Chapter 5 STAAR Alternate 2

Overview
Participation Requirements
Testing Requirements for Graduation
Test Development
Training
Test Administrations
Testing Accommodations
Student Success Initiative
Scores and Reports
Parent Brochure
Performance Standards
Scaling
Equating
Reliability
Validity
Measures of Student Progress
Sampling
Test Results

Overview

State of Texas Assessments of Academic Readiness (STAAR®) Alternate 2 is an assessment based on alternate academic achievement standards and is designed for students with the most significant cognitive disabilities. The assessment was designed for this student population to meet federal requirements mandated under No Child Left Behind (NCLB). STAAR Alternate 2 is not a traditional paper or multiple-choice test. Instead, it involves test administrators observing students as they respond to standardized, state-developed assessment items that link to the grade-level Texas Essential Knowledge and Skills (TEKS). Teachers evaluate student performance based on standard scoring instructions embedded into each item on the STAAR Alternate 2 and submit student results through the Texas Assessment Management System.

The assessments included in STAAR Alternate 2 are shown in Table 5.1. STAAR Alternate 2 was administered during the window of February 9, 2015, through February 27, 2015, for all tested subject areas and grades.

Table 5.1. 2014–2015 STAAR Alternate 2 Assessments

Grade	Assessed Subject Area/Course				
3	Mathematics	Reading			
4	Mathematics	Reading	Writing		
5	Mathematics	Reading		Science	
6	Mathematics	Reading			
7	Mathematics	Reading	Writing		
8	Mathematics	Reading		Science	Social Studies
High School	Algebra I	English I English II		Biology	U.S. History

Participation Requirements

STAAR Alternate 2 has specific participation requirements that an admission, review, and dismissal (ARD) committee must carefully consider when recommending these assessments for students receiving special education services. Prior to reviewing the eligibility criteria for STAAR Alternate 2, the ARD committee must understand all assessment options, including the characteristics of each assessment and the potential implications of each assessment choice.

If STAAR Alternate 2 is being considered, the ARD committee must review the four criteria below and indicate whether the description is applicable to the student. For a student to be eligible to participate in STAAR Alternate 2, the answer to all four questions below must be "Yes." If the answer to any question is "No," the student is not eligible to participate in STAAR Alternate 2 and must participate in one of the other state assessments. Each "Yes" answer must be justified by evidence that the student meets the criterion.

1. **Does the student have a significant cognitive disability?** A significant cognitive disability is determined by the ARD committee and must be based on evaluation information performed by a qualified evaluation team. The significant cognitive disability must affect the student's intellectual potential and be documented as such in the student's individualized education program (IEP). A student with a significant cognitive disability has limited potential to reach grade-level expectations; whereas, a student with a learning disability has the potential to reach grade-level expectations, but has difficulty doing so due to his or her disability.

2. **Does the student require specialized supports to access the grade-level curriculum and environment?** Federal regulations mandate that all students have access to, and are assessed on, grade-level curriculum. To access the state-mandated grade-level or course curriculum, the TEKS, a student with a significant cognitive disability needs specialized academic instruction as well as support throughout the day in areas such as expressing his or her needs, getting from place to place, eating lunch, negotiating social situations, and/or taking care of personal needs.
3. **Does the student require intensive, individualized instruction in a variety of instructional settings?** The student needs specialized academic instruction and techniques over a period of time to ensure that he or she can learn, retain information, and transfer skills to other settings.
4. **Does the student access and participate in the grade-level TEKS through prerequisite skills?** Access to the grade-level curriculum is mandated by the federal government. A student with a significant cognitive disability requires access to the TEKS through prerequisite skills that are linked to the grade-level curriculum.

Testing Requirements for Graduation

With the passage of House Bill (HB) 3, the relationship between high school courses and participation in the STAAR Alternate 2 end-of-course (EOC) assessments is linked to a student's graduation plan. HB 5 reduced the total number of tests to five: Algebra I, English I, English II, biology, and U.S. history. However, the ARD committee makes final determinations on the graduation requirements for students receiving special education services.

Test Development

STAAR Alternate 2 follows the same test development procedures as other STAAR assessments. However, the test development process does reflect the unique characteristics of STAAR Alternate 2, specifically its reliance upon scripted items and the learning styles of the STAAR Alternate 2 population.

Assessment Content

Like other STAAR assessments, STAAR Alternate 2 is linked to grade-level TEKS and student expectations for STAAR. The preliminary task in developing the alternate assessment was to link the assessment to the curriculum content and expectations. The Texas Education Agency (TEA) worked with experts in test development, special education, and content to develop curriculum frameworks and vertical alignment documents. The curriculum frameworks list the grade-level TEKS and the associated prerequisite skills for each grade and subject area. The vertical alignment documents link skills and knowledge across grades within the same subject area. After the initial creation of the curriculum frameworks and vertical alignment documents, TEA sought additional input from educator committees and a statewide steering committee that

included state assessment experts, parents, advocacy group representatives, related service providers, administrators, and Texas regional Education Service Center (ESC) professionals.

The next step in developing the alternate assessment was to generate essence statements that summarize the TEKS and student expectations and link the expectations to the prerequisite skills and assessment performance categories. Typically, each grade and subject area assessment contains 10–20 essence statements. From these, essence statements are identified for inclusion in the STAAR Alternate 2 assessment each year. The 2014–2015 assessed essence statements were made available to teachers in spring 2014 to allow time for instructional planning and developing standards-based IEPs for the following school year.

