Chapter 5 STAAR Modified

Overview

STAAR Modified is an alternate assessment based on modified academic achievement standards for students receiving special education services who meet participation requirements. STAAR Modified has been designed to meet federal requirements mandated under the NCLB. According to federal regulations, all students, including those receiving special education services, must be assessed on grade-level curriculum. STAAR Modified covers the same grade-level/course content as STAAR, but STAAR Modified assessments have been changed in format (larger font size, fewer items per page, etc.) and test design (shorter test blueprint, fewer answer choices, simpler vocabulary and sentence structure, etc.)

STAAR Modified is administered in English and includes assessments at grades 3–8 as well as 12 EOC assessments at the high school level. The initial administration of STAAR Modified occurred in spring 2012. However, not all of the STAAR Modified assessments were available at that time. House Bill (HB) 3 legislation called for the STAAR assessments to replace the TAKS assessments for students in grades 9 and below

starting with the 2011–2012 school year. As a result, the STAAR Modified assessments were available for students in grades 3–9 in 2012. However, students in grade 10 and above continued to take the TAKS–M assessments if appropriate. Consequently, the STAAR Modified EOC assessments for courses typically taken in grade 10 and above will be phased in over time. Refer to chapter 8, "Texas Assessment of Knowledge and Skills (TAKS), TAKS (Accommodated), TAKS–Modified (TAKS–M) and Linguistically Accommodated Testing (LAT)," for detailed information about the TAKS program.

The plan for implementing STAAR Modified is summarized below. Refer to the Test Administration section for additional information about the various types of the STAAR Modified administrations.

Spring 2012

- STAAR Modified special operational administrations: all 3–8 assessments; English I reading, English I writing, Algebra I, biology, world geography
- STAAR Modified stand-alone field tests: English II reading, English II writing, geometry

Spring 2013

- STAAR Modified operational administrations with embedded field-test items: 3–8 assessments; English I reading, English I writing, English II reading, English II writing, Algebra I, geometry, biology, world geography
- STAAR Modified special operational administration: world history

Spring 2014

- STAAR Modified operational administrations with embedded field-test items: 3–8 assessments; English I reading, English I writing, English II reading, English II writing, Algebra I, geometry, biology, world geography, world history
- STAAR Modified special operational administrations: English III reading, English III writing, U.S. history

Spring 2015

- STAAR Modified operational administrations with embedded field-test item: all 3–8 assessments; English I reading, English I writing, English II reading, English II writing, English III reading, English III writing, Algebra I, geometry, biology, world geography, world history, U.S. history

The STAAR Modified assessments administered in 2011–2012 are shown by grade and course in Table 5.1.

Table 5.1. 2011–2012 STAAR Modified Assessments

2011–2012 STAAR Modified Assessments	
Grade	**Assessed Content Area/Course**
Grade 3	mathematics and reading
Grade 4	writing, mathematics, and reading
Grade 5	mathematics, reading, and science
Grade 6	mathematics and reading
Grade 7	writing, mathematics, and reading
Grade 8	mathematics, reading, science, and social studies
High School (End-of-Course)	English I reading, English I writing, English II reading, English II writing, Algebra I, geometry, biology, world geography

Participation Requirements

Federal and state laws require the ARD committee for each student to make decisions about a student's placement in the Texas assessment program. The first consideration is the general assessment, STAAR. If the general assessment is not appropriate, the ARD committee then considers the participation guidelines for STAAR Modified. The decision to administer STAAR Modified is neither based solely on disability category or placement setting, nor is it determined administratively; rather, the decision lies with the ARD committee based on the participation requirements.

Students receiving special education services are eligible to take STAAR Modified in one of more content areas if the answer to all three eligibility questions below is "Yes."

1. Do the student's present level of academic achievement and functional performance (PLAAFP) statements in the Individualized Education Program (IEP) lead the ARD committee to conclude that the student is multiple years behind grade-level or course expectations and will not progress at the same rate and level of rigor as his or her non-disabled peers?
2. Does the student's IEP contain standards-based (TEKS-based) goals indicating the modified content the student requires in order to access the grade-level or course curriculum?
3. Does the student require direct and intensive instruction in order to acquire, maintain, and transfer skills to other contexts?

Refer to the STAAR Modified Participation Requirements page on TEA's Student Assessment Division website for more information. Any student who meets participation requirements for STAAR Modified may take the STAAR Modified assessments. However, only 2 percent of the tested population can count as proficient for the purpose of Adequate Yearly Progress (AYP) performance calculations.

Testing Requirements for Graduation

Students taking the STAAR Modified EOC assessments are on the MHSP because they are receiving modified instruction. For those students eligible to take STAAR Modified EOC, passing the assessments is not a requirement for graduation. A student's ARD committee determines individual requirements for graduation.

Test Development

The test development process for STAAR Modified follows, as closely as possible, the procedures used for development of other assessment programs in Texas, though coupled with the additional requirements specific to STAAR Modified.

The STAAR Modified blueprints mirror the STAAR blueprints and reflect that the students taking STAAR Modified are assessed on the same grade-level curriculum as general education students. The number of items on the STAAR Modified blueprints was reduced from the number of items on the STAAR blueprints by 20 percent. The STAAR Modified blueprints contain similar percentages of readiness and supporting items as are found in the STAAR blueprints.

Issues of validity, reliability, fairness, accessibility, and consistency in meaning are carefully considered as part of the item modification and review processes. As STAAR Modified items are developed and reviewed, attention is given to the standards of fairness and the principles of alignment and universal design. According to the principles of universal design, every item must have precisely defined constructs and demonstrate maximum legibility, maximum readability, and maximum comprehensibility. Similarly, each item must be adaptable for purposes of accommodations, be accessible and non-biased, and take special populations into consideration.

Using results from a literature review of modifications that would be appropriate for students with disabilities, TEA modified existing STAAR items. Modification guidelines were developed for each subject/content area to ensure that the modifications did not affect the construct of the items, and the item modifications would be consistent across development years.

In addition to strictly adhering to these modification guidelines, TEA convenes internal and external educator item review meetings to further ensure that the modified items meet the criteria listed above. Texas educators—general and special education classroom teachers, curriculum specialists, administrators, and regional ESC staff—play a vital role in the STAAR Modified development process. The participation of these education professionals enables TEA to develop high-quality alternate assessment instruments that accurately reflect the TEKS curriculum.

