Chapter **8** # End-of-Course Assessments

Overview

On December 16, 2005, Governor Rick Perry issued Executive Order RP53, which called for increased college readiness programs in Texas public schools and authorized "the development of a series of voluntary end-of-course assessments in Science, Mathematics, and other subjects, currently assessed by the 11th grade Texas Assessment of Knowledge and Skills, to measure student performance…." As a result of Executive Order RP53, the Texas Education Agency's (TEA's) Student Assessment Division began the development of end-of-course (EOC) assessments in geometry, biology, chemistry, physics, and U.S. history and reestablished the development of Algebra I.

In May 2007, the Texas Legislature enacted Senate Bill (SB) 1031, expanding the role of the 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 the new high school graduation requirements, beginning with the freshman class of 2011–2012. The bill required the development of six additional EOC assessments:

- Algebra II

- English I

- English II

- English III

- world geography
- world history

Additionally, SB 1031 required that EOC assessments include items to measure college readiness. Performance at the highest cut score will indicate a strong application of knowledge and skills, and 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.

In June 2009, the Texas Legislature enacted House Bill (HB) 3. Among its provisions, the new law requires the development of a college readiness measure in the Algebra II and English III EOC. In the process of implementing HB 3, TEA will set the standards for all 12 EOC assessments prior to the first mandatory administrations in 2011–2012 for graduation purposes. The standards will be externally validated by means of several studies that will ensure the rigor and alignment of the EOC assessments.

TEA's Student Assessment Division is planning the implementation of the new legislation. Table 44 reflects the proposed schedule to field-test and implement the 12 EOC assessments over the course of the next several years.

Table 44. EOC Assessments—Implementation Plan

EOC Assessment	Spring 2007	Spring 2008	Spring 2009	Spring 2010	Spring 2011	Spring 2012	Spring 2013
Algebra I	Operational	→→→	→→→	→→→	→→→	→→→	→→→
Geometry	Field Test	Operational	→→→	→→→	→→→	→→→	→→→
Biology	Field Test	Operational	→→→	→→→	→→→	→→→	→→→
Chemistry		Field Test	Operational	→→→	→→→	→→→	→→→
US History		Field Test	Operational	→→→	→→→	→→→	→→→
Physics			Field Test	Operational	→→→	→→→	→→→
World Geography			Field Test	Operational	→→→	→→→	→→→
English I				Field test	Operational	→→→	→→→
Algebra II				Field test	Operational	→→→	→→→
English II					Field test	Operational	→→→
World History					Field test	Operational	→→→
English III					Field test	Operational	→→→

EOC Participation Requirements

Currently, any student enrolled in and completing a course in the spring semester for which an EOC assessment is offered is eligible to participate in testing, regardless of grade level. In the 2009–2010 school year, mandatory field-test sampling occurred to support the development of the Algebra II and English I assessments and to support studies being conducted in Algebra I, geometry, biology, chemistry, and world geography. In addition to the sampling, participation in operational tests—those already implemented—was voluntary in 2009–2010 for each district. If a district chose to participate in testing, it had the flexibility to select participation on a district, campus, and individual student basis.

Test Development

Maintaining a student assessment system of the highest quality involves completing a set of tasks during the test development process. The procedures described in chapter 2 outline the steps used to develop a framework for each EOC assessment and explain the ongoing development. Because it is believed that an equitable and accurate measure of learning can be achieved only if development is a shared responsibility, TEA involves educators at each step of the development process.

Recent EOC development activities are summarized below:

- Using the established test development processes, items were developed and field-tested in the following courses: Algebra I, geometry, Algebra II, biology, chemistry, physics, English I, world geography, and U.S. history.

- In preparation for the move to a high-stakes assessment program in 2012, TEA invited advisory groups made up of curriculum specialists, teachers, and professors to meet and provide input and guidance about which of the content standards eligible for assessment were critical for student success and should be emphasized on the assessments. From this input, TEA developed a set of readiness and supporting standards designed to focus the assessment at each grade and course, and to provide a vertical link between the assessments from grade to grade and course to course. From this, a draft set of assessable student expectations and a draft test blueprint were developed for most of the new EOC assessments. Due to the timing of social studies Texas Essential Knowledge and Skills (TEKS) revisions, the social studies advisories were held at a later date and, therefore, draft student expectations and blueprints for social studies courses were initially developed in advance of advisory input. These documents will be revised in 2011 based on advisory input.

