

## Chapter 13: 2015 STAAR 3–8 Mathematics Standard Setting

This chapter summarizes the procedures and results of the standard setting conducted in July 2015 for the redesigned STAAR grades 3–8 mathematics assessments. It includes the following sections:

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

### Background

In April 2012, the State Board of Education (SBOE) revised the Texas Essential Knowledge and Skills (TEKS) for mathematics. The SBOE specified that districts implement the new curriculum standards in K–8 classrooms in the 2014–2015 school year. Districts will implement the revised high school curriculum standards in 2015–2016.

The Student Assessment Division of the Texas Education Agency (TEA), in coordination with the Curriculum Division, convened advisory committees in October 2013 to make recommendations about the implementation of the new TEKS in the STAAR mathematics assessments.

To reflect changes in the TEKS, TEA has redesigned the STAAR grades 3–8 mathematics assessments. Information about the redesigned STAAR grades 3–8 mathematics assessments (e.g., test blueprints) can be found on TEA’s website at: <http://tea.texas.gov/student.assessment/staar/math/>.

Content analyses indicated that the degree of content overlap between the new and old STAAR grades 3–8 mathematics assessments was not extensive enough to keep the performance standards that were set after the initial administration in 2012. As such, new performance standards for the redesigned STAAR grades 3–8 mathematics had to be established.

The evidence-based standard-setting approach (O’Malley, Keng, & Miles, 2012) was used to establish the performance standards for the redesigned STAAR grades 3–8 mathematics assessments. This approach combines considerations regarding policy, the TEKS curriculum 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 grades 3–8 mathematics assessments, in 2012. Refer to Chapter 2: Overview of the STAAR Standard-Setting Process for a detailed description of the evidence-based standard-setting approach.

## Validity and Linking Studies

This section provides a summary of results of the validity and linking studies conducted for the redesigned STAAR grades 3–8 mathematics standard-setting process. The following studies were conducted:

- Bridge Studies
- Linking Studies
- External Validity Studies
- Vertical Scale Study

Information that related scores on the STAAR assessments with outside assessments was presented in two ways: in relation to the borderline student (the student whose academic performance is barely indicative of a given performance level) and in relation to the typical student (the student whose academic performance is solidly representative of a given performance level). The borderline student and typical student were defined by where the cut scores fell after each judgment round. Refer to Chapter 3: Validity and Linking Studies for a complete description of the methods used in the validity and linking studies, and Chapter 7: Standard-Setting Committees for a description of how information was presented to the panelists.

### BRIDGE STUDIES

The bridge studies were designed to empirically link student performance on the original STAAR grades 3–8 mathematics assessments and the redesigned STAAR grades 3–8 mathematics assessments. The study produced estimated locations of the original Level II (phase-in 1 and final) and Level III cut scores on the new STAAR grades 3–8 mathematics scales. These results provided information about how the new performance standards compare to the original standards.

The STAAR mathematics bridge study consisted of three stages:

1. **Content overlap analysis:** This analysis was used to determine that there was shared content between the two versions of the assessments so that the prior STAAR mathematics performance standards could be mapped onto the new STAAR mathematics assessments.
2. **Empirical analysis:** The second stage of analysis statistically mapped the previous STAAR performance standards to the new assessments using student performance data. This was possible through the use of common items appearing on both the prior version and the new version of the STAAR mathematics assessments. Common items were field-tested as part of the previous mathematics assessment and were included as part of the new mathematics assessment.
3. **Impact data analysis:** The third stage involved evaluating the percent of students attaining each performance standard on the previous STAAR mathematics assessments in relation to student performance on the new STAAR mathematics assessments

(referred to as impact data). The impact data analysis provided supplemental information to support the results from the empirical analyses.

More information on the bridge studies for the STAAR grades 3–8 mathematics assessments can be found on TEA’s website at:

<http://tea.texas.gov/student.assessment/reports/>

**LINKING STUDIES**

The linking studies establish empirical links between STAAR assessments in adjacent grades in the same content area. In this case, links were estimated between consecutive grade levels (e.g., grade 4 to grade 5) and from grade 8 mathematics to Algebra I. In these studies, regression-based linking using logistic regression was employed. Logistic regression analyses provided the probability of attaining a particular score on a subsequent test within a content area given a student’s performance on a STAAR assessment. The results of such studies can be used to inform the alignment of performance standards across assessments. For example, it was found that borderline students who scored at the grade 8 mathematics Level II standard had a 46% chance of meeting the Level II standard on Algebra 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 grade 8 mathematics assessment and other assessments measuring similar constructs that are administered nationally. Two such studies were conducted. STAAR grade 8 mathematics performance was linked to RediStep and EXPLORE performance. The link to RediStep provides concurrent validity evidence that the STAAR grade 8 mathematics cut score relates to the College Board’s College and Career Readiness Benchmark for 8th Grade.

**Table 13.2: Estimated STAAR Grade 8 Mathematics Standards Associated with External Assessment Benchmarks**

Assessment	Standard	Probability of Borderline Student	Probability of Typical Student	Benchmark
Grade 8	Level II: Satisfactory	0.53	0.76	RediStep College and Career Readiness Benchmark for 8th Grade (3.7)
	Level III: Advanced	0.93	0.97	
	Level II: Satisfactory	0.32	0.58	ACT EXPLORE College Readiness Benchmark (19)
	Level III: Advanced	0.85	0.92	

Similarly, STAAR grade 8 mathematics performance was linked to ACT EXPLORE performance, which provides concurrent validity evidence that the STAAR grade 8 mathematics cut scores relate to the ACT EXPLORE College Readiness Benchmarks for College Algebra. Using the first

row of Table 13.2 as an example, external validity study results can be interpreted as follows: “Borderline students who score at the Level II: Satisfactory standard on STAAR grade 8 mathematics will have a 53% chance of meeting the RediStep College and Career Readiness Benchmark for 8th Graders.”

### **VERTICAL SCALE STUDY**

Under Texas Education Code (TEC) §39.036, the TEA is required to develop a vertical scale for assessing student performance in grades 3–8 mathematics. Because the vertical scale for mathematics empirically links student performance on STAAR grades 3–8 mathematics assessments, once the grade 8 neighborhood was established, the neighborhoods for STAAR mathematics for grades 3–7 were informed using the alignment of the vertical scale across grades. Vertical scale alignment was also provided to panelists as feedback between rounds of judgements. Refer to Chapter 6: Policy Committee and Neighborhood Development and Chapter 7: Standard-Setting Committees for more detail about neighborhoods and the panelist feedback provided to the standard-setting committees.