Training

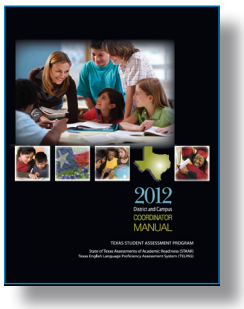
The test administration training for STAAR Modified is the same as the training available for STAAR and can be found in the Training section of chapter 4, "State of Texas Assessments of Academic Readiness (STAAR)." For each test administration in the 2011–2012 school year, ESC personnel and district coordinators were given a district testing coordinator packet that contained all the information and materials necessary for overseeing test administrations, including copies of the coordinator and test administrator manuals. Separate packets and manuals were provided for STAAR Modified grades 3–5, grades 6–8, and EOC.

Training opportunities were conducted via the Texas Education Telecommunication Network (TETN), 2012 ESC Training, and the 2012 Texas Assessment Conference.

The presentations addressed these topics:

- accommodation policies and procedures
- required documentation for assessment and accommodation decisions
- modification guidelines and processes
- information regarding making assessment decisions in ARD committee meetings

District and Campus Coordinator Manual



The *2012 District and Campus Coordinator Manual (DCCM)* explains the responsibilities of district and campus testing coordinators for the STAAR program. This manual contains preparation and administration procedures for every program for the 2012 calendar year. Separate test administrator manuals are distributed to districts prior to the first assessment administration for each grade.

Test Administrations

In spring 2012 the STAAR Modified assessments had two types of administrations: stand-alone field-test administrations and special operational administrations.

A stand-alone field test administration is one in which all items included on the test have not been previously field-tested. Students do not receive a test score based on their performance on a stand-alone field test. The purpose is to collect statistical information about items that will help evaluate whether the items should be eligible to appear on future operational assessments.

A special operational administration is one in which all items included on the test have not been previously field-tested. However, students do receive a test score based on their performance on the assessment. To give students their scores on a special operational administration, items are analyzed immediately after the administration, and item statistics are evaluated to determine which items are appropriate to contribute to a student's score. These items are also deemed suitable for use on future operational assessments.

The use of stand-alone field tests and special operational administrations was necessary because spring 2012 was the first time that STAAR Modified was administered. In subsequent administrations of the STAAR Modified assessments, the operational administration with embedded field-test design will be implemented. In this design, the operational items are common across all test forms and count toward an individual student's score, but each field-test item appears on only a small number of test forms (typically one form) and does not count toward students' scores. The embedded field-test design is preferable because test takers are unable to distinguish between the field-test items and the operational items on each test form, which results in student responses that are more reliable.

Refer to the Overview section for information about which type of administration was implemented for each STAAR Modified assessment in spring 2012 as well as the implementation plan for future STAAR Modified administrations. More than 260,000 STAAR Modified assessments were administered in 2011–2012. The number of STAAR Modified assessments administered is shown in Table 5.2.

Table 5.2. STAAR Modified Assessments Administered in 2011–2012

STAAR Modified Assessments Administered in 2011–2012	
STAAR Modified Assessment	**Assessments Administered**
Grade 3 reading	10,680
Grade 3 mathematics	9,322
Grade 4 reading	13,162
Grade 4 mathematics	12,101
Grade 4 writing	13,629
Grade 5 reading	14,084
Grade 5 mathematics	13,350
Grade 5 science	11,763
Grade 6 reading	13,265
Grade 6 mathematics	13,388
Grade 7 reading	12,604
Grade 7 mathematics	13,062
Grade 7 writing	13,082
Grade 8 reading	11,938
Grade 8 mathematics	12,721
Grade 8 social studies	10,660
Grade 8 science	11,146
English I reading	10,913
English I writing	10,893
English II reading*	7,235
English II writing*	7,234
Algebra I	11,713
Geometry*	6,629
Biology	8,931
World geography	9,161

*Stand-alone field-test administration

Educational Materials Required for Testing

Dictionaries And Thesauruses

English-language dictionaries and thesauruses must be provided to students for the writing assessment at grade 7, the reading assessments at grades 6–8, and the English I and II reading and writing assessments.

There must be at least one dictionary for every five students; it is also recommended for there be one thesaurus for every five students, if possible. Students may also use a combination dictionary/thesaurus. An ESL dictionary that uses simple English and

pictures to define words may be provided for ELLs. Both paper and electronic dictionaries are permitted, though electronic dictionaries must not allow access to the Internet.

Specific information regarding dictionary policies for each STAAR assessment, including STAAR Modified, can be found on the STAAR Resources page on TEA's Student Assessment Division website.

Calculators

Calculators must be provided to students for the STAAR Modified Algebra I, geometry, and biology assessments. Students may use their own calculators instead of those provided by the district. Districts may provide students with more than one calculator during the assessment but must provide, at a minimum, the following:

- a graphing calculator for each student taking Algebra I and geometry
- a calculator for every five students taking biology

Any calculator may be used to fulfill these minimum requirements except for those that include a computer algebra system (CAS) or that allow access to the Internet. In addition, an electronic device that has a calculator as an application cannot be used (e.g., a cell phone or smartphone).

All calculator memory must be cleared to factory default both before and after testing. Any programs or applications must be removed or disabled prior to testing.

Calculators may not be provided to a student taking the grades 3–8 mathematics assessments or the grades 5 and 8 science assessments unless the student meets the criteria for such an accommodation.

Specific information regarding calculator policies for each STAAR assessment can be found on the STAAR Resources page on TEA's Student Assessment Division website.

Testing Accommodations

Accommodations are practices and procedures that provide equitable access to grade-level or course curriculum during instruction and assessment. The decision to use a particular accommodation with a student is made on an individual basis and takes into consideration both the needs of the student and whether the student routinely receives the accommodation in classroom instruction and testing. Further information on testing accommodations can be found in chapter 4, "State of Texas Assessments of Academic Readiness (STAAR)." Specific information about each accommodation can be found on the Accommodations Resource page on TEA's Student Assessment Division website.

Accommodations for Students with Disabilities

For a student who receives special education or Section 504 services, the decision to allow the student to use accommodations during the statewide assessments is made by the student's ARD committee or Section 504 placement committee. In those rare instances where a student does not receive special education services but does meet the eligibility criteria due to a disabling condition, the decision to allow accommodations on the statewide assessments is made by the appropriate team of people at the campus level, such as the RtI team or student assistance team. Further information on the types of accommodations available for students with disabilities can be found in chapter 4, "State of Texas Assessments of Academic Readiness (STAAR)."

