Chapter 4 State of Texas Assessments of Academic Readiness (STAAR)

Overview

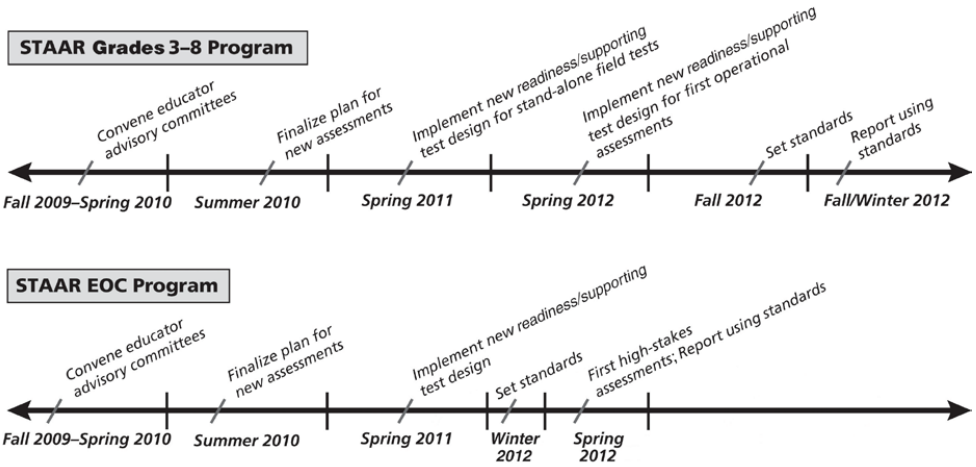
In May 2007 the 80th Texas Legislature enacted SB 1031, which expanded the role of the EOC assessment program. The bill phased out the TAKS assessments for grades 9–11 and replaced them with the EOC assessments as a component of new high school graduation requirements, beginning with the incoming freshman class of 2011–2012. Additionally, SB 1031 required the EOC assessments to include items that measure students' postsecondary readiness.

In 2009 the 81st Texas Legislature passed HB 3, which called for a unified comprehensive assessment program—STAAR to replace the existing TAKS program for all grades and courses beginning in spring 2012 for students in 9th grade and below. TEA, in collaboration with the THECB and Texas educators, developed the STAAR assessment system. The EOC assessments that were previously in development to fulfill the requirements of SB 1031 were updated to meet the additional legislative requirements for STAAR. The legislative requirements of HB 3 include the following:

- increasing the rigor and relevance of both standards and assessments;
- assessing postsecondary readiness;
- developing an aligned system of assessments that connect performance from the postsecondary readiness standards in Algebra II and English III down to grade 3 mathematics and reading, from course to course and from grade to grade; and
- developing a progress measure designed to provide an early-warning indicator for students who are not on track to meet the passing standard, who might not be successful in the next grade or course, who might not be ready for advanced coursework in mathematics and English in high school, or who might not be ready for postsecondary work in mathematics and English.

Beginning in fall 2009, TEA began the development and implementation of STAAR on the schedule shown in Figure 4.1.

Figure 4.1. STAAR Development and Implementation Schedule



TEA set the performance standards for the STAAR EOC assessments prior to the spring 2012 test administration. Extensive research studies were conducted to support the standard-setting process and to ensure that the STAAR EOC assessments meet requirements for rigor and alignment. Because of the aligned nature of the program, the performance standards for the STAAR 3–8 assessments could not be set until after the approval and implementation of the standards for STAAR EOC. Standards for the STAAR EOC assessments were approved and implemented in spring 2012, while standards for all other STAAR assessments were approved and implemented in January 2013.

Students who were in grade 9 for the first time in the 2011–2012 school year are required to take the STAAR assessments for courses in which they are enrolled. Students who were in grades 10 and above in the 2011–2012 school year or who repeated grade 9 in the 2011–2012 school year will graduate under TAKS requirements. Students in grades 3–8 who received instruction in

courses above their enrolled grade were required to take the STAAR assessments for those courses in which they received instruction if the content covered the entire curriculum for that course.

In addition, for 2011–2012 only, nine STAAR EOC assessments—English II reading, English II writing, English III reading, English III writing, Algebra II, chemistry, physics, world history, and U.S. history—were administered to TAKS students on selected campuses who were enrolled in these courses. For these students, participation in the STAAR EOC assessments was mandatory, though they still have to graduate under TAKS requirements. Refer to the Sampling section for more information.

STAAR

STAAR is an assessment designed to measure the extent to which a student has learned and is able to apply the knowledge and skills defined in the state-mandated curriculum, the TEKS. Every item on every STAAR test is directly aligned to the TEKS currently in effect for the grade and content area being tested. In grades 3–8, students are tested in mathematics and reading. In addition, students are tested in writing at grades 4 and 7, science at grades 5 and 8, and social studies at grade 8. The STAAR EOC assessments are available for students in English I (reading and writing), English II (reading and writing), English III (reading and writing), Algebra I, geometry, Algebra II, biology, chemistry, physics, world geography, world history, and U.S. history.

STAAR Spanish

The STAAR Spanish assessments are designed to measure the academic skills of ELLs who receive academic instruction in Spanish while they learn English. The STAAR Spanish assessments are offered at grades 3–5, including grades 3–5 reading and mathematics, grade 4 writing, and grade 5 science. The English and Spanish versions of STAAR have the same test blueprint and assess the same student expectations in the TEKS.

STAAR L

STAAR L is a linguistically accommodated English version of the STAAR mathematics, science, and social studies assessments. STAAR L is provided for ELLs who meet eligibility requirements to receive a substantial degree of linguistic accommodation in these content-area assessments.

The assessments required by grade level and course for STAAR, STAAR Spanish, and STAAR L are given in Table 4.1.

Table 4.1. 2011–2012 STAAR Assessments

2011–2012 STAAR Assessments	
Grade/Course	**Content Area**
Grade 3 (English, Spanish, and L*)	mathematics and reading
Grade 4 (English, Spanish, and L*)	writing, mathematics, and reading
Grade 5 (English, Spanish, and L*)	mathematics, reading, and science
Grade 6 (English and L*)	mathematics and reading
Grade 7 (English and L*)	writing, mathematics, and reading
Grade 8 (English and L*)	mathematics, reading, science, and social studies
End-of-Course Assessments (English and L*)	English I reading, English I writing, English II reading, English II writing, English III reading, English III writing, Algebra I, geometry, Algebra II, biology, chemistry, physics, world geography, world history, U.S. history

*The STAAR L assessments are available for mathematics, science, and social studies.

Participation Requirements

According to federal regulations, all students must be assessed on grade-level curriculum. As such, all students are expected to participate in the Texas assessment program. For ELLs, test participation decisions are made in accordance with agency guidelines by each student's language proficiency assessment committee (LPAC). Figure 4.2 provides a summary of the guidelines LPACs use in determining the appropriate STAAR assessment for an ELL. Additional information is available on the LPAC Resources page on TEA's Student Assessment Division website.

Participation requirements for STAAR Modified and STAAR Alternate are described in chapter 5 and chapter 6 respectively.

Figure 4.2. ELL Participation in STAAR

ELL Participation in STAAR	
STAAR	■ General statewide assessments ■ Taken by ELLs not administered STAAR Spanish and STAAR L ■ Some linguistic accommodations permitted
STAAR Spanish	■ Available for students in grades 3–5 for whom a Spanish version of STAAR most appropriately measures their academic progress ■ Not permitted for a student whose parent or guardian has declined bilingual/English as a second language (ESL) program services ■ Linguistic accommodations not permitted
STAAR L*	■ Linguistically accommodated STAAR mathematics, science, and social studies assessments ■ Available for ELLs who • are not most appropriately assessed with STAAR Spanish, AND • have not yet attained a TELPAS advanced high reading rating in grade 2 or above, AND • have been enrolled in U.S. schools for 3 years or less starting with first grade (5 years or less if a qualifying unschooled asylee or refugee) ■ Not permitted for an ELL whose parent or guardian has declined bilingual/ESL program services ■ More substantial linguistic accommodations allowed than for STAAR

*For the STAAR L EOC assessments, eligibility can be carried over from spring to the July and December administrations.

Testing Requirements for Graduation

Students who were enrolled as a first-time 9th grader or below in the 2011–2012 school year are required to take the STAAR EOC assessments rather than TAKS as their testing requirement for graduation and are no longer required to take TAKS. With the passage of HB 3, the relationship among high school courses, the STAAR EOC assessments, and performance on those assessments is now linked to a student's graduation program. The following provisions are mandated by HB 3.

- In order to graduate, a student must achieve a cumulative score that is at least equal to the product of the number of the STAAR EOC assessments taken in each foundation content area (English language arts, mathematics, science, and social studies) and a scale score that indicates satisfactory performance (Level II: Satisfactory Academic Performance).
- A student must achieve a minimum score on an assessment, as determined by the commissioner of education, for that score to count toward the student's cumulative score. If a student does not achieve the minimum score, the student must retake the assessment.
- For students on the Minimum High School Program (MHSP), the cumulative score requirement is based on the number of courses that are specifically required on the MHSP.

