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

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 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 program. 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 curriculum standards and assessments;

- assessing postsecondary readiness;

- developing an aligned system of assessments that connect performance from the postsecondary readiness standards in STAAR 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
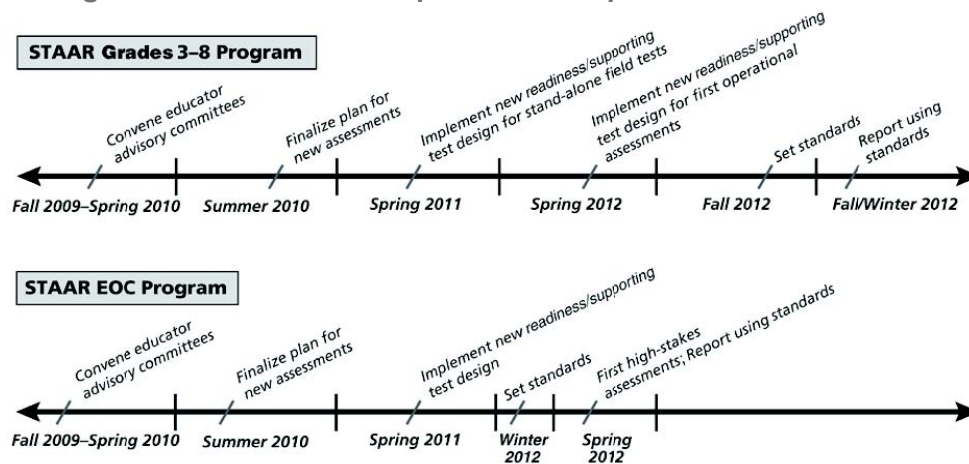
The performance standards for the STAAR EOC assessments were established prior to the spring 2012 test administration. Extensive research studies were conducted to support the standard-setting process and to ensure that the STAAR EOC assessments met requirements for rigor and alignment. Because of the aligned nature of the program, the performance standards for the STAAR grades 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 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, English I, English II, Biology, and U.S. History. Only the five STAAR EOC assessments required by HB 5 were administered in 2013–2014 and 2014–2015. To meet the requirements of HB 5, the English I and English II assessments were redesigned as combined reading and writing assessments; those redesigned assessments were first administered in spring 2014.

In the 2014–2015 school year, 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. The revised Texas Essential Knowledge and Skills (TEKS) in mathematics, which were adopted by the State Board of Education (SBOE) in April 2012, were first implemented in classrooms and in the grades 3–8 mathematics assessment program in the 2014–2015 school year. In July 2015, TEA convened standard-setting committees that recommended performance standards for STAAR grades 3–8 mathematics that aligned to the new mathematics TEKS.

STAAR

STAAR is an assessment program designed to measure the extent to which a student has learned and is able to apply the knowledge and skills defined in the TEKS, the state-mandated curriculum. Every item is directly aligned to the TEKS currently in effect for the grade/subject or course 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. As of spring 2014, only STAAR EOC assessments for Algebra I, Algebra II, English I, English II, Biology and U.S. History were available to students. STAAR Algebra II and English III are available for districts to administer as optional assessments.

Prior to spring 2017, STAAR A and STAAR L were available test versions for eligible students. In spring 2017, accommodation policies were consolidated for all special populations, and the STAAR general assessment with available embedded supports replaced STAAR A and STAAR L for all students who meet eligibility requirements. These embedded designated supports or accommodations are available for all online test forms and can be either content supports, language and vocabulary supports, or text-to-speech (TTS).

STAAR Spanish

The STAAR Spanish assessments are available to students for whom the language proficiency assessment committee (LPAC) decides the Spanish assessments are the most appropriate way to measure their mastery of skills or to 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 Spanish versions of STAAR assessments have the same test blueprint and assess the same TEKS student expectations for reading, mathematics, and science. Because of differences in the English and Spanish languages, the writing student expectations assessed are similar but not the same.

STAAR Alternate 2

STAAR Alternate 2 is an alternate assessment based on alternate academic achievement standards for students receiving special education services and who have a significant cognitive disability. STAAR Alternate 2 is described in chapter 5, "STAAR Alternate 2."

Testing Requirements for Graduation

Students enrolled in grade 9 or below for the first time in the 2011–2012 school year or later 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 STAAR English I and STAAR English II graduation requirements using performance on the separate English reading and writing assessments. The performance level, Satisfactory, was renamed *Approaches Grade Level* in the 2016–2017 school year.

In 2015, the 84th Texas Legislature passed SB 149, which revised the state's assessment graduation requirements. Beginning in May 2015, an eligible student may receive a Texas high school diploma by means of an individual graduation committee (IGC) if the student fails to achieve Satisfactory Academic Performance on no more than two STAAR EOC assessments. Eligibility criteria for an individual graduation committee are located in the Texas Education Code (TEC) §28.0258.

The ARD committee makes educational decisions for a student receiving special education services, including decisions related to state assessments and graduation requirements as described in Texas Administrative Code (TAC) §89.1070.

Test Development

Maintaining a student assessment program 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 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 composed 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/subject and course and that provide a vertical link between the assessments from grade to grade and 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 STAAR English I, STAAR English II, and STAAR 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 comprised of Texas writing educators to review the field-test results and validate the rubrics. Rubrics for short-answer reading, expository writing, 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 subsequently approved and published.

- The 2018 STAAR assessments were constructed to conform to the assessment blueprints and all established test construction guidelines for STAAR. Most items placed on the 2018 STAAR grades 3–8 assessments were developed and field-tested through embedded field testing on operational STAAR assessments in 2012–2015 and through stand-alone field tests for grades 4 and 7 writing in 2011. 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 2017.

- Item development activities continue annually to develop STAAR, to support the annual 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 their data results are reviewed. Field-tested items that meet established criteria are added to the bank.

- An additional design aspect of STAAR 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. For example, open-ended items that require students to derive a numerical answer independently are developed for mathematics and grade 8 science courses. Process skills in mathematics, science, and social studies are assessed in context, not in isolation, which allows for a more integrated and authentic assessment of these content areas. In reading, greater emphasis is given to critical analysis than to literal understanding. In writing, prompts have been developed to support expository, persuasive, and analytical writing.

■ STAAR test items are developed so that they can be delivered in both paper and online test formats. A plan was created to evaluate the comparability of assessment items delivered in paper mode and those delivered in online mode. The plan 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. The results suggested a mode effect (differences in difficulty for online versus paper tests) for English I and English II. Information about English I and English II comparability analyses can be found in the Technical Digests from previous years.

Item Development Approach for STAAR Spanish

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

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 in accordance with the educator review process described in chapter 2, "Building a High-Quality Assessment System."

STAAR English–Spanish Alignment

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

■ The development and review processes for the reading and writing assessments in English and Spanish are parallel, meaning that item reviews for both English and Spanish include judgments related to each item's alignment to the TEKS student expectations. 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. The mathematics and science assessments in Spanish are transadapted from their corresponding English assessments.