After determining the instructional accommodation(s) that are effective for a student with disabilities, the educator should investigate whether each accommodation is allowed on a statewide assessment.

Dyslexia Accommodations

Accommodations are available for students who have dyslexia and other reading disabilities on the STAAR Modified reading assessments in grades 3 through high school. These accommodations include:

- having all items and answer choices read aloud to a student, and
- extending the four-hour testing time over an entire school day.

The needs of the student should be carefully considered when determining the most appropriate accommodations for the STAAR Modified assessments.

Oral Administration

Oral administration is an accommodation that allows for test items and answer choices for reading, mathematics, science, and/or social studies tests to be read aloud or signed to eligible students taking the STAAR Modified assessments.

Students with Visual Impairments

Test administrators receive specific instructions for testing visually impaired students using large-print or braille test booklets. Districts are required to indicate on the answer document whether the student used a large-print or braille version of a test. Large-print test booklets and braille test booklets are available for all STAAR Modified operational administrations. However, large-print and braille test booklets are not available for field-test administrations.

Linguistic Accommodations

Linguistic accommodations are language supports that make grade level academic assessments in English more accessible to ELLs. The Texas English Language Proficiency Standards (ELPS) require all teachers to linguistically accommodate the

instruction of ELLs in their classes in a manner that is commensurate with the students' English language proficiency levels. The policies for the STAAR linguistic accommodations support these ELPS requirements.

More information about the STAAR linguistic accommodations can be found in chapter 4, "State of Texas Assessments of Academic Readiness (STAAR)" and is available in the document titled Linguistic Accommodations for ELLs Participating in the STAAR Program on the Accommodations Resources page on TEA's Student Assessment Division website.

Additional information regarding accommodations, including eligibility requirements, is available on the Accommodation Resources page on TEA's Student Assessment Division website.

Student Success Initiative

The SSI provides a system of academic support to help students achieve on grade level in reading and mathematics. The SSI incorporates a grade-advancement component adopted by the Texas Legislature in 1999. Because there were no performance standards in place for grades 3–8 in the 2011–2012 school year, SSI requirements could not be applied. Therefore, SSI retest opportunities were not offered in May or June of 2012.

For students who take the STAAR Modified assessments and do not achieve at least Level II: Satisfactory Academic Performance, the ARD committee determines if students must retest. The ARD committee also makes decisions regarding promotion or retention.

Scores and Reports

There are a variety of reports that show a student's performance on the assessments in the STAAR program. Refer to the information below for details about the types of scores given on reports and the types of reports available.

Description of Scores

Scores for the STAAR Modified assessments consist of the number of items answered correctly (raw scores), scale scores, and reporting category performance information.

Raw Score

The number of items that a student answers correctly on a STAAR Modified assessment is the student's raw score. The raw score can be interpreted only in terms of a specific set of test items on that test form. However, because the difficulty of items might vary among test forms over time, differences in

student performance across tests or administrations cannot be compared using raw scores alone. To compare student scores across different test forms and different administrations, 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: Advanced Academic Performance. (Performance-level cut scores are discussed in the Performance Standards section of this chapter.) Performance standards for the STAAR Modified assessments were not set until fall 2012; therefore, only raw scores were available after the spring 2012 administrations. Scale scores for all STAAR Modified assessments were made available in January 2013 after standards were applied.

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 in a gifted and talented program can be compared with the average scores of the other Hispanic students, the other gifted and talented students, all the students on a campus, or any combination of these aggregations at that grade.

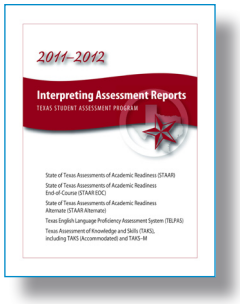
Additional Performance Information

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. Finally, individual student test scores are also used in conjunction with other performance indicators to assist in making placement decisions. While scores can contribute to decisions regarding placement, educational planning for a student should take into account as much of student information as possible.

Report Formats

Two types of reports are provided for the various testing programs: standard and optional. Standard reports are provided automatically to districts. Information contained in 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.

Standard and optional reports were provided in January 2013 for all STAAR Modified grades and courses that had operational assessments. The reports became available once performance standards were established for these assessments.



For more information about scoring and reporting for STAAR Modified, refer to the TEA publication *Interpreting Assessment Reports*.

Use of Test Results

Test results can be used to evaluate the performance of a group over time. Average scale scores and the percentage of students meeting the Level II and Level III standards can be analyzed by grade and content area across administrations to give insight into whether student performance is improving across years. For example, the average scale score for students who receive special education services and who took the STAAR Modified grade 4 writing test can be compared over time.

Test results can also be used to compare the performance of different demographic or program groups. The STAAR Modified scores can be analyzed within the same content area of any single administration to determine which demographic or program group had the highest average scale score, or which group had the lowest percentage meeting the Level II standard, or which group had the highest percentage achieving Level III performance, etc. Other scores can be used to help evaluate the academic performances of demographic or program groups in core academic areas. For example, reporting category data can help campuses and districts identify areas of potential academic weakness for a group of students. The same methodology can be applied to an entire campus or district. Test results for groups of students can be used when evaluating instruction or programs that require average-score or year-to-year comparisons. Because the tests are designed to measure content areas within the required state curriculum, the consideration of test results by content area and by reporting category might be helpful when evaluating curriculum and instruction. In addition, all test scores can be compared with regional and statewide performance within the same content area for any administration.

Generalizations from test results can be made to the specific content domain being measured on the test. However, because each test measures a finite set of skills with a limited set of items, any generalizations about student achievement derived solely from a particular assessment should be made with great care and with full reference to the fact that the conclusions were based only on that assessment. Instruction and program evaluations should take into account as much information as possible to provide a more complete picture of performance.

Parent Brochures

TEA's Student Assessment Division produces a brochure titled "Understanding the Confidential Student Report—A Guide for Parents" to help parents understand their child's STAAR 3–8 test results. This brochure provides a brief summary of the STAAR program, including STAAR Modified, and explains information contained on a CSR so that parents can understand their child's test report. The brochure, available in both English and Spanish, was provided to districts in January 2013 for distribution with individual student STAAR Modified performance results. For STAAR Modified EOC, an explanation of the test results is printed on the CSR for each individual assessment.