STAAR Alternate Redesign

As a result of House Bill 5 of the 83rd Texas Legislative Session, TEA redesigned the STAAR Alternate assessment. To meet requirements of the legislation and maintain an appropriate assessment for students with significant cognitive disabilities, an item-based approach to the assessment was implemented for the redesigned STAAR Alternate 2. The issues of validity, reliability, fairness, accessibility, and consistency in meaning were carefully considered. In addition, the principles of universal design were incorporated in the early stages of test development in order to develop accessible, non-biased items. Consideration was also given to students' response modes, which allow students to show what they know during the assessment.

After prototype items were developed, cognitive labs were conducted to gather information on student performance, engagement, and interaction with the redesigned STAAR Alternate 2 items. Test administrators were interviewed regarding the proposed test design and the feasibility of the assessment for students. The next step was a pilot test to gather further student performance data and survey test administrators regarding the STAAR Alternate 2 test items. The data from the cognitive labs, pilot tests, and test administrator surveys were used to develop items for the 2015 operational assessment.

Assessment Item Criteria

In addition to the procedures outlined in chapter 2, "Building a High-Quality Assessment System," and described above, nationally accepted criteria provided guidance during the development of the STAAR Alternate 2 items. Specifically, the following criteria were directly referenced during development of the redesigned STAAR Alternate 2.

- Standard 4.1 of the Standards for Educational and Psychological Testing (AERA, APA, NCME, 2014), which states:

  > Test specifications should describe the purpose(s) of the test, the definition of the construct or domain measured, the intended examinee population, and interpretations for intended uses. The specifications should include a

rationale supporting the interpretations and uses of test results for the intended purpose(s) (p.85).

- Standard 4.8 of the Standards for Educational and Psychological Testing (AERA, APA, NCME, 2014), which states:

  The test review process should include empirical analyses and/or the use of expert judges to review items and scoring criteria. When expert judges are used, their qualifications, relevant experiences, and demographic characteristics should be documented, along with the instructions and training in the item review process that the judges receive (p. 88).

- Universal design, with particular attention given to (1) students' response modes, allowing students to show what they know and can do; (2) differentiated supports and materials, allowing students to access the content of the assessment; and (3) multiple means of engagement to allow students more time to complete the task, meaningful activities, and context (Center for Applied Special Technology, 2002). According to the principles of universal design, each item has precisely defined constructs, has maximum legibility, has maximum readability and comprehensibility, is amenable to accommodations, is accessible and non-biased, and takes into consideration special populations.

Review of Items

During the item development process for the redesigned STAAR Alternate 2, educator committees met to complete reviews of every item. The committees were comprised of educators from across Texas, specifically special education experts, special education classroom teachers (including teachers from the Texas School for the Blind and Visually Impaired and the Texas School for the Deaf), teachers of English language learners, and general education teachers.

The educator committees focused on the relationship between the grade-level content and the items. Each committee member completes an item judgment form with the following questions in regard to each item.

- Does this item measure the reporting category/student expectation/essence statement/prerequisite skill it was designed to measure?
- Is this item an appropriate measure of the TEKS student expectation/essence statement/prerequisite skill?
- Is this item free from bias on the basis of students' personal characteristics such as gender, ethnicity, or disability?
- Would you expect students in your district to have received sufficient instruction by the end of the grade/course to enable them to answer this item correctly?

Feedback from the educator committees was used to revise the STAAR Alternate 2 items as needed.

Training

Resources were provided by TEA outlining administration procedures, sample items, and online activities prior to the testing window. It was recommended that all personnel who planned to administer STAAR Alternate 2 review these resources prior to the test administration window. In addition, a preview window was offered so that school personnel could review the actual assessment items and apply any accommodations appropriate for their students prior to the test administration window.

Test Administrations

More than 91,000 STAAR Alternate 2 assessments were administered in 2014–2015 to approximately 37,000 students. Table 5.2 further describes the 2014–2015 STAAR Alternate 2 administrations by grade and subject area.

Table 5.2. Students Tested in 2014–2015 STAAR Alternate 2 Assessments

STAAR Alternate 2 Assessments	Students Tested
Grade 3 mathematics	4,767
Grade 3 reading	4,769
Grade 4 mathematics	4,600
Grade 4 reading	4,596
Grade 4 writing	4,596
Grade 5 mathematics	4,620
Grade 5 reading	4,618
Grade 5 science	4,623
Grade 6 mathematics	4,304
Grade 6 reading	4,305
Grade 7 mathematics	4,084
Grade 7 reading	4,081
Grade 7 writing	4,078
Grade 8 mathematics	4,037
Grade 8 reading	4,039
Grade 8 science	4,037
Grade 8 social studies	4,038
Algebra I	3,584
English I	3,575
English II	3,329
Biology	3,454
U.S. history	3,009

Administration Procedures

The STAAR Alternate 2 assessment process is designed with scripted test administrator presentation instructions that mirror instructional techniques for a student with a significant cognitive disability. The essence statements, upon which the 2014–2015 STAAR Alternate 2 items were based, were made available in spring 2014 so that they could be included in students' IEPs, ARD committee meetings, and other planning related to the 2014–2015 school year. The STAAR Alternate 2 student and teacher booklets were made available during a preview window that opened when districts received materials (between January 19 and 23, 2015) and the test administration window from February 9, 2015, to February 27, 2015, which included the time necessary to administer the assessment to all students and enter results in TestNav. Because of the heterogeneity of the population of students who take STAAR Alternate 2, teacher assists are built into the assessment and its administration, allowing teachers to accommodate the items to fit the individual needs of each student. Test administrators take the following steps as part of the administration of the assessment.