■ 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.

■ The reading and writing assessments in English and Spanish are constructed concurrently and in coordination, adhering to the same test construction guidelines with regard to the range of item content and cognitive complexity. The mathematics and science assessments in English are constructed first and then transadapted to Spanish.

■ Each year 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 grades 3–8 were established, standard-setting panels reviewed both the English and Spanish transadapted tests with the goal of establishing comparable performance standards.

Training

Test administration procedures provided in the Test Security Supplement, the District and Campus Coordinator Manual, and the appropriate test administrator manuals must be followed so that all individuals 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 their campus testing coordinators, who are responsible for training any campus personnel participating in the administration of the assessments.

During the 2017–2018 school year, ESC personnel and district coordinators were also given a coordinator packet, which contained all the information and materials necessary for overseeing test administrations, including copies of the appropriate manuals.
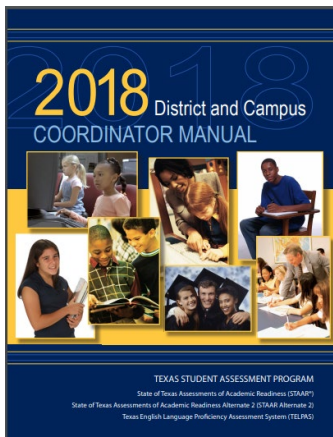
District and Campus Coordinator Manual and Test Administrator Manuals

The *2018 District and Campus Coordinator Manual* guided district and campus coordinators through their responsibilities as they oversaw the administration of the Texas assessment program. This manual contained preparation and administration procedures for every assessment program for the 2018 calendar year and was made available prior to the annual training of ESC, district, and campus coordinators. Separate test administrator manuals were provided to districts prior to the spring administrations for STAAR grades 3–5, grades 6–8, and EOC assessments.

Test Security Supplement

TEA's Student Assessment Division developed the *2018 Test Security Supplement* to promote a secure testing program and to help districts implement the requirements for STAAR, STAAR Alternate 2, and the Texas English Language Proficiency Assessment System (TELPAS). This document is not intended to replace any procedures or instructions contained in the *2018 District and Campus Coordinator Manual* or *test administrator manuals*. It is provided to testing personnel as a supplement to further support incident-free test administrations.

Test Administrations

In 2017–2018, STAAR grades 3–8 and EOC assessments were administered. The primary assessments for STAAR grades 3–8 and EOC were administered in spring 2018 — grades 4 and 7 writing in April; grades 5 and 8 mathematics and reading in April; English I and II in April; grades 3–8 in May; Algebra I, Algebra II, English III, Biology, and U.S. History in May. STAAR EOC assessments were also administered in December 2017 and June 2018. STAAR grades 5 and 8 mathematics and reading retests were administered in May and June 2018.

In spring 2018, STAAR English grades 3–8 and EOC assessments were offered on paper and online. STAAR Spanish grades 3–5 assessments were offered as paper tests only. During the 2017–2018 school year, over 9.7 million STAAR assessments, including STAAR and STAAR Spanish, were administered. The number of students tested for each STAAR assessment in 2017–2018 is shown in Table 4.1.

Table 4.1. STAAR Assessments Administered in 2017–2018

STAAR Assessment	Assessments Administered
Grade 3 English mathematics	386,467
Grade 3 English reading	368,263

STAAR Assessment	Assessments Administered
Grade 3 Spanish mathematics	16,668
Grade 3 Spanish reading	34,831
Grade 4 English mathematics	397,924
Grade 4 English reading	382,864
Grade 4 English writing	370,396
Grade 4 Spanish mathematics	9,171
Grade 4 Spanish reading	24,124
Grade 4 Spanish writing	24,571
Grade 5 English mathematics	476,728
Grade 5 English reading	504,593
Grade 5 English science	399,294
Grade 5 Spanish mathematics	8,207
Grade 5 Spanish reading	20,080
Grade 5 Spanish science	7,754
Grade 6 mathematics	387,665
Grade 6 reading	394,958
Grade 7 mathematics	350,802
Grade 7 reading	394,726
Grade 7 writing	386,417
Grade 8 mathematics	422,642
Grade 8 reading	499,431
Grade 8 science	386,971
Grade 8 social studies	389,004
Algebra I	509,957
Algebra II	11,965
English I	700,112
English II	620,151
English III	9,179
Biology	478,922
U.S. History	397,085

NOTE: For grades 5 and 8 mathematics and reading tests, the assessments administered include the sum total of the assessments from April, May, and June 2018 administrations; for EOC assessments (except for Algebra II and English III), the assessments administered include the sum total of the December 2017, April 2018, May 2018, and June 2018 administrations; and for the other tests, only the spring 2018 administration was included.

The Online Test Delivery System

In 2017–2018, STAAR online assessments were administered using the STAAR Assessment Management System, which includes the STAAR Online Testing Platform. 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 groups,

- which students are assigned to each group, and

- access to embedded supports.

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, details on delivery and reporting, and a list of frequently asked questions, is available on the Texas Assessment Management System website.

Make-up Testing

In the STAAR program, make-up testing opportunities for students who are absent are available during the testing window for all grades/subjects and courses. For the 2017–2018 school year, make-up testing opportunities were available for all administrations.

Out-of-District Testing

For STAAR grades 5 and 8 mathematics and reading retest administrations and for STAAR EOC administrations, students who are unable to test at their home district are allowed to test out-of-district (OOD). For example, a student from Houston who spends the summer in Dallas and wants to test in Dallas could register to test OOD. To make the receiving district aware of the student's intent to test and to ensure that the receiving district receives precoded test materials, OOD students are required to complete registration prior to the close of the out-of-district/out-of-school registration period. OOD students who do not register may show up at a registered OOD test site and test as walk-ins on the day of testing at the discretion of the receiving district and the ability of the test site to accommodate them. Students must present photo identification at the test administration site on the day of the test. OOD testing also applies to students who are part of the Texas Tech University or University of Texas high school programs.

A district or campus must accommodate the request of an OOD student to participate in STAAR administrations if that district or campus is testing one or more local students on the applicable test and if the student has registered in advance for OOD testing. Districts can use test materials from the district overage received with their combined shipment of materials or place an order for additional materials through the STAAR Assessment Management System.

Out-of-School Testing

Beginning in the 2014–2015 school year, out-of-school (OOS) testing was available for STAAR EOC examinees who were no longer enrolled in school. This allowed examinees who had not passed a STAAR EOC assessment and were no longer enrolled in school, but had otherwise completed requirements for graduation, to take an assessment during a test administration at a participating district.