More information on the new mathematics vertical scale is available in the June 2015

**Mathematics Vertical Scale Technical Report** located on TEA’s website at:

<http://tea.texas.gov/student.assessment/reports/>

### **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 grades 3–8 mathematics were developed previously, as described in Chapter 5: Performance Level Descriptors. These PLDs were revised by TEA content specialists in advance of standard setting to account for the changes in curriculum standards. During the STAAR grades 3–8 mathematics standard-setting meetings, the panelists reviewed and made recommendations for revisions to the new PLDs. After the meetings, TEA finalized the STAAR grades 3–8 mathematics PLDs, which are located on TEA’s website at:

<http://tea.texas.gov/student.assessment/staar/math/>.

### **Standard-Setting Committees**

#### **COMMITTEE COMPOSITION**

On July 14–16 and July 21–23, 2015, standard-setting committees consisting of Texas educators were convened in Austin to recommend performance standards for the redesigned STAAR grades 3–8 mathematics assessments. The demographics and educational experiences of the committees are summarized in Tables 13.3 and 13.4.

## COMMITTEE MEETING PROCEEDINGS

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

- 1) reviewed the test questions;
- 2) became familiar with the proficiency level descriptors (PLDs) and recommended revisions to the PLDs;
- 3) created borderline student descriptors for the assessments; and
- 4) applied an item-mapping procedure to make cut-score recommendations (Lewis, Mitzel, Green, & Patz, 1999<sup>1</sup>).

Prior to the standard-setting meetings, reasonable ranges (“neighborhoods”) for the performance standards were determined by considering student performance data; results from empirical validity, linking, and vertical scale studies conducted on the STAAR grades 3–8 mathematics assessments; the content of the questions falling within each range; and the percentage of questions students would need to get correct 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.

---

<sup>1</sup> Lewis, D. M., Mitzel, H. C., Green, D. R., & Patz, R. J. (1999). *The bookmark standard setting procedure*. Monterey, CA: McGraw-Hill.

**Table 13.3: Demographics and Educational Experiences of the STAAR Grades 6–8 Mathematics 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
<b>Current</b>	<b>Administrator</b>	0	0	2	0	0	2
	<b>Teacher</b>	0	2	3	4	5	14
	<b>Other</b>	0	1	1	0	2	4
	<b>Total</b>	<b>0</b>	<b>3</b>	<b>6</b>	<b>4</b>	<b>7</b>	<b>20</b>

**Gender Distribution**

Gender	N-Count
Female	11
Male	9

**Race Distribution**

Race	N-Count
American Indian or Alaskan Native	0
Asian	0
Black or African American	2
Native Hawaiian or Pacific Islander	0
White	17
Did Not Respond	1

**Experience with Student Populations**

Student Population	N-Count
General Education	17
Special Education	18
English Language Learners	18
Low Socioeconomic Status	17

**Ethnicity Distribution**

Ethnicity	N-Count
Hispanic or Latino	5
Not Hispanic or Latino	12
Did Not Respond	3

**District Type**

Type	N-Count
Metro	1
Suburban	8
Rural	10
Did Not Respond	1

**District Size**

Type	N-Count
Large	4
Medium	10
Small	5
Did Not Respond	1

**District Socioeconomic Status**

Type	N-Count
High	0
Moderate	8
Low	11
Did Not Respond	1

**Table 13.4: Demographics and Educational Experiences of the STAAR Grades 3–5 Mathematics 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
<b>Current</b>	<b>Administrator</b>	0	0	0	1	0	1
	<b>Teacher</b>	1	3	3	4	3	14
	<b>Other</b>	0	0	0	1	3	4
	<b>Total</b>	<b>1</b>	<b>3</b>	<b>3</b>	<b>6</b>	<b>6</b>	<b>19</b>

**Gender Distribution**

Gender	N-Count
Female	17
Male	2

**Race Distribution**

Race	N-Count
American Indian or Alaskan Native	0
Asian	0
Black or African American	2
Native Hawaiian or Pacific Islander	0
White	17

**Experience with Student Populations**

Student Population	N-Count
General Education	19
Special Education	18
English Language Learners	19
Low Socioeconomic Status	19

**Ethnicity Distribution**

Ethnicity	N-Count
Hispanic or Latino	10
Not Hispanic or Latino	9

**District Type**

Type	N-Count
Metro	1
Suburban	9
Rural	8
Did Not Respond	1

**District Size**

Type	N-Count
Large	4
Medium	10
Small	4
Did Not Respond	1

**District Socioeconomic Status**

Type	N-Count
High	2
Moderate	6
Low	10
Did Not Respond	1

The standard-setting meetings were conducted using the process described in Chapter 7: Standard-Setting Committees, except that the standard-setting committee panelists had the opportunity to make recommendations on revisions to the PLDs before creating the borderline student descriptors.

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 within performance categories. With these things in mind, the committee members recommended performance standards (or cut scores) for each STAAR grade 3–8 mathematics 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.

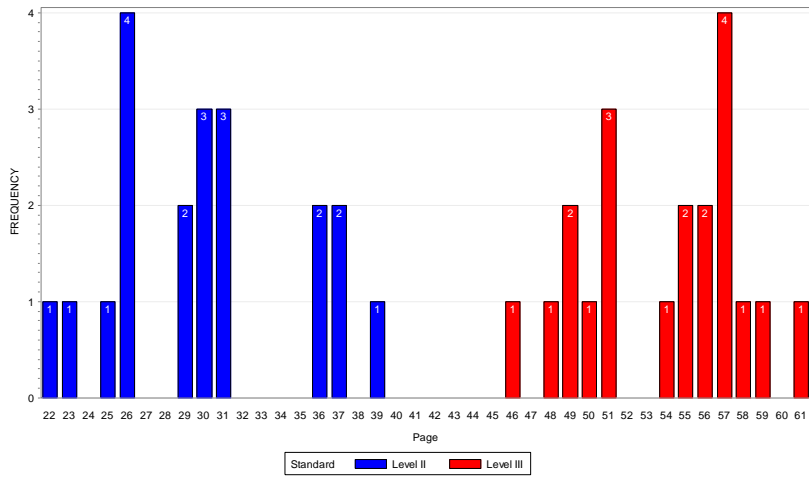
#### **COMMITTEE MEETING RESULTS**

This section includes several sets of tables and figures that describe the results of the STAAR grades 3–8 mathematics standard-setting committee meetings. Similar tables and figures for other assessments are located in Appendices 11, 12, 13, 14, and 15.

