Chapter 2
Test Design and Setting Student Performance Standards for State of Texas Assessments of Academic Readiness (STAAR) Grades 3–8 and STAAR End-of-Course (EOC)

Test Design

One of the primary goals of the STAAR program is to increase the rigor of the assessments so that students have the academic knowledge and skills they need to meet the challenges of the 21st century. As was the case with previous state assessment programs in Texas, the STAAR program will continue to assess the statewide content standards, the Texas Essential Knowledge and Skills (TEKS). However, the test design for STAAR is fundamentally different from past state assessments. STAAR assessments are being developed using three major design attributes: focus, clarity, and depth.

Design Attribute: Focus

By focusing on the TEKS that are most critical to assess, STAAR will better measure the academic performance of students as they progress from elementary to middle to high school. In an effort to structure STAAR assessments so that they are more focused, TEA has made a distinction between "readiness" and "supporting" standards from the TEKS content standards eligible for assessment. Based on feedback from Texas educators (from both K–12 and higher education), a set of readiness standards has been identified for each subject and grade or course drawn from the TEKS content standards eligible for assessment. These readiness standards will be emphasized annually in the STAAR assessments. The content standards that were deemed to be supporting are still an important part of instruction and are eligible for assessment. However, the supporting standards may not all be tested each year.

The following table compares readiness and supporting standards.

Comparison of Readiness and Supporting Standards

Readiness Standards	Supporting Standards
General characteristics	
• are essential for success in the current grade or course • are important for preparedness for the next grade or course • support college and career readiness • necessitate in-depth instruction • address significant content and concepts	• introduced in the current grade or course but may be emphasized in a subsequent year • reinforced in the current grade or course but may be emphasized in a previous year • play a role in preparing students for the next grade or course but not a central role • address more narrowly defined content and concepts
• Subject-specific characteristics	
• For Reading, Writing, and English Language Arts: • focus on specific reading genres (fiction and expository) and on writing for particular purposes • For Mathematics: • emphasize the integration and application of mathematical skills • For Science: • emphasize the integration and application of major scientific concepts • For Social Studies: • emphasize landmark historical events and foundational geographic concepts • emphasize unifying historical and geographical themes	• For Reading, Writing, and English Language Arts: • may apply to other reading genres (poetry, drama, literary nonfiction, and persuasive) • For Mathematics: • focus on skills that underlie more significant mathematical concepts • For Science: • focus on content that supports fundamental scientific principles • For Social Studies: • focus on discrete historical facts, events, or individual people, as well as more detail-oriented geographical facts and concepts

Design Attribute: Clarity

The TAKS program was originally designed to assess a wide range of knowledge and skills, resulting in an assessment system that covered a breadth of content standards. The STAAR program is designed to focus assessments on readiness standards and course-specific content standards. This design will provide school districts, teachers, and students clarity regarding what will be assessed and how the assessed content standards are preparing students for their next step—the following grade, course, or college and career. TEA will continue to provide educators with information about each assessment to identify readiness and supporting standards, clearly reflect the relationship between the TEKS and the STAAR assessment program, explain the role of readiness and supporting standards on the tests, and provide sample items from the new assessments. As new information regarding the STAAR program becomes available, TEA will alert district personnel via broadcast e-mails. Currently, information about the STAAR program can be found on the TEA website at http://www.tea.state.tx.us/student.assessment/staar/.

Another aspect of clarity in the STAAR program is that the majority of the assessments will test content studied that year, as opposed to testing content studied over multiple years. Doing so will strengthen the alignment between what is taught and what is tested for a given course of study. While STAAR mathematics, reading, writing, and social studies assessments in grades 3–8 will continue to address only those TEKS taught in the given subject and grade, the content of other STAAR assessments will change in the following ways:

- Although the new science assessments for grades 5 and 8 will continue to address TEKS from multiple grade levels, these tests will focus on the science TEKS for those respective grades. The science assessments at these two grades will emphasize the 5th and 8th grade content standards that best prepare students for the next grade or course (i.e., biology, chemistry, physics). In addition, these assessments will include content standards from two lower grades (i.e., grades 3 and 4 or grades 6 and 7) that support students' success on future science assessments. In contrast, the current Texas Assessment of Knowledge and Skills (TAKS) assessments uniformly address TEKS from multiple grade levels without any specific emphasis.
- The new end-of-course assessments will address only the TEKS for a given course, as opposed to the high school level TAKS assessments, which address TEKS from multiple courses.

Design Attribute: Depth

A primary feature of STAAR's test design is a focus on preparedness for success in subsequent grades or courses and, ultimately, for college and career. This requires the tests to emphasize depth rather than breadth in assessing student expectations. A number of changes have been implemented in STAAR to allow skills to be tested in a deeper way.

- Tests will contain a greater number of items that have a higher cognitive complexity level.
- Items will be developed to more closely match the cognitive complexity level evident in the TEKS.
- In reading, greater emphasis will be given to critical analysis than literal understanding.
- In writing, students will be required to write two essays rather than one. The writing prompts will support analytical, persuasive, and expository writing in addition to literary writing.
- In social studies, science, and mathematics, process skills will be assessed in context, not in isolation, which will allow for a more integrated and authentic assessment of these content areas.
- In science and mathematics, the number of open-ended (griddable) items will increase to allow students more opportunity to derive an answer independently.

Increased Rigor

With greater focus, clarity, and depth in assessment, it is possible to develop a more rigorous testing program. The following table summarizes how rigor will be emphasized in the program at the individual question level, at the total test level, and through the performance standards. Additional information regarding rigor in the STAAR program can be found in Chapter 1.

Increased Rigor in the STAAR Program

General Characteristics of STAAR That Will Contribute to Rigor

• The rigor of items will be increased by
  o assessing content and skills at a greater depth and higher level of cognitive complexity
  o assessing more than one student expectation in an item
• The rigor of the tests will be increased by
  o assessing more focused student expectations but doing so multiple times and in more complex ways
  o including a greater number of rigorous items on the test, thereby increasing the overall test difficulty
• Performance standards will be increased by
  o using empirical data to link performance in specific courses to college and career readiness
  o using empirical studies to compare student performance on the new assessments with other national assessments
  o reviewing performance standards at least once every three years and, if necessary, adjusting them to maintain a high level of rigor
  o expectations for student performance on STAAR will be raised to achieve the goal of graduating students who are college and career ready

Test and Item Specifications

Test Specifications

Test specifications provide the underlying structure for the assessments, supporting how the assessments will be designed, constructed, administered, and scored. Tests will be constructed to match a test blueprint that identifies the total number of questions on each test, with a majority of test questions addressing readiness standards from the content standards. Each STAAR assessment will consist primarily of multiple-choice questions addressing the content standards for the grade or course.