Educational Materials Required for Testing

DICTIONARIES

Dictionaries must be available to **all** students taking:

- STAAR grades 3–8 reading tests
- STAAR grades 4 and 7 writing tests, including revising and editing
- STAAR Spanish grades 3–5 reading tests
- STAAR Spanish grade 4 writing test, including revising and editing
- STAAR English I, English II, and English III tests

The following types of dictionaries are allowable:

- standard monolingual dictionaries in English or the language most appropriate for the student
- dictionary/thesaurus combinations
- bilingual dictionaries* (word-to-word translations; no definitions or examples)
- ESL dictionaries* (definition of an English word using simplified English)
- sign language dictionaries
- picture dictionary

Both paper and electronic dictionaries are permitted. However, electronic dictionaries that provide access to the Internet or have photographic capabilities are **NOT** allowed. For electronic dictionaries that are hand-held devices, test administrators must ensure that any features that allow note taking or uploading of files have been cleared of their contents both before and after the test administration.

While students are working through the tests listed above, they must have access to a dictionary. Students should use the same type of dictionary they routinely use during classroom instruction and classroom testing to the extent allowable. The school may provide dictionaries, or students may bring them from home. Dictionaries may be provided in the language that is most appropriate for the student. However, the dictionary must be commercially produced. Teacher-made or student-made dictionaries are not allowed. The minimum schools need is one dictionary for every five

students testing, but the state's recommendation is one for every three students or, optimally, one for each student.

Although thesauruses are not required, they are allowable on all the tests listed above, either in combination with a dictionary or as a separate resource. If districts make thesauruses available to students during testing, it is recommended that there be one thesaurus for every five students.

*Bilingual and ESL dictionaries should be provided in accordance with individual student needs based on how much students use them in instruction and classroom testing. While there is no requirement regarding a minimum number of bilingual or ESL dictionaries schools must provide, for ELs who depend heavily on a dictionary in language arts instruction, it is recommended that there be one dictionary for each student. Additionally, it is important for LPACs to consider the degree to which an EL student relies on a dictionary during language arts instruction or testing when making exit decisions at the end of the year.

CALCULATORS

Calculators are required for the STAAR grade 8 mathematics, grade 8 science, Algebra I, Algebra II, and Biology assessments. Students are permitted to use their own calculators instead of those provided by the district. Students are permitted to use more than one calculator during the assessment. At minimum, districts are required to provide the following:

- A graphing calculator for each student taking grade 8 mathematics, grade 8 science, Algebra I, or Algebra II assessments. Students must have a graphing calculator to use throughout the entire test (both paper and online versions). Districts may satisfy this calculator requirement by providing students with either a hand-held graphing calculator or a graphing calculator application available on a tablet.

- A calculator for every five students taking Biology.

Any calculator can be used to fulfill the minimum requirements listed above except for those that include a computer algebra system (CAS), allow access to the Internet, or have photographic capabilities.

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

Calculators are not provided to students taking the STAAR grades 3–7 mathematics assessments or the STAAR grade 5 science assessment unless a student meets the eligibility criteria for such an accommodation. This includes the STAAR Spanish assessments.

Understanding STAAR Accessibility

The goal of STAAR accessibility is to ensure that each student can interact appropriately with the content, presentation, and response modes of the state assessment. To meet this goal, STAAR accessibility features and designated supports are designed to allow all test takers to demonstrate their knowledge of the content being tested without the barriers of assessment format, non-tested language, or response type. The various features and designated supports made available on STAAR paper and online tests are also designed to be the same or similar to those accommodations commonly used during classroom instruction.

STAAR is available on paper and online as indicated for the grades and subjects listed in Table 4.2.

Table 4.2. STAAR Assessments in Paper and Online Formats

Program	Grade	Subject	Paper Available	Online Available
STAAR	3–8 & EOC	all	√	√
STAAR Spanish	3–5	all	√	
STAAR Alternate 2	3–8 & EOC	all	√	

Accessibility within the STAAR Program

For STAAR and STAAR Spanish, accessibility falls into three categories: accessibility features, locally-approved designated supports, and designated supports requiring TEA approval. These are available on paper and online. A student may take STAAR online for one subject and on paper for another, depending on appropriateness and/or accessibility. Refer to the individual policy documents on the Accommodation Resources page for comprehensive information about the procedures, materials, applicable grade levels, subjects and eligibility criteria within each of these categories.

Accessibility Features

Accessibility features may be provided to students based on their needs. In general, these procedures and materials are available to any student who regularly benefits from their use during instruction; however, a student cannot be required to use them during testing. Coordinators are responsible for ensuring that test administrators understand the proper implementation of these procedures and use of these materials. In some cases, a student may need to complete the test in a separate setting in order to eliminate distractions to other students and to ensure that the security and confidentiality of the test are maintained.

Locally-Approved Designated Supports

Locally-approved designated supports include accommodations that may be made available to students who meet eligibility criteria. The appropriate team of people at the campus level (e.g., RTI team, LPAC, Section 504 committee, ARD committee)

determines eligibility as indicated in each policy document. The decision to use a designated support during a state assessment should be made on an individual student basis and take into consideration the needs of the student and whether the student routinely receives the support during classroom instruction and classroom testing. In addition, the support has been proven to be effective in meeting the student's specific needs, as evidenced by student scores or teacher observations.

Designated Supports Requiring TEA Approval

These designated supports require the submission of an Accommodation Request Form to TEA. The appropriate team of people at the campus level as indicated in each policy document determines whether the student meets all the specific eligibility criteria and, if so, submits an Accommodation Request Form to TEA. Accommodation Request Forms must be received by TEA according to the posted deadlines. Late requests will NOT be processed unless circumstances involving the student change after the deadline (e.g., newly enrolled student, medical emergency, updated admission, review, and dismissal [ARD] committee decisions). The request must be approved by TEA before a student can use the designated support on the state assessment. This type of support must be documented in the student's paperwork as "pending TEA approval." If a request is denied, the campus should be prepared to meet the student's needs with accessibility features or any of the locally-approved designated supports.

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 grade 6. Additionally, students are required to pass the STAAR grade 8 mathematics and reading assessments to be promoted to grade 9.

Students in grades 5 and 8 are provided three opportunities (April, May, and June) to pass the STAAR mathematics and reading assessments. If a student does not pass an assessment, the school is required to provide the student additional instruction after each testing opportunity, and the student is required to participate in that instruction. Parents are notified if their child does not pass a STAAR assessment that is required for promotion.

If a student does not pass after the second testing opportunity, a grade placement committee (GPC) is formed as part of the SSI requirements. The GPC, which consists of the principal, teacher, and parent or guardian, is required to create an instructional plan based on the individual needs of the student. If a student does not demonstrate proficiency on the mathematics or reading assessment on the third testing opportunity, the student can advance to or be placed in the next grade level only if (1) the student's parent, guardian, or designee appeals the automatic retention, (2) he or she completes

all accelerated instruction required by the GPC, and (3) the GPC determines, by unanimous decision, that the student is 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 is 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 SSI is available on the Student Success Initiative webpage on TEA's Student Assessment Division website.

Scores and Reports

There are a variety of reports that show a student's performance on the STAAR assessments. 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 include the number of items answered correctly (raw scores), the 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. 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. Raw scores must be converted to scale scores to make comparisons of student performance.