1. Select any allowable accommodations that may be appropriate for each student. As part of this step, test administrators accommodate the materials and determine supports that are needed for individual students during the assessment.
2. Administer the items following the standardized presentation and scoring instructions.
3. Observe and score student performance.
4. Enter scoring information into an online transcription form in TestNav. Test administrators enter the scores so that each student's performance can be scored.

A student with a severe medical or cognitive impairment may not be able to complete a part of the assessment. For these exceptions, ARD committees can determine if a student's assessment may be coded as a Medical Exception or as No Authentic Academic Response. For both exceptions, the ARD committee makes the determination after reviewing medical and educational records. The decision is documented in the student's IEP along with evidence to support the determination. A decision not to assess a student should be rare. Descriptions of the two categories are provided below.

MEDICAL EXCEPTIONS

Students who are medically fragile and cannot attend to or tolerate any academic interaction can qualify for a medical exception for the following circumstances.

- The student is in the final stages of a terminal or degenerative illness.
- The student is receiving extensive short-term medical treatment due to a medical emergency or serious injury in an accident.
- The student is unable to interact with peers or staff without risk of infection or contamination to himself/herself or others.

- The student is receiving non-academic homebound services due to medical issues and does not receive academic instruction.

No Authentic Academic Response

Students who are not able to respond authentically to any verbal, visual, or tactile stimuli during academic instruction due to level of cognition rather than a medical condition can qualify for a No Authentic Academic Response (NAAR) exception for the following circumstances.

- The student does not show any observable reaction to a specific stimuli.
- The student exhibits only startle responses.
- The student tracks or fixates on objects at random and not for a purpose.
- The student moves or responds only to internal stimuli.
- The student vocalizes intermittently regardless of changes in the environment around him or her.
- The student with multiple impairments is unable to receive any visual, auditory, or tactile information during the assessment.

Test administrators are able to deliver the assessment and submit assessment scores at any time during the STAAR Alternate 2 assessment window.

Testing Accommodations

STAAR Alternate 2 is a standardized assessment intended to be appropriate for eligible students in its original intact form. However, ARD committees and test administrators may elect to provide appropriate allowable accommodations to some students whose disability precludes them from participating meaningfully in a two-dimensional standardized assessment. Accommodations may only be used if they are routinely provided in classroom instruction and listed in the student's IEP. Accommodations provided during classroom instruction and testing may differ from accommodations allowed for use on statewide assessments. Certain accommodations used in the classroom would invalidate the content being assessed or compromise the security and integrity of the test. For this reason, not all accommodations suitable for instruction are allowed during the statewide assessments. The accommodations in Figure 5.1 below are examples that can be considered for STAAR Alternate 2 along with guidelines on how they should be applied.

Figure 5.1. STAAR Alternate 2 Allowable Accommodations

Accommodations to the Two-Dimensional Stimulus Images	
Allowable Accommodation	**Guidelines**
• Color or highlight stimulus images or answer choices. • Place color overlays on images or text. • Photocopy and cut out stimulus images from the test booklet (can be affixed to appropriate presentation media, e.g., easels, poster board, card stock, etc.). • Pair images or text in the student's booklet with photographs of the same objects, real objects of the same content, or picture representations. • Attach textured materials to images in the student's booklet. • Demonstrate concepts or relationships in images. • Raise or darken the outline of drawings in stimulus images. • Enlarge images with magnification devices, photocopying, or computer magnification programs. • Add braille labels to images or provide text in braille. • Describe images for students with visual impairments.	• The accommodation must be presented uniformly so that the correct answer is not emphasized over the other answer choices. • If using separate paper, the answer choices must be placed in the same order (top/left; middle/middle; bottom/right). • All demonstrations must include only what was presented in the stimulus. • If photographs or real objects are placed over images, each answer choice must have a comparable photograph or real object. • Any replacements, photographs, or objects must be as close to the original as possible. • Description of images can include only details of what can be seen in the image without comments about the overall impression of the image.
Accommodations to Limit Number of Images Shown at One Time	
Allowable Accommodation	**Guidelines**
• Provide the stimulus on separate paper presented one at a time. • Cover or isolate each image until it is addressed.	• All images must be presented in the same order or configuration as shown in the test booklet. • All cover-up techniques must be uniformly applied to all images within an item.
Accommodations to Language Used in the Test Administration Instructions	
Allowable Accommodation	**Guidelines**
• Use routine picture representations for key words in verbal directions to the student. • Reread sections of the text as requested by the student.	• With the exception of words of encouragement, no additional information other than what is visually presented, stated in text, or supplied in the test administrator instructions can be provided.

Student Success Initiative

The Student Success Initiative (SSI) provides a system of academic support to help students achieve success on grade level in mathematics and reading. SSI incorporates a grade-advancement component adopted by the Texas Legislature in 1999. Students who participate in STAAR Alternate 2 are not subject to the SSI requirements. Each student's grade promotion decision is determined by the student's ARD committee rather than being based on STAAR Alternate 2 performance.

