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

Overview

Participation Requirements

Testing Requirements for Graduation

Test Development

Training

Test Administrations

Testing Accommodations

Student Success Initiative

Scores and Reports

Parent Brochures

Released Tests

Performance Standards

Scaling

Equating

Reliability

Validity

Measures of Student Progress

Sampling

Test Results

Overview

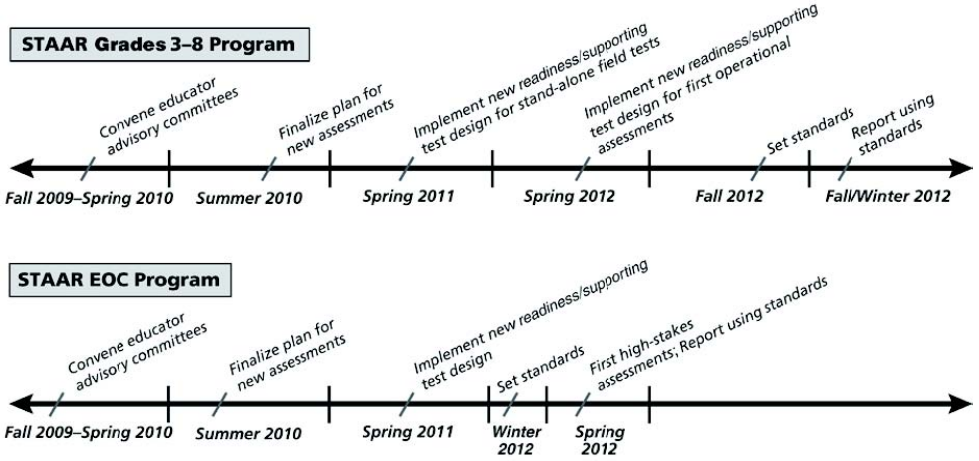
In May 2007, the 80th Texas Legislature enacted Senate Bill (SB) 1031, which expanded the role of the end-of-course (EOC) assessment program. The bill phased out the Texas Assessment of Knowledge and Skills (TAKS) assessments for grades 9–11 and replaced them with the EOC assessments as a component of 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 House Bill (HB) 3, which called for a unified, comprehensive assessment program, the State of Texas Assessments of Academic Readiness (STAAR), to replace the existing TAKS program for all grades and courses beginning in spring 2012 for students first enrolled in grade 9 or below. The Texas Education Agency (TEA), in collaboration with the Texas Higher Education Coordinating Board (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 included

- 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 of the standards for STAAR EOC assessments. 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.

In the 2013–2014 school year, the majority of students who were in grades 9 through 11 were required to take the STAAR EOC assessments for courses in which they were enrolled. Students in grade 12 or who repeated grade 11 in the 2013–2014 school year graduated 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 2013–2014, grade 4 English and Spanish prompt studies were administered to grade 4 students on selected campuses. For these students, participation in the prompt study assessments was mandatory, given that data from a sample of students, representative of the state, are critical to the state's ability to build high-quality assessments. Refer to the "Sampling" section for more information.

In June 2013, the 83rd Texas Legislature passed HB 5, which reduced the number of STAAR EOC assessments needed for graduation from fifteen to five—Algebra I, biology, English I, English II, and U.S. history. Only the five STAAR EOC assessments required by HB 5 were administered in 2013–2014. To meet the requirements of HB 5, the English I and English II assessments were redesigned as combined reading and writing assessments and were first administered in spring 2014[1].

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 Texas Essential Knowledge and Skills (TEKS). Every item on every STAAR assessment 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. During the 2013–2014 school year, STAAR EOC assessments were available to students in Algebra I, biology, English I (reading and writing), English II (reading and writing), and U.S. history.

STAAR Spanish

The STAAR Spanish assessments are designed to measure the academic skills of students who receive academic instruction in Spanish while they learn English. The STAAR Spanish assessments are offered at grades 3–5, including grades 3–5 mathematics and reading, grade 4 writing, and grade 5 science. The English and

[1] Separate English reading and writing tests were administered in July and December 2013.

Spanish versions of STAAR assessments have the same test blueprint and assess the same TEKS student expectations.

STAAR L

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

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

Table 4.1. 2013–2014 STAAR Assessments

Grade/Course	Content Area
Grade 3 (English, Spanish, and L*)	mathematics and reading
Grade 4 (English, Spanish, and L*)	mathematics, reading, and writing
Grade 5 (English, Spanish, and L*)	mathematics, reading, and science
Grade 6 (English and L*)	mathematics and reading
Grade 7 (English and L*)	mathematics, reading, and writing
Grade 8 (English and L*)	mathematics, reading, science, and social studies
End-of-Course Assessments (English and L*)	Algebra I, biology, English I, English II, and U.S. history

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

Participation Requirements

According to federal regulations, all students must be assessed on grade-level curricula. Consequently, 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 chapters 5 and 6, respectively.

Figure 4.2. ELL Participation in STAAR

ELL Participation in STAAR	
STAAR	■ General state assessment ■ Taken by ELLs not administered an assessment below ■ 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 an ELL whose parent or guardian has declined bilingual/ESL program services ■ Linguistic accommodations not permitted because students testing in native language
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 school 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 permitted 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 in grade 9 or below for the first time in the 2011–2012 school year are required to take the STAAR EOC assessments rather than TAKS as their testing requirement for graduation. With the passage of HB 3, performance on the STAAR EOC assessments was linked to a student's graduation program. The provisions of HB 3 were modified by HB 5 passed in June 2013. Under HB 5, in order to graduate, a student must score at or above the Level II: Satisfactory Academic Performance standard on five STAAR EOC assessments: Algebra I, English I, English II, biology, and U.S. history. Prior to spring 2014, students were eligible to meet their English I and II graduation requirements using performance on the separate English reading and writing assessments.

Students eligible to take STAAR Modified or STAAR Alternate are not required to pass EOC assessments in order to graduate. 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. For detailed information regarding each step of the test-development process, refer to chapter 2, "Building a High-Quality Assessment System." Chapter 2 outlines the processes that are used to develop a framework for each STAAR assessment and explains ongoing test development. TEA relies greatly on input from educators to ensure that all measures of learning for Texas public school students are equitable and accurate.

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 were eligible for assessment and which were critical for student success and therefore 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 and 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- 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 2014 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 2014 STAAR 3–8 assessments were developed and field-tested through stand-alone field tests for grades 4 and 7 writing in 2011 and through embedded field testing on operational STAAR assessments

in 2012 and 2013. For STAAR EOC assessments, the items eligible for placement were developed and field-tested through both stand-alone field testing and embedded field testing in operational EOC assessments administered in 2008 through 2013.

- Item-development activities continue annually to create multiple versions of STAAR, to support the periodic release of test questions and test forms, and to build the item bank to a depth necessary to support the program. These items are subsequently field-tested, and the field-test data are reviewed. Field-tested items that meet established 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 align closely with the cognitive complexity evident in the TEKS. Items developed for STAAR meet these established guidelines. For example, additional open-ended items that require students to derive an answer independently are developed for science and mathematics courses. For social studies, science, and mathematics, items that measure process skills in context rather than in isolation are developed. In reading, greater emphasis is given to critical analysis than to literal understanding. In writing, prompts are developed to support personal narrative, expository, persuasive, and analytical writing.
- STAAR EOC test items are also developed so that they can be delivered in both paper and online test formats. A plan was created to evaluate the comparability of EOC assessment 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, since such issues might create a different student experience when testing online than when testing on paper. Studies were conducted between 2009 and 2014 to evaluate the performance of EOC assessment items in both paper and online formats.