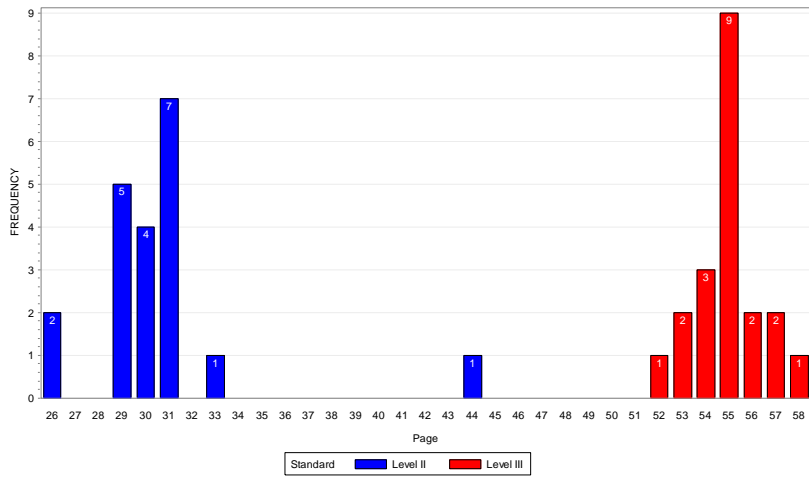
Figures 13.1–13.6 show the frequency distributions of the recommended cuts (i.e., bookmarked page numbers) after each round of judgment. The spread of the judgments provides a sense of overall agreement among the committee members. Table 13.5 summarizes the committee members' judgments by providing the 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 13.6. This includes the minimum, maximum, mean, standard deviation, and median of the standard-setting panelists' cut score recommendations (based on the Ordered Item Booklet page number) during each judgment round of the committee meetings.

Figure 13.7 shows 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. The impact data were computed using student performance on the spring 2015 administration of the redesigned STAAR grades 3–8 mathematics assessments.

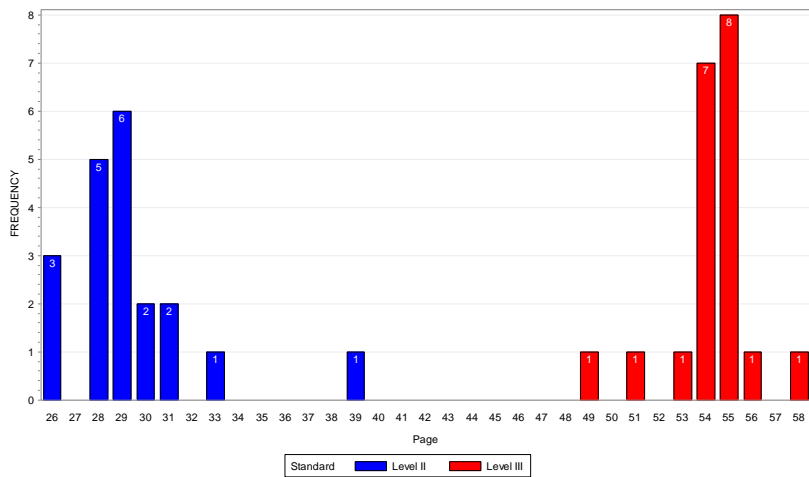
**Round 1 Panelist Agreement Data  
STAAR Grade 08 English Mathematics**



**Round 2 Panelist Agreement Data  
STAAR Grade 08 English Mathematics**

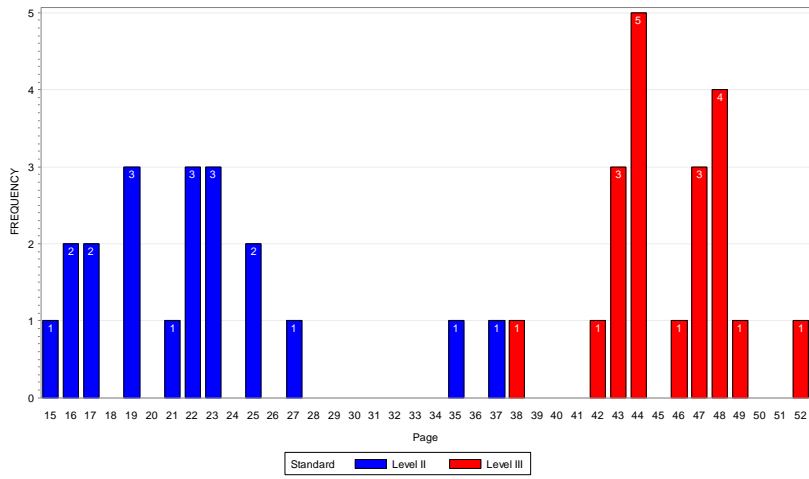


**Round 3 Panelist Agreement Data  
STAAR Grade 08 English Mathematics**

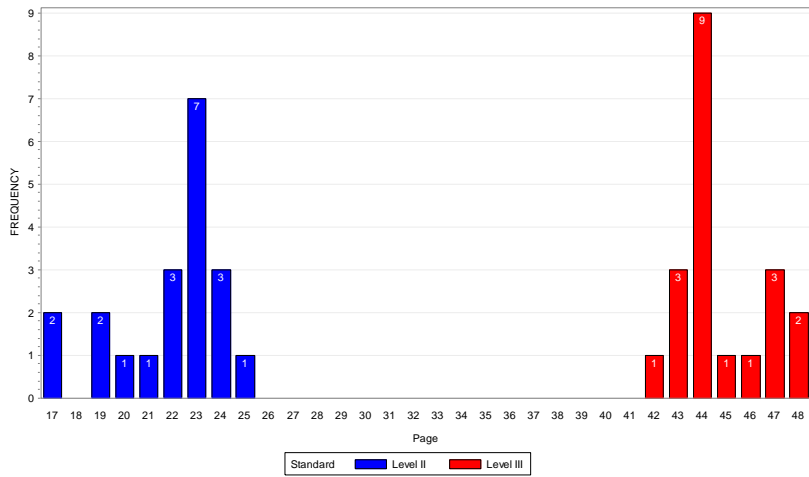


**Figure 13.1. Standard-Setting Panelists' Agreement Data for Grade 8 Mathematics.**

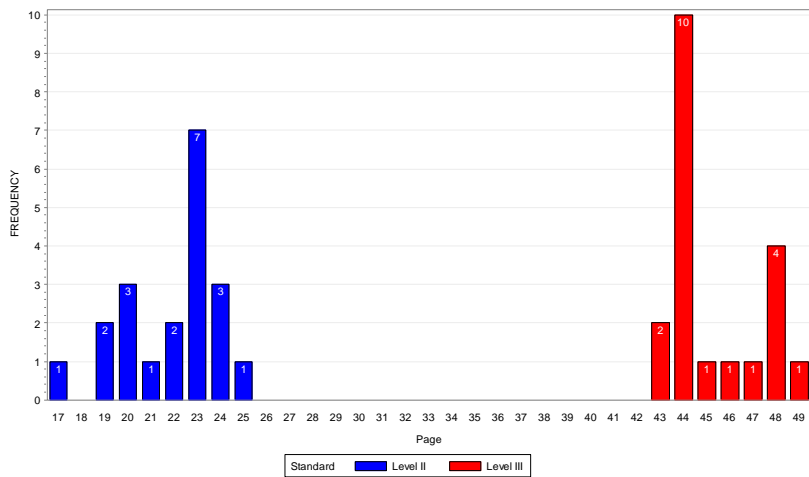
**Round 1 Panelist Agreement Data  
STAAR Grade 07 English Mathematics**