Scores and Reports

Scoring STAAR Alternate 2 Assessments

STAAR Alternate 2 is scored polytomously using a standard scoring rubric that follows the same process of item administration across all items and is applied to the student performance evaluation information that test administrators submit electronically. Each item is scored according to the level of independence with which a student responds to an item as follows:

- If a student responds correctly to the first presentation of an item, he or she receives a score point of 2. If the student does not respond or responds incorrectly, the item is presented again with allowable teacher assists.
- If the student responds correctly to the second presentation of the item, he or she receives a score point of 1.
- If the student does not respond or responds incorrectly to the second presentation, he or she receives a score point of 0.

Each item is scored in the same manner. Item scores range from 0 to 2. There are 20 scored items per test, resulting in a total test score range of 0 to 40 points.

Description of Scores

Scores for the STAAR Alternate 2 assessments consist of the number of points earned (raw scores), scale scores, and the resulting performance level associated with the student's score.

Raw Score

The number of points that a student earns on a STAAR Alternate 2 assessment is the student's raw score. The raw score can be interpreted only in terms of the specific set of test items on that test form. However, because the difficulty of items might vary among test forms over time, raw scores alone cannot be used to compare performance across tests or administrations. To make these comparisons of student performance, raw scores must be converted to scale scores.

Scale Score

A scale score is a conversion of the raw score onto a scale that is common to all test forms for that assessment. Scale scores allow for direct comparisons of student performance between specific sets of test items from different test administrations.

The scale score is used to determine whether a student attained Level II: Satisfactory Academic Performance or Level III: Accomplished Academic Performance. (Performance-level cut scores are discussed in the Performance Standards section of this chapter.) Along with raw scores, scale scores for all STAAR Alternate 2 assessments are reported following each test administration.

Scale scores are also used to compare the performance of an individual student with the performance of a demographic group, a program group, an entire campus, or a

district at a particular grade. For example, the scores for a Hispanic student can be compared with the average scores of other Hispanic students, all students on campus, or any combination of these aggregations at that grade.

ADDITIONAL PERFORMANCE INFORMATION

Other scores can provide information about a student's relative strengths or weaknesses in core academic areas. For example, reporting category level data can identify areas where a student might be having difficulty. This identification can help campuses plan the most effective instructional intervention.

Report Formats

Two types of reports are provided for the various testing programs, standard and optional. Standard reports are provided automatically to districts, and the information in the standard reports satisfies mandatory reporting requirements. To receive optional reports that detail student performance data in additional formats, a district must select the corresponding optional reports in the *Administration Details* screen in the Assessment Management System. Generally, districts are required to pay a nominal fee for each optional report requested.

For more information about scoring and reporting for STAAR Alternate 2, refer to the TEA publication Interpreting Assessment Reports located on TEA's Student Assessment Division website.

Use of Test Results

Reports of STAAR Alternate 2 students are used in

- helping parents monitor the progress their children make;
- informing instructional planning for individual students;
- reporting results to local school boards, school professionals, and the community;
- evaluating programs, resources, and staffing patterns; and
- evaluating district effectiveness in their instructional programs.

Parent Brochure

TEA's Student Assessment Division produces the brochure Understanding Your Child's Confidential Student Report. The brochure includes a sample CSR with explanations of each element of the report to help parents better understand their child's score report. Reporting categories for each subject area assessed with STAAR Alternate 2 are summarized. The guide, developed in both English and Spanish, is provided on TEA's Student Assessment Division website.

Performance Standards

Performance standards relate levels of test performance directly to what students are expected to learn as described in the statewide curriculum.

Performance Levels and Policy Definitions

For the STAAR Alternate 2 assessments, the performance levels are

- Level I: Developing Academic Performance,
- Level II: Satisfactory Academic Performance, and
- Level III: Accomplished Academic Performance.

More detailed descriptions of these performance levels, known as policy definitions, are given below.

Level I: Developing Academic Performance

Performance in this category indicates that students are insufficiently prepared for the next grade or course and need additional instructional support for accessing the curriculum through prerequisite skills. Students are able to acknowledge concepts, but they demonstrate a minimal or inconsistent understanding of the knowledge and skills that are linked to content measured in this grade or course. Even with continued support, students in this category are in need of significant intervention to show progress in the next grade or course.

Level II: Satisfactory Academic Performance

Performance in this category indicates that students are sufficiently prepared for the next grade or course with instructional supports for accessing the curriculum through prerequisite skills. Students demonstrate sufficient understanding of the knowledge and skills that are linked to content measured at this grade or course. Students exhibit the ability to determine relationships, integrate multiple pieces of information, extend details, identify concepts, and match concepts that are similar. With continued support, students in this category have a reasonable likelihood of showing progress in the next grade or course.

Level III: Accomplished Academic Performance

Performance in this category indicates that students are well prepared for the next grade or course with instructional supports for accessing the curriculum through prerequisite skills. Students demonstrate a strong understanding of the knowledge and skills that are linked to content measured at this grade or course. Students exhibit the ability to use higher-level thinking and more complex skills, which includes making inferences, comparisons, and solving multi-step problems. With support, students in this category have a high likelihood of showing progress in the next grade or course.

Standard Setting Process for STAAR Alternate 2

Standards were set for STAAR Alternate 2 in spring 2015. Standard setting for STAAR Alternate 2 involved a process of combining considerations regarding policy, the TEKS content standards, educator knowledge about what students should know and be able to do, and information about how student performance on state assessments aligns with student performance on other assessments. TEA used an evidence-based standard setting approach (O'Malley, Keng, & Miles, 2012) for the STAAR Alternate 2 program. Using this approach, TEA defined and implemented a nine-step process to establish performance standards for all the STAAR Alternate 2 grades 3–8 and EOC assessments. The nine steps were as follows:

1. Conduct validity and linking studies.
2. Develop performance labels and policy definitions.
3. Convene a policy committee and develop reasonable ranges for performance standards.
4. Develop grade- and course-specific performance level descriptors (PLDs).
5. Convene standard setting committees.
6. Review performance standards for reasonableness.
7. Approve performance standards.
8. Implement performance standards.
9. Review performance standards.