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 *Did Not Meet Grade Level*, *Approaches Grade Level*, *Meets Grade Level*, or *Masters Grade Level*. (Performance-level cut scores are discussed in the Performance Standards section of this chapter.) Scale scores for all STAAR 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.

Assessment Reports
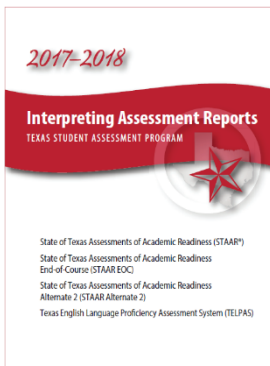
Standard reports are provided automatically to districts for the various testing programs. Information contained in standard reports satisfies mandatory reporting requirements. To receive additional copies of the STAAR Report Card or Confidential Student Label, the district must submit a request through the Assessment Management System in the *Ancillary Services* screen. Generally, districts are required to pay a nominal fee for each additional 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 *Approaches Grade Level*, *Meets Grade Level*, and *Masters Grade Level* performance standards can be analyzed by grade and content area across administrations to give insight into whether student performance is improving across years. For example, the average scale score for students who 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. 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 *Approaches Grade Level* performance standard, which group had the highest percentage achieving the *Meets Grade Level* 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 districts and campuses identify areas of potential academic weakness for a group of students. The same methodology can be applied to an entire district or campus. 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

statewide and regional 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 caution 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.

Understanding the Test Scores

TEA's Student Assessment Division produces interactive resources on both TEA's website (i.e., Interpreting Assessment Reports) and the Texas Assessment Management System website to assist school personnel in understanding and interpreting student performance data as required by Section 39.030(b) of the Texas Education Code and to help parents understand their child's STAAR grades 3–8 test results on the STAAR Report Card. In addition, parents can access their child's STAAR test results in the secure Student Portal using unique identifying information.

Released Tests

TEA released the primary administration (paper and online) of the following 2018 STAAR assessments with and without accommodations:

- 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 English II

- Biology

- U.S. History

The released tests included STAAR Spanish versions of the grades 3–5 assessments. All the STAAR released assessments listed above can be found on the STAAR Released Test Questions page of TEA's Student Assessment Division website.

Item Analysis Reports

Confidential Student Item Analysis Reports and Item Analysis Summary Reports for STAAR assessments released in their entirety were provided to districts. Item Analysis Summary Reports were provided at the statewide, region, district, and campus 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.

STAAR item analysis summary reports were made available on TEA's website and in the Assessment Management System for the following test administrations:

- April 2018 STAAR grades 4 and 7 writing; grades 5 and 8 reading; and grades 5 and 8 mathematics

- May 2018 STAAR grades 3, 4, 6, and 7 mathematics; grades 3, 4, 6 and 7 reading; grades 5 and 8 science; and grade 8 social studies

- spring 2018 STAAR Algebra I, English I, English II, Biology, and U.S. History EOC assessments

Students and districts may find it helpful to use the Item Analysis Summary Report in conjunction with the 2018 STAAR released tests. For STAAR grades 3–5, English and Spanish results are reported separately. For additional information, refer to the 2018 Interpreting Assessment Reports page of the TEA website.

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 STAAR and STAAR Spanish, the labels for the performance categories are as follows:

- Did Not Meet Grade Level

- Approaches Grade Level

- Meets Grade Level

- Masters Grade Level

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

DID NOT MEET GRADE LEVEL

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

APPROACHES GRADE LEVEL

Performance in this category indicates that students are likely to succeed in the next grade or course with targeted academic intervention. Students in this category generally demonstrate the ability to apply the assessed knowledge and skills in familiar contexts.

MEETS GRADE LEVEL

Performance in this category indicates that students have a high likelihood of success in the next grade or course but may still need some short-term, targeted academic intervention. Students in this category generally demonstrate the ability to think critically and apply the assessed knowledge and skills in familiar contexts.

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

MASTERS GRADE LEVEL

Performance in this category indicates that students are expected to succeed in the next grade or course with little or no academic intervention. Students in this category demonstrate the ability to think critically and apply the assessed knowledge and skills in varied contexts, both familiar and unfamiliar.

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

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 subjects). 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 program. The nine steps are as follows:

1. Conduct validity and linking studies.

2. Develop performance labels and policy definitions.

3. Develop grade- and course-specific performance level descriptors (PLDs).

4. Convene a policy committee and develop performance standard ranges.

5. Convene standard-setting committees.

6. Review performance standards for reasonableness.

7. Approve performance standards.

8. Implement performance standards.

9. Review performance standards.

Tables 4.4 and 4.5 provide high-level descriptions and timelines for these nine steps as implemented in the STAAR EOC and 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 "STAAR Standard Setting Technical Report," available on the STAAR Performance Standards page of TEA's Student Assessment Division website.

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

Standard-Setting Step	Description	Timeline
1. Conduct validity and linking studies	External validity evidence was collected to inform standard-setting and to support interpretations of the performance standards. Scores on each assessment are linked to performance on other assessments in the same content area.	Studies started in spring 2009 and will continue throughout the program.
2. Develop performance labels and policy definitions	A committee was convened jointly by TEA and THECB to recommend performance categories, performance category labels, and general policy definitions for each performance category.	September 2010
3. Develop course-specific performance level descriptors (PLDs)	Committees composed primarily of educators developed performance level descriptors (PLDs) as an aligned system, describing a reasonable progression of skills within each content area (mathematics, English, science, and social studies).	November 2011
4. Convene a policy committee and develop reasonable ranges for performance standards	A 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	Mathematics and English: Feb. 22–24, 2012; Science

Standard-Setting Step	Description	Timeline
	policy committee to recommend cut scores for each STAAR EOC assessment.	and Social Studies: Feb. 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 HB 5, which removed the requirement to convene standards review panels. However, TEA and the Commissioner of Education review statewide performance relative to the standards after each administration to help inform decisions about the appropriate schedule for the phase-in of standards.

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

Standard-Setting Step	Description	Timeline
1. Conduct validity and linking studies	External validity evidence was collected to inform standard setting and support interpretations of the performance standards. Scores on each assessment are linked across grades to performance on other assessments in the same content area.	Studies started in spring 2011 and will continue throughout the program.
2. Develop performance labels and policy definitions	A committee was convened jointly by TEA and 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 subject area (mathematics, reading, writing, science, and social studies).	June 2012
4. Develop reasonable ranges for performance standards	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 subject areas.	October 2012

Standard-Setting Step	Description	Timeline
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 HB 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 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 ELs) 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 careful consideration of the alignment of performance standards with the assessments in the same content area, relevant information from the STAAR 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 STAAR 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 and 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 and Spanish versions of a STAAR assessment.

To meet the requirements of HB 5, passed by the 83rd Texas Legislature in June 2013, the STAAR English I, STAAR English II, and STAAR English III assessments were redesigned so 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 STAAR English I, STAAR English II, and STAAR English III.

In July 2015, TEA convened standard-setting committees that recommended performance standards for STAAR grades 3–8 mathematics that aligned to the new mathematics TEKS first implemented in 2014–2015.