**Round 2 Panelist Agreement Data  
STAAR Grade 07 English Mathematics**

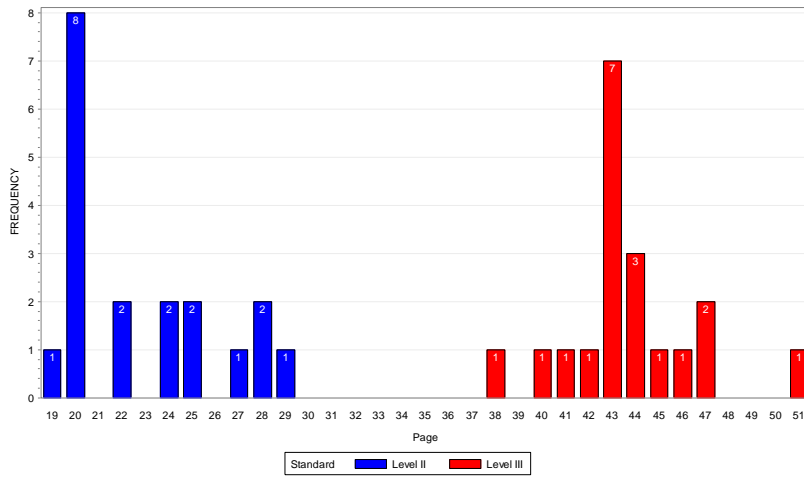


**Round 3 Panelist Agreement Data  
STAAR Grade 07 English Mathematics**

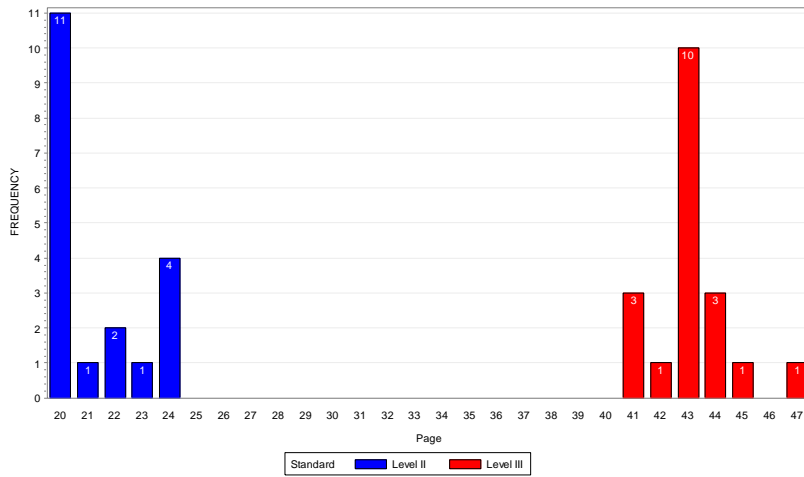


**Figure 13.2. Standard-Setting Panelists' Agreement Data for Grade 7 Mathematics.**

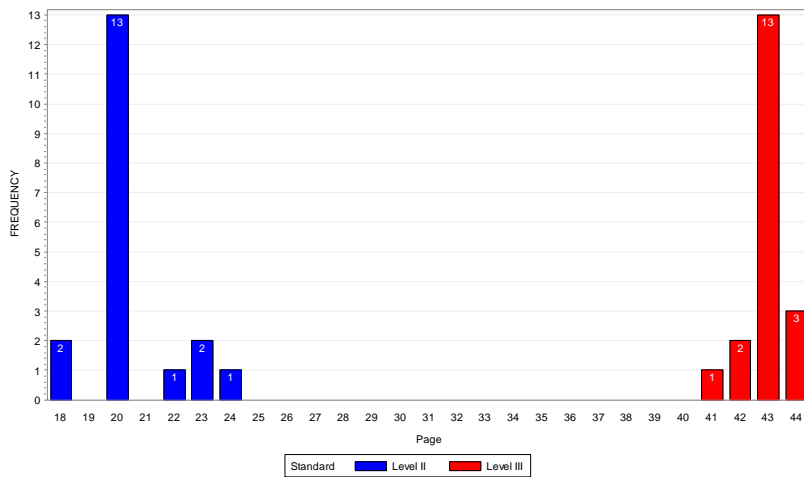
**Round 1 Panelist Agreement Data  
STAAR Grade 06 English Mathematics**