- For the Recommended High School Program (RHSP), students must meet the satisfactory performance standard (Level II: Satisfactory Academic Performance) on the Algebra II and English III assessments in addition to the cumulative score requirement.
- For the Distinguished Achievement Program (DAP), students must meet the postsecondary readiness performance standard (final recommended Level II: Satisfactory Academic Performance) on the Algebra II and English III assessments in addition to the cumulative score requirement. Note that for those students eligible to take STAAR Modified or STAAR Alternate, passing the EOC assessments is not a requirement for graduation. A student's admission, review, and dismissal (ARD) committee determines individual requirements for graduation.

Test Development

Maintaining a student assessment system of the highest quality involves many steps during the test-development process. The procedures described in chapter 2, "Building a High-Quality Assessment System," outline the processes that are used to develop a framework for each STAAR assessment and explain the ongoing development. TEA relies greatly on input from educators and assessment specialists to ensure that all measures of learning for Texas public school students are equitable and accurate.

For detailed information regarding each step of the test development process, refer to chapter 2, "Building a High-Quality Assessment System."

STAAR test development activities are summarized below.

- In 2010 and 2011 TEA convened advisory groups comprised of curriculum specialists, teachers, and professors who provided input and guidance about which of the TEKS standards eligible for assessment were critical for student success and should be emphasized on the assessments. From this information, TEA developed a set of readiness and supporting standards that focus the assessment at each grade and course and that provide a vertical link between the assessments from grade to grade or from course to course. From these standards a draft set of TEKS student expectations eligible for assessment on STAAR, as well as a draft test blueprint, were developed for each grade/subject and course assessed. The drafts of these assessed curricula and test blueprints were subsequently approved and published.
- With input from both national-level and state-level writing experts, scoring rubrics were developed to guide the scoring of English I, English II, and English III short answer reading items and written compositions and grades 4 and 7 written compositions. After the draft rubrics were used to evaluate field-test responses, TEA convened rubric validation committees made up of Texas writing educators to review the field-test results and validate the rubrics. Rubrics for reading short answer,

expository, and literary writing were validated in summer 2010. Rubrics for persuasive and analytical writing as well as grades 4 and 7 personal narrative and expository writing were validated in summer 2011. These draft rubrics were also subsequently approved and published.

- The 2012 STAAR assessments were constructed to conform to the assessment blueprints and all established test construction guidelines for STAAR. The items eligible for placement on the 2012 STAAR assessments were developed and field-tested through placement on TAKS 3–8 assessments administered in 2008 through 2011, through stand-alone field tests for grades 4 and 7 writing in 2011, and through both EOC stand-alone field tests and EOC operational assessments administered in 2008 through 2011.
- Item-development activities continue annually to build item bank depth and to support the number of test forms that are necessary for the program. These items are subsequently field-tested and the field-test data are reviewed. Field-tested items that meet established predetermined criteria are added to the bank.
- An additional design aspect is a focus on student preparedness for success in subsequent grades and courses, and ultimately for postsecondary readiness. This ongoing focus is supported by the development of items that have a high level of cognitive complexity and that closely align with the cognitive complexity evident in the TEKS. Items developed in 2011–2012 meet these established guidelines. For example, additional open-ended items that require students to derive an answer independently were developed for science and mathematics courses. For social studies, science, and mathematics, items that measure process skills in context rather than in isolation were developed. In reading, greater emphasis was given to critical analysis rather than literal understanding. In writing, prompts were developed to support expository, analytical, and persuasive writing.
- STAAR EOC test items were also developed so that they could be delivered in both paper and online test formats. A plan was created to evaluate the comparability of EOC items delivered in paper mode and those delivered in online mode. The plan also sought to minimize potential formatting issues, such as a student having to scroll down a screen through a long item, that might create a different student experience when testing online as compared with testing on paper. Studies were conducted between 2009 and 2012 to evaluate the performance of EOC items on both paper and online.

Item Development Approach for STAAR Spanish

The STAAR Spanish assessments include passages and items for reading and writing that are originally written in Spanish as well as transadapted items for mathematics and science—that is, items for mathematics and science translated from English and adapted as necessary to ensure cultural and linguistic accessibility. Items and passages for the STAAR Spanish writing and reading assessments are originally written in Spanish due to differences in English and Spanish grammar, mechanics, and usage.

TEA staff, Texas educators, and Spanish language experts work to develop the STAAR Spanish test materials. Spanish bilingual educators then review all transadapted and original test items before field testing in accordance with the educator review process used for the STAAR tests and described in chapter 2, "Building a High-Quality Assessment System."

Spanish–English STAAR Alignment

Alignment of the STAAR English and Spanish tests is reinforced by the following practices.

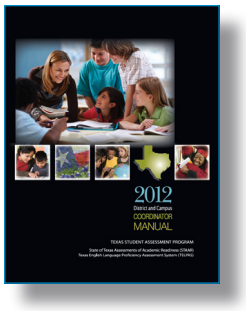
- The development and review processes for the English and Spanish tests are parallel, meaning that item reviews for both English and Spanish include judgments related to each item's alignment to the TEKS content standards. Also, field-test data reviews for English and Spanish items include review of item statistics based on actual student performance. These safeguards ensure that only psychometrically sound items are selected for inclusion in the item bank.
- Item writing and review processes for transadapted items ensure that items in each language are linguistically and culturally appropriate and that the interpretations of grade-level performance expectations will be the same.
- The test blueprints for the English and Spanish tests are also the same, including the number of items that assess each reporting category and the number of items on the test as a whole.
- The English and Spanish tests are constructed in concert, adhering to the same test-construction guidelines with regard to the range of item content and cognitive complexity.
- Each year TEA and the STAAR English and STAAR Spanish development staff participate in item-review meetings held for new English test items. Discussions at these meetings focus on the best ways to assess the TEKS, further enhancing the alignment between the English and Spanish tests. Finally, when the performance standards for STAAR 3–8 were established in 2012, standard-setting panels reviewed both the English and the Spanish transadapted tests with the goal of establishing comparable performance standards.

Training

Test administration procedures must be followed exactly and consistently so that all individuals eligible for testing have an equal opportunity to demonstrate their academic knowledge and skills. To reach this aim, TEA develops instructional materials and trains all testing personnel across the state on proper test administration procedures. Preparation for test administration begins every year with a training session for testing coordinators and other personnel from each of Texas's 20 regional ESCs as well as district testing coordinators from some of the state's largest districts. Then, using materials and information provided in the TEA training session, the ESC regional testing coordinators train the district coordinators in their respective regions. District coordinators then train the campus testing coordinators, who will be responsible for training any campus personnel participating in the administration of the tests.

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 grades 3–5, STAAR grades 6–8, and STAAR EOC.

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 and field-test administrator manuals are distributed to districts prior to the first assessment administration for each grade.

Test Administrations

All STAAR assessments were administered in spring (March, April, and May) 2012. In addition, the STAAR EOC retests were made available for the first time in summer (July) 2012. During the 2011–2012 school year, over 8,250,000 STAAR assessments, including STAAR 3–8, STAAR EOC, and STAAR L, were administered. Districts administered the STAAR assessments to eligible students as shown in Table 4.2.

Table 4.2. STAAR Assessments Administered in 2011–2012*

STAAR Assessment	Assessments Administered
Grade 3 English reading	327,936
Grade 3 English mathematics	337,305
Grade 3 Spanish reading	36,318
Grade 3 Spanish mathematics	19,859
Grade 3 STAAR L mathematics	8,653
Grade 4 English reading	334,484
Grade 4 English mathematics	346,249
Grade 4 English writing	332,417
Grade 4 Spanish reading	23,249
Grade 4 Spanish mathematics	10,824
Grade 4 Spanish writing	24,453
Grade 4 STAAR L mathematics	2,007
Grade 5 English reading	348,806
Grade 5 English mathematics	353,030
Grade 5 English science	354,628
Grade 5 Spanish reading	9,986
Grade 5 Spanish mathematics	3,631
Grade 5 Spanish science	4,064
Grade 5 STAAR L mathematics	2,073
Grade 5 STAAR L science	2,310
Grade 6 reading	354,387
Grade 6 mathematics	344,977
Grade 6 STAAR L mathematics	4,059
Grade 7 reading	347,911
Grade 7 mathematics	323,015
Grade 7 writing	347,294
Grade 7 STAAR L mathematics	4,377
Grade 8 reading	340,860

STAAR Assessment	Assessments Administered
Grade 8 mathematics	312,342
Grade 8 social studies	336,762
Grade 8 science	336,661
Grade 8 STAAR L mathematics	4,348
Grade 8 STAAR L social studies	4,703
Grade 8 STAAR L science	4,685
English I reading	388,593
English I writing	413,946
English II reading	27,678
English II writing	28,275
English III reading	42,343
English III writing	43,075
Algebra I	358,963
STAAR L Algebra I	6,009
Geometry	85,461
STAAR L Geometry**	135
Algebra II	38,072
STAAR L Algebra II**	39
Biology	336,699
STAAR L Biology	4,977
Chemistry	48,215
STAAR L Chemistry**	19
Physics	56,774
STAAR L Physics **	14
World geography	348,376
STAAR L World geography	6,038
World history	30,112
STAAR L World history**	182
U.S. history	40,942
STAAR L U.S. history**	49

*EOC totals include both spring and summer 2012 administrations.
**N-count for this STAAR L EOC assessment is low because no sampling took place for this assessment and it is not a course typically taken by students in grade 9 or below. As such, few ELLs eligible for STAAR L took this upper-grade level EOC assessment.