Table 5.3 provides high-level descriptions and timelines for the steps in the STAAR Alternate 2 standard setting process.

Table 5.3. Overview of the STAAR Alternate 2 Standard Setting Process

Standard Setting Step	Description	Timeline
1. Conduct empirical studies	Analyses of pilot data as well as analysis of score distributions	Fall 2015
2. Develop performance labels and policy definitions	A committee was convened jointly by TEA and the Texas Higher Education Coordinating Board (THECB) to recommend performance categories, performance category labels, and general policy definitions for each performance category. The STAAR Alternate 2 performance labels and policy definitions were adapted from those created by the committee.	September 2010
3. Develop reasonable ranges for performance standards	The committee considered the policy implications of performance standards, empirical study results, and content recommendations to identify reasonable ranges for performance standards (neighborhoods).	Fall 2015
4. Develop grade and course PLDs	TEA and Pearson created draft specific PLDs and educator committees reviewed and edited the PLDs. A goal of the development and review of the specific PLDs was to create an aligned system describing a reasonable progression of skills within each subject area (mathematics, reading, science, and social studies).	January 2015
5. Convene standard setting committees	Committees consisting of general education and special education experts with experience in grades 3–12 used performance labels, policy definitions, specific PLDs, and predetermined ranges within which to recommend cut scores for each STAAR Alternate 2 assessment. These committees also provided comments to assist TEA with finalizing the specific PLDs.	April 2015
6. Review performance standards for reasonableness	TEA reviewed the recommendations across subject areas.	April 2015
7. Approve performance standards	The commissioner of education approved the STAAR Alternate 2 performance standards.	April 2015
8. Implement performance standards	Once established, performance standards were reported to students for the spring 2015 administration.	May 2015
9. Review performance standards	Performance standards are reviewed at least once every three years.*	If applicable

*In June 2013, the 83rd Texas Legislature enacted HB 5, which removed the requirement to review performance standards (Step 9). TEA may review the performance standards if deemed applicable.

More details about each of the steps in the STAAR Alternate 2 standard setting process are provided in the STAAR Alternate 2 Standard Setting Technical Report available on the STAAR Alternate 2 Standard Setting Information page of TEA's Student Assessment Division website.

Standard Setting Committees

The goal of each standard setting committee was to recommend two cut scores that would define the three performance levels for each of the STAAR Alternate 2 assessments. The standard setting committees were made up of K–12 educators. When selecting standard setting committee members, TEA placed an emphasis on experience with the population of students for whom STAAR Alternate 2 is appropriate, as well as content knowledge and classroom experience. Standard setting committees also included educators who had ELL and general education expertise.

In April 2015, educator committees were convened to recommend performance standards for all STAAR Alternate 2 assessments. Committees reviewed STAAR Alternate 2 test booklets, policy definitions, and PLDs. The panelists also received training in the evidence-based standard setting process that incorporated aspects of the extended Angoff, where panelists make judgments about the score needed on each item to demonstrate proficiency (Angoff, 1971; Hambleton & Plake, 1995). Committee members were provided reasonable ranges within which performance standards should be set. The ranges were determined using a content review of items, policy definitions, PLDs, and impact data. With this information in mind, committee members were asked to provide recommendations for where the cut scores should be placed in order to create the three STAAR Alternate 2 performance levels. Three rounds of recommendations were provided, with time for discussion and feedback between rounds. Committee members also participated in an articulation round where they could look at the third round recommendations across grades/courses and suggest adjustments. TEA used the third round and articulation recommendations in making final decisions about the performance standards. Performance standards from the April 2015 standard setting meetings were used to report students' scores in spring 2015.

Outcome of Standard Setting

The purpose of the standard setting process is to establish cut scores that reflect the level of performance a student must demonstrate in order to be classified into a performance level on each STAAR Alternate 2 assessment. These performance standards were approved by the commissioner of education in April 2015. Table 5.4 presents the approved performance standards for STAAR Alternate 2. TEA may review the performance standards if deemed applicable.

Table 5.4. STAAR Alternate 2 Performance Standards

Assessment	Final Recommended Level II	Final Recommended Level III
Grade 3 reading	300	381
Grade 4 reading	300	384
Grade 5 reading	300	387
Grade 6 reading	300	371
Grade 7 reading	300	371
Grade 8 reading	300	379
Grade 4 writing	300	363
Grade 7 writing	300	359
English I	300	367
English II	300	366
Grade 3 mathematics	300	375
Grade 4 mathematics	300	387
Grade 5 mathematics	300	379
Grade 6 mathematics	300	373
Grade 7 mathematics	300	375
Grade 8 mathematics	300	365
Algebra I	300	361
Grade 5 science	300	387
Grade 8 science	300	382
Biology	300	383
Grade 8 social studies	300	372
U.S. history	300	368

Scaling

Scaling is a statistical procedure that places raw scores on a common scoring metric in order to make test scores comparable across test administrations. As with previous Texas assessment programs, the STAAR Alternate 2 program uses the Rasch Partial-Credit Model (RPCM) to place test items on the same scale across administrations for a given assessment. Once performance standards have been set for an assessment, the Rasch scale is then transformed to a more user-friendly metric to facilitate interpretation of the test scores. Details of the RPCM scaling method used in Texas are provided in chapter 3, “Standard Technical Processes.”