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. Transadaptation involves translating items from English and adapting them as necessary to ensure cultural and linguistic accessibility. Passages and items for the reading and writing tests are developed originally in Spanish. This approach allows the Spanish language arts curriculum to be assessed in a more authentic and meaningful manner.

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 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 assessments 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 are the same for English and Spanish.
- The test blueprints for the English and Spanish assessments are 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 assessments 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 newly developed 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 assessments. 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.

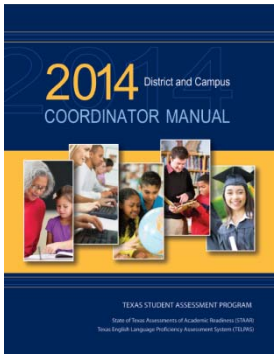
Training

Test administration procedures must be followed in accordance with the directions provided in the Test Security Supplement, the District and Campus Coordinator Manual, and the appropriate test administrator manuals so that all individuals eligible for testing have an equal opportunity to demonstrate their academic knowledge and skills. To help achieve this goal, TEA develops instructional materials to support the training of all testing personnel across the state on proper test-administration procedures. Preparation for test administration begins every year with a training-of-

trainers session for testing coordinators and other personnel from each of the twenty Texas regional education service centers (ESCs), as well as district testing coordinators from the state's 25 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 assessments.

For each test administration in the 2013–2014 school year, ESC personnel and district coordinators were given a district testing coordinator packet, which contained all the information and materials necessary for overseeing test administrations, and copies of the coordinator and test administrator manuals.

District and Campus Coordinator Manual and Test Administrator Manuals



The 2014 District and Campus Coordinator Manual guides district and campus testing coordinators through their responsibilities as they oversee the administration of the Texas assessment program. This manual is made available in time for the annual training of ESC, district, and campus testing coordinators. This manual contains preparation and administration procedures for every assessment program for the 2014 calendar year. Separate test administrator manuals are provided to districts prior to the spring administrations for STAAR grades 3-5, STAAR grades 6-8, and EOC.

Test Administrations

All primary STAAR assessments were administered in spring (March–May) 2014. In addition, STAAR EOC assessments were made available in fall (December) 2013 and summer (July) 2014. STAAR grades 5 and 8 mathematics and reading retests were made available in May and June 2014.

STAAR 3–8 assessments (including STAAR Spanish) were offered as paper tests only, except for grades 4 and 7 mathematics and reading and grade 8 science and social studies, which were also available as online tests for standardized oral administration (SOA). STAAR EOC assessments in 2013–2014 were offered in online and paper modes, while STAAR L assessments were administered primarily online. During the 2013–2014 school year, over 9.3 million STAAR assessments, including STAAR 3–8, STAAR EOC, and STAAR L, were administered. The number of students tested for each STAAR assessment in 2013–2014 is shown in Table 4.2.

Table 4.2. STAAR Assessments Administered in 2013–2014

STAAR Assessment	Assessments Administered
Grade 3 English mathematics	351,877
Grade 3 English reading	336,392
Grade 3 Spanish mathematics	18,774
Grade 3 Spanish reading	37,364
Grade 3 STAAR L mathematics	4,981
Grade 4 English mathematics	353,913
Grade 4 English reading	338,865
Grade 4 English writing	337,238
Grade 4 Spanish mathematics	9,913
Grade 4 Spanish reading	25,122
Grade 4 Spanish writing	26,208
Grade 4 STAAR L mathematics	1,748
Grade 5 English mathematics	480,825
Grade 5 English reading	466,689
Grade 5 English science	356,231
Grade 5 Spanish mathematics	7,155
Grade 5 Spanish reading	16,783
Grade 5 Spanish science	6,571
Grade 5 STAAR L mathematics	3,742
Grade 5 STAAR L science	2,025
Grade 6 mathematics	348,024
Grade 6 reading	355,709
Grade 6 STAAR L mathematics	4,409
Grade 7 mathematics	337,118
Grade 7 reading	365,015
Grade 7 writing	364,747
Grade 7 STAAR L mathematics	5,127

STAAR Assessment	Assessments Administered
Grade 8 mathematics	402,529
Grade 8 reading	439,713
Grade 8 science	355,028
Grade 8 social studies	357,123
Grade 8 STAAR L mathematics	10,181
Grade 8 STAAR L science	5,062
Grade 8 STAAR L social studies	5,058
Algebra I	504,983
STAAR L Algebra I	13,366
English I reading*	117,112
English I writing*	163,035
English I	564,679
English II reading*	56,309
English II writing*	99,195
English II	453,210
Biology	433,630
STAAR L Biology	12,713
U.S. history	340,988
STAAR L U.S. history	2,409

* The STAAR English I reading, English I writing, English II reading, and English II writing assessments were administered as separate assessments for the last time in December 2013.

The Online Test Delivery System

STAAR online assessments are administered 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 test status while the test session is still in progress.

Additional information about the Assessment Management System, such as an overview of the system, minimum system requirements, information on delivery and reporting, and a list of frequently asked questions, is available on the Texas Assessment Management System 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 2013–2014 school year, make-up testing opportunities were available for all administrations.

Out-of-District Testing

For the summer STAAR 3–8 administrations and for all STAAR EOC administrations, students who are unable to test at their home district's designated test site are allowed to test out-of-district. Out-of-district students are required to complete the out-of-district registration form prior to the test administration date. Students must present photo identification at the test administration site on the day of the test. 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 University or University of Texas high school programs.

A campus or district must accommodate the request of an out-of-district student to participate in STAAR administrations 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 and II 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 English as a second language (ESL) dictionary, which uses simple English language as well as 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 Reading Resources page on TEA's Student Assessment Division website.

CALCULATORS

Calculators were provided to students for the STAAR Algebra I and biology assessments. Students were permitted to use their own calculators instead of those provided by the district. Students were permitted to use more than one calculator during the assessment. At minimum, districts were required to provide the following:

- a graphing calculator for each student taking Algebra I (including STAAR L and STAAR Modified)
- a calculator for every five students taking biology (including STAAR L and STAAR Modified)

Any calculator could be used to fulfill the minimum requirements listed above except for those that included a computer algebra system (CAS), that allowed access to the Internet, or that had photographic capabilities. The use of an electronic device that had a calculator as an application (e.g., a cell phone) was not permitted.

All calculator memory had to be cleared to factory default both before and after testing. Any programs or applications had to be removed or disabled prior to testing. Test administrators were instructed to contact a calculator's manufacturer for specific assistance in effectively preparing a calculator for use during testing.

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

Testing Accommodations

Accommodations are practices and procedures that provide equitable access to grade-level or course curriculum during instruction and assessment. Students receiving testing accommodations include students with disabilities, and ELLs who are eligible for linguistic accommodations.

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 student's needs. Not all accommodations that are appropriate for instructional use are appropriate for use on a standardized assessment.