All STAAR EOC assessments in 2011–2012 were offered in online and paper modes during the spring and summer administrations. Table 4.3 shows the number of STAAR EOC online and paper tests that were administered in 2011–2012.

Table 4.3. 2011–2012 Administration by Mode (Online and Paper)

Test Administration	Number of Online Tests Administered	Number of Paper Tests Administered
Spring Administrations (March/May 2012)		
English I reading	15,967	318,861
English I writing	16,045	318,899
English II reading	101	27,412
English II writing	98	27,800
English III reading	50	42,290
English III writing	51	43,020
Algebra I	36,457	297,132
Geometry	8,186	76,093
Algebra II	2,500	35,467
Biology	25,912	293,160
Chemistry	3,119	45,027
Physics	5,584	51,105
World Geography	30,326	290,645
World History	3,630	24,996
U.S. History	5,086	35,595
Total	**153,112**	**1,927,451**
Summer Administration (July 2012)		
English I reading	9,545	44,220
English I writing	13,359	65,643
English II reading	13	152
English II writing	37	340
English III reading	0	3
English III writing	1	3
Algebra I	6,445	18,929
Geometry	305	877
Algebra II	14	91
Biology	4,357	13,270
Chemistry	5	64
Physics	66	19
World Geography	6,271	21,134
World History	85	1,401
U.S. History	83	178
Total	**40,586**	**166,324**
Total 2011–2012 STAAR EOC Assessments	**193,698**	**2,093,775**

The Online Test Delivery System

The STAAR EOC online tests are delivered using the Texas Assessment Management System, delivered through Pearson Access. This system provides secure online tools for authoring tests, delivering tests, and reporting students' results. The Assessment Management System meets the stringent security requirements of the Texas assessment program and protects the integrity of test items and student data.

The Assessment Management System enables test administrators to control

- which test is administered,
- when it is administered,
- the number of testing sessions, and
- which students are assigned to each session.

Using the Assessment Management System's *Test Session Management* screen, a test administrator can monitor each student's current status while the test session is still in progress.

Additional information about the Assessment Management System, such as an overview of the system, information on delivery and reporting, and a list of frequently asked questions, is available on Pearson's Texas State Assessments website.

Make-up Testing

For the STAAR program, make-up testing opportunities for students who are absent are available for all grades, content areas, and courses during the testing window. For the 2011–2012 school year, make-up testing opportunities were available for all administrations, including the summer administrations.

Out-of-District Testing

For the summer STAAR EOC assessments, students who are unable to test at their home district's designated test site are allowed to test out of district. Enrolled out-of-district students are required to complete the out-of-district registration form and present photo identification. For example, a student from Houston who spends the summer in Dallas and who wants to test in Dallas could register to test out-of-district. Out-of-district testing also applies to students who are part of the Texas Tech or University of Texas high school programs.

A campus or district must accommodate the request of an out-of-district student to participate in the summer STAAR EOC administration if that campus or district is testing one or more local students on the applicable test and if the student has registered in advance for out-of-district testing.

Educational Materials Required for Testing

Dictionaries And Thesauruses

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

There must be at least one dictionary for every five students; it is also recommended that there be one thesaurus for every five students, if possible. Students may also use a combination dictionary/thesaurus. An ESL dictionary, which uses simple English and pictures to define words, or a bilingual dictionary. 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 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 Algebra I, geometry, Algebra II, biology, chemistry, and physics 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 (including STAAR L)
  - Geometry (including STAAR L)
  - Algebra II (including STAAR L)
- A graphing or scientific calculator for each student taking
  - Chemistry (including STAAR L)
  - Physics (including STAAR L)
- A calculator for every five students taking
  - Biology (including STAAR L)

Any calculator may be used to fulfill the minimum requirements listed above 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 may not 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. For specific assistance in effectively preparing a calculator for use during testing, contact that calculator's manufacturer.

Calculators may not be provided to a student taking the STAAR grades 3–8 mathematics assessments or the STAAR grades 5 and 8 science assessments unless the student meets the eligibility criteria for such an accommodation. This includes the STAAR Spanish, STAAR L, and the STAAR Modified versions of the assessments.

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. This includes general education students who have special needs, ELLs who are eligible for linguistic accommodations, as well as students with disabilities who receive special education services.

Accommodations should not result in reduced learning expectations. Therefore, the use of accommodations should not replace the teaching of content-area knowledge and skills as outlined in the state curriculum for each grade. Certain accommodations might be more useful or appropriate than others, depending on factors such as the content area being assessed, the student's age, and the degree of the student's needs. Not all accommodations that are appropriate for instructional use are appropriate for use on a standardized assessment.

Accommodations are provided on an individual basis and take into consideration the needs of each student. It would be neither appropriate nor effective to try to provide "one size fits all" accommodations to students. For example, one student with a visual impairment might benefit from large-print instructional materials whereas another student with a visual impairment might benefit more from a magnification device. In most cases, accommodations are appropriate for a specific student and should not be provided to an entire group of students, such as those in the same class or disability category.

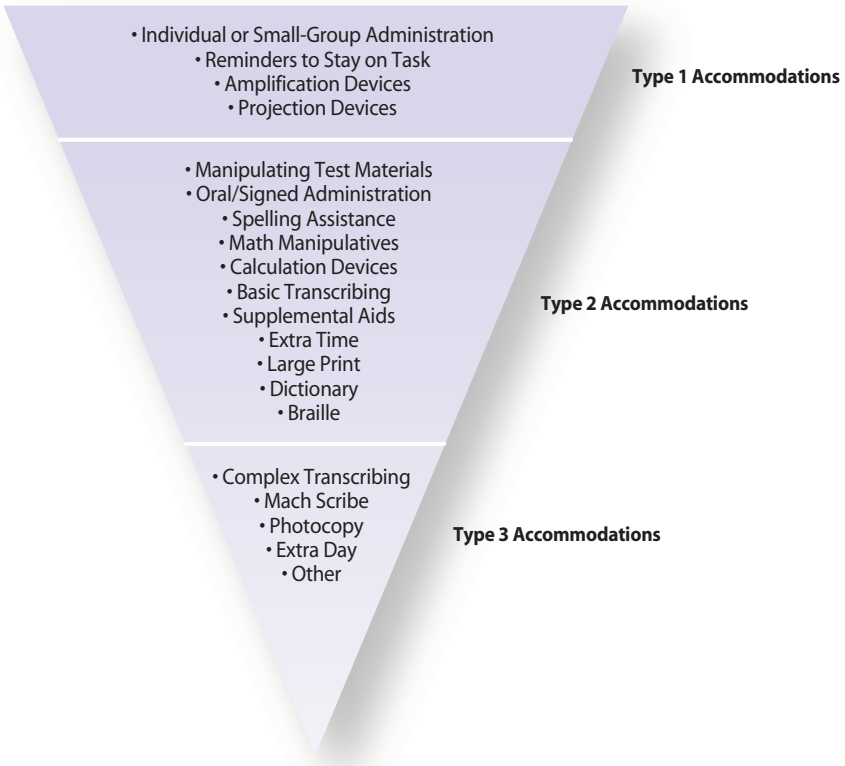
Accommodations for Students with Disabilities

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 meets 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 Response to Intervention (RtI) team or student assistance team.

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. The pyramid in Figure 4.3 presents

accommodations for students with disabilities by type in accordance with the specificity of the eligibility criteria and the need for TEA approval. Note that not all accommodations are applicable to all assessments. Specific information about each accommodation can be found on the Accommodation Resources page on TEA's Student Assessment Division website.

Figure 4.3. Accommodations for Students with Disabilities



Dyslexia Accommodations for Reading Assessments

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

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

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

Oral Administration

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

Students with Visual Impairments

Test administrators receive specific instructions for the use of large-print or braille test booklets when testing students who have visual impairments. 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 are available for all STAAR assessments. Braille test booklets are available for all STAAR operational administrations.

Linguistic Accommodations

Linguistic accommodations are language supports that make grade level academic assessments in English more accessible to ELLs. The Texas ELPS require all teachers to linguistically accommodate the instruction of ELLs in their classes commensurate with the students' English language proficiency levels. Policies for the STAAR linguistic accommodations support these ELPS requirements.

The information in Figure 4.4 shows the ways in which the language needs of ELLs are addressed in STAAR.