Performance Standards

Performance standards relate levels of test performance directly to what students are expected to learn as defined in the statewide curriculum. This is done by establishing cut scores that distinguish between performance levels or categories. Standard setting is the process of establishing these cut scores that define the performance levels for an assessment.

Performance Levels and Policy Definitions

For the STAAR Modified assessments, the performance levels are:

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

More detailed descriptions, known as policy definitions, of each performance level are as follows:

Level III: Advanced Academic Performance*

Performance in this category indicates that students are well prepared for the next grade or course with instructional modifications such as simplified language and concepts. They demonstrate the ability to understand and apply the assessed knowledge and skills. Students in this category have a high likelihood of success in the next grade or course with continued modifications and support.

* For English III, this level of performance also indicates that students are prepared for postsecondary endeavors with support.

Level II: Satisfactory Academic Performance

Performance in this category indicates that students are sufficiently prepared for the next grade or course with instructional modifications such as simplified language and concepts. They generally demonstrate the ability to understand and apply the assessed knowledge and skills in familiar contexts. Students in this category have a reasonable likelihood of success in the next grade or course with continued modifications and support.

Level I: Unsatisfactory Academic Performance

Performance in this category indicates that students are inadequately prepared for the next grade or course even with instructional modifications such as simplified language and concepts. They demonstrate an insufficient understanding of the assessed knowledge and skills. Students in this category are unlikely to succeed in the next grade or course without significant and/or additional modifications and increased support.

Standard-Setting Process for STAAR Modified

Standard setting for STAAR Modified took into consideration a variety of factors such as policy, TEKS content standards, educator knowledge about what students should know and be able to do, and information about how student performance on statewide assessments aligns with performance on other assessments.

TEA used an evidence-based standard-setting approach (O'Malley, Keng, & Miles, 2012) for the STAAR program. Using this approach, TEA defined and implemented a nine-step process to establish performance standards for STAAR Modified. The nine steps are:

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

Tables 5.3 and 5.4 provide high-level descriptions and timelines for these nine steps as implemented in the STAAR Modified EOC and STAAR Modified 3–8 standard-setting processes, respectively.

Additional detail about each step in the STAAR standard-setting process is given in the STAAR Modified Standard Setting Technical Report, available on the STAAR Resources page on TEA's Student Assessment Division website.

Table 5.3. Overview of the STAAR Modified EOC Standard-Setting Process

Overview of the STAAR Modified EOC Standard-Setting Process		
Standard-Setting Step	**Description**	**Timeline**
1. Conduct empirical studies	Scores on each assessment are linked to performance on other assessments in the same content area (when available).	Studies started in spring 2012 and will continue throughout the program.
2. Develop performance labels and policy definitions	A committee was convened jointly by TEA and the THECB to recommend performance categories, performance category labels, and general policy definitions for each performance category for the STAAR program. This information was adapted to apply to the STAAR Modified program.	September 2010
3. Convene a policy committee	Committee considers policy implications of performance standards and empirical study results and makes recommendations to identify reasonable ranges ("neighborhoods") for the STAAR EOC cut scores . This information is used to inform the reasonable ranges for the STAAR Modified cut scores.	February 1–2, 2012
4. Develop grade-/course-specific performance level descriptors (PLDs)	Committees consisting primarily of educators started with the general STAAR PLDs to develop STAAR Modified PLDs as an aligned system, describing a reasonable progression of skills within each content area (English, mathematics, science, and social studies).	June 2012
5. Convene standard-setting committees	Committees consisting of K–12 educators used the performance labels, policy definitions, PLDs, and reasonable ranges to recommend cut scores for each STAAR Modified EOC assessment.	August 8–14, 2012
6. Review performance standards for reasonableness	TEA reviews the recommendations across content areas.	September 2012
7. Approve performance standards	The commissioner of education approves the performance standards.	December 2012
8. Implement performance standards	Performance standards are reported to students for the spring 2012 administration with phase-in standards applied.	January 2013
9. Review performance standards	Performance standards are reviewed at least once every three years.	Fall 2014

Table 5.4. Overview of the STAAR Modified 3–8 Standard Setting Process

Overview of the STAAR Modified 3–8 Standard Setting Process		
Standard-Setting Step	**Description**	**Timeline**
1. Conduct empirical studies	Scores on each assessment are linked to performance on other assessments in the same content area (when available).	Studies started in spring 2012 and will continue throughout the program.
2. Develop performance labels and policy definitions	Committee is convened jointly by TEA and the THECB to recommend performance categories, performance category labels, and general policy definitions for each performance category for the STAAR program. This information was adapted to apply to the STAAR Modified program.	September 2010
3. Develop performance standard ranges	STAAR Modified EOC performance standards and empirical study results are used to identify reasonable ranges ("neighborhoods") for the cut scores for Levels II and III.	August 2012
4. Develop grade/subject specific performance level descriptors (PLDs)	Committees consisting primarily of educators started with the general STAAR PLDs to develop STAAR Modified PLDs as an aligned system, describing a reasonable progression of skills within each content area (English, mathematics, science, and social studies).	September 2012
5. Convene standard-setting committees	Committees consisting of K–12 educators used the performance labels, policy definitions, PLDs, and reasonable ranges to recommend cut scores for each STAAR Modified assessment.	October 29–November 9, 2012
6. Review performance standards for reasonableness	TEA reviews the cut-score recommendations across grades and subjects.	November 2012
7. Approve performance standards	The commissioner of education approves performance standards.	December 2012
8. Implement performance standards	Performance standards are reported to students for the spring 2012 administration with phase-in standards applied.	January 2013
9. Review performance standards	Performance standards are reviewed at least once every three years.	Fall 2014

Standard-Setting Committees

The task of each standard-setting committee was to recommend the two cut scores that would define the three performance levels for each of the STAAR Modified assessments. The standard-setting committees were made up of K–12 educators. Each committee included general education teachers who were experts in both the assessed content and the curriculum and special education teachers having expertise with students who take the STAAR Modified assessments.

Committee members were provided with reasonable ranges within which performance standards should be set. The ranges were determined after a careful consideration of the alignment of performance standards with the STAAR Modified EOC and 3–8 assessments in the same content area, the relevant information from the policy committee, and the results of various empirical studies. (The studies provided research-based anchors for setting performance standards that were meaningful and rigorous.)

In August 2012, TEA convened standard-setting committees that recommended performance standards for the STAAR Modified Algebra I, geometry, biology, English I reading, English II reading, English I writing, English II writing, and world geography assessments.