**Round 2 Panelist Agreement Data  
STAAR Grade 06 English Mathematics**

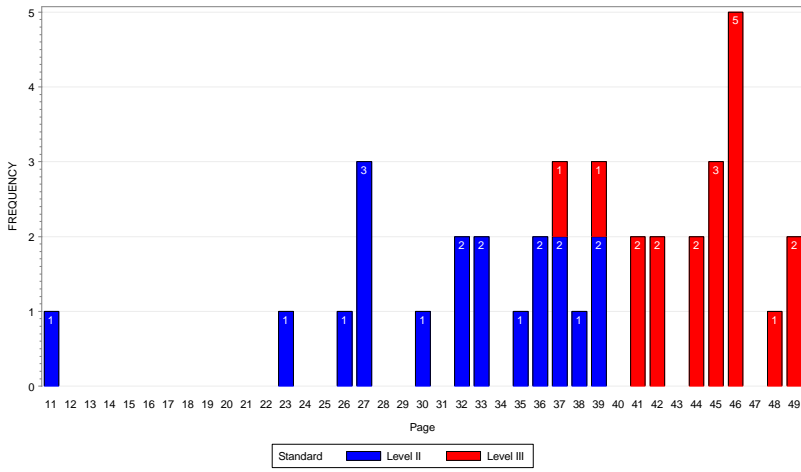


**Round 3 Panelist Agreement Data  
STAAR Grade 06 English Mathematics**

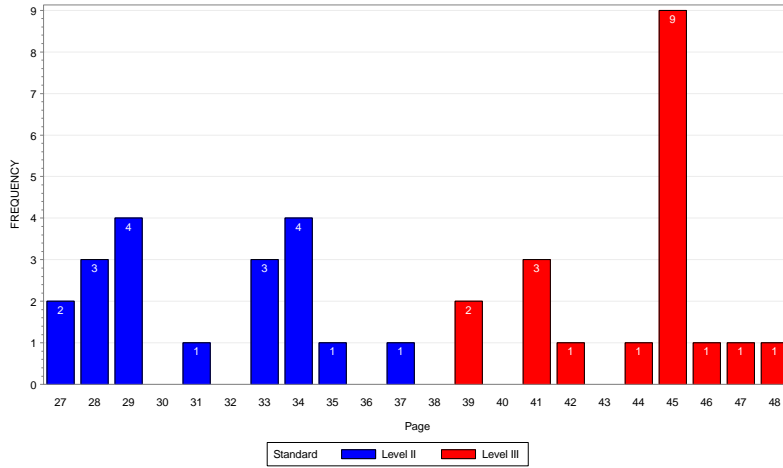


**Figure 13.3. Standard-Setting Panelists' Agreement Data for Grade 6 Mathematics.**

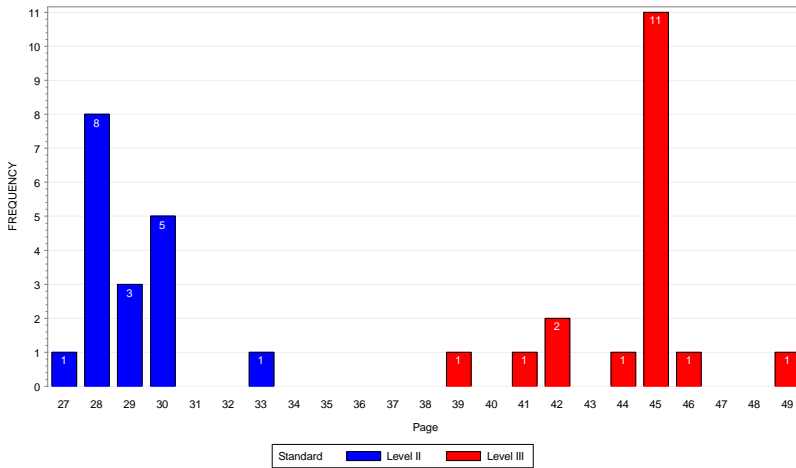
**Round 1 Panelist Agreement Data  
STAAR Grade 05 English Mathematics**



**Round 2 Panelist Agreement Data  
STAAR Grade 05 English Mathematics**

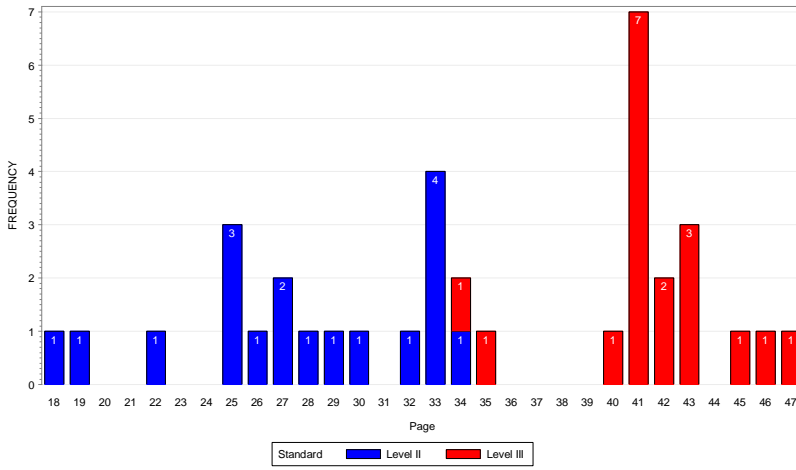


**Round 3 Panelist Agreement Data  
STAAR Grade 05 English Mathematics**

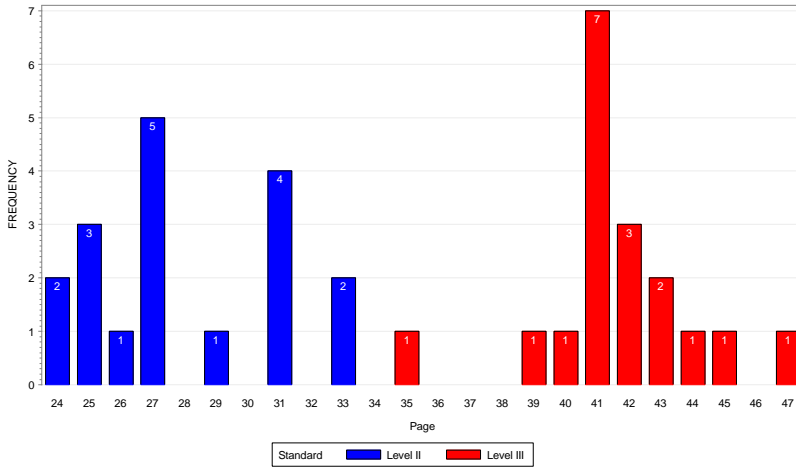


**Figure 13.4. Standard-Setting Panelists' Agreement Data for Grade 5 Mathematics.**

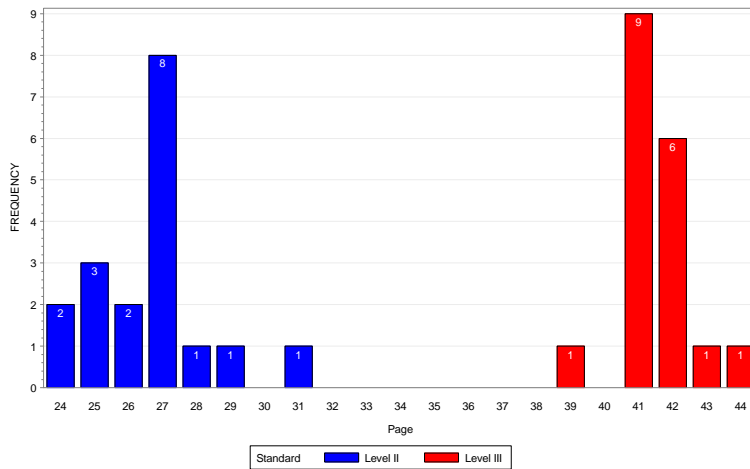
**Round 1 Panelist Agreement Data  
STAAR Grade 04 English Mathematics**