Phase-In of Performance Standards

A phase-in period was implemented for the STAAR Level II: Satisfactory Academic Performance standards 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.

Phase-in 1 performance standards for Level II were in effect for the 2011–2012, 2012–2013, 2013–2014, and 2014–2015 school years. In 2015–2016, the Commissioner of Education decided to replace the phase-in approach with a standard progression approach from 2015–2016 through 2021–2022, the year final standards were scheduled to be in place. Instead of larger jumps to more rigorous performance standards every few years, this progression approach would mean smaller predictable increases every year through the 2021–2022 school year.

The STAAR phase-in performance standard that students are assigned for all EOC assessments is based on the standard in place when they took an EOC assessment for the first time. For example, the 2012–2015 standard for Level II: Satisfactory Academic Performance was in effect in spring 2014. Therefore, students who first took STAAR Algebra I in spring 2014 are held to the 2012–2015 standard for Algebra I, Algebra II, English I, English II, English III, Biology, and U.S. History.

Phase-in performance standards were frozen and renamed in spring 2017.

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. Initial performance standards for all STAAR assessments were recommended by standard-setting committees and approved by the Commissioner of Education in April and December 2012.

Performance Standards in Use

The four performance level labels were changed in the spring 2017—from Level I, Level II (phase-in), Level II (recommended), and Level III to *Did Not Meet Grade Level*, *Approaches Grade Level*, *Meets Grade Level*, and *Masters Grade Level* —as part of the initiative to make the score reporting more family friendly. These four performance levels and labels are in use since spring 2017 for STAAR and STAAR Spanish.

Tables 4.6 and 4.7 show the *Approaches Grade Level*, *Meets Grade Level*, and *Masters Grade Level* performance standards in scale scores for the STAAR grades 3–8 and EOC assessments (including STAAR Spanish).

Table 4.6. STAAR 3–8 Performance Standards

Assessment	Approaches Grade Level (Scale score)	Meets Grade Level (Scale score)	Masters Grade Level (Scale score)
Grade 3 English Mathematics	1360	1486	1596
Grade 4 English Mathematics	1467	1589	1670
Grade 5 English Mathematics	1500	1625	1724
Grade 6 Mathematics	1536	1653	1772
Grade 7 Mathematics	1575	1688	1798
Grade 8 Mathematics	1595	1700	1854
Grade 3 English Reading	1345	1468	1555
Grade 4 English Reading	1434	1550	1633
Grade 5 English Reading	1470	1582	1667
Grade 6 Reading	1517	1629	1718
Grade 7 Reading	1567	1674	1753
Grade 8 Reading	1587	1700	1783
Grade 4 English Writing	3550	4000	4612
Grade 7 Writing	3550	4000	4602
Grade 5 English Science	3550	4000	4402
Grade 8 Science	3550	4000	4406
Grade 8 Social Studies	3550	4000	4268
Grade 3 Spanish Mathematics	1360	1486	1596
Grade 4 Spanish Mathematics	1467	1589	1670
Grade 5 Spanish Mathematics	1500	1625	1724
Grade 3 Spanish Reading	1318	1444	1532
Grade 4 Spanish Reading	1413	1539	1636
Grade 5 Spanish Reading	1461	1582	1701
Grade 4 Spanish Writing	3550	4000	4543
Grade 5 Spanish Science	3550	4000	4402

Table 4.7. STAAR EOC Performance Standards

Assessment	Approaches Grade Level 2012–2015 (Scale Score)	Approaches Grade Level (Scale Score)	Meets Grade Level (Scale Score)	Masters Grade Level (Scale Score)
Algebra I	3500	3550	4000	4333
Algebra II		3550	4000	4411
English I	3750	3775	4000	4691
English II	3750	3775	4000	4831

Assessment	Approaches Grade Level 2012–2015 (Scale Score)	Approaches Grade Level (Scale Score)	Meets Grade Level (Scale Score)	Masters Grade Level (Scale Score)
English III		3775	4000	4546
Biology	3500	3550	4000	4576
U.S. History	3500	3550	4000	4440

Scaling

Scaling is a statistical procedure that places raw scores on a common scoring metric 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 grade 4 writing

- STAAR grade 7 writing

- STAAR English grade 5 science

- STAAR grade 8 science and 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 *Meets Grade Level* performance standard. In addition, the standard deviation for those scales was set to 500.

It is important to note that although *Approaches Grade Level* and *Meets Grade Level* scale score values are fixed across horizontally scaled assessments, *Masters Grade Level* scale score values vary across STAAR assessments. For a given assessment, the performance standards remain 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 following equation:

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

where SS_θ is the scale score for a Rasch proficiency level estimate (θ). A and B in Equation (1) are referred to as the horizontal scaling constants. These same transformations will be applied each year to the Rasch proficiency level estimates (θ) for that year's set of test items. Values for the horizontal scaling constants are provided in Tables 4.8 and 4.9 for the horizontally scaled STAAR grades 3–8 and EOC assessments, respectively.

Table 4.8. Horizontal Scaling Constants for STAAR 3–8

Assessment			A	B
Grade	Language	Content Area		
4	English	Writing	522.0322	3300.4769
7	English	Writing	505.0725	3227.2391
5	English	Science	439.1622	2673.7302
8	English	Science	516.7927	3051.6854
8	English	Social Studies	468.3839	2849.6491
4	Spanish	Writing	574.7681	3230.9603
5	Spanish	Science	439.1622	2673.7302

Table 4.9. Horizontal Scaling Constants for STAAR EOC

Assessment	A	B
Algebra I	441.1057	3448.1767
Algebra II	491.1456	3898.8240
English I	453.3627	3550.2642
English II	472.9035	3482.6436
English III	526.3158	3743.1579
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 mathematics

- STAAR Spanish grades 3–5 mathematics

- STAAR English grades 3–8 reading

- 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 writing, science, or social studies at the elementary and middle school levels.

For grade 8 mathematics and English reading vertical scales, a scale score of 1700 represents the *Meets Grade Level*. The standard deviation for the grades 3–8 mathematics vertical scales was set to 150.

For the STAAR English and Spanish grade 5 reading assessments, a scale score of 1582 represents the *Meets Grade Level*. The standard deviation for the reading vertical scale was set to 150.

It is important to note that although *Meets Grade Level* scale score values are fixed for the highest grade in the vertical scale, the *Meets Grade Level* scale score values for the other assessments in the vertical scale are different. However, these *Meets Grade Level* scale score values, as well as the *Approaches Grade Level* and *Masters Grade Level* scale score values, remain constant over time.

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

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

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 4.10. Once established, these same transformations are applied each year to the proficiency level estimates for that year's set of test questions.