In October and November 2012, TEA convened standard-setting committees that recommended performance standards for STAAR Modified grades 3–8 mathematics, grades 3–8 reading, grades 5 and 8 science, grade 8 social studies, and grades 4 and 7 writing.

Standard-setting committees for STAAR Modified world history, U.S. history, English III reading, and English III writing will be convened after each assessment has been administered for the first time.

Phase-in of Performance Standards

A phase-in period has been implemented for the STAAR Modified performance standards in order to provide school districts with sufficient time to adjust instruction, to provide new professional development, to increase teacher effectiveness, and to close knowledge gaps. A four-year, two-step phase-in for Level II: Satisfactory Academic Performance is in place for all STAAR Modified 3–8 and EOC assessments. There is no phase-in for the Level III performance standard for any of the STAAR Modified assessments.

The STAAR Modified EOC phase-in periods for performance standards are on a student-by-student basis by content area (mathematics, English, science, and social studies). The phase-in standard to which students are held depends on when the students begin testing in a content area, and it applies to all assessments in that content area. For example, for students who first took Algebra I in spring 2012, the first phase-in standard for Level II: Satisfactory Performance applies to both mathematics assessments (Algebra I and geometry).

The STAAR Modified 3–8 phase-in standards for Level II performance began with the 2012 test administration. Phase-in 1 standards for Level II will be in effect for the 2011–2012 and 2012–2013 school years, and phase-in 2 standards will be in effect for the 2013–2014 and 2014–2015 school years. The final recommended Level II standards will be in place for the STAAR Modified 3–8 assessments beginning in the 2015–2016 school year. There is no phase-in for the Level III performance standards for STAAR Modified 3–8.

Figure 5.1 illustrates how the phase-in of standards applies to several cohorts of students who have taken STAAR Modified or will be taking STAAR Modified in the English/reading content area. Phase-in performance standards apply to the STAAR Modified assessments shown in bold. Final recommended performance standards apply to the STAAR assessments not shown in **bold**. The vertical-dashed blue and green lines mark the beginning of the first and second phase-in periods, respectively, for Level II. The vertical-dashed orange line signals the implementation of the final recommended performance standards for Level II.

Figure 5.1. Phase-in of STAAR Modified Performance Standards Across

Level II* Phase-in for All STAAR Modified Assessments							
Cohort	2011–2012	2012–2013	2013–2014	2014–2015	2015–2016	2016–2017	2017–2018
1	Grade 9 English I	Grade 10 English II	Grade 11 English III				
2	Grade 8 Reading	Grade 9 English I	Grade 10 English II	Grade 11 English III			
3	Grade 7 Reading	Grade 8 Reading	Grade 9 English I	Grade 10 English II	Grade 11 English III		
4	Grade 6 Reading	Grade 7 Reading	Grade 8 Reading	Grade 9 English I	Grade 10 English II	Grade 11 English III	
5	Grade 5 Reading	Grade 6 Reading	Grade 7 Reading	Grade 8 Reading	Grade 9 English I	Grade 10 English II	Grade 11 English III
6	Grade 4 Reading	Grade 5 Reading	Grade 6 Reading	Grade 7 Reading	Grade 8 Reading	Grade 9 English I	Grade 10 English II
7	Grade 3 Reading	Grade 4 Reading	Grade 5 Reading	Grade 6 Reading	Grade 7 Reading	Grade 8 Reading	Grade 9 English I

*The Level II phase-in example used above will be applied to all STAAR Modified assessments.

Outcome of Standard Setting

The standard-setting process elicited recommended cut scores that reflect the level of performance a student must achieve for each performance category of the STAAR Modified assessments. Performance standards for all STAAR Modified assessments were recommended by standard-setting committees and approved by the commissioner of education in December 2012. Tables 5.5 and 5.6 show the approved performance standards in scale score units for the STAAR Modified 3–8 and EOC assessments, respectively.

Table 5.5. STAAR Modified 3–8 Performance Standards

STAAR Modified 3–8 Performance Standards				
Assessment	**Phase-in 1 Level II**	**Phase-in 2 Level II**	**Final Recommended Level II**	**Final Recommended Level III**
Grade 3 Mathematics	2800	2900	3000	3578
Grade 4 Mathematics	2800	2900	3000	3526
Grade 5 Mathematics	2800	2900	3000	3691
Grade 6 Mathematics	2800	2900	3000	3462
Grade 7 Mathematics	2800	2900	3000	3551
Grade 8 Mathematics	2800	2900	3000	3577
Grade 3 Reading	2800	2900	3000	3306
Grade 4 Reading	2800	2900	3000	3238
Grade 5 Reading	2800	2900	3000	3312
Grade 6 Reading	2800	2900	3000	3316
Grade 7 Reading	2800	2900	3000	3368
Grade 8 Reading	2800	2900	3000	3436
Grade 4 Writing	2800	2900	3000	3349
Grade 7 Writing	2800	2900	3000	3422
Grade 5 Science	2800	2900	3000	3234
Grade 8 science	2800	2900	3000	3509
Grade 8 Social Studies	2800	2900	3000	3348

Table 5.6. STAAR Modified EOC Performance Standards

STAAR Modified EOC Performance Standards				
Assessment	**Phase-in 1 Level II**	**Phase-in 2 Level II**	**Final Recommended Level II**	**Final Recommended Level III**
English I Reading	1430	1450	1500	1672
English II Reading	1430	1450	1500	1652
English I Writing	1430	1450	1500	1612
English II Writing	1430	1450	1500	1604
Algebra I	2800	2900	3000	3470
Geometry	2800	2900	3000	3743
Biology	2800	2900	3000	3500
World Geography	2800	2900	3000	3354

Review of Performance Standards

State statute TEC §39.0242 requires performance standards for the STAAR program to be reviewed at least once every three years. To maintain compliance with state statute, the performance standards will need to be reviewed by 2015. The current plan is to review the STAAR Modified performance standards in fall 2014.

Scaling

Scaling is a statistical procedure that places raw scores on a common scoring metric in order to make test scores easier to interpret and easier to compare across test administrations. As with previous Texas assessment programs, the STAAR program uses the RPCM to place test items on the same scale across administrations for a given STAAR Modified assessment. Once performance standards have been set for an assessment, its Rasch scale is then transformed to the more user-friendly metric of a reporting scale in order to further 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 Modified 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 all STAAR Modified mathematics, science, and social studies assessments, and STAAR Modified grades 4 and 7 writing assessments, a scale score of 3000 represents the final recommended Level II performance standard. In addition, the standard deviation for those scales was set at 200.