STAAR Grades 3–8

All STAAR grades 3–8 assessments will be offered in paper-and-pencil format. Each STAAR grades 3–8 assessment will consist primarily of multiple-choice questions addressing the content standards for the grade level and subject. All mathematics assessments and the grades 5 and 8 science assessments will include open-ended items that are machine scorable, referred to as griddable items, in which the answer is generated by the student instead of being selected from a set of options. In this

format, a student records a numerical response using several columns of response bubbles. In addition, TEA is considering dictionary and calculator use on some STAAR assessments at grades 3–8, and these decisions will be communicated to districts as soon as they are finalized.

The writing assessments for grades 4 and 7 will be administered over the course of two days (the STAAR Writing Test Design documents can be found on the TEA website at http://www.tea.state.tx.us/student.assessment/staar/) and will consist of multiple-choice questions addressing revising and editing skills and two one-page written compositions. This design differs from TAKS in that the TAKS writing assessment at grades 4 and 7 was administered over a one-day period and required only one personal narrative essay. For grade 7 only, a third written composition and a small number of multiple-choice questions as field-test items will be embedded. This represents a major change from TAKS, since the TAKS design did not allow the embedded field-testing of writing prompts and thus required an annual stand-alone field test.

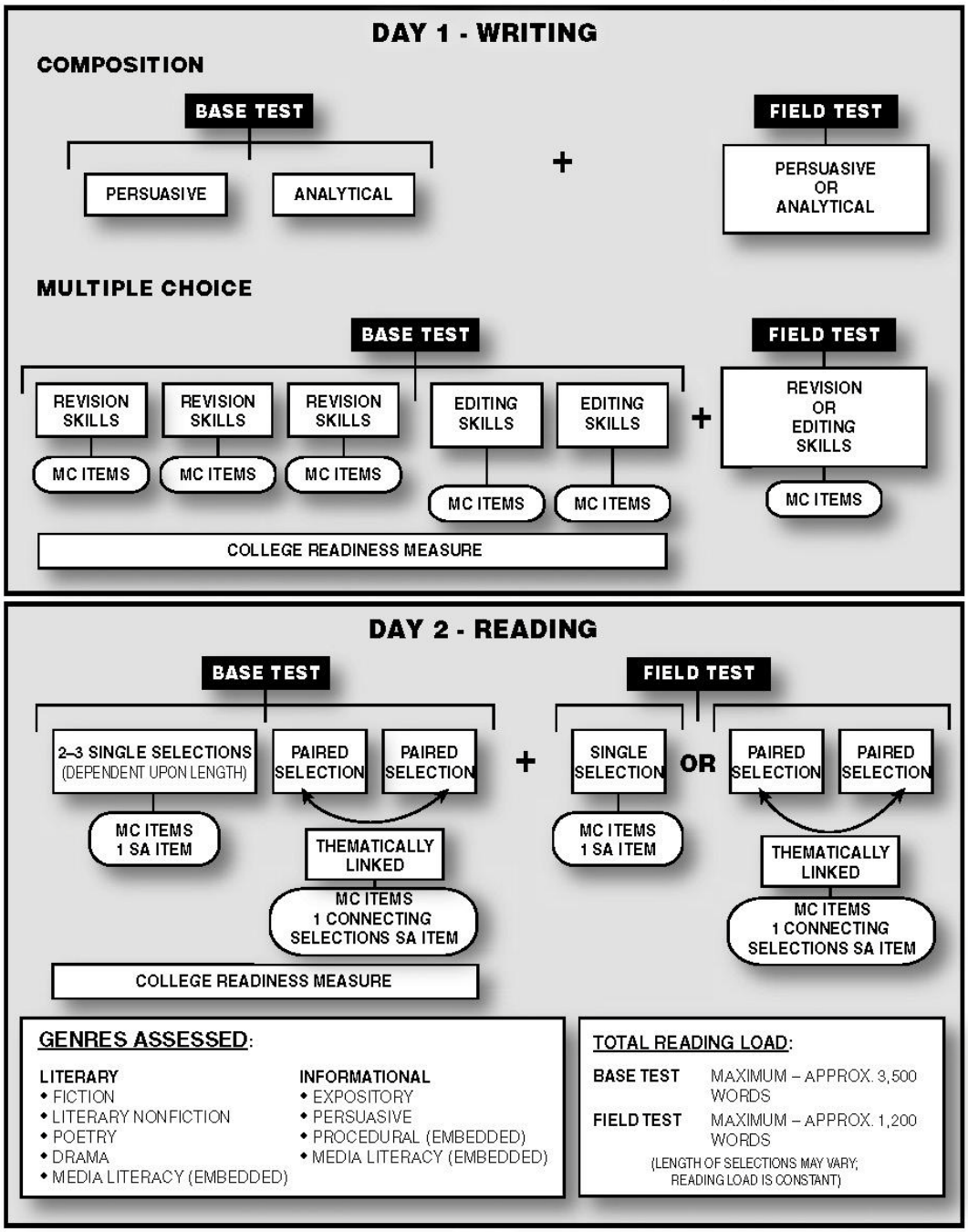
The STAAR grades 3–8 reading assessment will consist of multiple-choice questions related to reading selections drawn from a variety of published and commissioned pieces. For grades 4–8, three to four stand-alone selections and one pair of thematically linked selections will be included on each test; grade 3 reading will not include paired selections. For thematically linked reading selections, the test will incorporate questions that require students to demonstrate an understanding of the connections between the two texts. The selections on the reading assessments will be genre-based and will include both literary (fiction, literary nonfiction, poetry, drama, media literacy) and informational (expository, persuasive, procedural, media literacy) texts.