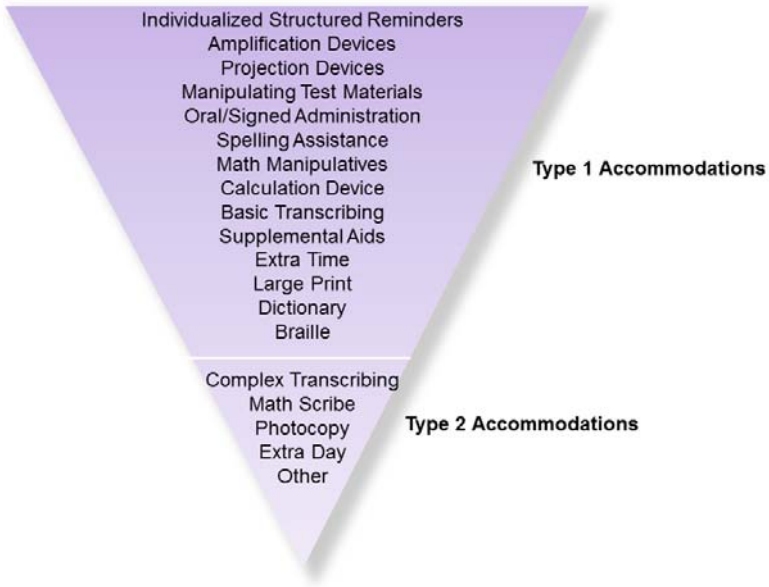
Accommodations are provided on an individual basis and take into consideration the needs of each student. It would be neither appropriate nor effective to try to provide "one size fits all" accommodations to students. For example, one student with a visual impairment might benefit from large-print instructional materials, whereas another student with a visual impairment might benefit more from a magnification device. In most cases, accommodations are selected for a specific student based on their appropriateness/effectiveness 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 state assessments is made by the student's ARD committee or Section 504 placement committee. In those rare instances in which a student does not receive special education or Section 504 services but meets the eligibility criteria, the decision to allow accommodations on the state 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 must investigate whether each accommodation is allowed on a state assessment. The triangle 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 Assessment

Accommodations are available for students with dyslexia and other reading disabilities on the STAAR reading assessments in grade 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 must be carefully considered when determining the most appropriate accommodations for the STAAR assessments.

Students with Visual Impairments

Test administrators receive specific instructions for testing visually impaired students using large print or braille test booklets. Districts are required to indicate on the answer document whether the student used a large-print or braille version of a test. Large-print and braille test booklets are available for all STAAR operational assessments.

Linguistic Accommodations

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

	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
English Language Arts	**Reading and Writing:** ■ Dictionaries of various types* ■ Extra time (same day) **Writing:** ■ Clarification in English of word meaning in writing prompts **English I–II:** ■ Accommodations allowable for reading and writing ■ 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

* 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 STAAR 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.

In 2012, STAAR L accommodations were provided to students by test administrators with a plan to evaluate computerized linguistic accommodations for use in 2013. A study of the clarification accommodations was conducted during the May 2012 operational STAAR EOC assessment. Results of the study indicated that ELL students

used the accommodations more frequently than non-ELL students and that ELL students reported finding the accommodations more helpful than non-ELL students found them to be. Most students reported no difficulty using the online accommodations. In addition, performance data suggested that the accommodations were not impacting student responses in unexpected ways.

Based on these results, STAAR L assessments transitioned to an online interface in spring 2013. The online interface provides students with access to computer-based linguistic accommodations when they take mathematics, science, and social studies tests in grades 3 and above, including EOC assessments. The linguistic accommodations provided through the computerized assessment include the following:

- Clarification: As students test, they are able to click on pre-identified text to see definitions, synonyms, paraphrases, pictures, and animations, as needed, to improve comprehension of words and language structures likely to be unfamiliar to students at the targeted proficiency levels. Content terms assessed are not eligible for clarification. Two interface design features are used to provide this accommodation.
    - Pop-ups/Word Tool: Using the mouse cursor, students can click on pre-identified words and phrases, which open a window showing the clarification (definition, synonym, picture, animation).
    - Stickers/Sentence Tool: For larger portions of text, students can click on an icon and see a paraphrase that unrolls like a sticker on top of the original text. Stickers are used in items in which language is more appropriately clarified through a paraphrase or when the number of individual pop-ups needed to clarify the meaning of the text would impede comprehension or appear too cluttered on the computer screen. Stickers are used less frequently than pop-ups.
- Read-aloud: Interface functionality enables students to hear words read aloud. This accommodation may be beneficial for ELLs at lower proficiency levels who have difficulty pronouncing English words and decoding them as they read due to phonological differences between their native language and English.

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

Oral Administration

Oral administration is an accommodation that allows for test questions and answer choices for mathematics, reading, science, and/or social studies tests to be read aloud or signed to eligible students taking the STAAR assessments. Two levels of reading support are available: reading questions and answer choices at the word-by-word level at the student's request, and reading all questions and answer choices in a continuous reading format. Required reference materials (e.g., a dictionary) and allowable

supplemental aids (e.g., a list of grammar rules or a math chart) may be read to students, but the test administrator cannot interpret or help apply the information contained within the aids. STAAR may be administered orally to individual students or to a group of students, depending on student needs. All references to reading support during an oral administration also apply to signing during a signed administration.

A student is eligible for an oral administration if he or she routinely and effectively uses this accommodation during classroom instruction and testing and meets at least one of the following criteria.

- The student receives special education services and is identified with dyslexia or has evidence of reading difficulties.
- The student receives Section 504 services and is identified with dyslexia or has evidence of reading difficulties.
- The student does not receive special education or Section 504 services but is identified with dyslexia.

The ARD committee or Section 504 committee decides whether a student exhibits evidence of a reading difficulty, which is defined as a problem with reading that may be caused by a learning disability or other conditions such as ADHD, an emotional or behavioral disability, or processing or memory issues.

STANDARDIZED ORAL ADMINISTRATION (SOA)

In spring 2013 a small-scale trial of computer-delivered SOA was offered to a sample of students taking STAAR grade 4 mathematics, grade 8 science, or grade 8 social studies. The reasons for introducing a computer-based solution to oral administration was twofold: to reduce district testing costs and teacher burden and to improve the standardization of the oral administration, thereby reducing variance in test scores caused by possible differences in the quality of the oral administration when delivered by a live test administrator. The SOA is designed to provide a similar level of reading support as that provided by a teacher during an oral administration. The support includes text-to-speech that uses synchronized text highlighting and student control over the type of audio playback, the audio volume, and the speed of speech.

Cognitive labs, a small-scale trial administration, and post-administration surveys were used to evaluate the quality of SOA and to identify possible difficulties facing students in the SOA computer interface. Feedback from students and educators was positive, and none of the findings suggested that the basic process for providing SOA required substantial change. Based on feedback from these studies, SOA was offered across the state on an on-request basis in spring 2014 for the grades 4 and 7 mathematics and reading and grade 8 science and social studies assessments.

Student Success Initiative

The Student Success Initiative (SSI) provides a system of academic support to help students achieve success in mathematics and reading. The SSI was adopted by the Texas Legislature in 1999 to ensure that all students receive the instruction and support they need to be academically successful in mathematics and reading.

Under the SSI grade-advancement requirements, students are required to pass the STAAR grade 5 mathematics and reading assessments to be promoted to sixth grade. Additionally, students are required to pass the STAAR grade 8 mathematics and reading assessments to be promoted to ninth grade.

