to STAAR 3–8 to determine the reasonableness of the system of standards recommended. No adjustments were made to the panel-recommended performance standards during the reasonableness review. For STAAR English III, the cut scores recommended by the committee were considered preliminary. A review of the English III cut score recommendations is currently planned for after the first optional administration of STAAR English III in spring 2016, when the impact data can be estimated using a broader sample of students.

A visual representation of the estimated impact data (percentage of students that would have been classified under each performance category with the cut scores) based on student performance on the spring 2013 administration of the STAAR English assessments is provided in Figure 12.5. As noted previously, students took separate reading and writing tests for each STAAR English assessment on separate days in spring 2013. Student performance data were combined to estimate impact data for a combined assessment.

SCALE-SCORE SYSTEM AND PHASE-IN RECOMMENDATIONS

Once the final recommended cut scores for Level II and Level III were established, the scale-score systems and phase-in cuts (for Level II) were established for the STAAR English I and II assessments. The English III scale-score system will not be established until the assessment is

administered in spring 2016 and the recommended cut scores are reviewed. For each STAAR English scale, a score value of 4000 was used to designate the Recommended Level II cut score, and the standard deviation (SD) of the scale was set to a value of 500. This was done to match the scale-score systems for the STAAR EOC assessments in the other content areas.

Initially, Level II Phase-in 1 and 2 cut scores were set at 0.5 SD and 0.2 SDs below the Recommended Level II respectively (corresponding to scale scores of 3750 and 3900, respectively). Other SD options were considered for the phase-in cut scores (e.g., 1.0 and 0.5 SDs, which is the phase-in used for the other EOC assessments). However, using 0.5 and 0.2 SDs matched the phase-in rule for the original STAAR English assessments and therefore provides a more consistent transition between the original and the redesigned English assessments.

This decision in spring 2014 was superseded in summer 2014 by the decision to transition the STAAR program to a three-step phase-in process for Level II. For STAAR English assessments, the Level II phase-in cuts correspond to 0.5, 0.35, and 0.15 standard deviations below the Recommended Level II standard. The resulting cut scores for the STAAR English I and II assessments are summarized in Table 12.11.

Table 12.11: Recommended and Phase-in STAAR English Cut Scores

STAAR Assessment	Level II Phase-in 1	Level II Phase-in 2	Level II Phase-in 3	Level II Recommended	Level III
English I	3750	3825	3925	4000	4691
English II	3750	3825	3925	4000	4831

Approval of Cut Scores

On February 25, 2014, the Texas commissioner of education approved the phase-in 1 and recommended cut scores for the STAAR English I and English II assessments (provided in Table 12.11) for use with the spring 2014 administration and future administrations of these two assessments.