STAAR EOC

STAAR EOC assessments are offered in both online and paper formats. As with STAAR grades 3–8, all mathematics assessments and some science assessments will include griddable items. In addition, the current policies for calculator use for EOC assessments will continue for STAAR. Calculators will be required for all mathematics and science EOC assessments as was announced in the letter TEA sent to school districts in September 2009.

The English I, II, and III assessments are designed as two-day assessments. The following English III test design is provided as an example of how the assessment is administered across two days (the English I, II Test Design documents can be found on the TEA website at http://www.tea.state.tx.us/student.assessment/staar/).

STAAR English III Test Design



Writing, administered on day one, consists of two one-page written compositions and multiple-choice questions addressing revising and editing skills. Reading, administered on day two, primarily consists of multiple-choice questions related to reading selections drawn from published pieces. Each test contains two to three stand-alone reading selections and one pair of thematically linked selections. Test questions for the thematically linked selections will require students to demonstrate an understanding of the connections between the two texts. In addition, the tests include two open-ended questions to which students provide a short written response. The selections on the English assessments are genre-based and include both literary (fiction, literary nonfiction, poetry, drama, media literacy) and informational (expository, persuasive, procedural, media literacy) texts. This design differs from TAKS in that TAKS ELA at grade 10 and exit level was a one-day assessment that consisted of thematically linked "triplets" of texts with an integrated personal response writing sample for the reading component and multiple-choice items for the revising and editing component. Although the TAKS design was an authentic reflection of classroom instruction, the level of rigor associated with college and career readiness was not the focus of this test. STAAR English I, II, and III assessments, however, were designed with this focus in mind. In addition, access to dictionaries will be required for English I, II, and III. Dictionary use on other EOC assessments is currently being considered, and this decision will be communicated to districts as soon as it is finalized.

Field-Test Items as Part of Test Specifications

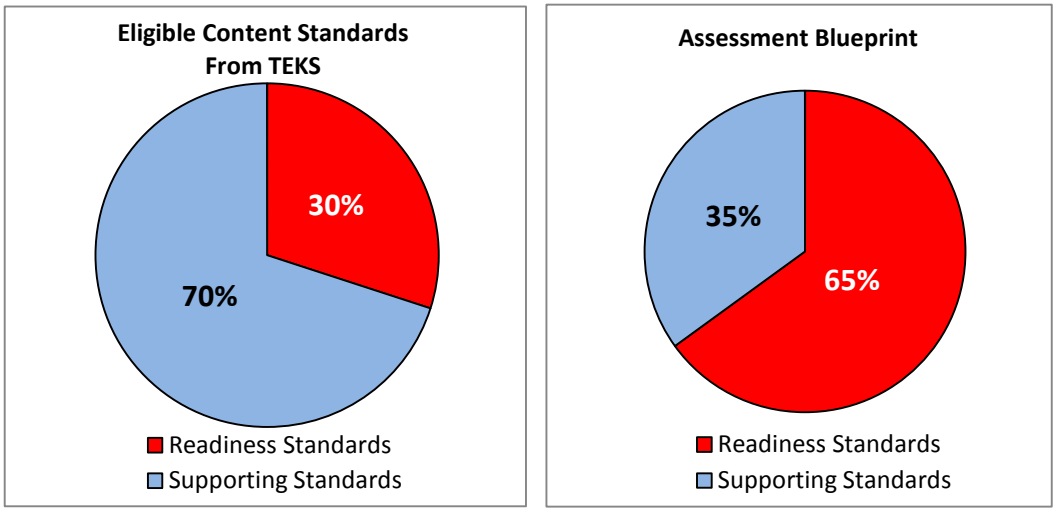
All STAAR assessments incorporate embedded field-test items. It is necessary to field-test items in order to gather item-level student performance data so that it can be determined how well the items will perform for the intended purpose. Student performance on field-test items does not contribute to a student's score. Up to eight multiple-choice items are embedded within each test. The STAAR English I, II, and III reading tests include one additional field-test reading selection (or one pair of thematically linked reading selections), multiple-choice field-test questions, and one open-ended short-answer field-test question. The STAAR EOC English I, II, and III writing tests include a field-test written composition prompt and multiple-choice field-test questions for revising and editing. Griddable field-test items are embedded in mathematics and science tests, as appropriate. The STAAR grades 3–8 reading tests include one additional reading selection (or one pair of thematically linked reading selections) and a set of multiple-choice field-test questions. As noted, the grade 7 writing test includes embedded multiple-choice questions for revising and editing and one field-test written composition prompt. For more information regarding field-testing, see Chapter 1.

Item Specifications

Item specifications provide guidance to the professional item writers who develop test questions for the STAAR program. The specifications offer guidelines for assessment strategies and include descriptions and samples of the kinds of items appropriate for each content standard. Item specifications for reading tests include acceptable ranges for selection length and guidelines for readability. The STAAR item specifications are in the process of being finalized.

Blueprints

Test blueprints specify the set of reporting categories (formerly referred to as objectives in TAKS) and student expectations to be measured on an assessment, as well as the number of items to be tested for each reporting category. The following graphic shows the relative relationship between the readiness and supporting standards in the TEKS content standards and the readiness and supporting standards that are assessed each year. The STAAR test blueprints are designed so that a larger number of test items measure student expectations designated as readiness standards. For more information about the blueprints, including example blueprints, see Chapter 1.



Alignment of the Assessments with the Content Standards

Alignment is central to the validity of the new STAAR student assessment system. STAAR will provide useful information for valid accountability decisions and educational improvement only to the extent that all components of the system are aligned. It is important to determine the extent to which STAAR adequately measures the knowledge and skills specified in the TEKS and the extent to which STAAR includes items that cover the full range of achievement standards, particularly at the highest achievement level.

Demonstrating that every item on STAAR can be matched to one or more content standards in the TEKS is necessary but not sufficient to ensure alignment. In addition to the content match, evidence of alignment also addresses the degree to which STAAR reflects the full range and breadth of the content standards as well as the degree of cognitive complexity evident in the standards.