Reporting Scales

Scale scores for STAAR Alternate 2 assessments are reported on a horizontal scale. Horizontal scale scores allow for direct comparisons of student performance between specific sets of test items from different test administrations for a specific grade and

subject. Refer to chapter 3, "Standard Technical Processes," for detailed information about the scaling process for the different types of reporting scales.

HORIZONTAL REPORTING SCALES

For STAAR Alternate 2 assessments, a scale score of 300 represents the final recommended Level II performance standard. The standard deviation is 60.

It is important to note that although Level II scale score values are fixed across horizontally scaled assessments, Level III scale score values vary across STAAR Alternate 2 assessments. For a given assessment, the Level III scale score value remains constant over time.

The STAAR scale scores represent linear transformations of Rasch proficiency level estimates (θ). Specifically, the transformation is made by first multiplying θ by a slope constant (A) and then adding an intercept constant (B). This operation is described by the equation below:

$$SS_\theta = A \times \theta + B \qquad (1)$$

where SS_θ is the scale score for a Rasch proficiency level estimate (θ). A and B in Equation (1) are referred to as the horizontal scaling constants. These same transformations will be applied each year to the Rasch proficiency level estimates (θ) for that year's set of test items. Values for the horizontal scaling constants are provided in Table 5.5 for the STAAR Alternate 2 grades 3–8 and EOC assessments.

Table 5.5. Horizontal Scaling Constants for STAAR Alternate 2

Assessment	*A*	*B*
Grade 3 reading	43.5388	283.9777
Grade 4 reading	45.6246	277.9633
Grade 5 reading	49.4951	276.0444
Grade 6 reading	45.0369	277.0312
Grade 7 reading	45.2817	278.5818
Grade 8 reading	42.5894	277.6406
Grade 4 writing	49.1207	286.3444
Grade 7 writing	45.6246	276.914
English I	46.1127	288.1951
English II	46.9087	292.0724
Grade 3 mathematics	43.9599	297.2305
Grade 4 mathematics	42.3406	297.9677
Grade 5 mathematics	42.9221	293.4758
Grade 6 mathematics	47.3082	293.8972
Grade 7 mathematics	45.0653	292.6994
Grade 8 mathematics	45.9897	283.5357
Algebra I	46.1042	287.8285
Grade 5 science	43.8943	291.6601
Grade 8 science	38.5892	298.495
Biology	38.2614	293.1129
Grade 8 social studies	41.4662	282.7501
U.S. history	41.3565	283.7055

Equating

Overview

Used in conjunction with the scaling process, equating is the statistical process that takes into account the slight differences in difficulty across test forms and administrations and allows for the scores to be placed onto a common scale. By using statistical methods, TEA equates the results of different tests so that scale scores across test forms and testing administrations can be compared. In the 2014–2015 school year, TEA conducted initial calibrations of the new STAAR Alternate 2 assessments that included a base-test review of the special administration, live calibrations, and field-test equating. In future years, these activities can include pre-equating, post-equating, and field-test equating. Refer to chapter 3, "Standard Technical Processes," for detailed information about equating.

Base Test Review

All the STAAR Alternate 2 base-test items were newly developed items in 2014–2015, meaning they had not been previously field tested. Prior to providing students' test scores based on these newly developed items, an immediate review of each item's

statistics was conducted after the administration to evaluate the items' performance. The process of reviewing base test item statistics is called a base-test review. In this process:

- item statistics for base-test items are reviewed;
- base-test items that do not meet predetermined statistical criteria are identified;
- statistically identified base-test items will undergo a content/data review; and
- if an identified base test item is determined to not be an acceptable item, then all students are given full credit for that item (i.e., 2 points).

The results of the base-test review were provided to TEA. No items were deemed eligible for full credit.

Live Calibrations

The live calibration process uses data from the operational test administration to estimate item difficulties for items that have not been previously field-tested. During a live calibration, the initial scale of the item bank for an assessment is created and future equating activities will use that initial scale.

Field-Test Equating

To replenish the item bank as new tests are created each year, newly developed items must be field-tested and equated to the item bank scale, as described in the technical details and procedures in chapter 3, "Standard Technical Processes." Whenever possible, embedded designs are used to field test new items so that test takers are unable to distinguish between the field-test items and operational items on each test form. STAAR Alternate 2 uses this design to embed a cluster of items on every form. This results in student performance data that are more stable.

Reliability

Reliability refers to the expectation that repeated administrations of the same test should generate consistent results. Reliability is a critical technical characteristic of any measurement instrument because unreliable scores cannot be interpreted as valid indicators of students' knowledge and skills.

During the 2014–2015 school year, reliability for the STAAR Alternate 2 test scores was estimated using statistical measures such as internal consistency, classical standard error of measurement, conditional standard error of measurement, and classification accuracy. Refer to chapter 3, "Standard Technical Processes," for detailed information about reliability.

Internal Consistency

Internal consistency is a measure of the consistency with which students respond to the items within a test. For STAAR Alternate 2, coefficient alpha was used to estimate reliability.