For all STAAR EOC English assessments, a scale score of 1500 represents the final recommended Level II performance standard. The standard deviation for the STAAR EOC English scales was set at 100.

It is important to note that although Level II scale score values are fixed across horizontally scaled assessments within content areas, Level III scale score values vary across all STAAR assessments. However, these Level III scale score values will stay constant over time (not accounting for the phase in of standards).

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

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

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 for that year's set of test items.

Values for the horizontal scaling constants are provided in Tables 5.7 and 5.8 for the STAAR Modified 3–8 and EOC assessments, respectively.

Table 5.7. Horizontal Scaling Constants for STAAR Modified 3–8

Horizontal Scaling Constants for STAAR Modified 3–8		
STAAR Assessment	**A**	**B**
Grade 3 reading	244.2943	2877.1200
Grade 3 mathematics	317.5113	2864.1052
Grade 4 reading	232.5136	2797.9457
Grade 4 mathematics	304.9280	2914.9251
Grade 4 writing	386.5413	2919.2129
Grade 5 reading	275.1252	2862.7125
Grade 5 mathematics	389.5123	2988.7041
Grade 5 science	260.2186	2679.6709
Grade 6 reading	281.7367	2852.3700
Grade 6 mathematics	446.3956	3011.6063
Grade 7 reading	278.1452	2880.1194
Grade 7 mathematics	459.9354	3040.9343
Grade 7 writing	408.0328	2862.4930
Grade 8 reading	253.5742	2857.9984
Grade 8 mathematics	457.1662	3030.1730
Grade 8 science	328.2878	2817.4720
Grade 8 social studies	301.6569	2853.0931

Table 5.8. Horizontal Scaling Constants for STAAR EOC

Horizontal Scaling Constants for STAAR Modified EOC		
STAAR Assessment*	**A**	**B**
English I Reading	156.6402	1485.9024
English II Reading	151.3773	1478.0503
English I Writing	126.1548	1411.6916
English II Writing	150.1727	1450.1427
Algebra I	405.0300	2972.8630
Geometry	493.9173	3069.6423
Biology	341.6655	2847.2755
World Geography	340.4011	2946.5570

* Scaling constants for English III reading, English III writing, world history, and U.S. history will be determined once performance standards are set for these assessments.

Equating

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 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 2011–2012 school year, the STAAR Modified equating activities included post-equating and field-test equating. Refer to chapter 3, "Standard Technical Processes," for detailed information about equating.

Post-Equating

The post-equating process uses data from an operational test administration to estimate item difficulties and place them onto the scale of the item bank. Because spring 2012 was the initial operational administration of the STAAR Modified assessments, the scale of the item bank was set through the equating process for each assessment.

In the 2011–2012 school year, post-equating for STAAR Modified assessments used a conventional common-item non-equivalent groups design with concurrent calibration to place all items across the multiple test forms within a grade and content area onto a common scale. This method of calibration allows all items within a grade and content area to be calibrated at the same time, thus obtaining only one Rasch difficulty value for each item. This differs from form-by-form calibration, which is a calibration method that can produce multiple Rasch difficulty values for an item if that item appears on multiple test forms. Because all STAAR Modified items within a grade and content area were concurrently calibrated, all items within that grade and content area were placed on the same underlying Rasch scale. In addition, by conducting one single concurrent calibration run rather than multiple calibration runs, no form-by-form equating constant was computed. The scale resulting from each concurrent calibration is the scale of the item bank for the given STAAR Modified assessment.

During the 2011–2012 school year, post-equating was conducted for all STAAR Modified assessments that were given as special operational administrations. A special STAAR Modified operational administration is one in which each item on the test form has not been previously field-tested. Unlike a stand-alone field test, students do receive a test score based on their performance on the assessment. To provide student scores on the special operational administration, statistical characteristics of all the items were closely examined immediately after they were equated. Only items with sound psychometric properties were used to generate the RSSS conversion table for each STAAR Modified test form.

Field-Test Equating

To replenish the item bank as new tests are developed each year, items must be field-tested and equated to the scale of the item bank. Because spring 2012 was the initial operational administration of the STAAR Modified assessments, the scale of the item bank was set through the equating process for each assessment. The concurrent calibration method described in the previous

section was used to place all field-test items onto the same scale for all STAAR Modified assessments administered in spring 2012, including stand-alone field tests and special operational administrations.

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 the operational items on each test form. This results in student performance data that are more reliable. In subsequent operational administrations of the STAAR Modified assessments, the embedded field-test design will be used to collect data on field-test items. The equating procedure described in the technical details and procedures in chapter 3, "Standard Technical Processes," for embedded field tests will be used to put field-test items on the item bank scale for each STAAR Modified assessment.

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 instruments cannot be used to make valid interpretations.

During the 2011–2012 school year, the reliability of the STAAR Modified scores was estimated using statistical methods 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

For the STAAR Modified assessments administered as operational tests in spring 2012, the internal consistency estimates ranged from 0.55 to 0.86. 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 reading form 1, for example, the reliability for the total group was 0.81, for female only was 0.79, for male only was 0.82, for African American only was 0.82, for Hispanic only was 0.78, and for white only was 0.83.

Because internal consistency estimates typically decrease as the number of test items decrease, internal consistency estimates made at the reporting category level can be noticeably different from those made at the level of the full test. In spring 2012, the internal consistency estimates at the reporting category level were generally lower than at the total score level, indicating that, as expected, interpretations of student reporting category scores are not as reliable as those based on the complete test. For example, the STAAR Modified grade 5 mathematics reporting category "Numbers, Operations, and Quantitative Reasoning" contains 14 items. The estimated reliability for

the scores in this reporting category on form 2 was 0.62. Therefore, the lower reliability at the reporting category level should be taken into account when making interpretations of the scores at this level.

Estimates of internal consistency at the overall level as well as by reporting categories and for student groups for all spring 2012 STAAR Modified assessments are provided in Appendix C.

Classical Standard Error of Measurement (SEM)

For the STAAR Modified assessments administered operationally in spring 2012, SEM values range from approximately 2 to 4 raw score points across grades and content areas. The SEM values for the operational STAAR assessments administered in spring 2012 are provided in Appendix C.