The state gathers significant evidence to ensure that the tests are closely aligned to the grade-level content standards. The systematic and well-documented test development process used for STAAR includes annual item review committees composed of educators who represent the 20 regions of the state. These educators review every item for alignment to the content standards and to the sub-content

areas and discuss and reassign the content standard and sub-content area being assessed, as needed. Item judgments are collected for every item related to each item's alignment to content standards in response to the question "Does this item measure the reporting category/student expectation it was designed to measure?" Summaries of the committees' judgments related to each item's alignment to specific content standards and sub-content areas clearly demonstrate alignment between the STAAR tests and the content standards. The summaries are maintained as Item Content Committee Review Reports for every grade and subject for STAAR.

Every item chosen for inclusion on a STAAR test has undergone extensive review by TEA, its testing contractor, and approximately 40 independent Texas educators (20 in item review and 20 in data review) in terms of its alignment to the specific content standard and sub-content area. Because of the thoroughness of this content alignment, TEA is confident that STAAR reflects the knowledge and skills in the TEKS. It should be noted that there are plans by the State Board of Education (SBOE) for additional TEKS revisions. These revisions will have an overall impact on the alignment of the assessments with the content standards. If revisions are approved by the SBOE, TEA will work through a process similar to the one noted above to verify that STAAR items and the revised TEKS are aligned. In addition to the alignment process described above, current federal regulations require an independent alignment study as part of the peer review process. For more information about the peer review process, see Chapter 15.

STAAR Resources for Educators

TEA will provide educators with information about each assessment to identify readiness and supporting standards, clearly reflect the relationship between the TEKS and the STAAR assessment program, explain the role of readiness and supporting standards on the tests, and provide sample items from the new assessments. Some of this information is already posted on the TEA website (http://www.tea.state.tx.us/student.assessment/staar/), with additional information being added as it becomes available. The new STAAR resources for educators will include

- an overview of the subject or course within the context of the assessment;
- the TEKS that are eligible for the assessment, their grouping under reporting categories, and the identification of readiness and supporting standards;
- the test blueprint (the number of items under each reporting category and the number of items on the test as a whole, as well as information regarding the relative emphasis placed on readiness or supporting standards);
- additional information about each reporting category that will help educators understand how it is assessed; and
- sample items that demonstrate some of the ways in which content standards are assessed.

In addition, resources will be provided to state education leaders and school district personnel during the transition from TAKS to STAAR. Such resources include presentations at statewide

assessment and content-area conferences, statewide training sessions, and specialized training sessions (such as WebEx and teleconference meetings).

Release of Tests and Items

In addition to the information for educators noted above, TEA understands the need to release test items to school districts as they continue to prepare their students for the STAAR program and become more familiar with the new program. Current Texas Education Code requires the Texas Education Agency to release the primary form of the state assessment for every grade and subject tested every three years, but there is also a separate state statute [Texas Education Code (TEC) §39.025(f)(3)] that allows TEA to override this policy when a new assessment program is being developed and implemented. There are many reasons why items are not released during this time period, most notably to ensure that there are sufficient test items in the item bank to construct future tests so that, from the beginning of the program, the content standards can be assessed in the most valid and reliable way. In addition, a strong item bank is essential for the long-term life of the STAAR program.

Release Timeline

In August 2011, selected items that illustrate the new approach being used for the STAAR assessment program will be posted on the TEA website. As previously stated, in 2012 and 2013, the first few years that STAAR will be administered as high-stakes assessments, no release of tests will occur. In these years, it is possible that TEA will be able to release an additional small set of items. In 2014, TEA is planning the first full release of primary test forms of STAAR to meet the needs of educators and to fulfill current legislative requirements.

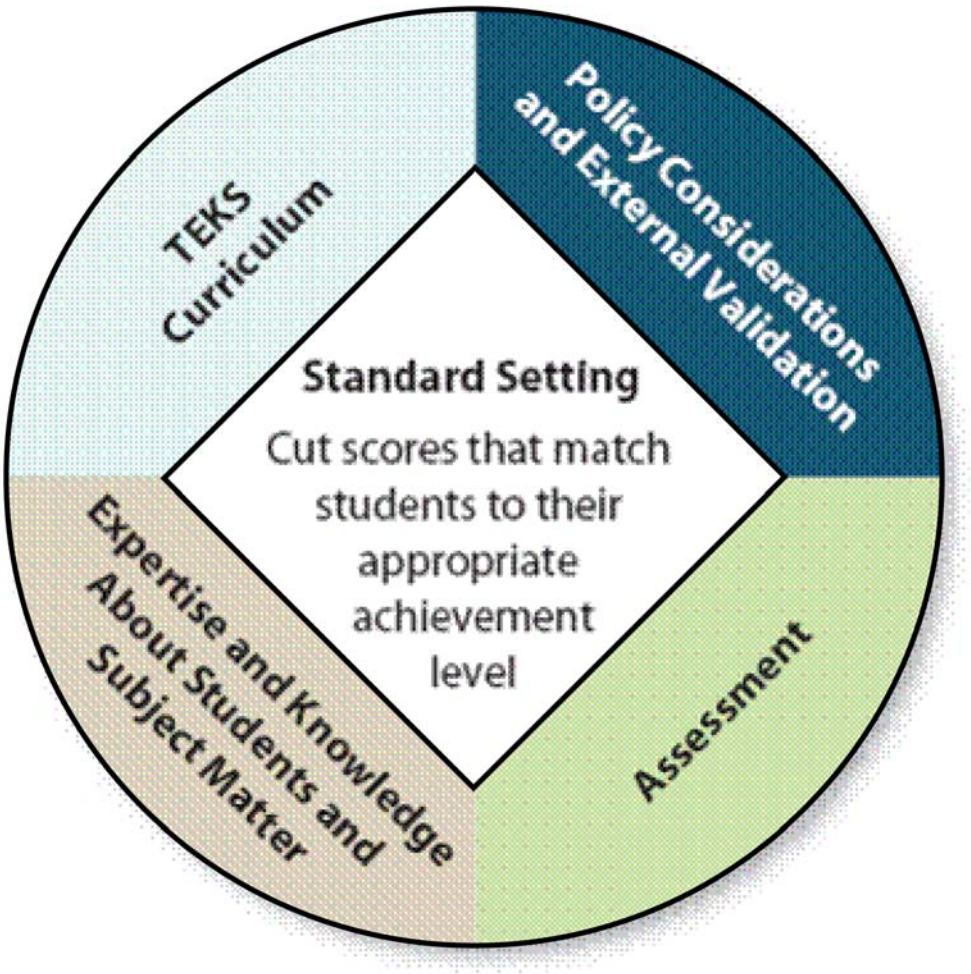
The release timeline is interdependent with field-test plans for the STAAR program. With STAAR, stand-alone field testing will be phased out and field-test items will be embedded in operational forms. If in the future it is determined that items are to be released more frequently, more items will need to be developed to replenish the item bank. This will require additional field testing on an annual basis either as a stand-alone field-testing model or by embedding more field-test items in operational forms which will increase overall test length. A change to the release plan will also increase the overall cost to the assessment program as more items will need to be developed and field tested.