Figure 4.4. Allowable Linguistic Accommodations for ELLs Taking STAAR

Allowable Linguistic Accommodations for ELLs Taking STAAR			
	STAAR (English)	**STAAR L**	**STAAR Modified**
Mathematics **Science** **Social Studies**	■ Bilingual dictionary ■ Extra time (same day)	■ Bilingual dictionary ■ Extra time (same day) (Clarification in English of word meaning and reading aloud of text are provided in the online interface for all students taking STAAR L)***	■ Bilingual dictionary ■ Extra time (same day) ■ Clarification in English of word meaning ■ Reading aloud of text ■ Oral translation ■ Bilingual glossary
Reading* **Writing***	**Reading and Writing:** ■ Dictionaries of various types** ■ Extra time (same day) **Writing:** ■ Clarification in English of word meaning in writing prompts **English I–III Reading:** ■ Clarification in English of word meaning in short answer questions	Not Applicable	■ Dictionaries of various types** ■ Extra time (same day) ■ Reading aloud of eligible text ■ Clarification in English of word meaning ■ Oral translation

*Unless otherwise indicated, "reading" and "writing" refer to the reading and writing components of the English I, English II, and English II assessments as well as the 3–8 assessments.

**The STAAR Dictionary Policy for reading and writing in grade 6 and above includes use of standard English, ESL (simplified English), and bilingual dictionaries for all students. For grades 3–5 reading and grade 4 writing, use of dictionaries is permitted as a linguistic accommodation for eligible ELLs.

***In addition to examining the eligibility criteria for STAA L, LPACs should consider whether a student routinely needs and uses clarification in English of word meaning and/or reading aloud of text. Students who do not need either accommodation should take STAAR, not STAAR L.

Decisions concerning accommodations must be made and documented by the LPAC. In the case of an ELL receiving special education services, decisions are made by the student's ARD committee in conjunction with the LPAC. The decisions must be based on the individual needs of the student and whether such accommodations are used routinely in instruction and testing.

More information about the STAAR linguistic accommodations is available in the document titled "Linguistic Accommodations for ELLs Participating in the STAAR Program" on the Accommodation Resources page on TEA's Student Assessment Division website.

Student Success Initiative

The Student Success Initiative (SSI) provides a system of academic support to help students achieve on grade level success 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 the 2011–2012 school year, districts made promotion/retention decisions based on the same academic information that was used to make these decisions in non-SSI grades (i.e., the recommendation of the student's teacher and the student's grade in the content area). As stated in TEC §28.021(a), "a student may be promoted only on the basis of academic achievement or demonstrated proficiency of the subject matter of the course or grade level."

Scores and Reports

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

Description of Scores

Scores for the STAAR assessments consist of the number of items answered correctly (raw scores), scale scores, and the resulting performance level associated with the student's score.

Raw Score

The number of items that a student answers correctly on a STAAR 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: Advanced Academic Performance. (Performance-level cut scores are discussed in the Performance Standards section of this chapter.) Scale scores for the STAAR EOC assessments were reported following the May 2012

and July 2012 test administrations. However, because performance standards for STAAR 3–8 were not set until fall 2012, only raw scores were reported after the spring 2012 administrations. Scale scores for the grades 3–8 STAAR assessments were first made available in January 2013 after standards were applied.

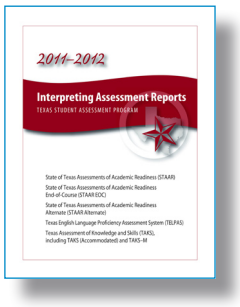
Scale scores are also used to compare the performance of an individual student with the performance of a demographic group, a program group, an entire campus, or a district at a particular grade. For example, the scores for a Hispanic student in a gifted and talented program can be compared with the 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 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, districts 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, 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 standard and the Level III standard can be analyzed by grade and content area across administrations to give insight into whether student performance is improving across years. For example in the future, after additional years of STAAR administrations, the average scale score for students who receive special education services and who took the STAAR 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 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 test should be made with great care and with full reference to the fact that the conclusions were based only on that test. Instruction and program evaluations should take into account as much information as possible, rather than on test scores alone, 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 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 performance results. An explanation about the STAAR L assessments was included in the brochure. For the STAAR EOC assessments, an explanation of the test results is printed on the CSR for each individual assessment. A flyer is also available in Spanish for distribution with the individual CSRs.

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 3–8 and EOC assessments (including STAAR Spanish and STAAR L), 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. They demonstrate the ability to think critically and apply the assessed knowledge and skills in varied contexts, both familiar and unfamiliar. Students in this category have a high likelihood of success in the next grade or course with little or no academic intervention.

* For Algebra II and English III, this level of performance also indicates students are well prepared for postsecondary success.

Level II: Satisfactory Academic Performance**

Performance in this category indicates that students are sufficiently prepared for the next grade or course. They generally demonstrate the ability to think critically 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 but might need short-term, targeted academic intervention.

** For Algebra II and English III, this level of performance also indicates students are sufficiently prepared for postsecondary success.

Level I: Unsatisfactory Academic Performance

Performance in this category indicates that students are inadequately prepared for the next grade or course. They do not demonstrate a sufficient understanding of the assessed knowledge and skills. Students in this category are unlikely to succeed in the next grade or course without significant, ongoing academic intervention.

Standard-Setting Process for STAAR

Standard setting for STAAR 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. Standard-setting committees made up of diverse groups of stakeholders carefully considered the interaction of these elements for each STAAR assessment.

The goal of the STAAR program is to have a comprehensive assessment system with curriculum standards and performance standards that are vertically aligned within a content area (i.e., the curriculum and performance standards link from the high school courses back to the middle school and elementary school grades and content areas). As a result, the STAAR performance standards were set for the STAAR EOC assessments first, the middle school assessments next, and the elementary school assessments last.

TEA used an evidence-based standard-setting approach (O'Malley, Keng, & Miles, 2012) for the STAAR program. This evidence-based standard-setting approach incorporated elements of a traditional standard-setting framework (e.g., performance level descriptors, item-mapping methods, etc.) and supported that framework with empirical-based research studies and policy considerations. Using this approach, TEA defined and implemented a nine-step process to establish performance standards for the STAAR 3–8 and EOC assessments program. The nine steps are:

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

7. Approve performance standards
8. Implement performance standards
9. Review performance standards

Tables 4.4 and 4.5 provide high-level descriptions and timelines for these nine steps as implemented in the STAAR EOC and the STAAR 3–8 standard-setting processes, respectively.

Additional detail about each step in the STAAR standard-setting process is given in the State of Texas Assessments of Academic Readiness (STAAR) Standard Setting Technical Report, available on the STAAR Resources page of TEA's Student Assessment Division website.

Table 4.4. Overview of the STAAR EOC Standard-Setting Process

Standard-Setting Step	Description	Timeline
1. Conduct validity and linking studies	External validity evidence was collected to inform standard setting and to support interpretations of the performance standards. Scores on each assessment are linked to performance on other assessments in the same content area.	Studies started in spring 2009 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.	September 2010
3. Develop grade/course specific performance level descriptors (PLDs)	Committees comprised primarily of educators developed performance level descriptors (PLDs) as an aligned system, describing a reasonable progression of skills within each content area (English, mathematics, science, and social studies).	November 2011
4. Convene a policy committee	The committee considered the policy implications of performance standards and empirical study results and made recommendations to identify reasonable ranges ("neighborhoods") for the cut scores.	February 1–2, 2012
5. Convene standard-setting committees	Committees comprised of K–12 educators and higher education faculty used the performance labels, policy definitions, PLDs, and neighborhoods set by the policy committee to recommend cut scores for each STAAR EOC assessment.	Mathematics and English: February 22–24, 2012 Science and Social Studies: February 29–March 2, 2012
6. Review performance standards for reasonableness	TEA and THECB reviewed the cut-score recommendations across content areas.	March 2012
7. Approve performance standards	The commissioner of education approved the performance standards for satisfactory academic performance and advanced academic performance.*	April 2012
8. Implement performance standards	Performance standards were reported to students for the spring 2012 administration with phase-in standards applied.	May 2012
9. Review performance standards	Performance standards will be reviewed at least once every three years.	Fall 2014

*Minimum scores were also established empirically below the satisfactory and advanced academic performance standards and approved by the commissioner of education.

Table 4.5. Overview of the STAAR 3–8 Standard-Setting Process

Standard-Setting Step	Description	Timeline
1. Conduct validity and linking studies	External validity evidence was collected to inform standard setting and support interpretations of the performance standards. Scores on each assessment are linked across grades to performance on other assessments in the same content area.	Studies started in spring 2011 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.	September 2010
3. Develop grade/subject specific performance level descriptors (PLDs)	Committees comprised primarily of educators developed performance level descriptors as an aligned system, describing a reasonable progression of skills within a content area (reading, writing, mathematics, science, and social studies).	June 2012
4. Develop performance standard ranges	EOC performance standards and empirical study results were used to identify reasonable ranges ("neighborhoods") for the cut scores for Levels II and III.	July 2012
5. Convene standard-setting committees	Committees comprised of K–12 educators used the performance labels, policy definitions, PLDs, and neighborhoods to recommend cut scores for each STAAR assessment.	October 2–12, 2012
6. Review performance standards for reasonableness	TEA reviewed the cut-score recommendations across grades and content areas.	October 2012
7. Approve performance standards	The commissioner of education approved the performance standards.	December 2012
8. Implement performance standards	Performance standards were 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 STAAR 3–8 and EOC assessment. The standard-setting committees were made up of K–12 and higher education professionals. Each committee included general education teachers who were experts in both the assessed content and the curriculum (i.e., the TEKS). Educators who represented special populations (students with disabilities and English language learners) were also part of the standard-setting committees.