In 2013–2014, students in grades 5 and 8 had three opportunities (April, May, and June) to pass the STAAR mathematics and reading assessments. If a student did not pass one or both assessments, the school was required to give the student additional instruction after each testing opportunity, and the student had to participate in that instruction. Parents were notified if their child did not pass a STAAR assessment that is required for promotion.

If a student did not pass after the second testing opportunity, a grade placement committee (GPC) was formed. The GPC, which consists of the principal, teacher, and parent or guardian, was required to create an instructional plan based on the individual needs of the student. If a student did not demonstrate proficiency on one or both of these assessments on the third testing opportunity, the student could advance to or be placed in the next grade level only if (1) he or she completed all accelerated instruction required by the GPC, and (2) the GPC determined, by unanimous decision, that the student was likely to perform on grade level by the end of the next school year given additional accelerated instruction during the course of the year. In making promotion decisions, the GPC was required to consider the recommendation of the student's teacher, the student's grades, the student's state assessment scores, and any other relevant academic information.

More information about the Student Success Initiative is available on TEA's Student Assessment Division website.

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.) Along with raw scores, scale scores for all of the STAAR 3–8 and STAAR EOC assessments are reported following each test administration.

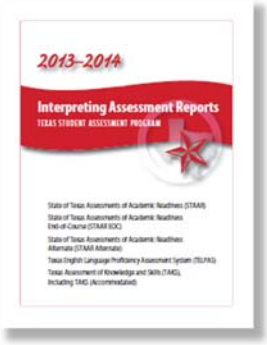
Scale scores are also used to compare the performance of an individual student with the performance of a demographic group, a program group, an entire campus, or a district at a particular grade. For example, the scores for a Hispanic student in a gifted and talented program can be compared with the average scores of other Hispanic students, other gifted and talented students, all students on campus, or any combination of these aggregations at that grade.

Additional Performance Information

Other scores can provide information about a student's relative strengths or weaknesses in core academic areas. For example, reporting category-level data can identify areas where a student might be having difficulty. This identification can help campuses plan the most effective instructional intervention. 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 performance standard and the Level III performance 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, which group had the lowest percentage meeting the Level II performance standard, which group had the highest percentage achieving the Level III performance standard, 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 instructional programs. 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 area 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 relying on test scores alone, to provide a more complete picture of performance.

Parent Brochures

TEA's Student Assessment Division produces a brochure titled "Understanding Your Child's Confidential Student Report (CSR): 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. The brochure, available in both English and Spanish, was provided to districts in May 2014 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.

Released Tests

TEA released the primary administration of the following 2014 general STAAR assessments:

- grades 3–8 mathematics and reading
- grades 4 and 7 writing
- grades 5 and 8 science
- grade 8 social studies
- Algebra I
- English I and II
- biology
- U.S. history

The released tests also included STAAR Spanish grades 3–5 versions of the assessments. All the released STAAR assessments listed above can be found on TEA's website.

Item Analysis Reports

Individual Student Item Analysis Reports and Item Analysis Summary Reports for STAAR were provided to districts. Item Analysis Summary Reports were provided at the campus, district, region, and statewide levels. These reports showed the item number, the reporting category measured by the item, and the percentage of students selecting each possible answer choice for that test item. The summary reports were provided for all students.

These reports were also made available in PDF format on the Texas Assessment Management System. Item analysis reports are available for the following test administrations:

- 2014 April STAAR grades 4 and 7 writing, grades 5 and 8 mathematics, and grades 5 and 8 reading
- 2014 spring STAAR grades 3–8
- 2014 spring STAAR Algebra I, biology, English I, English II, and U.S. history

Students and districts may find it helpful to use the Item Analysis Summary Report in conjunction with the 2014 released STAAR tests. For STAAR grades 3–5, English and Spanish results are reported separately. For additional information, refer to the 2013–2014 Interpreting Assessment Reports.

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. Standard setting is the process of establishing the 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 I: Unsatisfactory Academic Performance
- Level II: Satisfactory Academic Performance
- Level III: Advanced Academic Performance

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

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.

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.

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.

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 state 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) 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 as follows:

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 4.3 and 4.4 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. Although HB 5 legislation removed the requirement for the review of performance standards at least once every three years, step nine is included in the process because this step was considered when setting performance standards for STAAR.

Additional details 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 Performance Standards page of TEA's Student Assessment Division website.

Table 4.3. 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/ subject 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. Develop performance standard ranges	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 are reviewed at least once every three years.**	If applicable

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

** In June 2013, the 83rd Texas Legislature enacted House Bill 5, which removed the requirement to review performance standards (Step 9). Prior to this legislation, Step 9 was scheduled for fall 2014. TEA may review the performance standards, if deemed applicable.

Table 4.4. 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.*	If applicable

* In June 2013, the 83rd Texas Legislature enacted House Bill 5 which removed the requirement to review performance standards (Step 9).

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 ELLs) 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 Algebra I, geometry, Algebra II, English I reading, English II reading, English III reading, English I writing, English II writing, English III writing, 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 4 and 7 writing, grades 5 and 8 science, and grade 8 social studies. Separate committees were convened, when applicable, for the English version and Spanish version of a STAAR assessment.

To meet the requirements of HB 5, passed by the 83rd Texas Legislature in June 2013, the STAAR English assessments were redesigned such that both reading and writing skills would be measured with a single instrument on a single day. Because of this change, TEA convened a standard setting committee in January 2014 that recommended performance standards for English I, II, and III.

Phase-In of Performance Standards

A phase-in period has been implemented for the STAAR Level II: Satisfactory Academic 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. The commissioner of education determined the appropriate timeline for phasing in the performance standards.

During 2013–2014, a two-step phase-in for Level II: Satisfactory Academic Performance was in place for all STAAR 3–8 and EOC assessments. Phase-in 1 performance standards for Level II were in effect for the 2011–2012, 2012–2013, and 2013–2014 school years. Following a review of trends in STAAR performance between

2012 and 2014, the commissioner of education elected to extend phase-in 1 through the 2014–2015 school year. The commissioner also decided to transition to a three-step phase-in, with phase-in 2 performance standards scheduled to begin in 2015–2016, phase-in 3 in 2018–2019, and final recommended in 2021–2022.

The STAAR phase-in performance standard students are assigned for all five EOC assessments is based on the phase-in standard in place when they took an EOC assessment for the first time. For example, the phase-in 1 standard for Level II: Satisfactory Academic Performance was in effect in spring 2014. Therefore, students who first took Algebra I in spring 2014 are held to the phase-in 1 standard for Algebra I, English I, English II, biology, and U.S. history.

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

Table 4.5. STAAR 3–8 Performance Standards Applied in 2013-2014

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

Assessment	Phase-in 1 Level II	Phase-in 2 Level II*	Final Recommended Level II	Final Recommended Level III
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 4.5 shows the values associated with phase-in 2 Level II that were in effect during 2013–2014, when a two-step phase-in was planned. In fall 2014, the values associated with phase-in 2 Level II were changed when a three-step phase-in was introduced.