General Overview of Setting Performance Standards

A critical aspect of any statewide testing program is the establishment of performance levels that provide a frame of reference for interpreting test scores. Once an assessment is given, students, parents, educators, administrators, and policymakers want to know, in clear language, how students performed on that assessment. In general, performance standards relate test performance directly to the student expectations expressed in the state curriculum in terms of what students are expected to

learn by the completion of each grade level or for a specific course. Performance standards describe the level of competence students are expected to exhibit at specific grades/courses as they progress through the educational system.

As Texas moves toward implementing the STAAR program, which includes indicators of college and career readiness, a standard-setting method that is more evidence-based will be used. Standard setting for the STAAR program involves a process of combining considerations about policy, the TEKS content standards, educator knowledge about what students should know and be able to do, and information about how student performance on the statewide assessments aligns with performance on other assessments. Standard-setting advisory panels composed of diverse groups of stakeholders consider the interaction of these elements for each assessment. The following graphic illustrates the critical elements of standard setting.



The timing and details of the processes used for STAAR grades 3–8 and STAAR EOC will differ because of different legislative requirements and the timing of available assessment data. See the following table for more information regarding the timing and details of standard-setting activities.

Preliminary Plan for the Standard-Setting Process for STAAR

Standard-Setting Process	TAKS	STAAR	STAAR Timeline
1. Conduct validity and linking studies	N/A	External validity evidence will be collected to inform standard setting and support interpretations of the standards. Scores on the assessments will be linked to past and future performance in the same content area	Studies begun in spring 2009 and will continue throughout the program.
2. Develop performance labels and policy definitions	Committee convened by Texas Education Agency (TEA)	Committee convened jointly by TEA and Texas Higher Education Coordinating Board	September 2010
3. Develop specific performance-level descriptors for each grade, subject, and course	Were developed separately during each standard-setting committee meeting	To be developed prior to the standard-setting committee meetings as an aligned system describing an appropriate progression of skills	March 2011
4. Standard-Setting Committee	Membership primarily of K–12 educators	Increased representation of members with higher education and policy backgrounds in addition to K–12 educators	February 2012*
5. Policy Review Committee	N/A	Considers policy implications and alignment across content areas	March 2012*
6. Approval of Performance Standards	Approved by State Board of Education	Approved by Commissioner of Education (and Commissioner of Higher Education for college readiness standards)	February 2012*
7. Implementation of Performance Standards	Phase-in based on standard error of measurement	Phase-in process TBD	May 2012*
8. First review of performance standards	Completed after major changes to the program	Completed on a pre-determined schedule at least every three years	Fall 2013*

* These dates are for the STAAR end-of-course program. These steps for the STAAR 3–8 program will occur the following school year.

Performance Standard Requirements

Any grades 3–8 assessment used for state or federal accountability needs to have at least two cut scores, or performance standards—one that distinguishes between basic and proficient achievement levels (referred to as the proficient cut score) and one that distinguishes between the proficient and advanced achievement levels (referred to as the advanced cut score).

Current state legislation mandates setting several performance standards on each STAAR EOC assessment. For all twelve assessments, there should be a cut score that indicates satisfactory performance. There should be a minimum score set below but within a reasonable range of the satisfactory score, which will be used to determine whether a student's score on a particular STAAR EOC assessment may count toward his or her cumulative score in that content area. The minimum score will be set empirically. Performance at the highest cut score will indicate a strong application of knowledge and skills, and the results from performance at this level will be interpreted differently depending on the EOC assessment. For example, this highest cut will indicate college readiness for Algebra II and English III. It will indicate advanced course readiness for Algebra I, English I, and English II, and it will indicate advanced performance for the remaining courses. For more information regarding advanced-course readiness, see Chapter 3.

The following table gives a summary of the performance standards that will be set for the STAAR EOC program.

Summary of Performance Standards for the STAAR EOC Program

STAAR EOC Assessments	Satisfactory Performance*	Advanced Performance
Algebra I	√	√ (Advanced Course Readiness)
Geometry	√	√
Algebra II	√	√ (College Readiness)
English I	√	√ (Advanced Course Readiness)
English II	√	√ (Advanced Course Readiness)
English III	√	√ (College Readiness)
Biology	√	√ (College Readiness-TBD)
Chemistry	√	√ (College Readiness-TBD)
Physics	√	√ (College Readiness-TBD)
World Geography	√	√
World History	√	√
U.S. History	√	√ (College Readiness-TBD)