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 EOC and 3–8 assessments in the same content area, the relevant information from the STAAR EOC 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 February 2012, TEA convened standard-setting committees that recommended performance standards for English I reading, English II reading, English III reading, English I writing, English II writing, English III writing, Algebra I, geometry, Algebra II, biology, chemistry, physics, world geography, world history, and U.S. history.

In October 2012, TEA convened standard-setting committees that recommended performance standards for STAAR grades 3–8 mathematics, grades 3–8 reading, grades 5 and 8 science, grade 8 social studies, and grades 4 and 7 writing. Separate committees were convened, when applicable, for the English-version and Spanish-version of a STAAR assessment.

Phase-in of Performance Standards

A phase-in period has been implemented for the STAAR 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 3–8 and EOC assessments.

In addition, the STAAR English III reading, English III writing, and Algebra II assessments have a two-year, one-step phase-in for Level III: Advanced Academic Performance. The phase-in for Level III allows an appropriate amount of time for students and school districts to adjust to the new assessment requirements.

The STAAR 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 all three mathematics assessments (Algebra I, geometry, and Algebra II).

The STAAR 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 3–8 assessments beginning in the 2015–2016 school year. There is no phase-in for the Level III performance standards for STAAR 3–8.

Figure 4.5 illustrates how the phase-in of standards applies to several cohorts of students who have taken STAAR or will be taking STAAR in the mathematics content area. Phase-in performance standards apply to the STAAR 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 and Level III.

Figure 4.5. Phase-in of STAAR Performance Standards Across Multiple Cohorts of Students

Figure 4.5. Phase-in of STAAR Performance Standards Across Multiple Cohorts of Students

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

Level III** Phase-in for STAAR Algebra II, English III Reading, and English III Writing							
Cohort	**2011–2012**	**2012–2013**	**2013–2014**	**2014–2015**	**2015–2016**	**2016–2017**	**2017–2018**
1	**Grade 9 Algebra I**	**Grade 10 Geometry**	**Grade 11 Algebra II**				
2	Grade 8 Mathematics	**Grade 9 Algebra I**	**Grade 10 Geometry**	**Grade 11 Algebra II**			
3	Grade 7 Mathematics	Grade 8 Mathematics	**Grade 9 Algebra I**	**Grade 10 Geometry**	**Grade 11 Algebra II**		
4	Grade 6 Mathematics	Grade 7 Mathematics	Grade 8 Reading	**Grade 9 Algebra I**	**Grade 10 Geometry**	**Grade 11 Algebra II**	
5	Grade 5 Mathematics	Grade 6 Mathematics	Grade 7 Mathematics	Grade 8 Mathematics	**Grade 9 Algebra I**	**Grade 10 Geometry**	**Grade 11 Algebra II**

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

**The Level III phase-in example used above will be applied only to Algebra II, English III reading, and English III writing. There is no phase-in of Level III for the other STAAR 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 assessments. Performance standards for all STAAR assessments were recommended by standard-setting committees and approved by the commissioner of education in April and December 2012.

Tables 4.6 and 4.7 show the approved performance standards in scale score units for the STAAR 3–8 and EOC assessments (including STAAR L).

Table 4.6. STAAR 3–8 Performance Standards

STAAR 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 English Mathematics	1392	1460	1529	1615
Grade 4 English Mathematics	1471	1535	1599	1677
Grade 5 English Mathematics	1489	1558	1627	1710
Grade 6 Mathematics	1509	1584	1658	1762
Grade 7 Mathematics	1551	1615	1678	1798
Grade 8 Mathematics	1583	1641	1700	1863
Grade 3 English Reading	1331	1400	1468	1555
Grade 4 English Reading	1422	1486	1550	1633
Grade 5 English Reading	1458	1520	1582	1667
Grade 6 Reading	1504	1567	1629	1718
Grade 7 Reading	1556	1615	1674	1753
Grade 8 Reading	1575	1637	1700	1783
Grade 4 English Writing	3500	3750	4000	4612
Grade 7 Writing	3500	3750	4000	4602
Grade 5 English Science	3500	3750	4000	4402
Grade 8 science	3500	3750	4000	4406
Grade 8 Social Studies	3500	3750	4000	4268
Grade 3 Spanish Mathematics	1392	1460	1529	1615
Grade 4 Spanish Mathematics	1471	1535	1599	1677
Grade 5 Spanish Mathematics	1489	1558	1627	1710
Grade 3 Spanish Reading	1304	1374	1444	1532
Grade 4 Spanish Reading	1398	1469	1539	1636
Grade 5 Spanish Reading	1447	1515	1582	1701
Grade 4 Spanish Writing	3500	3750	4000	4543
Grade 5 Spanish Science	3500	3750	4000	4402

Table 4.7. STAAR EOC Performance Standards

STAAR EOC Performance Standards								
Assessment	**Phase-in 1 Minimum**	**Phase-in 1 Level II**	**Phase-in 2 Minimum**	**Phase-in 2 Level II**	**Final Recommended Minimum**	**Final Recommended Level II**	**Phase-in Level III**	**Final Recommended Level III**
English I Reading	1813	1875	1887	1950	1936	2000	N/A	2304
English II Reading	1806	1875	1880	1950	1929	2000	N/A	2328
English III Reading	1808	1875	1882	1950	1932	2000	2135	2356
English I Writing	1798	1875	1872	1950	1921	2000	N/A	2476
English II Writing	1807	1875	1880	1950	1928	2000	N/A	2408
English III Writing	1808	1875	1881	1950	1929	2000	2155	2300
Algebra I	3371	3500	3626	3750	3872	4000	N/A	4333
Algebra II	3350	3500	3604	3750	3852	4000	4080	4411
Geometry	3362	3500	3619	3750	3868	4000	N/A	4397
Biology	3367	3500	3621	3750	3868	4000	N/A	4576
Chemistry	3348	3500	3600	3750	3846	4000	N/A	4607
Physics	3346	3500	3600	3750	3848	4000	N/A	4499
World Geography	3383	3500	3632	3750	3874	4000	N/A	4404
World History	3326	3500	3576	3750	3822	4000	N/A	4634
U.S. History	3372	3500	3624	3750	3869	4000	N/A	4440

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 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 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 assessments are reported on either a horizontal scale or a vertical scale. Horizontal scale scores allow for direct comparisons of student performance between specific sets of test items from different test administrations. Vertical scale scores allow for direct comparisons of student scores across grades within a content area. Student increases in vertical scale scores provide information on the year-to-year growth of students. Refer to chapter 3, "Standard Technical Processes," for detailed information about the scaling process for the different types of reporting scales.

Horizontal Reporting Scales

The following STAAR assessments are reported on horizontal scales:

- STAAR English grades 4 and 7 writing
- STAAR English grades 5 and 8 science
- STAAR English grade 8 social studies
- STAAR Spanish grade 4 writing
- STAAR Spanish grade 5 science
- All STAAR EOC assessments

For all STAAR science and social studies assessments, STAAR EOC mathematics assessments, and STAAR grades 4 and 7 writing assessments, a scale score of 4000 represents the final recommended Level II performance standard. In addition, the standard deviation for those scales was set to 500.

For all STAAR EOC English assessments, a scale score of 2000 represents the final recommended Level II performance standard. This scale score value is half as large as the corresponding scale score values in other content areas because there are six STAAR EOC English assessments that are combined to create a cumulative score and only three STAAR EOC assessments in the other content areas. The standard deviation for the STAAR EOC English scales was set to 250. Refer to the Cumulative Score for Graduation section in this chapter for additional information about the cumulative score requirement for graduation in the STAAR EOC program.

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 remain 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 4.8 and 4.9 for the horizontally-scaled STAAR 3–8 and STAAR EOC assessments, respectively.

Table 4.8. Horizontal Scaling Constants for STAAR 3–8

Horizontal Scaling Constants for STAAR 3–8				
STAAR Assessment			*A*	*B*
Grade	Language	Content Area		
4	English	Writing	522.0322	3300.4769
7	English	Writing	505.0725	3227.2391
5	English	Science	439.1622	2673.7302
8	English	Science	516.7927	3051.6854
8	English	Social Studies	468.3839	2849.6491
4	Spanish	Writing	574.7681	3230.9603
5	Spanish	Science	439.1622	2673.7302

Table 4.9. Horizontal Scaling Constants for STAAR EOC

Horizontal Scaling Constants for STAAR EOC		
STAAR Assessment	*A*	*B*
English I reading	209.9857	1740.6680
English II reading	241.8717	1835.2850
English III reading	239.6718	1857.6350
English I writing	243.3990	1642.6900
English II writing	216.0176	1637.7380
English III writing	212.0993	1697.9710
Algebra I	441.1057	3448.1767
Algebra II	491.1456	3898.8240
Geometry	448.9410	3700.5560
Biology	452.5972	3558.7180
Chemistry	511.4592	3668.5740
Physics	504.8445	3790.4900
World geography	467.5174	3490.4060
World history	694.0511	3652.9740
U.S. history	494.3999	3577.7820

Vertical Reporting Scales

As required by TEC §39.036, TEA developed vertical scales for assessing student performance in grades 3–8 for reading and mathematics. Vertical scales were developed for the following grades and content areas in STAAR:

- STAAR English grades 3–8 reading
- STAAR English grades 3–8 mathematics
- STAAR Spanish grades 3–5 reading

The vertical scale established for the English versions of grades 3–5 mathematics was also applied to the Spanish versions of grades 3–5 mathematics, because the Spanish versions of the mathematics tests are transadapted from the English test forms. A vertical scale is not required for science, social studies, or writing at the elementary and middle school levels.