Table 4.10. Vertical Scale Score Transformation and Scaling Constants for STAAR 3–8 Mathematics and Reading

Assessment			A	B	V_g
Grade	Language	Content Area			
3	English/Spanish	Mathematics	107.9372	1619.0471	-2.4791
4	English/Spanish	Mathematics	107.9372	1619.0471	-1.4486
5	English/Spanish	Mathematics	107.9372	1619.0471	-0.9990
6	English	Mathematics	107.9372	1619.0471	-0.1322
7	English	Mathematics	107.9372	1619.0471	0.0870
8	English	Mathematics	107.9372	1619.0471	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

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 test forms so that scale scores across test forms and testing administrations can be compared. TEA's equating activities on the STAAR assessments include 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 places 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.

The pre-equating model applies to all STAAR Algebra I, Algebra II, Biology, and U.S. History assessments. In addition, STAAR forms administered in the summer or fall (except for STAAR English I and English II), are only pre-equated either because the testing population is too small (for example, for the STAAR assessments with embedded supports) or because the sample of students taking the test is not representative of the general population.

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 generally conducted for all primary administrations of the STAAR grades 3–8, English I, English II, and English III general assessments.

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.

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 2017–2018 school year, field-test equating was conducted for the STAAR assessments through an embedded field-test design. Multiple-choice (and griddable items when applicable) field-test items were embedded in each STAAR assessment.

Comparability Analyses

When tests are administered both online and on paper, 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 equating process 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. These studies started with the spring 2012 administration of the STAAR English I reading and writing assessments and continued with the STAAR English II reading and writing assessments in spring 2013. Information about STAAR English I and STAAR English II comparability analyses can be found in the Technical Digests from previous years.

Comparability analyses were conducted for the redesigned STAAR English I and STAAR English II assessments in spring 2014 and for STAAR English III in spring 2016 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 STAAR English I and STAAR English II and for the spring 2016 administration of STAAR English III. These adjustments involved using a raw score equivalency table, which maps each raw score on the online test to an equivalent raw score on the paper test to produce equivalent scale scores for the online assessment.

Comparability analyses were conducted for the revised STAAR English I, STAAR English II, and STAAR English III tests in the spring 2017 administration. The comparability analysis results showed test mode effects on the three tests. Therefore, for each test the post-equating process was employed to equate the online scale to the paper scale using the common multiple-choice items as anchors. For the details on the common-item non-equivalent groups equating design, see chapter 3, "Standard Technical Processes."

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.

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 2018, the internal consistency estimates for the total group ranged from 0.82 to 0.93; all EOC tests had reliability estimates higher than 0.89, while the grade 5 Spanish science test had the lowest reliability estimate of 0.82. For the different student groups, estimates were found to be similar. For the STAAR English grade 8 mathematics test, for example, the reliability for the total group was 0.90, for females 0.90, for males 0.91, for African Americans 0.89, for Hispanics 0.90, and for whites 0.90.

Because internal consistency estimates typically decrease as the number of test items decreases, internal consistency estimates for individual reporting categories can be noticeably lower than those for the full assessment. 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 "Data Analysis and Personal Financial Literacy" contains only four items. The estimated reliability for the scores in this reporting category was 0.54. Therefore, the lower reliability at the reporting category level should be considered when making interpretations of the scores at this level.

Estimates of internal consistency at the overall level, as well as at the level of reporting categories for student groups for spring 2018 STAAR paper (and online for English I, English II, and English III) primary assessments (including STAAR Spanish), are provided in Appendix B.

Classical Standard Error of Measurement

Classical standard error of measurement (SEM) represents the amount of variance in a score that results from random factors other than what the assessment is intended to measure. The SEM is helpful for quantifying the margin of uncertainty that occurs on every test. For the primary STAAR English and STAAR Spanish assessments in spring 2018, SEM values are between 2 to 3 raw score points for grades 3–8 tests and between 3 to 4 raw score points for EOC tests. The SEM values for the primary STAAR assessments (including STAAR Spanish) administered in spring 2018 are provided in Appendix B.

Conditional Standard Error of Measurement

It is important to note that the SEM index provides only an estimate of the average test score error for all students regardless of their individual levels of proficiency. By comparison, the 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 2017–2018 school year, CSEM values for vertically scaled assessments were approximately 30 to 60 scale score points in the middle of the scale score ranges. For the remaining STAAR horizontally scaled tests, the CSEM values were approximately 110 to 260 scale score points in the middle of the scale score ranges. CSEM values for all spring 2018 STAAR administrations (including 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 2017–2018 school year range from 72.6 to 78.5 percent for grades 3–8 assessments and from 75.8 to 88.2 for EOC assessments. Classification accuracy rates for all spring 2018 STAAR primary administrations (including 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 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 development of 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 continue to be refined through the test development process, as various STAAR item review educator committees share their feedback 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, including STAAR Spanish, 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 committee meetings are regularly held to support the validity of test content: item review committees and content validation committees. Item review committees are made up of Texas 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. Though these committees do not edit or revise items, they can recommend that certain items be replaced on operational assessments. Content validation is conducted 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 per item type and administration mode. Across STAAR, three types of responses are required of students: multiple-choice items on all assessments, gridded-response items on mathematics and science assessments, and constructed-response items (e.g., written composition) on grades 4 and 7 writing, English I, English II, and English III assessments.

Multiple-choice items are developed so that students must apply what they have learned about the content, thereby supporting an accurate measurement of the construct being assessed.

Gridded-response items require students to determine a numerical answer and then record 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 responses are used for the STAAR reading and writing assessments. Passage-based items are 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 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 confirming that response processes do not result in an advantage or disadvantage for any student group. 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 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 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 written compositions that includes training and qualification requirements for readers; ongoing monitoring during scoring; adjudication and resolution processes for student responses that do not meet the exact/adjacent scoring requirements; and rescoring of responses for which concerns have been raised by districts, campuses, or teachers regarding the assigned score. A more comprehensive description of the scoring process for constructed-response items is available in chapter 2, "Building a High-Quality Assessment System."

Score reliability for every STAAR assessment is generated and evaluated in terms of reader agreement rates and the commonly-used kappa with quadratic weights (Fleiss & Cohen, 1973). The written compositions are scored with the adjacent agreement model. The exact agreement rate, the adjacent agreement rate, and the total agreement rate (exact and adjacent) between the first reader's score and the second reader's score are generated (see Table 4.11). When the first reader's score and the second reader's score are not in exact or adjacent agreement, the written composition student response is adjudicated by a scoring leader.

Table 4.11. Summary of Reader Agreement (Reliability)
for Spring 2018 STAAR Written Composition

Test	Mode	Number of Responses	Agreement Rate (%) after Two Ratings			Quadratic Weighted Kappa
			Exact	Adjacent	Exact + Adjacent	
Grade 4 writing (English)		369,077	60%	38%	98%	0.65
Grade 4 writing (Spanish)		24,494	59%	39%	98%	0.64
Grade 7 writing		385,026	62%	36%	98%	0.64
English I	Paper	384,424	59%	39%	98%	0.62
	Online	88,390	64%	34%	98%	0.71
English II	Paper	367,974	60%	39%	99%	0.62
	Online	76,241	61%	37%	98%	0.72
English III	Paper	7,478	59%	39%	98%	0.67
	Online	1,609	63%	34%	97%	0.73

Validity is evaluated through validity papers, which are student responses from the field-test and current administrations that are representative of different levels of writing performance based on the scoring rubrics. Validity papers are identified by scoring leaders 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. Validity agreement rate in Table 4.12 is expressed in terms of exact agreement between the score assigned by a given reader and the "true" score approved by TEA.