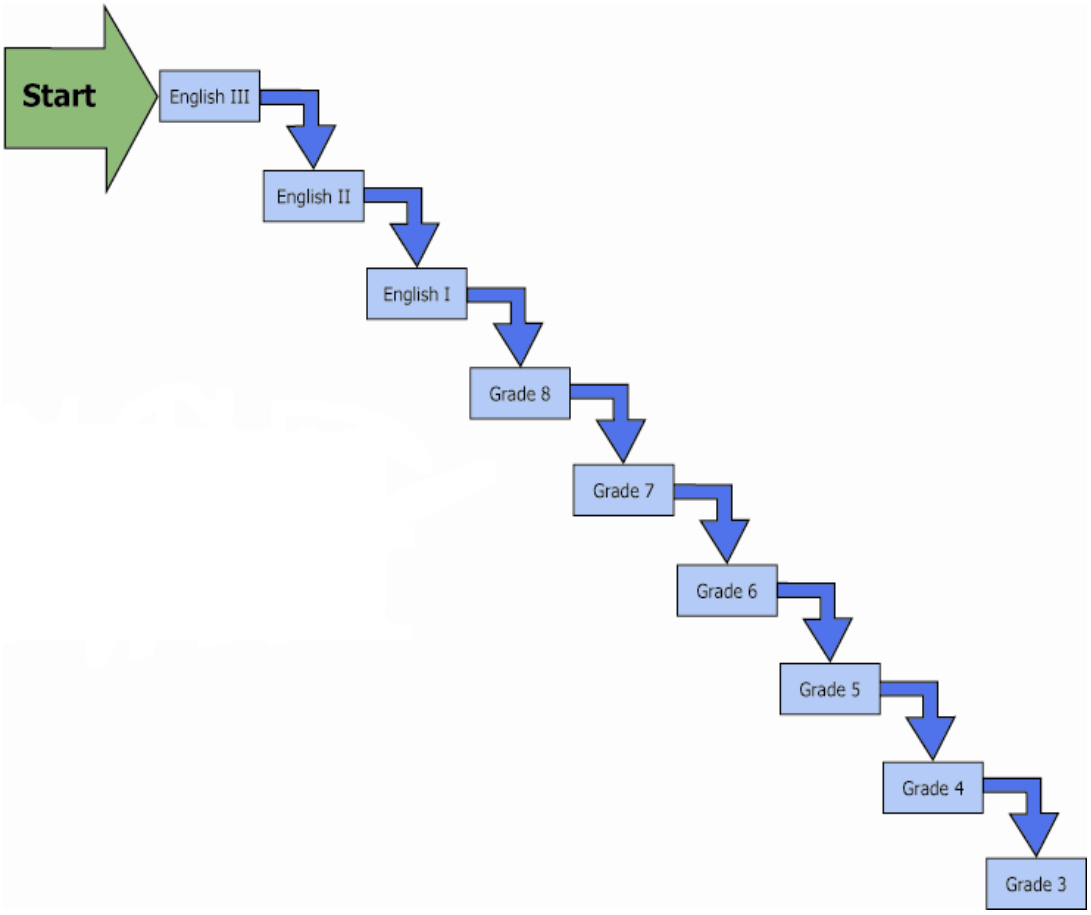
* A minimum score within a reasonable range of the satisfactory score will be set empirically.

Proposed Standard-Setting Process

To achieve alignment for the STAAR program, performance standards will be established first for Algebra II and English III, and will be used to establish performance links down the subject/grade levels all the way to grade three, which is the earliest grade that Texas assesses mathematics and reading.

The following illustration demonstrates how performance standards will be linked between higher-level courses and lower-level grades, starting with the highest-level course, English III, and moving down to grade 3 reading.

Process for Setting Performance Standards from English III Backwards to Grade 3 Reading Using Empirical Data



An eight-step process is currently planned for setting the performance standards on the STAAR assessments. The eight steps include the following:

1. Conducting external validity and linking studies
2. Developing performance-category labels and policy definitions
3. Developing preliminary specific performance-level descriptors (PLDs) for each grade or course
4. Convening standard-setting committees
5. Convening a policy-review committee
6. Approving performance standards
7. Implementing performance standards
8. Reviewing performance standards

A description of each step in the standard-setting process, including the planned timelines, is provided on the following pages.

Step 1: Conducting external validity and linking studies

STAAR Grades 3–8

State law mandates that research studies be conducted to evaluate the link between passing standards on the STAAR grades 3–8 assessments across grade levels in mathematics and reading. Performance standards should be set based on information from the results of these linking studies. HB 3 also requires that studies be conducted to evaluate the correlation between passing standards for grade 8 to EOC assessments in mathematics and reading. The details of these and other planned studies are discussed later in this section.

STAAR EOC

HB 3 mandates that research studies be conducted to evaluate the link between scores on the STAAR EOC assessments within the English content area and between scores on the algebra assessments. The performance standards should be set based on information from the results of these linking studies. HB 3 also requires the establishment of college- and career-readiness performance standards for English III and Algebra II assessments. It also includes the evaluation of potential college- and career-readiness performance standards in the science and social studies content areas based on studies that correlate performance on STAAR EOC assessments and college and career readiness. Additional studies required by HB 3 include the comparability of EOC assessments to national and international assessments and to military and/or workforce success. More details about the studies that will be conducted are included later in this section.

Step 2: Developing performance-category labels and policy definitions

A committee of stakeholders convened at the end of September 2010 to develop performance-category labels and policy definitions that are to be used across all STAAR assessments. The panel consisted of representatives from a variety of education and policy groups. Because standards are required to be linked between STAAR grades 3–8 and STAAR EOC, a single committee was convened to make recommendations for the entire STAAR assessment program.

The Performance Descriptor Advisory Committee meeting was conducted over two days. This committee was charged with the following.

- Reach consensus on recommendations for the names of the performance labels (categories of performance) for student achievement on the assessments (general, modified, and alternate).
- Make recommendations for key words/phrases to be used in drafting the policy definitions that will define student performance within each category.
- Assume that the state assessment system will be implemented under current federal and state statute, both of which require a minimum of three performance levels.

A detailed report outlining the activities and recommendations from the Performance Descriptor Advisory Committee meeting can be found in Appendix A.

Step 3: Developing preliminary specific PLDs for each grade or course

Educator committees with specific content knowledge (K–12 and higher education) will then develop preliminary specific PLDs for STAAR reading and writing, mathematics, science, and social studies content areas.

The PLDs developed by these committees will extend the policy definitions of the performance-category levels to the specific grade/course and content areas. The committee members will be asked to conceptualize more clearly the various labels in terms of specific content-based behaviors. For example, what specific knowledge and skills are needed for a student to be college or career ready in Algebra II? What is expected of a student who performs satisfactorily in science at grade 5 or in high school physics? PLDs for each performance level will describe what students at that level can do and to what degree. Because each successive performance level will assume students possess the knowledge and skills at the lower level(s), the PLDs will describe only the behaviors that are new or that differentiate students in the higher performance level.

Emphasis will also be placed on developing specific PLDs that are not only appropriate for the performance categories within the subject or course but will also align well with the corresponding categories in the other STAAR EOC and STAAR grades 3–8 assessments in the same content area. For example, the specific PLDs that describe satisfactory performance on the English I assessment should be logical steps up from the analogous specific PLDs in grade 8 reading and should represent a reasonable progression into the corresponding PLDs in English II. The committees will be expected to develop the PLDs independently. However, guidance and examples will be provided by TEA to assist in the process. The preliminary specific PLDs recommended by each committee will be reviewed, and the PLDs will continue to be refined and finalized through an iterative process during the blended standard-setting committee meetings (Step 4 below). Note that Step 3 for STAAR EOC and STAAR grades 3–8 will occur at different times. The plan and timeline chart found earlier in this section outlines preliminary plans for standard setting and highlights key differences between the STAAR and TAKS programs.