Conditional Standard Error of Measurement (CSEM)

It is important to note that the SEM index provides an estimate of the average test score error for all students regardless of their individual levels of proficiency. By comparison, the CSEM provides a reliability estimate at each score point on a test. More specifically, CSEM is an estimate of the average measurement error that is conditional on the proficiency or scale score estimate. CSEM values for all operational spring 2012 STAAR administrations are provided in Appendix C.

Classification Accuracy

Classification accuracy provides an estimate of the accuracy of student classifications into performance categories based on assessment results. Classification accuracy rates for all spring 2012 operational STAAR Modified administrations are provided in Appendix C.

Validity

The results of the STAAR Modified assessments are used to guide educational planning related to the knowledge and skills that students are acquiring in each academic content area. Texas collects validity evidence annually to support the many uses of the STAAR Modified scores. Texas follows national standards of best practice to continue to build its body of validity evidence for the STAAR Modified assessments. TEA also receives ongoing input from the Texas Technical Advisory Committee with regard to plans for collecting validity evidence for the Texas assessment program. The sections that follow describe how different types of validity evidence were collected for the STAAR Modified assessments during the 2011–2012 school year. Refer to chapter 3, "Standard Technical Processes," for more detailed information about validity.

Evidence Based on Test Content

The STAAR Modified assessments have been developed to align with content as defined by the TEKS. Content validity evidence is collected at all stages of the test-development process. Nationally established test-development processes for the Texas assessment program are followed while developing the STAAR Modified assessments in order to support the use of the STAAR Modified scores in making inferences about students' knowledge and understanding of the TEKS.

Because STAAR Modified is a version of STAAR, the test development processes for both assessments play an intricate role in building validity evidence. To achieve the highest level of content validity, the process of aligning both STAAR and STAAR Modified to the curriculum included review by numerous committees of Texas educators.

Relationship To The Statewide Curriculum

As part of the transition to a high-stakes assessment program in 2012, teachers, curriculum specialists, test-development specialists, college educators, and TEA staff worked together in advisory committees to identify appropriate assessment reporting categories for STAAR, including the STAAR Modified assessments. The input of the advisory committees is reflected in the assessed curricula and test blueprints.

Early in the development process, prototype items were developed for the STAAR Modified assessments. As part of the item-development process, advisory committees and TEA staff reviewed these prototypes to identify how well these items would measure the student expectations to which the items were aligned. These early reviews provided valuable suggestions for item development guidelines and item types. Item development guidelines continued to be refined throughout the test-development process, as various STAAR Modified item review committees shared their feedback in 2011–2012 about how the student expectations could be effectively assessed.

Educator Input

As part of the annual process of item development, committees of Texas educators meet to review the STAAR Modified items and confirm that each item appropriately measures the TEKS to which it is aligned. These item review committees also review the assessment items for content and bias. The committees are made up of Texas K-12 educators, and these committees revise and edit items, as appropriate, prior to test administrations. Item review committees are convened for all STAAR Modified assessments.

Test Developer Input

Item writers and reviewers follow test development guidelines that explain how content, aligned to given TEKS statements, should be measured. At each stage of development, writers and reviewers verify the alignment of the items with the assessed student expectations. When STAAR Modified was designed as the alternate

assessment based on modified achievement standards, special education content specialists developed detailed guidelines so the modifications made to the STAAR items were consistent. After the items were modified, educator committees for each content area at each grade level reviewed the original STAAR item and the STAAR Modified version of the item to make sure that the modified item still measured the same underlying skill as the original item. In this way, the alignment between the TEKS curriculum and the STAAR items carries through to the STAAR Modified items.

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 support the expectation that the way students respond to items on the STAAR Modified assessments reflects the accurate measurement of the construct.

Item Types

Student response processes on the STAAR Modified assessments vary according to item types. Across STAAR Modified, three types of items are administered to students: multiple-choice items on all assessments; gridded-response items on mathematics assessments; and written compositions on grades 4 and 7 and English I and II writing assessments.

When item types are initially modified for STAAR Modified, the items are reviewed by educator committees to help ensure that the modifications made the items accessible to the STAAR Modified student population.

Texas also gathers evidence to show that response processes do not advantage or disadvantage one or more student groups. This evidence comes from several sources. When item types were initially considered for inclusion in the STAAR Modified assessments, the item types were pilot-tested to study the way students engage with the various item presentations. After item types were determined to be appropriate for STAAR Modified, evidence about student responses is gathered annually through educator and expert reviews and analyses of individual student responses to these items. Every year, during educator reviews, educators evaluate whether item content for a given item type 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 point-biserial correlations, and differential item functioning are all evaluated with regard to the item type. For additional information refer the Item Analyses section of chapter 3, "Standard Technical Processes."

Scoring Process

For multiple-choice items, statistical keychecks are conducted for all STAAR Modified assessments during the equating process. Score reliability and validity indices are generated and evaluated for every STAAR Modified assessment (refer to the Item Analyses section of chapter 3, "Standard Technical Processes").

Tables 5.9 and 5.10 summarize reader agreement rates by grade and the validity results, respectively, for the STAAR Modified assessments administered in spring 2012 that included written compositions. The reader agreement rate is expressed in terms of absolute agreement between the first reader's score and the second reader's score. Validity is expressed in terms of exact agreement between the score assigned by a given reader and the "true" score approved by TEA.

Table 5.9. Summary of Reader Agreement (Reliability) for 2012 STAAR Modified

Summary of Reader Agreement (Reliability) for 2012 STAAR Modified				
STAAR Assessment	**Number of Responses Read**	**Agreement Rate (%) After 2 Readings**	**Number of Third Readings**	**Agreement Rate (%) After 3 Readings**
Grade 4 writing	13,852	79%	254	98%
Grade 7 writing	13,340	78%	284	98%
English I writing	11,268	80%	158	99%
English II writing	7,245	81%	60	99%

Table 5.10. Summary of Validity Packet Results for 2012 STAAR Modified

Summary of Validity Packet Results for 2012 STAAR Modified	
STAAR Assessment	**Agreement Rate (%)**
Grade 4 writing	78%
Grade 7 writing	91%
English I writing	75%
English II writing	69%

Evidence Based on Internal Structure

The internal consistency of the STAAR Modified assessments is evaluated every year using the KR20 for assessments that have only dichotomously scored items (i.e., multiple-choice and gridded-response items). For the STAAR Modified assessments that have a combination of multiple-choice items and written compositions (i.e., the writing assessments), internal consistency is evaluated using the stratified coefficient alpha. These internal consistency evaluations are made for all students and for reported student groups, such as female, male, African American, Hispanic, and white

students. Estimates of internal consistency are made for the full assessment as well as for each reporting category within a content area and can be found in the Reliability section of this chapter.