Table 4.6. STAAR EOC Performance Standards Applied in 2013-2014

Assessment	Phase-in 1 Level II	Phase-in 2 Level II*	Final Recommended Level II	Level III
Algebra I	3500	3750	4000	4333
English I	3750	3900	4000	4691
English II	3750	3900	4000	4831
Biology	3500	3750	4000	4576
U.S. History	3500	3750	4000	4440

*Table 4.6 shows the values associated with phase-in 2 Level II that were in effect during 2013–2014, when a two-step phase-in was planned. In fall 2014, the values associated with phase-in 2 Level II were changed when a three-step phase-in was introduced. Beginning in December 2014, the phase-in 2 and 3 performance standards for Algebra I, biology, and U.S. history will be 3650 and 3850, respectively. The phase-in 2 and 3 performance standards for English I and II will be 3825 and 3925, respectively

Review of Performance Standards

In June 2009, state statutes TEC §39.024 and TEC §39.0242 required that performance standards for the STAAR program be reviewed at least once every three years by gathering data and performing studies to evaluate the performance standards. Step 9 of the standard-setting process—review the STAAR performance standards—was planned for fall 2014. In June 2013, the 83rd Texas Legislature enacted House Bill 5, which removed the requirement to gather data and perform studies to evaluate the performance standards every three years.

Scaling

Scaling is a statistical procedure that places raw scores on a common scoring metric in order to make test scores comparable across test administrations. As with previous Texas assessment programs, the STAAR program uses the Rasch Partial-Credit Model (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, the Rasch scale is then transformed to a more user-friendly metric to facilitate interpretation of the test scores. Details of the RPCM scaling method used in Texas are provided in chapter 3, "Standard Technical Processes."

Reporting Scales

Scale scores for STAAR 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 assessments reported on a horizontal scale, 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.

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

The STAAR scale scores represent linear transformations of Rasch proficiency level estimate (θ). Specifically, the transformation is made by first multiplying θ by a slope

constant (*A*) and then adding an intercept constant (*B*). This operation is described by the equation below:

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

where SS_θ is the scale score for a Rasch proficiency level estimate (θ). *A* and *B* in Equation (1) are referred to as the horizontal scaling constants. These same transformations will be applied each year to the Rasch proficiency level estimates (θ) for that year's set of test items. Values for the horizontal scaling constants are provided in Tables 4.7 and 4.8 for the horizontally scaled STAAR 3–8 and STAAR EOC assessments, respectively.

Table 4.7. 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 4.8. Horizontal Scaling Constants for STAAR EOC

STAAR Assessment	*A*	*B*
Algebra I	441.1057	3448.1767
English I	453.3627	3550.2642
English II	472.9035	3482.6436
Biology	452.5972	3558.7180
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 mathematics and reading. 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 (θ) for vertical scale scores is described by the equation below:

$$SS_\theta = A \times (\theta - V_g) + B \qquad (2)$$

where SS_θ is the scale score for a Rasch proficiency level estimate (θ). A and B are vertical scale score transformation constants, and V_g is the vertical scaling constant for each grade for Equation (2). The values of A, B, and V_g for the vertically scaled STAAR assessments are provided in Table 4.9. Once established, these same transformations are applied each year to the proficiency level estimates for that year's set of test questions.

Table 4.9. Vertical Scale Score Transformation and Scaling Constants for STAAR 3–8 Mathematics and Reading

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	0.0000
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	0.0000
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	0.0000

Cumulative Score for Graduation

Although HB 5 removed the cumulative score requirement for graduation, TEA proposed using the minimum—and cumulative—score concepts during the transition from separate STAAR English I and English II reading and writing assessments to combined assessments as enacted by HB 5. The minimum—and cumulative—score concepts were used to determine whether students taking separate STAAR English I and English II reading and writing assessments in the 2011–2012 or 2012–2013 school years or fall 2013 met the English I and English II requirements for graduation introduced by HB 5 in June 2013.

The concepts were applied within each course, not across courses, and required students who took separate reading and writing assessments to meet three criteria:

- pass one assessment (either reading or writing)
- meet at least the minimum score on the other

- achieve a combined scale score of 3750 (the phase-in 1 standard), which represents the sum of the scale scores needed to reach Level II for reading (1875) and Level II for writing (1875)

Based on this proposal, students taking separate STAAR English I and English II reading and writing assessments in the 2011–2012 or 2012–2013 school years or fall 2013 who met the above criteria have met the graduation requirement under HB 5 for English I and English II.

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 2013–2014 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 administrations of the STAAR 3–8 assessments and the EOC English assessments administered in the spring.

STAAR post-equating is conducted on a sample of students. The requirements for the sample include a minimum sample size of 100,000 students, regional representation similar to the student population, ethnic distribution similar to the student population, and a stable raw score distribution. 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.

Only the test forms with the equating item sets are used in determining the equating constant that will place the base-test Rasch item difficulties on the Rasch scale common across administrations for an assessment. However, student data from all test forms are used in estimating the Rasch item difficulties for the base-test items.

Field-Test Equating

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

In the 2013–2014 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 EOC English assessments, there was an embedded field-test written composition prompt. For EOC English 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 assessments), the Standards for Educational and Psychological Testing (AERA, APA, NCME, 2014) 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. This started with the spring 2012 administration for the English I reading and writing assessments and continued with the English II reading and writing assessments in spring 2013. Information about past English I and II comparability analyses can be found in the past Technical Digests.

In spring 2014, comparability analyses were conducted for the redesigned English I and II assessments to determine whether the use of the same raw score to scale score (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. Both analyses suggested an effect of testing mode. Consequently, comparability adjustments were made to generate separate RSSS tables for online and paper tests for the spring, summer, and fall 2014 administrations of English I and II. 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 assessment. The same comparability adjustments will be used to generate future RSSS tables for the online English I and II assessments.

Reliability

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

During the 2013–2014 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 Kuder-Richardson 20 (KR20) statistic was used to estimate reliability. For tests involving a combination of dichotomous and polytomous constructed-response items, stratified coefficient alpha was used to estimate reliability.

As a general rule, reliability coefficients ranging from 0.70 to 0.79 are considered adequate, those from 0.80 to 0.89 are considered good, and those at 0.90 or above are

considered excellent. However, what is considered appropriate can vary in accordance with how assessment results are used.

For the primary STAAR English and STAAR Spanish assessments administered in spring 2014, the internal consistency estimates ranged from 0.85 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 females only was 0.91, for males only was 0.91, for African Americans only was 0.89, for Hispanics only was 0.90, and for whites only was 0.91.

Because internal consistency estimates typically decrease as the number of test items decrease, internal consistency estimates for individual reporting-categories can be noticeably lower than those for the full assessment. In spring 2014, the internal consistency estimates at the reporting-category level were generally lower than at the total-score level. Lower internal consistency estimates indicate that reporting category scores were not as reliable as those based on the full assessment. For example, the 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 was 0.55. Therefore, the lower reliability at the reporting-category level should to be taken into account when making interpretations of the scores at this level.