Step 4: Convening standard-setting committees

The STAAR standard-setting process should take into account the assessed TEKS content standards as well as policy considerations. The process will need to include recommendations from a blended committee comprised of three primary groups of constituents.

- Texas educators (K–12 and higher education)
- Policy experts (business community, workforce leaders, or other advocacy representatives)

- Individuals with dual expertise – education and policy. Some of these individuals will be asked to participate in the policy-review committee noted in Step 5 to provide continuity to the overall process.

These committees will be charged with recommending a cut score or a range of cut scores for each STAAR assessment and finalizing the specific PLDs. The committees could also make recommendations about possible phase-in options (for example, phase-in of performance standards over a period of time). It is anticipated that the satisfactory performance standard will be phased in over several years, but the highest performance standard (including the college- and career-readiness standard for Algebra II and English III) would not be phased in but applied as approved when STAAR becomes operational. The sequence of standard-setting activities expected to take place in each blended committee meeting includes the following.

- Each committee member will take an applicable STAAR assessment to experience the types of items, content, and depth of knowledge measured on STAAR.
- The committee will review the general and specific PLDs and be asked to consider more concretely what students in each performance category should know and be able to do.
- Committee members will look at results from the linking and external validity studies to see how performance on each STAAR assessment is related to that of other STAAR assessments and to national and international assessments in the same content area. The goal of this process is to identify score ranges in which it would be reasonable and meaningful to set the performance standards. Doing so will prevent committees from setting cut points on portions of the scale that are not empirically supported (for example, points on the scale that are below chance level).
- The committee will look at item content and recommend cut scores (or ranges of cut scores) for each assessment.
- Throughout the process, the committee will refine the PLDs as necessary so that there is solid alignment between the final committee cut score recommendations and the specific PLDs.

As with Step 3, Step 4 for STAAR EOC and STAAR grades 3–8 will occur at different times.

Step 5: Convening a policy-review committee

After the committees noted in Step 4 have met and made recommendations, a final policy-review committee will be convened to examine the recommendations made by the previous committees and determine the reasonableness of the standards across all assessments. The policy-review committee will consist of the following:

- educational policy experts (dual expertise) who participated in the standard-setting committees, noted in Step 4, to carry forward recommendations made by the blended committees; and
- new committee members who were not part of the blended committees.

This is an additional step in the standard-setting process that has been added for the STAAR program. As test scores serve an increased number of functions, it is important to consider the reasonableness and meaningfulness of the performance standards from a variety of perspectives. Reasonableness of the cut scores will be evaluated given the empirical data provided across content areas and for the entire STAAR program, including the empirical links from grade to grade in STAAR grades 3–8 assessments, from grade 8 to the English I and Algebra I assessments, and from course to course in the foundation content areas in high school. In addition the policy considerations such as accountability ratings and graduation impact data and the goals of the assessment program will be reviewed. This committee will also consider the recommended cut scores and possible phase-in plans for the standards. A single policy-review committee will be convened to evaluate the recommendations for all STAAR grades 3–8 assessments, and a separate committee will be convened for all twelve STAAR EOC assessments.

Step 6: Approving performance standards

The recommendations of the policy-review committee will be sent to the commissioner of education for review and approval. Both the commissioner of education and the commissioner of higher education will review the recommendations related to the college- and career-readiness performance standards. Approval of the performance standards and a potential phase-in plan for STAAR grades 3–8 will occur in December 2012 and in April 2012 for STAAR EOC.

Step 7: Implementing performance standards

Reports based on the new performance standards for STAAR grades 3–8 are scheduled to be first provided to students, campuses, and districts in late fall 2012 or early 2013. The new performance standards are also expected to be used in federal and state accountability systems beginning in 2013. Reports based on the new performance standards for STAAR EOC assessments are scheduled to be provided to students, campuses, and districts in May 2012. The new EOC performance standards are also expected to be used in federal and state accountability systems in 2013.

State accountability ratings will not be assigned in 2012. TEA will submit an Adequate Yearly Progress (AYP) Transition Plan to the U.S. Department of Education (USDE) in January 2011, for approval of the release schedule for the 2012 federal accountability ratings.

Step 8: Reviewing performance standards

Standards will be reviewed at least once every three years, as required by state statute. Additional impact and validity-study data collected after the initial standard-setting meetings will be presented during these reviews. This is an important step in the overall longevity of the program. As student performance increases because of improved instructional practices, the standards may need to be reviewed and then raised to continue to challenge the students of Texas to achieve a higher level of performance. In addition, over a three-year period, additional student data can be collected that can more accurately substantiate the correlation of student performance across grades and courses and to

postsecondary readiness. Timelines and descriptions of the STAAR grades 3–8 and STAAR EOC standard-setting process can be found earlier in this section.

Validity Studies

Empirical studies are a component of the implementation plan for STAAR. Test score interpretations and the uses of STAAR assessment data must be supported by validity evidence, such as that provided by correlating the STAAR assessments with other tests or measures of student performance. To help inform empirical studies and provide validity evidence based on test content, an analysis is also being conducted to compare the assessed content standards on the STAAR assessments with other external assessments.

Comparisons with National and International Assessments

Some of the studies planned to inform setting performance standards for the STAAR program are comparisons with national and international assessments, as it is important for Texas students to be competitive in the global economy. These studies will inform the performance standards for multiple content areas. For the initial standard setting for the STAAR program, performance of Texas students on the National Assessment of Educational Progress (NAEP) will be used, in conjunction with other data, to evaluate the reasonableness and rigor of the performance standards.