Table 4.12. Summary of Validity Results
for Spring 2018 STAAR Written Composition

STAAR Assessment	Exact Agreement Rate (%)
Grade 4 writing (English)	78%
Grade 4 writing (Spanish)	79%
Grade 7 writing	79%
English I	81%
English II	80%

ADMINISTRATION MODE

All English STAAR assessments are administered in both paper and online testing modes. Texas has conducted comparability studies (as described in the Comparability Analyses section) comparing student responses in paper and online versions for all STAAR assessments having more than 5% regular online test takers. In these studies, the comparability of scores was evaluated to determine whether an effect due to testing mode exists.

Evidence Based on Internal Structure

Texas collects evidence that shows the relationship of students' responses between items, within reporting categories of items, and within the full tests 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 written compositions (i.e., grade 4 writing, grade 7 writing, English I, English II, and English III 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 the following 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 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 2017–2018 STAAR administrations (including STAAR Spanish) 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. Student test results for STAAR are classified into one of four performance levels:

- Did Not Meet Grade Level

- Approaches Grade Level

- Meets Grade Level

- Masters Grade Level

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.

DISTRICT- OR CAMPUS-LEVEL

- STAAR performance results can be aggregated to provide one indicator of overall student proficiency at a district or campus.

- 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 district or campus.

- STAAR performance results can be aggregated across years to provide one indicator of overall student academic progress at a district or campus.

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. In 2017–2018, progress measures were calculated and reported for STAAR English mathematics and reading for grades 4–8, Spanish mathematics and reading for grades 4 and 5, and the Algebra I and English II assessments.

For STAAR, progress is measured as a student's gain score, meaning the difference between the scale score a student achieved in the prior school year and the scale score a student achieved in the current school year. These gain scores are then classified as *Limited*, *Expected*, and *Accelerated* in relation to progress targets.

2013–2014 Progress Measure Updates

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

CHANGES MADE TO INCLUDE MORE STUDENTS IN PROGRESS MEASURE CALCULATIONS

- Accelerated students/skipped grade(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 also provided for students who received accelerated instruction and skipped grades or courses. For example, a progress measure was reported if a student tested in grade 5 reading in the 2012–2013 school year and in grade 7 reading in the 2013–2014 school year.

- Changes related to different language-version tests

  Beginning in 2014, progress measures were reported for students who test in a different language in subsequent school years 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 school year and the current school year.

- Progress measure for EOC Assessments

  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, a progress measure is computed only on a student's first attempt of an EOC test.

■ Previous school 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 mathematics), the higher valid score between the primary administration and May administration is used for progress measure calculation. Retest scores for the June grades 5 and 8 administration from the previous school year are not used for progress measure calculations. If a student took the same EOC test multiple times during the previous school year, the score from the first administration of the previous school year test is used to calculate progress.

PROGRESS MEASURE UPDATES MADE IN RESPONSE TO LEGISLATIVE CHANGES IN 2013

■ English tests

Beginning in 2014, redesigned STAAR English tests were administered. Rather than having separate English reading and writing tests, these assessments were combined into one test per course (STAAR English I and STAAR 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 STAAR English I reading and STAAR English I writing tests in 2013, and progress could not be measured from those separate English I tests to the new combined STAAR English II test administered in 2014. For most students, progress for STAAR English II was measured for the first time in spring 2015, after the new combined English tests had been administered for two years.

2014–2015 Progress Measure Updates

Changes to various aspects of the STAAR program affected TEA's ability to report student progress on the assessments. The list below outlines adjustments to the STAAR progress measures reported for the 2014–2015 school year:

■ Students did not receive a STAAR progress measure for mathematics in grades 4–8 (Algebra I students still received a STAAR progress measure).

■ A STAAR progress measure was reported for the first time for grade 7 writing.

■ A STAAR progress measure was reported for eligible STAAR A students who took STAAR or STAAR L in the 2013–2014 school year.

■ Students did not receive a STAAR Alternate 2 progress measure.

2015–2016 Progress Measure Updates

There were some adjustments to the STAAR progress measures reported for the 2015–2016 school year:

■ Students did not receive a STAAR progress measure for grade 7 writing because grades 4 and 7 writing tests were shortened in the 2015–2016 school year.

■ A progress measure was available for STAAR Alternate 2 in 2016.

2016–2017 Progress Measure Updates

■ Beginning in spring 2017, the previous three-performance level system (i.e., Unsatisfactory, Satisfactory, and Advanced) was replaced with the four-performance level system (i.e., *Did Not Meet Grade Level*, *Approaches Grade Level*, *Meets Grade Level*, and *Masters Grade Level*).

■ The previous labels for STAAR progress measure (i.e., *Did Not Meet*, *Met*, and *Exceeded*) were replaced with *Limited*, *Expected*, and *Accelerated*.

■ STAAR A and STAAR L were no longer administered. The elimination of STAAR A and STAAR L changed the eligibility rules.

■ Blueprints for STAAR grades 3–8 tests, STAAR English I, STAAR English II, and STAAR English III were changed. Short-answer items were removed from the STAAR EOC tests (STAAR English I, STAAR English II, and STAAR English III) beginning in spring 2017. This had impacts on the maximum and chance scores in Table 4.13.

2017–2018 Progress Measure Updates

■ Beginning in the 2017–2018 school year, progress measure was also calculated for ELs. Progress for students who take STAAR or STAAR Spanish was measured in the same way for all students, including ELs. Prior to the 2017–2018 school year, qualifying ELs who tested in English received the ELL progress measures rather than the STAAR progress measures.

Progress Classifications

To interpret the gain scores, the scores are compared to progress targets to determine if a student made *Limited*, *Expected*, and *Accelerated* progress. 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 *Meets Grade Level* or above and having high performing students maintain *Masters Grade Level* performance.

Specifically, for students who achieved *Did Not Meet Grade Level, Approaches Grade Level, or Meets Grade Level* in the prior year, the *Expected* progress target is defined

as the distance between the *Meets Grade Level* performance standard on the prior-school-year test and the *Meets Grade Level* performance standard on the current-school-year test in the same content area. For students who achieved *Masters Grade Level* in the prior year, the progress target is based on the distance between the *Masters Grade Level* on the prior-school-year test and the *Masters Grade Level* in the current-school-year test in the same content area.

The *Accelerated* progress classification is a designation reserved for those students who have demonstrated significant growth over the course of the year, beyond that of the *Expected* progress target. The *Accelerated* progress target is defined as the distance between the *Meets Grade Level* on the prior-year test and the *Masters Grade Level* on the current-year test.