Estimates of internal consistency at the overall level, as well as at the level of reporting categories for student groups for primary spring 2014 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 random factors other than what the assessment is intended to measure. The SEM is helpful for quantifying the margin of uncertainty that occurs on every test. For the primary STAAR assessments in spring 2014, 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 2014 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. For the 2013–2014 school year, CSEM values for vertically scaled assessments were approximately 30 to 50 scale score points in the middle of the scale

score ranges. For the remaining STAAR horizontally scaled tests, the CSEM values were approximately 130 to 160 scale score points in the middle of the scale score ranges. CSEM values for all spring 2014 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 primary STAAR assessments during the 2013–2014 school year range from 84.3 to 94.8 percent. Classification accuracy rates for all primary spring 2014 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. When test scores are used to make inferences about student achievement, it is important that the assessment supports those inferences. In other words, the assessment should measure what it was intended to measure in order for any uses and interpretations about test results to be valid.

Texas follows national standards of best practice and collects validity evidence annually to support the interpretations and uses of the STAAR test scores. 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. 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. The sections that follow describe how different types of validity evidence were collected for the STAAR assessments. Refer to chapter 3, "Standard Technical Processes," for additional 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

The TEKS are designed to ensure that Texas students receive a solid education that will enable them to be successful in life, whether they choose to pursue higher

education or enter the workforce directly after graduation. Specifically, the TEKS are aligned to the Texas College and Career Readiness Standards (CCRS). The CCRS specify the knowledge and skills necessary to succeed in entry-level community college and university courses. The CCRS have been incorporated into the secondary TEKS to form a vertically articulated set of curriculum standards. STAAR focuses on fewer skills and addresses those skills in a deeper manner. This is accomplished through the identification of readiness and supporting standards in the TEKS and includes a larger number of items that assess readiness standards in the test blueprint. The STAAR assessments, therefore, focus on the TEKS that are most critical to success in the next grade or course.

EDUCATOR INPUT

As part of the transition from TAKS to STAAR, 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. In addition, prototype items were developed for the STAAR assessments early in the development process. The educator advisory committees reviewed these prototypes to identify how well the 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 educator committees shared their feedback in 2013–2014 about how the student expectations could be effectively assessed.

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 and content-validation committees. Item-review committees are made up of Texas K–12 educators, and these committees revise and edit items, as appropriate, prior to test administrations. Item-review committees are convened for all STAAR 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.

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 show that the way students respond to items on the STAAR assessments reflects 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 answer responses on English I and II assessments; and written compositions on grade 4 and 7 writing and English I and II 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 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 box. 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.

Written compositions require students to construct (i.e., write) an original response to a given prompt. This provides an authentic way to evaluate how well students 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 new item types or changes to the format of existing item types are considered for STAAR assessments, cognitive labs are used to study the way students engage with the various item presentations. After item types are determined to be appropriate for STAAR, evidence about student responses is gathered annually through educator and expert reviews and analyses of

individual student responses to these items. Every year, during item reviews, educators evaluate whether the content for a given item 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 key check 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 the "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 and 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 throughout the scoring project. An important feature of validity papers is that they 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. A more comprehensive description of the scoring process for

constructed-response items is available in chapter 2, "Building a High-Quality Assessment System."

Tables 4.10 and 4.11 summarize reader agreement rates by grade and the validity packet result, respectively, for those spring 2014 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 4.10. Summary of Reader Agreement (Reliability) for 2014 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)	678,636	65%	11,258	100%
Grade 4 writing (Spanish)	52,550	63%	1,359	100%
Grade 7 writing	734,222	63%	11,628	100%
English I short answer	962,242	73%	259,148	97%
English I written composition	494,379	67%	4,112	100%
English II short answer	789,888	67%	263,416	97%
English II written composition	407,517	63%	5,797	100%

Table 4.11. Summary of Validity Packet Results for 2014 STAAR

STAAR Assessment	Agreement Rate (%)	STAAR Assessment	Agreement Rate (%)	STAAR Assessment	Agreement Rate (%)
Grade 4 writing (English)	77%	English I short answer	88%	English I written composition	79%
Grade 4 writing (Spanish)	76%	English II short answer	85%	English II written composition	78%
Grade 7 writing	79%				

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 Analyses" section, for all STAAR EOC mathematics, science, and social studies assessments, no significant effects due to testing mode were observed. The exceptions are STAAR English I and II, for which different online score conversion tables were produced by modifying the respective paper raw score to scale table to adjust for the effect of testing mode.

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 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 and the English assessments), internal consistency is evaluated using 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 Appendix B. The "Reliability" section of this chapter provides a summary of these estimates.

Evidence Based on Relationships to Other Variables

Another method Texas uses to provide validity evidence for the STAAR assessments is analyzing the relationship between performance on STAAR and performance on other assessments, a process that supports what is referred to as criterion-related validity. Evidence can be collected to show that the empirical relationships are consistent with the expected relationships. Numerous research studies were conducted as part of the development of STAAR to evaluate the relationships between scores on the STAAR assessments and other related variables. These studies were used to inform the establishment of performance standards across the STAAR assessments and can be grouped into six categories:

- 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)
- 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 grade 8 mathematics and Algebra I)

- 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 U.S. history)
- Grade correlation studies, which link performance on the STAAR EOC assessments to course grades
- External validity studies, which link performance on the STAAR assessments to external measures (specifically: SAT, ACT, THEA, ACCUPLACER, Explore, and Readistep)
- 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 student performance on the primary 2013–2014 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 "State of Texas Assessments of Academic Readiness (STAAR) Standard Setting Technical Report," available on the STAAR Performance Standards page of TEA's Student Assessment Division website.
- Results for most of the EOC 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.

Evidence Based on the Consequences of Testing

Another method to provide validity evidence is by documenting the intended and unintended consequences of administering an assessment. The collection of consequential validity evidence typically occurs after a program has been in place for some time and on a regular basis. Surveys should be developed and administered to educators and stakeholders in the state to evaluate the impact of STAAR test results on classroom instruction, resource allocation, teacher development, and student academic choices and postsecondary endeavors.

Intended Interpretations and Uses of STAAR Test Scores

Given the important stakes associated with the STAAR assessment program, valid STAAR test scores are critical in supporting their intended interpretations and uses. For the STAAR assessments, the intended interpretations of test results are stated in the policy definitions of the performance levels. Recall that for STAAR, student test results are classified into one of three performance levels:

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

Refer to the "Performance Standards" section in this chapter for the policy definitions of the STAAR performance levels. Each performance level describes a student's knowledge and skills in a content area and a student's level of preparation for the next grade or course. Below are intended uses of STAAR test scores based on the policy definitions for the performance levels.

STUDENT-LEVEL

- Performance on STAAR is one indicator of a student's level of proficiency in a content area or specific course.
- Performance on STAAR is one indicator of a student's readiness for the next course or grade level in the same content area.
- Performance on STAAR is one indicator of a student's possible need for academic intervention.
- Performance on STAAR across years provides one indicator of a student's academic progress within a content area.
- Performance on STAAR may provide information about expected student performance on external assessments, such as the ACT or SAT, that measure similar knowledge and skills.

CAMPUS- OR DISTRICT-LEVEL

- STAAR performance results can be aggregated to provide one indicator of overall student proficiency at a campus or district.
- STAAR performance results can be aggregated to provide one indicator of overall student readiness (for the next course or grade level in the same content area) at a campus or district.
- STAAR performance results can be aggregated across years to provide one indicator of overall student academic progress at a campus or district.

Measures of Student Progress

Student progress measures provide information beyond performance levels by providing a comparison of performance over time. Whereas performance level information describes students' current achievement, progress measures describe students' achievement across multiple years.