Data from an international assessment will be evaluated when performance standards are reviewed. Data from the Trends in International Mathematics and Science Study (TIMSS) administrations, which is being administered in 2011 in conjunction with NAEP to create a TIMSS/NAEP link, will be used for reviewing the standards for STAAR at grades 3–8. However, because this international assessment will not be administered until 2011, data will not be available to TEA at the time of the initial standard-setting activities. Once the data are available, time will be needed to complete a Texas comparative study linking the TIMSS/NAEP performance information to performance on the STAAR assessments. The earliest that the findings from the Texas study are likely to be available will be in the 2013–2014 school year.

TEC §39.028 requires TEA to obtain nationally comparative results for the state assessment program. This requirement was met in the past through periodic administration of the Iowa Test of Basic Skills to representative samples of Texas students. Given the number of national and, in the near future, international assessments administered in the state, which provide a variety of sources of information about the performance of Texas students, TEA will propose a plan to use the studies legislatively mandated in HB 3 to fulfill the national comparative data study requirements.

Additional Studies

Some of the planned validity studies described on the following pages are specifically mandated in legislation, while others have been added in order to support the transition to the STAAR program. The research designs for these studies have been reviewed by the Texas Technical Advisory

Committee (TTAC). This committee is comprised of national assessment and psychometric experts who provide technical guidance for the Texas assessment program. The majority of these validity studies have been planned so that they can be used to inform the standard-setting process for the STAAR grades 3–8 and STAAR EOC assessments. It should be noted that the data collected initially may be limited because the testing program is still in the developmental stages, and students taking the assessments are not necessarily motivated to demonstrate their best performance. Over time, however, data more indicative of motivated student performance can be gathered to continue to refine the level of performance required at each grade or course to be prepared for the next grade or course and, ultimately, for postsecondary success. As the data become more refined, they will be used in the review of performance standards, which will occur at least once every three years.

Studies for STAAR Grades 3–8 Standard Setting

Timeline	STAAR 3–8 Study	Purpose
Initial Studies 2011–2012	Vertical scale	This provides for a measure of student progress between grades for reading and mathematics. This also provides information about the alignment of the standards such that there is an appropriate increase in performance standards across grade levels.
	Links to STAAR EOC	The relationship between performance on STAAR grade 8 and Algebra I and English I will be determined. This information can provide information for standard setting such that a student passing grade 8 is on track to pass Algebra I and English I EOC assessments.
	Links between grades	The relationship between consecutive grades within a subject area will be determined. This information can provide information for standard setting such that a student passing a lower grade is on track to pass at the next grade.
	Comparison with TAKS 3–8	Studies will compare the STAAR 3–8 assessments with the TAKS 3–8 assessments to evaluate the rigor of performance standards for STAAR against TAKS standards.
	Comparison with NAEP	Comparisons with performance on the National Assessment of Educational Progress help to evaluate the rigor of the state performance standards at grades 4 and 8 in comparison to NAEP standards.
Additional Studies 2013–2014	Comparison with EXPLORE and Readistep	EXPLORE and Readistep are college-readiness tests typically taken by students in grade 8. Comparisons with these tests help to evaluate the rigor of the performance standards for STAAR grade 8.
	Comparison with TIMSS	Comparisons with international assessments of reading, mathematics, and science help to evaluate the rigor of the performance standards in STAAR grades 4 and 8.

Studies for STAAR EOC Standard Setting That Inform All Cuts

Timeline	STAAR EOC Study	Purpose
Initial Studies 2010–2011	Links between courses	The relationship between performance on Algebra I and Algebra II will be determined. Likewise, performance will be compared between English I, English II, and English III. This information can inform standard setting such that a student passing a lower-level course is also on track to pass a higher-level course. These studies will also provide information for the advanced course readiness indicator.
	Comparison with high school TAKS	Studies will compare certain EOC assessments with the TAKS high school assessments to evaluate the rigor of performance standards for STAAR EOC against high school TAKS. These studies will help ensure that the expectations for student performance on STAAR are high enough to achieve the goal of graduating students who are college and career ready.
	Comparison with course performance	Studies will compare performance on EOC assessments with comparisons in the corresponding course to evaluate consistency between passing the assessment and passing the course.
	Comparison with NAEP	Comparison with performance on the National Assessment of Educational Progress helps to evaluate the rigor of the state performance standards in comparison with NAEP standards.
Additional Studies 2011–2013	Comparison with AP, IB, SAT subject tests	Studies will be conducted between the EOC assessments and AP, IB, and SAT subject tests so that scores on these assessments could substitute for scores on the EOC assessments.
	Comparison with PSAT and PLAN	Studies will be conducted between the EOC assessments and PSAT and PLAN so that scores on these assessments could substitute for scores on the EOC assessments.

Studies for STAAR EOC Standard Setting That Provide Information for the College- and Career-Readiness Cuts

Timeline	College- and Career-Readiness Study	Purpose
Initial Studies 2010–2011	Comparison with SAT and ACT	SAT and ACT are used nationally and internationally and are commonly taken by students applying to four-year colleges and universities. They are used for college admissions and are predictive of success in the first year of college.
	Comparison with ACCUPLACER and THEA	ACCUPLACER and THEA are commonly taken by Texas students entering community colleges. The tests are currently used for Texas Success Initiative (TSI) exemptions and typically provide information about whether a student needs remediation.
	College students take STAAR EOC	This provides a direct measure of college student performance on the EOC assessments. Comparisons can be made between students who were successful in the entry-level course and those who were not.
Additional Studies 2011–2013	Comparison with AP	AP test scores can be used to place out of entry-level college courses. Students receiving high AP scores should also be likely to meet the college- and career-readiness standard.
	Comparison with SAT subject tests	SAT subject tests are used for college admissions, particularly for selective colleges, and to place students out of entry-level college courses. Students scoring well on the SAT subject tests should also be likely to meet the college- and career-readiness standard.
	Comparison with COMPASS	This test is currently used for TSI exemptions and typically provides information about whether a student needs remediation.
	Comparison with success in the military	This study will provide information about the relationship between the college- and career-readiness standard and success in the military.
	Comparison with workforce certifications	This study will provide information about the relationship between the college- and career-readiness standard and performance in a workforce training, certification, or other credential program.
	Science and social studies analyses	Studies will determine if college- and career-readiness standards should be set on STAAR EOC assessments in science and social studies.
Additional Studies 2014–2015	Longitudinal studies	This study follows Texas high school students into college to evaluate how well EOC performance in high school can predict performance in entry-level college courses.

.