**Round 2 Panelist Agreement Data  
STAAR Grade 04 English Mathematics**

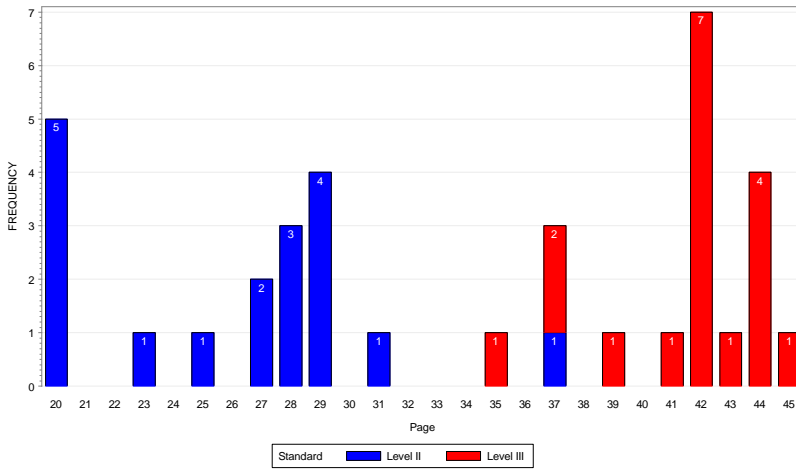


**Round 3 Panelist Agreement Data  
STAAR Grade 04 English Mathematics**

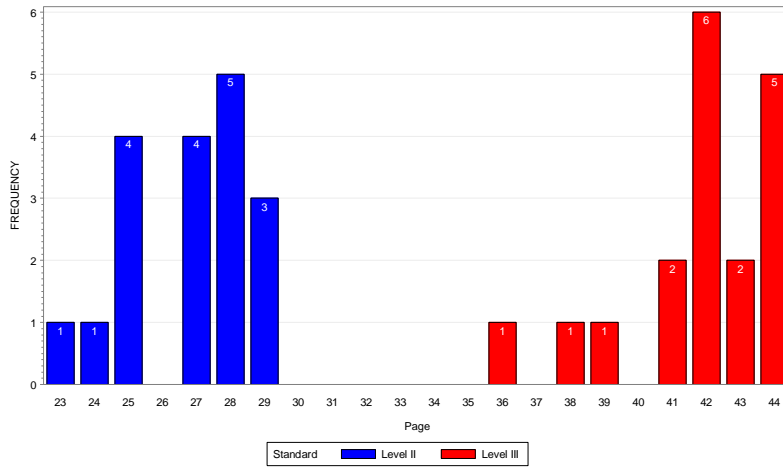


**Figure 13.5. Standard-Setting Panelists' Agreement Data for Grade 4 Mathematics.**

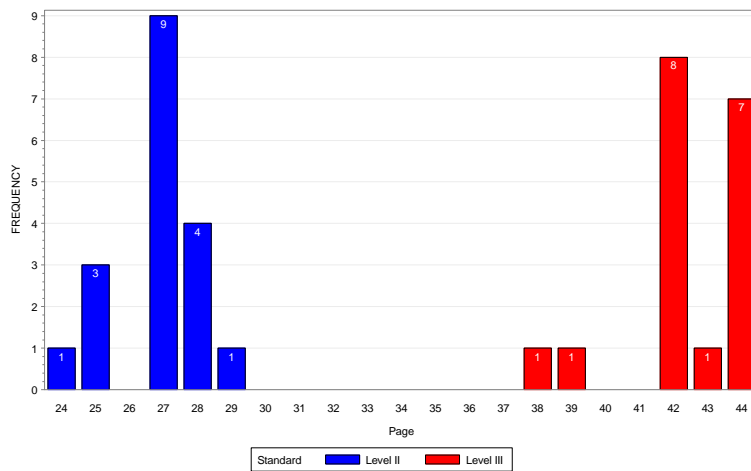
**Round 1 Panelist Agreement Data  
STAAR Grade 03 English Mathematics**



**Round 2 Panelist Agreement Data  
STAAR Grade 03 English Mathematics**



**Round 3 Panelist Agreement Data  
STAAR Grade 03 English Mathematics**



**Figure 13.6. Standard-Setting Panelists' Agreement Data for Grade 3 Mathematics.**

**Table 13.5: Summary of Cut-Score Recommendations, by OIB page, for STAAR Grades 3–8 Mathematics.**

	Grade 3		Grade 4		Grade 5		Grade 6		Grade 7		Grade 8	
	Level II	Level III	Level II	Level III	Level II	Level III	Level II	Level III	Level II	Level III	Level II	Level III
<b>Round 1</b>	28	42	28	41	33	45	22	43	22	45	30	55
<b>Round 2</b>	27	42	27	41	31	45	20	43	23	44	30	55
<b>Round 3</b>	27	42	27	41	29	45	20	43	23	44	29	55
<b>Reasonableness Review</b>	27	42	27	41	29	45	20	43	23	44	29	55

**Table 13.6: Summary of Standard-Setting Panelists' Judgments, by OIB page, for STAAR Grades 3–8 Mathematics.**

		Round	Minimum	Maximum	Mean	Standard Deviation	Median
Grade 3	Level II	1	20	37	26.1	4.76	28
		2	23	29	26.8	1.83	27
		3	24	29	26.8	1.29	27
	Level III	1	35	45	41.5	2.77	42
		2	36	44	41.8	2.23	42
		3	38	44	42.4	1.72	42
Grade 4	Level II	1	18	34	27.7	4.88	28
		2	24	33	27.9	3.02	27
		3	24	31	26.6	1.72	27
	Level III	1	34	47	41.5	3.19	41
		2	35	47	41.6	2.50	41
		3	39	44	41.5	1.04	41
Grade 5	Level II	1	11	39	31.5	6.91	33
		2	27	37	31.2	3.11	31
		3	27	33	28.9	1.39	29
	Level III	1	37	49	44.3	3.23	45
		2	39	48	43.8	2.57	45
		3	39	49	44.3	2.17	45
Grade 6	Level II	1	19	29	22.8	3.34	22
		2	20	24	21.3	1.69	20
		3	18	24	20.4	1.54	20
	Level III	1	38	51	43.7	2.81	43
		2	41	47	43.1	1.41	43
		3	41	44	42.9	0.71	43
Grade 7	Level II	1	15	37	22.2	5.80	22
		2	17	25	21.9	2.30	23
		3	17	25	21.9	2.10	23
	Level III	1	38	52	45.5	3.12	45
		2	42	48	44.8	1.77	44
		3	43	49	45.3	1.97	44
Grade 8	Level II	1	22	39	30.0	4.91	30
		2	26	44	30.6	3.58	30
		3	26	39	29.3	2.87	29
	Level III	1	46	61	53.9	4.16	55
		2	52	58	55.0	1.43	55
		3	49	58	54.3	1.80	55