Progress measures are legislatively mandated as an essential aspect of the Texas assessment 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).

In 2012–2013, STAAR progress measures were reported for the first time. Specifically, progress measures are computed for mathematics and reading, as well as some writing assessments. For the full list of the grades and content areas for which progress measures were reported in 2013–2014, as well as the planned implementation for upcoming years, see the "STAAR Progress Measures Implementation Schedule" on the STAAR Resources page of TEA's Student Assessment Division website.

For STAAR, progress is measured as a student's gain score, meaning the difference between the score a student achieved in the prior year and the score a student achieved in the current year. These gain scores are then classified as *Did Not Meet*, *Met*, or *Exceeded* in relation to progress targets.

2013–2014 Progress Measure Updates

There were some updates to the implementation of progress measures both to include more students in the calculation of progress measures and to respond to legislative changes made in 2013.

Changes Made to Include More Students in Progress Measure Calculations

- Accelerated students/skipped grades(s)

In 2013, students had to follow the typical grade/course sequence and test in consecutive grades/courses in order to receive a progress measure, with the exception of students who went from grade 7 mathematics to Algebra I. Beginning in 2014, progress measures were provided for all students who receive accelerated instruction and skip grades or courses. For example, a progress measure is reported if a student tested in grade 5 reading in 2013 and in grade 7 reading in 2014.

- Changes related to different language-version tests

The requirement that students test in the same language in subsequent years has changed for mathematics. Beginning in 2014, progress measures were reported for students who test in a different language for mathematics. Because Spanish and English mathematics tests consist of the same items (i.e., the Spanish items are transadapted versions of the English items) and because the Spanish and English mathematics scores are reported on the same scale, progress measures are generated despite the change from one language-version test to the other. Specifically, progress measures are reported for students who test in Spanish and then in English, as well as those who test in English and then in Spanish.

The same change was not made for reading because the Spanish and English tests contain different items, and scores are reported on separate scales. Therefore,

progress measures for reading are provided only for students who test in the same language in the prior year and the current year.

- Progress for EOC

In 2013, only the spring administrations of EOC assessments were used to compute progress measures. Beginning in 2014, all EOC administrations were considered for use within these calculations. However, progress is computed only on a student's first attempt of an EOC test.

- Previous year retests

Beginning in 2014, if a student retested in the previous year in grades 5 or 8 (e.g., a current grade 6 student who retested in grade 5 or a current Algebra I student who retested in grade 8 math), the higher valid score between the primary administration and May administration is used for progress-measure calculation. Previous year June administration retest scores for grades 5 and 8 are not used for progress-measure calculations. If a student takes the same EOC test multiple times during the previous year, the score from the first administration of the previous year test is used to calculate progress.

- ELL parent denials

If a student is designated as an ELL but the student's parents have not approved ELL services for instruction, the student will receive a STAAR progress measure instead of an ELL progress measure.

PROGRESS MEASURES UPDATES MADE IN RESPONSE TO LEGISLATIVE CHANGES IN 2013

- English tests

Beginning in 2014, redesigned EOC English tests were administered. Rather than having separate English reading and writing tests, these assessments were combined into one English test per course (English I and English II). This change to the assessments impacted STAAR progress measures. In order to measure progress across two assessments, the tests must measure the same construct or content area. For example, in 2013 progress was measured from grade 8 reading to English I reading. Beginning in 2014, however, the English I test included both reading and writing content; therefore, progress could not be measured from either grade 8 reading or grade 7 writing to the new combined English I test.

Additionally, students took separate English I reading and English I writing tests in 2013, and progress could not be measured from those separate English I tests to the new combined English II test administered in 2014. For most students, progress for English II will be measured for the first time in spring 2015, after the new combined English tests have been administered for two years.

Progress Classifications

In order to interpret the gain scores, they are compared to progress targets to determine if a student *Did Not Meet*, *Met* or *Exceeded* the progress expectation. The progress targets define the expectation of annual progress for each grade and content area. These progress targets are grounded in the STAAR performance standards and the goal of having all students achieve at or above Level II: Satisfactory Academic Performance.

Specifically, the *Met* progress target is defined as the distance between the final recommended performance standards from the prior year grade and the current year grade in the same content area. For students who achieved Level I or Level II performance in the prior year, the *Met* progress target is based on the distance between the final recommended Level II standards in the prior year and current year grades in the same content area. This definition is based on the goal that students in Level II will at least maintain Level II performance. For students who achieved Level III performance in the prior year, the progress target is based on the distance between the Level III standards in the prior year and current year grades in the same content area.

The *Exceeded* progress classification is a designation reserved for those students who have demonstrated significant growth over the course of the year, beyond that of the *Met* progress target. The *Exceeded* progress target is defined as the distance between the final recommended Level II standard in the prior year and the Level III standard in the current year.

Students with gain scores less than the *Met* progress target are classified as *Did Not Meet* progress. Student with gain scores greater than or equal to the *Met* progress target and less than or equal to the *Exceeded* progress target are classified as having *Met* the progress target. Students with gain scores greater than the *Exceeded* progress target are classified as having *Exceeded* the progress target.

Because the Level II and Level III performance standards are not the same across grades and content areas (i.e., they do not have the same numerical value), the *Met* and *Exceeded* progress targets differ from grade to grade and across content areas. Table 4.12 lists the STAAR progress measure targets.

Table 4.12. 2014 STAAR Progress Measure Targets

Current Year Test	Prior Year Test	*Met* Level I/II[1] Target	*Met* Level III[2] Target	*Exceeded*[3] Target	Top Score Range[4]	Chance Score Range[5]
Grade 4 Mathematics[6]	Grade 3 Mathematics	70	62	148	46–48	0–11
Grade 5 Mathematics[6]	Grade 4 Mathematics	28	33	111	48–50	0–11
Grade 6 Mathematics	Grade 5 Mathematics	31	52	135	50–52	0–12
Grade 7 Mathematics	Grade 6 Mathematics	20	36	140	52–54	0–12
Grade 8 Mathematics	Grade 7 Mathematics	22	65	185	54–56	0–13

Current Year Test	Prior Year Test	*Met* Level I/II[1] Target	*Met* Level III[2] Target	*Exceeded*[3] Target	Top Score Range[4]	Chance Score Range[5]
Algebra I	Grade 7 Mathematics	2322	2535	2655	52–54	0–12
Algebra I	Grade 8 Mathematics	2300	2470	2633	52–54	0–12
Grade 4 English Reading	Grade 3 English Reading	82	78	165	42–44	0–11
Grade 4 Spanish Reading	Grade 3 Spanish Reading	95	104	192	42–44	0–11
Grade 5 English Reading	Grade 4 English Reading	32	34	117	44–46	0–11
Grade 5 Spanish Reading	Grade 4 Spanish Reading	43	65	162	44–46	0–11
Grade 6 Reading	Grade 5 English Reading	47	51	136	46–48	0–12
Grade 7 Reading	Grade 6 Reading	45	35	124	48–50	0–12
Grade 8 Reading	Grade 7 Reading	26	30	109	50–52	0–13
English II	English I	0	140	831	90–92	0–18

NOTE: Negative progress targets result from the use of horizontal scales (all writing and EOC tests have horizontal scales) and the movement across scales (from grades 3–8 to EOC). For more information, please see question 6 in the STAAR Progress Measure Q & A document.