For the STAAR 3–8 mathematics and English reading vertical scales, a scale score of 1700 represents the final recommended Level II performance standard for the grade 8 assessment. The standard deviation for these scales was set to 150.

For the STAAR Spanish grade 5 reading assessment, a scale score of 1582 represents the final recommended Level II performance standard. This scale score is set to the same value as the Level II performance standard for the STAAR English grade 5 reading assessment. The standard deviation for the Spanish reading vertical scale was set to 150.

It is important to note that although Level II scale score values are fixed for the highest grade in the vertical scale, the Level II scale score values for the other assessments in the vertical scale are different. However, these Level II scale score values, as well as the Level III scale score values, remain constant over time (not accounting for the phase-in of standards).

The linear transformation of the underlying Rasch proficiency level estimate for vertical scale scores is described by the equation below:

$$SS_j = A \times (\theta_j - V_g) + B \qquad (2)$$

where SS_j is the scale score for student j, θ_j is the Rasch partial-credit model proficiency level estimate for student j, A and B are vertical scale score transformation constants, and V_g is the vertical scaling constant for each grade. The values of A, B, and V_g for the vertically-scaled STAAR assessments are provided in Table 4.10.

Table 4.10. Vertical Scale Score Transformation and Scaling Constants for STAAR

Vertical Scale Score Transformation and Scaling Constants for STAAR					
STAAR Assessment			*A*	*B*	V_g
Grade	Language	Content Area			
3	English/Spanish	Mathematics	111.0491	1535.0921	-2.7895
4	English/Spanish	Mathematics	111.0491	1535.0921	-2.1030
5	English/Spanish	Mathematics	111.0491	1535.0921	-1.6506
6	English	Mathematics	111.0491	1535.0921	-0.9130
7	English	Mathematics	111.0491	1535.0921	-0.4388
8	English	Mathematics	111.0491	1535.0921	0
3	English	Reading	113.7657	1523.3219	-2.0057
4	English	Reading	113.7657	1523.3219	-1.3854
5	English	Reading	113.7657	1523.3219	-1.0586
6	English	Reading	113.7657	1523.3219	-0.6679
7	English	Reading	113.7657	1523.3219	-0.2101
8	English	Reading	113.7657	1523.3219	0
3	Spanish	Reading	135.2141	1393.7820	-0.8513
4	Spanish	Reading	135.2141	1393.7820	-0.2700
5	Spanish	Reading	135.2141	1393.7820	0

Cumulative Score for Graduation

Students receive a scale score for each STAAR EOC assessment. A student's cumulative score for graduation is obtained by combining the individual test scores within each of the four foundation content areas (English reading/writing, mathematics, science, and social studies). In order to graduate, students must reach or exceed their cumulative score target, which is based on the Level II performance standard for each content area. The specific cumulative score target for each student varies depending upon the student's graduation plan and when the student started taking high school courses in Texas.

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 2011–2012 school year, TEA conducted equating activities on the STAAR

assessments that included pre-equating, post-equating, field-test equating, and comparability analysis. Refer to chapter 3, "Standard Technical Processes," for detailed information about equating.

Pre-Equating

The pre-equating process takes place prior to test administration. It links a newly developed test form onto the scale of the item bank through the use of a set of items that appeared previously on one or more test forms. This permits the difficulty level of the newly developed form to be closely determined even prior to its administration. Thus the anticipated raw scores that correspond to scale scores at performance standards can be identified. Pre-equating is conducted for all tests for which scale scores are reported as part of the test construction process. For some tests, post-equating is conducted as well. However, a "pre-equating only" model is used when a small or non-representative sample of students takes the assessment or when faster reporting of scores is a priority.

In STAAR, the pre-equating-only model applies to the mathematics, science, and social studies EOC assessments in an effort to provide faster reporting of scores. In addition, all STAAR assessments administered in the summer or fall, all STAAR L assessments, and all braille forms are only pre-equated either because the testing population is too small (for example, for the STAAR L assessments and braille tests) or because the sample of students taking the test is not representative of the general population (for example, STAAR assessments that are administered in the summer or fall).

Post-Equating

The post-equating process uses data from the operational test administration to re-estimate item difficulties and place them onto the scale of the item bank. For the STAAR assessments, post-equating uses a conventional common-item non-equivalent groups equating design that is described in greater detail in chapter 3, "Standard Technical Processes." Post-equating is conducted for all primary STAAR 3–8 assessments as well as the reading and writing EOC assessments that are administered in the spring.

The data used for post-equating in spring 2012 varied depending on the assessment. For reading and writing EOC assessments, in order to meet reporting deadlines, a representative sample of the population in excess of 100,000 students was used. Because performance standards were not set until fall 2012 for STAAR 3–8, the entire population of test takers was used for the STAAR 3–8 post-equating process.

However, in subsequent school years, it is expected that the English versions of the STAAR 3–8 assessments will use a sample for post-equating in order to meet reporting deadlines. The STAAR Spanish reading and Spanish grade 4 writing assessments, on the other hand, will include nearly the entire population of test takers because the population is relatively small.

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 and operational items on each test form. This results in student performance data that are more stable.

In the 2011–2012 school year, field-test equating was conducted for all the STAAR assessments through an embedded field-test design. In general, a number of multiple-choice field-test items were embedded in each STAAR assessment. Additionally, for grade 7 writing and writing EOC assessments, there was an embedded field-test written composition prompt. For reading EOC assessments there was also an embedded short answer question.

Comparability Analyses

When tests are administered both online and on paper (as is the case for STAAR EOC), the Standards for Educational and Psychological Testing (AERA, APA, NCME, 1999) require that comparability information regarding test scores that result from online and paper modes be collected and evaluated.

As part of the field-test equating process for STAAR EOC during the spring administrations in 2009, 2010, and 2011, comparability analyses were conducted for all EOC assessments. The results of these analyses showed no effect of testing mode for the mathematics, science, and social studies EOC assessments. However, some differences in student performance between online and paper modes were observed for the STAAR English assessments. As a result, additional studies to evaluate the comparability between online and paper testing modes were planned for the English assessments.

In spring 2012, comparability analyses were conducted for the English I assessments in both reading and writing to determine whether the use of the same RSSS conversion table for both testing modes (online and paper) was warranted. The matched sample comparability analyses method (MSCA) (Way, Davis, & Fitzpatrick, 2006) was used to evaluate comparability at the test level. Comparability analyses were not conducted for the English II and English III assessments due to small sample sizes for the online tests.

For English I reading, the comparability analysis suggested an effect of testing mode. Consequently, comparability adjustments were made to generate a separate RSSS table for the online test. These adjustments involved using a raw score equivalency table, which maps each raw score on the paper test to an equivalent raw score on the online test in order to produce equivalent scale scores for the online English I reading assessment. The same comparability

adjustments were also used to generate the RSSS table for the online English I reading assessment for the summer (July) 2012 administration, and these adjustments will be used for all future online English I reading assessments.

For the English I writing assessment, the comparability analyses found no effect of testing mode. Therefore, the same RSSS conversion table was used for both the paper and the online versions of the English I writing assessment given in the same administration for spring 2012 and summer 2012. For future English I writing assessments, one RSSS conversion table will continue to be used for both the paper and online versions.

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, reliability for the STAAR test score 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 tests involving dichotomously scored (i.e., multiple-choice and gridded response) items, the KR20 was used to calculate the reliability estimates. For tests involving a combination of dichotomous and polytomous constructed response items, the stratified coefficient alpha was used to calculate the reliability estimates.

As a general rule, reliability coefficients 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 primary STAAR English and STAAR Spanish assessments administered in spring 2012, the internal consistency estimates ranged from 0.81 to 0.93. 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 0.91, for female only was 0.91, for male only was 0.91, for African American only was 0.88, for Hispanic only was 0.89, and for white only was 0.91.

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 assessment. 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 full assessment. For example, the STAAR English grade 5 mathematics reporting category "Patterns, Relationships, and Algebraic Reasoning" contains six items. The estimated reliability for the scores in this reporting category is 0.56. 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 and for student groups for all primary Spring 2012 STAAR assessments (including STAAR L and STAAR Spanish) are provided in Appendix B.

Classical Standard Error of Measurement (SEM)

Classical standard error of measurement (SEM) represents the amount of variance in a score that results from 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 primary STAAR assessments in spring 2012, SEM values are generally between 2 to 4 raw score points. The SEM values for the primary STAAR assessments (including STAAR L and STAAR Spanish) administered in spring 2012, are provided in Appendix B.

Conditional Standard Error of Measurement (CSEM)

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. CSEM values for all spring 2012, STAAR administrations (including STAAR L and STAAR Spanish) are provided in Appendix B.