As a general rule, reliability coefficients ranging from 0.70 to 0.79 are considered adequate, those from 0.80 to 0.89 are considered good, and those at 0.90 or above are considered excellent. However, what is considered appropriate can vary in accordance with how assessment results are used.

For the STAAR Alternate 2 assessments administered in spring 2015, the internal consistency estimates ranged from 0.84 to 0.89. Internal consistency estimates across grades and content areas were found to be of a similarly high level, with no noticeable increases or decreases across grades or content areas. For the different student groups, estimates were found to be similar. For grade 8 mathematics, for example, the reliability for the total group was 0.86, for females only was 0.86, for males only was 0.87, for African Americans only was 0.88, for Hispanics only was 0.86, and for whites only was 0.86.

Because internal consistency estimates typically decrease as the number of test items decrease, internal consistency estimates for individual reporting categories can be noticeably lower than those for the full assessment. In spring 2015, the internal consistency estimates at the reporting category level were generally lower than at the total score level. Lower internal consistency estimates indicate that reporting category scores were not as reliable as those based on the full assessment. For example, the grade 5 mathematics reporting category “Numerical Representations and Relationships” contains four items. The estimated reliability for the scores in this reporting category was 0.53. Therefore, the lower reliability at the reporting category level should to be taken into account when making interpretations of the scores at this level.

Estimates of internal consistency at the overall level, as well as at the level of reporting categories for student groups for spring 2015 STAAR Alternate 2 assessments are provided in Appendix C.

Classical Standard Error of Measurement

Classical standard error of measurement (SEM) represents the amount of variance in a score that results from random factors other than what the assessment is intended to measure. The SEM is helpful for quantifying the margin of uncertainty that occurs on every test. For the STAAR Alternate 2 assessments administered in spring 2015, SEM values are generally between 1 to 3 raw score points. The SEM values are provided in Appendix C.

Conditional Standard Error of Measurement

It is important to note that the SEM index provides only an estimate of the average test score error for all students regardless of their individual levels of proficiency. By comparison, conditional standard error of measurement (CSEM) provides a reliability estimate at each score point on a test. More specifically, CSEM is an estimate of the average test score measurement error that is conditional on the proficiency or scale score estimate. For the 2014–2015 school year, CSEM values were approximately 13 to 17 scale score points in the middle of the scale score ranges. CSEM values for all spring 2015 STAAR Alternate 2 administrations are provided in Appendix C.

Classification Accuracy

Classification accuracy provides an estimate of the accuracy of student classifications into performance categories based on current test results. Classification accuracy rates for the STAAR Alternate 2 assessments during the 2014–2015 school year range from 79.7 to 86.3 percent. Classification accuracy rates for all spring 2015 STAAR Alternate 2 administrations are provided in Appendix C.

Validity

STAAR Alternate 2 scores are used to make inferences about student achievement. In support of these inferences, evidence is continually collected throughout the development and administration of STAAR Alternate 2 to demonstrate that the assessments measure the intended content. This validity evidence can be categorized as being based on test content, response processes, internal structure, relations to other variables, and the consequences of testing. This validity evidence supports multiple uses of test scores. Texas follows national standards of best practice to continue to build its body of validity evidence for all the STAAR assessments. The Texas Technical Advisory Committee (TTAC) provides ongoing input to TEA about STAAR Alternate 2 validity evidence. The following sections describe the validity evidence that has been collected for STAAR Alternate 2.

Evidence Based on Test Content

Validity evidence based on test content refers to evidence of the relationship between tested content and the construct the assessment is intended to measure. All STAAR assessments, including STAAR Alternate 2, have been designed to align with the content defined by the TEKS. The STAAR Alternate 2 test development process played an integral role in providing validity evidence based on test content for the assessment. The test development process and the evidence collected related to test content support the use of STAAR Alternate 2 scores in making inferences about students' knowledge and understanding of the TEKS.

Relationship to the Statewide Curriculum

At the inception of the STAAR Alternate 2 assessments, a steering committee was convened to review and provide feedback on the alignment of STAAR Alternate 2 tasks

to the TEKS. Educator reviews and focus group meetings continue to be a part of ongoing content development with revisions to the STAAR Alternate 2. Both focus groups and educator review meetings have occurred to review and provide feedback on alignment of items and content standards as well as to review and provide feedback on items themselves.

EDUCATOR INPUT

Professional judgments from educator review meetings provided additional content validity evidence. Educators from across the state reviewed the content of every item to validate that each item matched the appropriate content standard. The educator committees included special education experts, special education classroom teachers, teachers from the Texas School for the Blind and Visually Impaired and the Texas School for the Deaf, teachers of English language learners, and general education teachers.

As part of the review meetings, educators considered each item and were asked, "Does this item measure the reporting category, student expectation, essence statement, and prerequisite skills it was designed to measure?" To respond to this question, educators referenced resources such as the TEKS curriculum documents to verify the match of the reporting category, student expectation, essence statement, and prerequisite skills to each item. Across STAAR Alternate 2 items, educator review committees affirmed the relationship between the items and the TEKS. Additional committee input also confirmed that students are provided opportunities to learn the content before the assessment is administered.

Another important source of content validity is evidence related to bias. In order to be valid, an assessment must not only assess the intended content, but also be free of bias. To provide this validity evidence, educator committees were asked the following question regarding each assessment task: "Is this item free from bias on the basis of students' personal characteristics such as gender, ethnicity, or disability?" Committee members affirmed that STAAR Alternate 2 items are free from bias.