[1] *Met* Level I/II is the distance or difference between the final recommended Level II standards on the current-year and prior-year tests.

[2] *Met* Level III is the distance or difference between the Level III standards on the current-year and prior-year tests.

[3] *Exceeded* is the distance or difference between the current-year test Level III standard and the prior-year test final recommended Level II standard.

[4] Top Score Range is the range of the top three possible raw scores on the current-year test.

[5] Chance Score Range is the range of raw scores that could be reasonably attained through guessing alone.

[6] Applies for both English and Spanish mathematics.

Steps for calculating progress measures and progress targets for each STAAR grade and content area, including when students skip grade levels, can be found in the "Calculating Progress Measures" document on the STAAR Resources page of TEA's Student Assessment Division website.

Classification Exceptions

There are some places on the STAAR scale, specifically at the extreme high and low ends of the scale, where the application of the *Did Not Meet*, *Met*, and *Exceeded* definitions would not be appropriate. At the extreme ends of the scale, unlike the rest of the scale, answering one more question correctly results in large differences in scale scores. For this reason, several places on the scale have been identified as exceptions to the *Did Not Meet*, *Met*, and *Exceeded* definitions.

- All students scoring at the three highest raw scores in the current year will be classified as having *Exceeded* the progress target.
- Students who maintained Level III performance from the prior year to the current year will be classified as having *Met* or *Exceeded* the progress target. (*Did Not Meet* classification will not be applied to these students.)

- Students scoring at or below chance in the current year will be classified as *Did Not Meet* progress.

Chance represents the score that could be reasonably obtained by guessing alone. For the mathematics and reading assessments, chance is defined as 25 percent of the possible multiple-choice raw-score points, as these questions have four answer options. The EOC English assessments include multiple-choice questions, short answer questions, and essays that are then weighted and combined to compute the total-test score. Chance on these assessments is defined as 25 percent of the possible multiple-choice raw-score points, scores of zero on the short answer questions, plus the weighted value associated with summed scores of 2 on the essays (representing a rubric score of 1 from both readers).

The score values associated with these exceptions for each STAAR grade and content can be found in Table 4.12 and in the "Calculating Progress Measures" document on the STAAR Resources page of TEA's Student Assessment Division website.

Results

STAAR progress measure results from 2013–2014 are provided, by performance level, in Appendix B. For each grade and content area or course, the largest numbers of students were classified as either *Did Not Meet* or *Met*. In contrast, the *Exceeded* classification applied to a smaller number of students. This pattern was expected because, by definition, it requires significantly more progress to receive an *Exceeded* progress classification than to receive the *Met* progress classification.

STAAR On-Track Measure

While the STAAR progress measure accounts for performance from the prior year and the current year, it does not include any information about how the student is likely to perform in the future. Because this additional information may be helpful to students, teachers, and other stakeholders, TEA developed the STAAR on-track measure, which was reported for the first time in 2013–2014. Specifically, the on-track measure utilized the STAAR progress measure and extrapolated performance into future years to determine if a student is on track to achieve Level II performance in a later grade or course.

The STAAR on-track measure is reported for students in grades 4 through high school in mathematics and reading/English. In 2014, on-track measures were available only for mathematics in grades 4–8 and reading in grades 4–7. The STAAR on-track measure may also be available for Algebra I and English II in future years. The STAAR on-track measure is calculated and reported for all of the Level II standards. In 2013–2014, this included the phase-in 1, phase-in 2, and final recommended Level II standards.

On-track is measured only for students who take STAAR assessments. In 2014, there was no on-track measure for STAAR Modified, STAAR Alternate, or English language learners (ELLs) who are eligible for the ELL progress measure.

In order to calculate the STAAR on-track measure, three assessments covering the same content area must be available (previous year, current year, and target year). For example, on track can be calculated for grade 8 mathematics (current year assessment) because the previous year assessment was grade 7 mathematics and the target year assessment will be Algebra I; all three of these assessments measure mathematics. However, on track cannot be calculated for grade 8 reading because the next assessment in a traditional course-taking sequence will be English I, which includes both reading and writing content. Therefore, although the previous year assessment (grade 7 reading) and the current year assessment (grade 8 reading) cover the same content area, the target year assessment (English I) covers additional content and an on-track measure cannot be calculated. In addition, the STAAR on-track measure cannot be calculated for English I because the previous year assessment (grade 8 reading) covers different content.

Additional information about the STAAR on-track measure and the steps taken to calculate these measures can be found on the STAAR Resources page of TEA's Student Assessment Division website.

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 assessments while minimizing the burden placed on campuses and districts.

During the 2013–2014 school year, sampling was conducted for a prompt study on the STAAR grade 4 English writing assessment. Participation in the prompt study was mandatory for those selected by the sampling procedures.

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 who would be participating in the corresponding STAAR assessment. The estimated student counts for each campus were based on the number of students who were enrolled in each of the sampled courses in the 2012–2013 school year.

The following factors were considered in determining the STAAR grade 4 English writing samples for spring 2014.

- The sample was selected in order to be representative of the overall population of Texas fourth grade students taking the course in terms of ethnic composition and campus size.
- Each sampled campus was required to test all of its enrolled students testing in English writing in grade 4.

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 grade 4 for the selected assessment
- Campuses that have been rated as academically unacceptable (AU) according to the state accountability system for three or more years
- Campuses that failed to meet federal Adequate Yearly Progress (AYP) targets and were at stage 5 in 2012

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 2013–2014 STAAR grade 4 English writing prompt study sample is provided in Table 4.13.

Table 4.13. Sampling Summary for Spring 2014
Grade 4 English Writing Prompt Study

	Districts N-Count	Campuses N-Count	Estimated Sample N-Count	Maximum Students at Campus	Minimum Students at Campus
Statewide (Eligible for Sampling)	979	4,048	329,825	431	15
Sample (Percent of Statewide)	366 (37%)	917 (23%)	74,721 (23%)	333	15

Test Results

Appendix B provides scale score distributions and summary statistics, raw score to scale score conversion tables, as well as mean p-values and reliability estimates for all primary STAAR assessments, including STAAR Spanish and STAAR L, administered in spring 2014. Table 4.14 shows the spring 2014 pass rates for the STAAR assessments. For STAAR EOC assessments, pass rates are given for first-time testers.

Table 4.14. STAAR Spring 2014 Pass Rates (at the Level II Phase-in 1 Standard)

Content Area	Grade/Course	Pass Rate
Mathematics	Grade 3	70%
	Grade 4	70%
	Grade 5	79%
	Grade 6	78%
	Grade 7	67%
	Grade 8	79%
	Algebra I	86%
Mathematics (Spanish)	Grade 3	60%
	Grade 4	52%
	Grade 5	45%
Reading	Grade 3	76%
	Grade 4	74%
	Grade 5	76%
	Grade 6	77%
	Grade 7	75%
	Grade 8	82%
Reading (Spanish)	Grade 3	65%
	Grade 4	61%
	Grade 5	62%
Writing	Grade 4	73%
	Grade 7	70%
Writing (Spanish)	Grade 4	64%
Science	Grade 5	73%
	Grade 8	70%
	Biology	93%
Science (Spanish)	Grade 5	47%
Social Studies	Grade 8	61%
	US History	92%
English	English I	72%
	English II	73%