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 all primary spring 2012 STAAR administrations (including STAAR L and STAAR Spanish) are provided in Appendix B.

Validity

Validity refers to the extent to which a test measures what it is intended to measure. The results of the STAAR assessments are used to guide educational planning related to the knowledge and skills that students are acquiring in each academic content area. Validity evidence for an assessment can come from a variety of sources, including test content, response processes, internal structure, relationships with other variables, and analysis of the consequences of testing.

Texas collects validity evidence annually to support the many uses of the STAAR test scores. Texas follows national standards of best practice to continue to build its body of validity evidence for the STAAR assessments. The Texas Technical Advisory Committee (TTAC), a panel of national testing experts created specifically for the Texas assessment program, provides ongoing input to TEA about STAAR validity evidence. The sections that follow describe how different types of validity evidence were collected for the STAAR 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

Validity evidence based on test content refers to evidence of the relationship between tested content and the construct that the assessment is intended to measure. The STAAR 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 assessments in order to support the use of the STAAR scores in making inferences about students' knowledge and understanding of the TEKS.

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 the STAAR 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 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 through the test-development process, as various STAAR 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 items and confirm that each item appropriately measures the TEKS to which it is aligned. These committees also review items for content and bias. Two distinct types of educator committees are regularly convened to support the validity of test content. "Item review" committees are made up of Texas K-12 educators, and these committees revise and edit items, as appropriate, prior to test administrations. These item-review committees are convened for all STAAR assessments. "Content validation" committees, by comparison, are made up of university faculty who are experts in the relevant subject matter. Content validation committees review items to ensure that relevant content is being represented and assessed fairly and appropriately by test items. Though these committees do not edit or revise items, they can recommend that certain items be replaced on operational assessments. Content validation committees are convened for all STAAR EOC assessments: English I (reading and writing), English II (reading and writing), English III (reading and writing), Algebra I, geometry, Algebra II, biology, chemistry, physics, world geography, world history, and U.S. history.

Test Developer Input

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

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 assessments reflects the accurate measurement of the construct.

Item Types

Student response processes on the STAAR assessments vary according to item type and administration mode. Across STAAR, four types of responses are required of students: multiple-choice items on all assessments; gridded-response items on mathematics and science assessments; short constructed-response items on English I, II, and III reading assessments; and written compositions on grade 4, grade 7, and English I, II, and III writing assessments.

The multiple-choice format is frequently used because it most closely resembles what students typically experience in classroom testing. Multiple-choice items are developed so that students must recall and apply what they have learned about the course, thereby supporting an accurate measurement of the construct being assessed.

Gridded-response items require students to determine a numerical answer and then record (or, for online tests, type in) their answer using a griddable item response sheet. This item type facilitates the assessment of the students' knowledge and skills at an even deeper level by requiring students to generate answers independently without being influenced by given answer options. This item type is helpful in assessing the constructs of mathematics and science.

Constructed-response item types, such as short answer items and written compositions, require students to evaluate and synthesize information and then construct (i.e., write) an original response. This provides an authentic way to evaluate how well students understand what they have read and how well they can compose a written response within various genres of writing.

In addition, passage-based multiple-choice and short answer responses are used for the STAAR reading and writing assessments. Passage-based items can be an individual item or a group of items associated with a common stimulus, such as a literary selection or an informational passage. The use of passage-based items requires students to apply their knowledge and skills within the context introduced by the stimulus in order to respond correctly to the items. This skill is important in assessing the language arts constructs of reading and writing.

To support these mixed-format assessments, Texas gathers theoretical and empirical evidence that supports the expectation that the way students respond to test items does not add construct-irrelevant variance. 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 assessments, the item types were included in cognitive labs to study the way students engage with the various item presentations. Since item types were determined to be appropriate for STAAR, evidence about student responses has been 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 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 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.

For multiple-choice items, statistical key checks are conducted for all STAAR assessments (including STAAR L and STAAR Spanish) during the equating process. A statistical keycheck is a procedure in which the statistical properties of all items on every test form are computed. Items whose statistics do not meet predetermined

criteria (see Item Analyses section in chapter 3, "Standard Technical Processes") are flagged for further review by content experts to verify that they are correctly keyed and scored.

For constructed-response items, such as short answer items and written compositions, rubrics are used by human readers to score student responses. All score rubrics for the STAAR assessments have been validated by educator committees and content experts. In addition, TEA has implemented a rigorous scoring process for the constructed-response items that includes training and qualification requirements for readers; ongoing monitoring during scoring; adjudication and resolution processes for student responses that do not meet the perfect/adjacent scoring requirements; and rescoring of responses for which concerns have been raised regarding the assigned score by districts, campuses, or teachers. Score reliability and validity indices are also generated and evaluated for every STAAR assessment. In the context of scoring constructed-response items, reliability is also supported by reader agreement rates, or the correlation of scores from the first reading and the second reading of a student response. Validity is further evaluated through the use of validity papers, which are student responses from the current administrations that are representative of different levels of writing performance based on the scoring rubrics. Validity papers are identified by scoring supervisors and scoring directors and approved by the TEA English language arts and writing team. Then they are given to readers systematically on a daily basis throughout the scoring project. An important feature of validities is that these papers are not identifiable as such; in fact, they are indistinguishable from unscored student responses. Each reader's daily scores on validity papers are compared with the scores approved by TEA. Validity papers are used throughout the scoring project as a primary quality control measure, the purpose of which is to ensure that readers are scoring accurately and reliably on a daily basis and across time.

Tables 4.11 and 4.12 summarize reader agreement rates by grade and the validity packet result, respectively, for those spring 2012 STAAR assessments that included constructed-response items. 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 4.11. Summary of Reader Agreement (Reliability) for 2012 STAAR

STAAR Assessment	Number of Responses Read	Agreement Rate (%) After 2 Readings	Number of Third Readings	Agreement Rate (%) After 3 Readings
Grade 4 writing (English)	668,326	61%	15,480	98%
Grade 4 writing (Spanish)	49,050	63%	1,121	98%
Grade 7 writing	698,602	60%	16,026	98%
English I reading	643,644	70%	192,586	97%
English I writing	643,642	59%	15,787	99%
English II reading	60,104	74%	15,838	98%
English II writing	60,104	69%	1,121	99%
English III reading	96,354	76%	22,777	98%
English III writing	96,354	74%	2,085	99%

Table 4.12. Summary of Validity Packet Results for 2012 STAAR

STAAR Assessment	Agreement Rate (%)	STAAR Assessment	Agreement Rate (%)	STAAR Assessment	Agreement Rate (%)
Grade 4 writing (English)	75%	English I reading	87%	English I writing	77%
Grade 4 writing (Spanish)	72%	English II reading	91%	English II writing	78%
Grade 7 writing	77%	English III reading	84%	English III writing	76%

Administration Mode

All STAAR EOC assessments are administered in both paper and online testing modes. Texas has conducted comparability studies comparing student responses in online and paper versions for all STAAR EOC assessments. In these studies, the comparability of scores was evaluated to determine whether an effect due to testing mode exists.

As described in the Comparability Analysis section, for all STAAR EOC mathematics, science, and social studies assessments as well as the STAAR English I writing assessments, no significant effects due to testing mode were observed. The one exception is STAAR English I reading, for which a different online score conversion table was produced by modifying the paper RSSS table to adjust for the effect due to testing mode. Additional studies have been planned in order to continue the evaluation of comparability across modes for the English II and English III assessments in 2012–2013 and 2013–2014, respectively.

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 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 assessments that have a combination of multiple-choice items and short answer items or written compositions (i.e., the writing assessments and the English I, II, and III reading assessments), internal consistency is evaluated using the stratified coefficient alpha. 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 the Reliability section of this chapter.

Evidence Based on Relationships to Other Variables

Another method by which Texas provides validity evidence for the STAAR assessments is by analyzing the relationship between performance on STAAR and performance on other assessments, a process that supports what is referred to as criterion-related validity. By examining this relationship, evidence can be collected to show that those are consistent with the relationships expected based on empirical data. Numerous research studies were conducted during the 2011–2012 school year to evaluate the relationships between scores on the STAAR assessments and other related variables. These studies were initially used to inform the establishment of performance standards across the STAAR assessments and can be grouped into six categories:

1. STAAR-to-TAKS comparison studies, which link performance on the STAAR assessments to performance on TAKS assessments (for example, STAAR grade 7 mathematics and TAKS grade 7 mathematics)
2. STAAR linking studies, which link performance on the STAAR assessments across grade levels or courses in the same content areas (for example, grade 4 reading and grade 5 reading, or Algebra I and Algebra II)
3. STAAR correlation estimates, which evaluate the strength of the relationship (or the lack of one) between scores on the STAAR assessments across different content areas (for example, grade 4 mathematics and grade 4 reading, or biology and world geography)
4. Grade correlation studies, which link performance on the STAAR EOC assessments to course grades

5. External validity studies, which link performance on the STAAR assessments to external measures (specifically: SAT, ACT, THEA, ACCUPLACER, Explore, and Readistep)
6. College students taking STAAR studies, which link performance on the STAAR EOC assessments to college course grades

Results from all of these categories of studies are not given in this chapter, but detailed information can be found by referring to the following resources.