Evidence Based on Relationships to Other Variables

Another method by which Texas provides validity evidence for the STAAR Modified assessments is by analyzing the relationship between performance on a given STAAR Modified assessment and performance on another STAAR Modified or TAKS–M assessment, a process that supports what is referred to as criterion-related validity. By examining this relationship, evidence can be collected to show that the relationships are consistent with those expected based on empirical data.

Correlations between the STAAR Modified content area scale scores were calculated. As shown in Table 5.11, the correlations between content area scores ranged from 0.06 to 0.96. Most correlations are considered moderate correlations, which suggests that scores across content areas are related and neither redundant nor irrelevant. This is expected because the constructs being measured are both academic content areas but assess different types of knowledge and skills. Correlations between writing and science and writing and social studies are not included because students do not take these assessments in the same grade. Correlations between the STAAR Modified content-area scale scores at the same grade level can be found in Appendix C.

Table 5.11. Overall STAAR Modified Correlations Between Content Area Scores

Overall STAAR Modified Correlations Between Content Area Scores		
Content Areas Compared (Using Scale Scores)	**N-Count**	**Correlation**
Mathematics & Reading	72,381	0.30*
Science & Reading	28,594	0.28*
Social Studies & Reading	18,243	0.06*
Writing & Reading	34,374	0.96*
Science & Mathematics	28,718	0.46*
Social Studies & Mathematics	17,976	0.31*
Writing & Mathematics	30,531	0.35*
Social Studies & Science	18,108	0.55*

*Indicates that correlations were significant at the $p \leq .01$ level.

The correlations between the total score and the STAAR Modified reporting category scores were also calculated within grade and content area. Across all subjects and grades, the correlations between each reporting category and score ranged from 0.55 to 0.88. The magnitudes of these correlations were found to support theoretical relations between reporting categories and the

overall test. More specifically, the range of correlations within reading across all grades was 0.63 to 0.88. For mathematics, the range of correlations was 0.55 to 0.87. Science had a correlation range of 0.58 to 0.83, while social studies had a correlation range of 0.63 to 0.86. Lastly, the correlations of the total test score to the reporting category scores for writing ranged from 0.60 to 0.83.

Additional validity evidence was collected in the form of discriminant validity, which demonstrates that the STAAR Modified scores are unrelated to demographic variables (e.g., gender and ethnicity). Theoretically, student characteristics such as ethnicity and gender should not relate to their performance on the assessment; therefore, the lack of meaningful empirical relationships between these measures is expected.

To investigate the relationship between the STAAR Modified scores and demographic variables, correlations were computed specifically for gender and ethnicity. The correlation between the STAAR Modified scores and gender was 0.009, and the correlation between the STAAR Modified scores and ethnicity was 0.056. Both the gender and ethnicity correlations are very small and do not indicate a meaningful relationship between the STAAR Modified scores and either demographic variable.

In addition, research studies were conducted during the 2011–2012 school year to evaluate the relationships between scores on the STAAR Modified assessments and scores on the TAKS–M assessments (for example, the STAAR Modified grade 6 mathematics and TAKS–M grade 6 mathematics). These studies were initially used to inform the establishment of performance standards across the STAAR Modified assessments. Refer to chapter 3 of the *STAAR Modified Standard Setting Technical Report* available on the STAAR Resources page of TEA's Student Assessment Division website, for more information about the STAAR Modified-to-TAKS–M comparison studies.

Evidence Based on Consequences of Testing

TEA has developed and implemented a plan to formally document the evidence of the consequential validity of the STAAR Modified program. Surveys asking about the intended and unintended consequences resulting from the STAAR Modified assessments were administered to the standard-setting committee. Once analyzed, results from the STAAR Modified consequential validity surveys will be reported and used to help promote the continuous improvement of the STAAR Modified program.

Measures of Student Progress

Progress measures require multiple years of performance data, and were therefore not available for STAAR Modified in 2011–2012. However, development of these measures is underway. As part of these development efforts, many factors are being considered:

- Different models for measuring student progress (refer to the Measures of Student Progress section in chapter 3, "Standard Technical Processes") to determine which model could be used for STAAR Modified
- Content relationships among the STAAR Modified assessments to determine where progress measures are appropriate
- Federal and state requirements that determine how progress measures can be used for accountability
- Reporting options that allow information about progress measures to be communicated most effectively

In 2011–2012 input was sought from a number of advisory groups with regard to the development of STAAR progress measures, including measures for STAAR Modified. Several options for progress measures were presented to the TTAC who then provided recommendations and guidance for the development activities. Progress measures were also discussed with the ATAC and the APAC, which are groups made up of educators from various Texas campuses, districts, and ESCs, as well as parents, higher education representatives, business leaders, and legislative representatives. Input from these groups was requested at several points during development and will continue to be requested as the STAAR Modified progress measures are developed and refined. Whenever possible, pilot studies and empirical data are also being used to inform these development activities. Student progress information for STAAR Modified will be available for the first time in 2013.

Sampling

In 2011–2012 stand-alone census field testing was conducted for STAAR Modified. This method tests the entire student population in the specific grade and subject instead of testing a sample of the population. Census field testing is used for STAAR Modified because data are needed from all students to obtain a large enough number of students to conduct field-test analyses. Sampling was not needed for any audits or empirical research studies for STAAR Modified in 2011–2012.

Test Results

Appendix C provides scale score distributions and statistics and RSSS conversion tables, as well as mean p-values and reliability estimates by reporting category and content area, for all operational STAAR Modified assessments administered in spring 2012. Table 5.12 shows spring 2012 pass rates for STAAR Modified.

Table 5.12. STAAR Modified Spring 2012 Pass Rates (at the Phase-in 1 Standard)

Mathematics	Grade 3	63%
	Grade 4	64%
	Grade 5	66%
	Grade 6	58%
	Grade 7	64%
	Grade 8	56%
	Algebra I	41%
Reading	Grade 3	70%
	Grade 4	68%
	Grade 5	75%
	Grade 6	74%
	Grade 7	69%
	Grade 8	66%
	English I	67%
Writing	Grade 4	65%
	Grade 7	67%
	English I	59%
Science	Grade 5	56%
	Grade 8	66%
	Biology	49%
Social Studies	Grade 8	58%
	World Geography	62%