### STAAR Grades 3-8 Mathematics Impact Data Across Grades

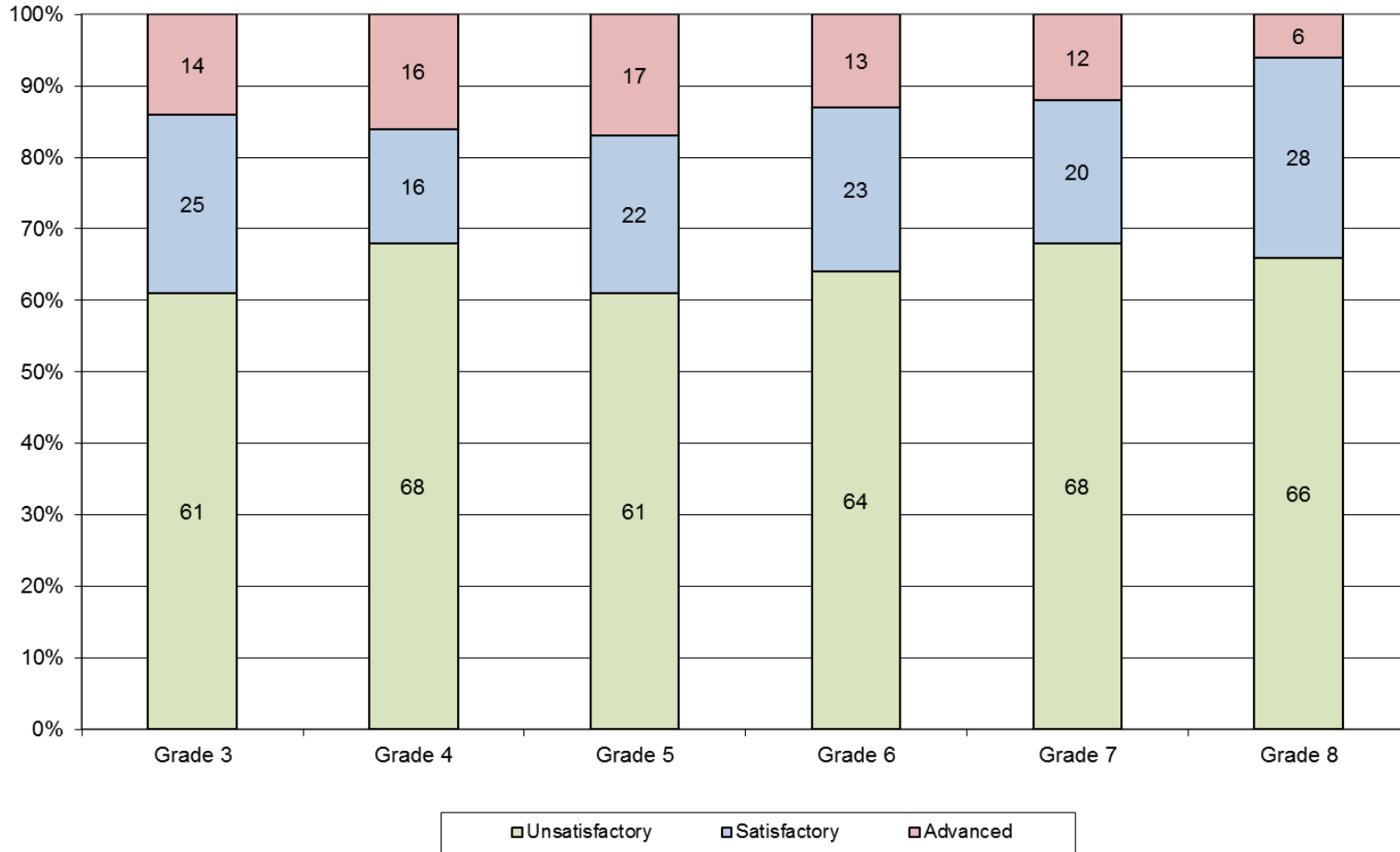


Figure 13.7. Estimated Impact Across Grades after Round 3 and Reasonableness Review.

## STANDARD-SETTING PROCESS EVALUATION

At the end of the standard-setting meetings, 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 grades 3–8 mathematics assessments. Panelists' responses to the evaluation form are summarized in Tables 13.7 for STAAR grades 6–8 mathematics and in Table 13.8 for STAAR grades 3–5 mathematics.

**Table 13.7: Standard-Setting Process Evaluation Summary Results for STAAR Grades 6–8 Mathematics.**

### 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%	37%	63%	0%
Discussion of the performance labels and the definitions	0%	0%	21%	79%	0%
Taking the actual assessments	0%	0%	11%	89%	0%
Overview of the item mapping procedure	0%	0%	26%	74%	0%
Practice exercise for the item-mapping procedure	0%	11%	16%	74%	0%
Feedback data provided in each round	0%	0%	16%	84%	0%
Discussion after each round	0%	0%	11%	89%	0%

### Section 2: Usefulness of Activities and Information

Activity or Information	Not Useful	Somewhat Useful	Useful	Very Useful	Omit
Performance Level Descriptors (PLDs)	0%	5%	42%	53%	0%
Training in the bookmark standard-setting method	0%	11%	16%	74%	0%
Feedback data provided after Round 1	0%	0%	16%	84%	0%
Feedback data provided after Round 2	0%	0%	16%	84%	0%
Presentation of data across grades	0%	0%	21%	79%	0%

### Section 3: Adequacy of Meeting Elements

Meeting Element	Not Adequate	Somewhat Adequate	Adequate	More Than Adequate	Omit
Training provided	0%	0%	21%	79%	0%
Amount of time spent training	0%	0%	32%	68%	0%
Feedback provided between rounds	0%	0%	21%	79%	0%
Facilities used for the session	0%	0%	11%	63%	26%
Total amount of time in groups to make judgments	0%	0%	32%	68%	0%
Number of rounds for the judgments	0%	5%	21%	74%	0%

**Section 4: Performance Level Descriptors**

Performance Category	Not Confident	Somewhat Confident	Confident	Very Confident	Omit
Level I: Unsatisfactory Academic Performance	0%	5%	58%	37%	0%
Level II: Satisfactory Academic Performance	0%	5%	42%	53%	0%
Level III: Advanced Academic Performance	5%	5%	47%	42%	0%

**Section 5: Cut-Score Recommendations**

Cut Score	Not Confident	Somewhat Confident	Confident	Very Confident	Omit
Level II: Satisfactory Academic Performance	0%	5%	47%	47%	0%
Level III: Advanced Academic Performance	5%	5%	53%	37%	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%	5%	95%	0%
Ask questions about the standards and how they will be used	0%	0%	16%	79%	5%
Ask questions about the process of making cut score recommendations	0%	0%	11%	89%	0%
Interact with your fellow panelists	0%	0%	11%	89%	0%

**Section 7: Respect**

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

**Table 13.8: Standard-Setting Process Evaluation Summary Results for STAAR Grades 3–5 Mathematics.**

**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%	17%	83%	0%
Discussion of the performance labels and the definitions	0%	0%	6%	94%	0%
Taking the actual assessments	0%	0%	0%	100%	0%
Overview of the item mapping procedure	0%	0%	11%	89%	0%
Practice exercise for the item-mapping procedure	0%	0%	17%	83%	0%
Feedback data provided in each round	0%	0%	0%	100%	0%
Discussion after each round	0%	0%	0%	100%	0%

**Section 2: Usefulness of Activities and Information**

Activity or Information	Not Useful	Somewhat Useful	Useful	Very Useful	Omit
Performance Level Descriptors (PLDs)	0%	0%	6%	94%	0%
Training in the bookmark standard-setting method	0%	0%	0%	100%	0%
Feedback data provided after Round 1	0%	0%	0%	100%	0%
Feedback data provided after Round 2	0%	0%	6%	94%	0%
Presentation of data across courses	0%	0%	0%	100%	0%

**Section 3: Adequacy of Meeting Elements**

Meeting Element	Not Adequate	Somewhat Adequate	Adequate	More Than Adequate	Omit
Training provided	0%	0%	11%	89%	0%
Amount of time spent training	0%	0%	22%	78%	0%
Feedback provided between rounds	0%	0%	17%	83%	0%
Facilities used for the session	0%	0%	11%	78%	11%
Total amount of time in groups to make judgments	0%	0%	11%	89%	0%
Number of rounds for the judgments	0%	0%	17%	83%	0%