- STAAR correlation estimates based on the student performance on the primary 2011–2012 STAAR administrations are provided in Appendix B.
- For a detailed discussion of the process by which research studies were planned and implemented to support the STAAR standard-setting process, refer to chapter 3 of the *STAAR Standard Setting Technical Report,* available on the STAAR Resources page of TEA's Student Assessment Division website.
- Results for most of the research studies are available in individual study profiles and in an overall study quality summary on the STAAR EOC External Validity Studies page of TEA's Student Assessment Division website.

The evaluation of relationships between STAAR performance and other measures will continue in future years. The 2011–2012 validity studies will be updated as additional data are collected.

Evidence Based on the Consequences of Testing

Another way to collect validity evidence is by documenting the intended and unintended consequences of administering an assessment. Based on the requirements in state statue, intended consequences of the STAAR assessment include these aims:

- STAAR assessments can serve as an indicator of whether students are ready to succeed in the next course or grade.
- STAAR assessments can be used to identify students in need of academic intervention or remediation.
- STAAR EOC assessments can be used to indicate the degree to which students are ready to graduate from high schools in Texas and to succeed in postsecondary endeavors such as college and careers.
- Performance on the STAAR assessments can be used to track the academic progress of students across years.
- Performance on the STAAR assessments can be used to gauge the educational effectiveness of campuses and districts.

The following validity evidence supports these intended consequences.

- One of the first steps in the STAAR standard-setting process was to develop performance labels and policy definitions. (Refer to the Performance Standards section in this chapter for the STAAR performance labels and policy definitions.) The STAAR performance labels and policy definitions, which were recommended by a committee of Texas educators and policy experts, contain specific language about the degree of academic readiness that students in each STAAR performance category possesses.
- All subsequent steps in the STAAR standard-setting process, including the development of the PLDs and the work performed by the policy committee and standard-setting committees, were grounded in these performance labels and policy definitions.
- Research studies used to support standard setting were designed so that students' performance on STAAR can be used to draw inferences about students' readiness to succeed in the next grade, course, or college and careers.
- STAAR is the testing requirements for high school graduation in Texas.
- State and federal accountability ratings for districts and campuses is based on the aggregate performance of its students on STAAR.
- Performance on STAAR is also used in determining distinction designations and is included in the performance reports for districts and campuses.

The collection of consequential validity evidence, particularly evidence related to the unintended consequences of the assessments, typically occurs after a program has been in place for some time and is intended to continue in future years. Examples of such additional evidence the following:

- Longitudinal data that show how students who took STAAR performed in subsequent grades and courses in the same content area and in college and careers
- Survey data that show the impact of STAAR on classroom instruction, resource allocation, and teacher development at both the secondary and postsecondary levels

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.

For STAAR, progress measures are legislatively mandated as an essential aspect of the program. Specifically, the STAAR progress measures must reflect annual improvement and indicate the progress required for students to perform satisfactorily in grades 5 and 8 and on the EOC assessments required for graduation (TEC §39.034).

Progress measures require multiple years of performance data and therefore were not available for STAAR in 2011–2012. However, development of these measures is well 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
- Content relationships among STAAR 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 the STAAR progress measures. Several options for progress measures were presented to the TTAC, a national group of psychometric experts, who then provided recommendations and guidance for the development activities. Progress measures were also discussed with the Accountability Technical Advisory Committee (ATAC) and the Accountability Policy Advisory Committee (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 progress measures continue to be developed and refined. Whenever possible, pilot studies and empirical data are also being used to inform these development activities. Student progress information for STAAR will be available for the first time in 2013.

Sampling

Sampling plays a critical role in the research and annual development activities that are necessary in supporting the Texas assessment program. The assessment program affects all students in Texas. A sample is a group of students smaller than the population that can be used to represent the overall population. Through the careful selection of student samples, TEA is able to gather reliable information about student performance on its tests while minimizing the burden placed on campuses and districts.

During the 2011–2012 school year, sampling was conducted for the following STAAR EOC assessments: English II reading and writing, English III reading and writing, Algebra II, chemistry, physics, world history, and U.S. history. Sampling was also conducted for

two STAAR L assessments: geometry and world history. Campuses sampled for the STAAR EOC assessments tested students on paper, and campuses sampled for the two STAAR L assessments tested students online. Participation in these STAAR assessments was mandatory for selected campuses.

Sampling Process

A stratified sampling design was used for the STAAR program in which the sampling unit was the campus but the observation unit was the student. Each campus was classified into one of five strata based on campus size or the estimated student count for each STAAR assessment. Because the campus was the sampling unit, it was necessary to obtain the student course enrollment from each campus to provide an estimate of the number of students that would be participating in the corresponding STAAR assessment. The estimated student counts for each campus were then based on the number of students who were enrolled in each of the sampled courses in the 2010–2011 school year.

The following factors were considered in determining each STAAR sample for spring 2012.

- The sample was selected in order to be representative of the overall population of Texas high school students taking the course in terms of ethnic composition and campus size.
- The sample was selected to include a minimum of 280 students per form from each major ethnic group and gender groups (i.e., African American, Hispanic, white, male, female).
- Campuses were not assigned to take more than three assessments, with English reading and English writing counting as two separate assessments.
- Campuses were assigned a maximum of two assessments per grade level, based on the typical course-taking sequence.
- Campuses sampled for the STAAR L world history test were not sampled for the STAAR world history test.
- Each sampled campus was required to test all of its enrolled students in the course, regardless of grade, up to a maximum of 300 students during the testing window.

To reduce the field-testing burden on districts and campuses, eligibility criteria were used to eliminate the following campuses from the sample:

- Campuses with fewer than 15 students enrolled in the course for the selected assessment
- Campuses that are part of the Juvenile Justice Alternative Education Program (JJAEP), Disciplinary Alternative Education Programs (DAEP), or Texas Youth Commission (TYC)

Additionally, the process for selecting campuses included these two steps.

1. One campus was randomly selected from the eligible campuses. The number of tests assigned to each campus and grade was taken into consideration so that no campus was assigned more than three assessments overall, and more than two assessments per grade level.
2. The number of students in the sample was evaluated relative to the target total number of students after the campus had been selected. Additional campuses were sampled until the target number of students was reached.

The final sample was determined after evaluating three key elements: comparability with the statewide ethnic percentages, the number of campuses, and the number of students. A summary of the number of campuses and students selected for the 2011–2012 STAAR samples is provided in Table 4.13.

Table 4.13. Sampling Summary for Spring 2012 STAAR EOC Assessments

Course	Paper			
	# Districts	# Campuses	Target *n*-Count	Sample *n*-Count
English II (reading and writing)	160	181	26,775	27,033
English III (reading and writing)	344	404	50,673	50,704
Algebra II	314	396	40,270	40,316
Chemistry	408	499	61,452	61,500
Physics	381	497	55,372	55,291
World history	215	243	28,651	28,705
US history	393	490	51,350	51,675
Total			**314,543**	**315,224**

Table 4.14. Sampling Summary for Spring 2012 STAAR L Assessments

Course	English Language Learners (Online)			
	# Districts	# Campuses	Target *n*-Count	Sample *n*-Count
Geometry	16	117	2,000	2,004
World history	17	119	2,000	2,005
Total			**4,000**	**4,009**
Course	**Non-English Language Learners (Online)**			
	# Districts	**# Campuses**	**Target *n*-Count**	**Sample *n*-Count**
Geometry	13	14	1,500	1,514
World history	14	14	1,500	1,500
Total			**3,000**	**3,014**

Test Results

Appendix B provides scale score distributions and summary statistics, RSSS conversion tables, as well as mean p-values and reliability estimates by reporting category and content area for all primary STAAR assessments, including STAAR L and STAAR Spanish, administered in spring 2012. Table 4.15 shows the spring 2012 pass rates for the STAAR assessments.

Table 4.15. STAAR Spring 2012 Pass Rates (at the Phase-in 1 Standard)

Mathematics	Grade 3	68%
	Grade 4	68%
	Grade 5	77%
	Grade 6	77%
	Grade 7	71%
	Grade 8	76%
	Algebra I	83%
	Geometry*	98%
	Algebra II*	65%
Mathematics (Spanish)	Grade 3	56%
	Grade 4	53%
	Grade 5	49%
Reading	Grade 3	76%
	Grade 4	77%
	Grade 5	77%
	Grade 6	75%
	Grade 7	76%
	Grade 8	80%
	English I	68%
	English II*	61%
	English III*	50%
Reading (Spanish)	Grade 3	65%
	Grade 4	60%
	Grade 5	69%
Writing	Grade 4	71%
	Grade 7	71%
	English I	55%
	English II*	46%
	English III*	38%
Writing (Spanish)	Grade 4	64%
Science	Grade 5	73%
	Grade 8	70%
	Biology	87%
	Chemistry*	54%
	Physics*	61%
Science (Spanish)	Grade 5	41%
Social Studies	Grade 8	59%
	World geography	81%
	World history*	43%
	US history*	63%

*Assessments typically taken in grade 10 and 11, STAAR was the testing requirement for students enrolled in grade 9 or below in 2011–2012.