TEST DEVELOPER INPUT

Item writers and reviewers, who include content experts and special education experts, follow test development guidelines and item specifications that explain how the content of the assessed TEKS should be measured. At each stage of development, writers and reviewers verify the alignment of the test items with the assessed reporting categories.

Evidence Based on Response Processes

Response processes refer to the cognitive behaviors that are required to respond to a test item. Texas collects evidence to show that the way students respond to items on the STAAR Alternate 2 assessments reflects accurate measurement of the construct.

ITEM TYPES

Texas gathers theoretical and empirical evidence that supports the expectation that the way students respond to test items does not add construct-irrelevant variance. When new item types or changes to the format of existing item types are considered for STAAR Alternate 2 assessments, cognitive labs are used to study the way students engage with the various item presentations. After prototype items are determined to be appropriate for STAAR Alternate 2, evidence about student responses is gathered annually through educator and expert reviews and analyses of individual student responses to these items. Every year, during item reviews, educators evaluate whether the content for a given item is being appropriately assessed and whether students will be able to accurately demonstrate their knowledge of the construct given the items' planned format. When items are field-tested, additional data are gathered about students' responses. Data such as item difficulty, item-total correlations, and item fit are all evaluated. For additional information, see the Item Analyses section of chapter 3, "Standard Technical Processes."

SCORING PROCESS

The process used to score items can provide additional validity evidence based on response processes. This type of validity evidence is predicated on accurate scoring. Within the Test Administrator Booklet, test administrators are provided exact scoring rules and scripted instructions for how to present every item to a student. Test administrators are provided a number of resources to prepare for a STAAR Alternate 2 test administration including a period of time directly prior to the testing window in which they can preview the test booklet to prepare for a valid test administration.

Evidence Based on Internal Structure

Texas collects evidence that shows the relationship between items and reporting categories in order to verify that the elements of an assessment conform to the intended test construct. Texas conducts annual internal consistency studies to gather evidence based on internal structure. The internal consistency of the STAAR Alternate 2 assessments is evaluated every year using coefficient alpha for assessments that have only polytomously scored items. These internal consistency evaluations are made for all students and for student groups such as female, male, African American, Hispanic, and white students. Estimates of internal consistency are made for the full test as well as for each reporting category within a content area and can be found in Appendix C. The Reliability section of this chapter provides a summary of these estimates.

Evidence Based on Relationships to Other Variables

Another method Texas uses to provide validity evidence for the STAAR Alternate 2 assessments is analyzing the relationship between performance on STAAR Alternate 2 and performance on other assessments, a process that supports what is referred to as criterion-related validity. Evidence can be collected to show that the empirical relationships are consistent with the expected relationships. STAAR Alternate 2

correlation estimates, which evaluate the strength of the relationship (or the lack of one) between scores on the STAAR Alternate 2 assessments across different content areas (for example, grade 4 mathematics and grade 4 reading, or biology and U.S. history) were calculated. Results from all these analyses are provided in Appendix C. As expected, higher correlations were found between similar content areas like reading and writing than less similar content areas like mathematics and social studies.

Evidence Based on Consequences of Testing

Another way of providing validity evidence is by documenting the intended and unintended consequences of administering an assessment. Some of the intended consequences of the STAAR Alternate 2 assessment, based on the requirements in federal and state statutes, are listed below.

- Students with the most severe cognitive disabilities can receive challenging instruction that is linked to state content standards.
- Students with the most severe cognitive disabilities can be included in state assessment programs.
- STAAR Alternate 2 can assess the achievement of students with the most severe cognitive disabilities.
- Performance on STAAR Alternate 2 assessments can be used to track the academic progress of students across years.

Measures of Student Progress

Student progress measures are able to provide information beyond performance level by considering performance over time. Whereas performance level information describes students' current achievement, progress measures describe students' achievement across multiple years. Changes to the STAAR Alternate 2 assessment resulting in the second version administered in 2014–2015 prevented progress measures from being calculated in 2014–2015 because the progress measure requires a minimum of two years of data.

Sampling

Sampling occurs in two ways for STAAR Alternate 2. First, for the test administration, campus assignment of forms uses a sampling process wherein a single form is assigned to each campus in such a way that every form has a sample of students responding that are representative of the state demographic makeup. This approach is used in order to ensure that each campus administers the same form to all students and teachers only have to prepare to administer a single form.

Sampling also typically occurs for STAAR Alternate 2 when audits are completed. No audits were completed in 2014–2015; therefore, no sampling was necessary.

Test Results

Appendix C provides scale score distributions and summary statistics, raw score to scale score conversion tables, as well as mean p-values and reliability estimates for all STAAR Alternate 2 assessments administered in spring 2015. Table 5.6 shows the spring 2015 pass rates for the STAAR Alternate 2 assessments.

Table 5.6. STAAR Alternate 2 Spring 2015 Pass Rates

Subject Area	Grade/Course	Pass Rate
Mathematics	Grade 3	87%
	Grade 4	88%
	Grade 5	86%
	Grade 6	87%
	Grade 7	87%
	Grade 8	81%
	Algebra I	83%
Reading/English Language Arts	Grade 3	82%
	Grade 4	82%
	Grade 5	82%
	Grade 6	81%
	Grade 7	81%
	Grade 8	83%
	English I	86%
	English II	85%
Writing	Grade 4	80%
	Grade 7	79%
Science	Grade 5	90%
	Grade 8	92%
	Biology	89%
Social Studies	Grade 8	86%
	U.S. history	85%