**Section 4: Performance Level Descriptors**

Performance Category	Not Confident	Somewhat Confident	Confident	Very Confident	Omit
Level I: Unsatisfactory Academic Performance	0%	0%	28%	72%	0%
Level II: Satisfactory Academic Performance	0%	0%	28%	72%	0%
Level III: Advanced Academic Performance	0%	0%	28%	72%	0%

**Section 5: Cut-Score Recommendations**

Cut Score	Not Confident	Somewhat Confident	Confident	Very Confident	Omit
Level II: Satisfactory Academic Performance	0%	6%	28%	67%	0%
Level III: Advanced Academic Performance	0%	6%	28%	67%	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%	11%	89%	0%
Ask questions about the standards and how they will be used	0%	0%	11%	72%	17%
Ask questions about the process of making cut score recommendations	0%	0%	6%	94%	0%
Interact with your fellow panelists	0%	0%	6%	94%	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 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” or “very confident.” All committee members thought that they were given adequate opportunity to express their opinions, ask questions, and interact with other committee members. Additionally, all panelists indicated that they believed their opinions and judgments were respected by others.

## Post-Standard-Setting Activities

### REASONABLENESS REVIEW

Following the standard-setting meetings, TEA conducted a reasonableness review of the STAAR grades 3–8 mathematics cut-score recommendations. The cut scores were reviewed in comparison to each other and to STAAR Algebra I to determine the reasonableness of the system of standards recommended. No adjustments were made to the panel-recommended performance standards during the reasonableness review.

### SCALE-SCORE SYSTEM AND PHASE-IN RECOMMENDATIONS

Once the final recommended cut scores for Level II and Level III were in place, the scale-score systems and phase-in cuts (for Level II) were established for the STAAR grades 3–8 mathematics assessments. Under Texas Education Code (TEC) §39.036, the TEA is required to develop a vertical scale for assessing student performance in grades 3–8 mathematics. For the STAAR grades 3–8 mathematics vertical scale, a score value of 1700 was used to designate the grade 8 mathematics Final Level II cut score, and the standard deviation (SD) across all grades of the vertical scale was set to a value of 150. This was done to match the vertical scale-score system for the STAAR grades 3–8 reading assessments and the previous STAAR grades 3–8 mathematics vertical scale-score system.

Figure 13.8 illustrates the Level II and Level III recommended cut scores for mathematics on the vertical scale. The STAAR mathematics vertical scale has the following characteristics.

- The range is approximately 700 to 2300 scale score points.
- The Level II cut score is 1700 for grade 8 mathematics.
- The Level II cut scores increase across grades.
- The Level III cut scores increase across grades.

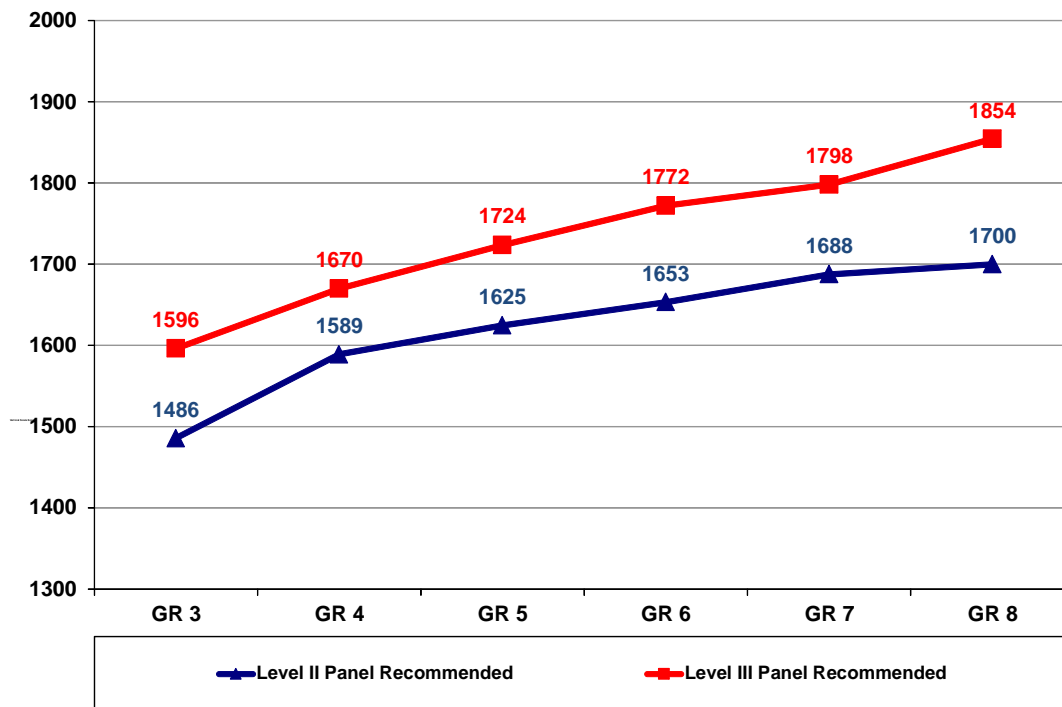


Figure 13.8 STAAR Grades 3–8 Mathematics Vertical Scale Scores

Initially in 2012, each STAAR grade 3–8 mathematics assessment set a Level II Phase-in 1 and 2 cut scores at 1.0 SD and 0.5 SDs below the Final Level II cut score respectively. In summer 2014, a three-step phase-in process for Level II was implemented. For both the old and new versions of the STAAR grades 3–8 mathematics assessments, the Level II phase-in cuts correspond to 1.0, 0.7, and 0.3 standard deviations below the Final Level II performance standard. The phase-in schedule for the new STAAR grades 3–8 mathematics performance standards remains the same as the phase-in schedule for all other STAAR assessments. The resulting cut scores for the STAAR grades 3–8 mathematics assessments are summarized in Table 13.9.

Table 13.9: Recommended Final and Phase-in Cut Scores for STAAR Grades 3–8 Mathematics

STAAR Mathematics Assessment	Level II Phase-in 1	Level II Phase-in 2	Level II Phase-in 3	Level II Final	Level III
Grade 3	1347	1388	1444	1486	1596
Grade 4	1453	1494	1548	1589	1670
Grade 5	1487	1528	1583	1625	1724
Grade 6	1523	1562	1614	1653	1772
Grade 7	1563	1600	1650	1688	1798
Grade 8	1583	1618	1665	1700	1854

### Approval of Cut Scores

On July 29, 2015, the Texas commissioner of education approved the phase-in and recommended cut scores for the STAAR grades 3–8 mathematics assessments (provided in Table 13.9) for use with the spring 2015 administration and future administrations of these assessments.