Students with gain scores less than the *Expected* progress target are classified as *Limited* progress. Students with gain scores greater than or equal to the *Expected* progress target and less than or equal to the *Accelerated* progress target are classified as having *Expected* progress. Students with gain scores greater than the *Accelerated* progress target are classified as having *Accelerated* progress.

Because the performance standards are not the same across grades and content areas (i.e., they do not have the same numerical value), the *Expected* and *Accelerated* progress targets differ from grade to grade and across content areas. Table 4.13 lists the STAAR progress measure targets.

Table 4.13. 2018 STAAR Progress Measure Targets

Current-School-Year Test	Prior-School-Year Test	Expected (Did Not Meet/ Approaches[1] Target)	Expected (Meets[2] Target)	Accelerated[3] Target	Top Score Range[4]	Chance Score Range[5]
Grade 4 mathematics[6]	Grade 3 mathematics	103	74	184	32–34	0–7
Grade 5 mathematics[6]	Grade 4 mathematics	36	54	135	34–36	0–8
Grade 6 mathematics	Grade 5 mathematics	28	48	147	36–38	0–8
Grade 7 mathematics	Grade 6 mathematics	35	26	145	38–40	0–9
Grade 8 mathematics	Grade 7 mathematics	12	56	166	40–42	0–9
Algebra I	Grade 7 mathematics	2312	2535	2645	52–54	0–12
Algebra I	Grade 8 mathematics	2300	2479	2633	52–54	0–12
Grade 4 English reading	Grade 3 English reading	82	78	165	34–36	0–9

Current-School-Year Test	Prior-School-Year Test	Expected (Did Not Meet/Approaches[1] Target)	Expected (Meets[2] Target)	Accelerated[3] Target	Top Score Range[4]	Chance Score Range[5]
Grade 4 Spanish reading	Grade 3 Spanish reading	95	104	192	34–36	0–9
Grade 5 English reading	Grade 4 English reading	32	34	117	36–38	0–9
Grade 5 Spanish reading	Grade 4 Spanish reading	43	65	162	36–38	0–9
Grade 6 reading	Grade 5 English reading	47	51	136	38–40	0–10
Grade 7 reading	Grade 6 reading	45	35	124	40–42	0–10
Grade 8 reading	Grade 7 reading	26	30	109	42–44	0–11
English II	English I	0	140	831	66–68	0–17

NOTE:

[1] Expected Did Not Meet/Approaches is the distance or difference between the *Meets Grade Level* performance standard (scale score cut) on the prior-year test and the *Meets Grade Level* performance standard (scale score cut) on the current-year test.

[2] Expected Meets is the distance or difference between the *Masters Grade Level* standard (scale score cut) on the prior-year test and the *Masters Grade Level* performance standard (scale score cut) on the current-year test.

[3] Accelerated is the distance or difference between the *Meets Grade Level* performance standard (scale score cut) on the prior-year test and the *Masters Grade Level* performance standard (scale score cut) on the current-year test.

[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. For STAAR mathematics, reading, and Algebra I, chance is defined as ¼ of the multiple-choice questions (i.e., not including griddable questions). Chance on STAAR English II is defined as ¼ of the possible multiple-choice raw-score points, plus 4 to account for the essay chance score (the score weight of 2 times the essay chance score of 2). Chance score is rounded to the nearest integer.

[6] Applies for both STAAR English and Spanish mathematics.

Additional information such as steps for calculating progress measures and progress targets for each STAAR grade and content area, including when students skip grade levels, can be found on the Progress Measures 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 *Limited*, *Expected*, and *Accelerated* 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 *Limited*, *Expected*, and *Accelerated* definitions.

■ All students scoring at the three highest raw scores in the current year will be classified as having *Accelerated* progress.

- Students who maintained *Masters Grade Level* from the prior year to the current year will be classified as having *Expected* or *Accelerated* progress. (*Limited* classification will not be applied to these students.)

- Students scoring at or below chance in the current year will be classified as having *Limited* 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 STAAR English I and STAAR English II assessments include multiple-choice 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, plus the weighted value (2x) associated with summed scores of 2 on the essays (representing a rubric score of 1 from both raters).

The score values associated with these exceptions for each STAAR grade and content can be found in Table 4.13 and in the "Calculating Progress Measures" document on the Progress Measures page of TEA's Student Assessment Division website.

Results

STAAR progress measure results from 2017–2018 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 *Limited* or *Expected*. In contrast, the *Accelerated* classification applied to a smaller number of students. This pattern was expected because, by definition, it requires significantly more progress to receive an *Accelerated* progress classification than to receive the *Expected* 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 used the STAAR progress measure and extrapolated performance into future years to determine if a student is on track to achieve *Meets Grade Level* in a later grade or course.

In 2018, on-track measures were available for mathematics (English and Spanish) in grades 4–8, English reading in grades 4–7, and Spanish reading in grade 4. The STAAR on-track measure is not available for EOC tests.

On-track is measured only for students who take STAAR assessments (including STAAR Spanish). There is no on-track measure for STAAR Alternate 2.

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 7 reading (current year assessment)

because the previous year assessment was grade 6 reading and the target year assessment will be grade 8 reading; all three of these assessments measure reading. 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 Progress Measures 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. 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 gathers reliable information about student performance on its assessments while minimizing the burden placed on districts and campuses.

During the 2017–2018 school year, the stratified random sampling method was used to select a representative student sample for the standalone writing prompt study to be conducted in February 2019 for grade 4 writing in English and Spanish, grade 7 writing, English I, and English II. The spring 2018 statewide student data were used for the sampling. Campuses were the smallest sampling units, which were grouped (i.e., stratified) according to students' performance; then, a random sample was selected from each group. For more details about sampling, refer to chapter 3, "Standard Technical Processes."

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, administered in spring 2018. Table 4.14 shows the spring 2018 pass rates for the STAAR assessments, including STAAR Spanish. For STAAR EOC assessments, pass rates are given for first-time testers.

Table 4.14. STAAR Spring 2018 Pass Rates

Content Area	Grade/Course	Pass Rate
Mathematics	Grade 3	77%
	Grade 4	78%
	Grade 5	84%
	Grade 6	76%
	Grade 7	71%
	Grade 8	78%
Mathematics (Spanish)	Grade 3	70%
	Grade 4	61%
	Grade 5	63%
Mathematics	Algebra I	88%
	Algebra II	77%
Reading	Grade 3	76%
	Grade 4	72%
	Grade 5	78%
	Grade 6	66%
	Grade 7	72%
	Grade 8	76%
Reading (Spanish)	Grade 3	73%
	Grade 4	61%
	Grade 5	76%
Writing	Grade 4	61%
	Grade 7	67%
Writing (Spanish)	Grade 4	63%
Science	Grade 5	75%
	Grade 8	74%
	Biology	90%
Science (Spanish)	Grade 5	52%
Social Studies	Grade 8	64%
	U.S. History	93%
English	English I	71%
	English II	75%
	English III	72%