- For English II, English III, and world history, a set of prototype items were written to align with the draft assessed curriculum in each course. The prototype items were reviewed by TEA staff and educators at the advisory committee meetings and were used in developing a set of item specifications for the item writers and item reviewers in drafting the first submission of test items for English II, English III, and world history.

■ An additional design aspect of the new assessments is a focus on preparedness for success in subsequent grades and courses, and ultimately for post-secondary education and career. This focus is reflected in the development of items in 2009–2010 that have a higher level of cognitive complexity and that closely align with the cognitive complexity evident in the TEKS. Additional open-ended items that require students to derive an answer independently have been developed for science and mathematics courses. In social studies and science courses, items that measure process skills in context rather than in isolation have been developed. In reading, greater emphasis will be given to critical analysis rather than literal understanding. In writing, prompts have been developed to support expository, analytical, and persuasive writing.

■ Test items were developed so that they could be delivered in both paper and online test formats. A plan was created to establish the comparability of items delivered in paper mode to items delivered in online mode. The plan also sought to minimize formatting issues during item development, such as scrolling, that might create a different student experience when testing online as opposed to testing on paper. The following 2010 EOC assessments were delivered in both paper and online modes, and a comparability study was conducted to identify field-test items that performed differently between the two modes: geometry, Algebra II, biology, chemistry, and English I.

■ In January 2008, TEA, in conjunction with the Texas Higher Education Coordinating Board (THECB), adopted a set of College and Career Readiness Standards (CCRS) that identified what students should be able to demonstrate in order to be successful during their first year in college. Since that time, the CCRS have been fully incorporated into the revised TEKS for mathematics, science, English language arts, and social studies. TEA and THECB staff, high school and higher education faculty, and national experts with experience in defining college and career readiness worked together to identify the TEKS in Algebra II and English III that are critical for college and career readiness and align to the CCRS. The critical college and career readiness skills within the TEKS were validated by external committees of educators and will be used for assessment, teacher preparation, professional development, and instructional materials.

Test Administrations

Each EOC assessment measures a student's mastery of the TEKS for that specific course. Because the assessments are designed to be administered at the end of the course, the majority of the student expectations are eligible for testing. Students may be given an EOC assessment upon completing the course of study.

All 2009–2010 EOC operational tests and field tests were offered as online and paper administrations with the exception of the Algebra I, physics, and U.S. history operational tests, which were offered online only.

Further information about the online system, including an overview of the system, information on delivery and reporting, and a list of frequently asked questions, is available in the Texas Assessment Management System.

Table 45. EOC Assessments Administered Online and On Paper in 2009–2010

Test Administration	Tests Administered Online	Tests Administered on Paper
Algebra I (fall 2009 study)	9,959	N/A
Algebra II Field Test	34,642	28,200
English I Writing Field Test	30,755	25,240
English I Reading Field Test	29,721	24,499
Algebra I	101,887	N/A
Geometry	81,777	55,840
Biology	88,351	63,896
Chemistry	76,456	52,614
Physics	25,241	N/A
World Geography	62,616	26,698
U.S. History	37,349	N/A

Testing Accommodations

For students who met the eligibility criteria, an oral administration of the Algebra I, geometry, biology, chemistry, physics, world geography, and U.S. history EOC assessments was allowed during spring 2010. An oral administration was not available for the field tests in Algebra II or English I. Directions for test administrators conducting an oral administration were included in the test administrator manuals.

Accommodation Request Forms were not required for EOC assessments in the 2009–2010 school year; the use of accommodations was determined at the local level.

Scores and Reports

The various reports available for each EOC assessment are described in this section.

Description of Scores

For a detailed description of how test scores are derived, refer to chapter 2.

RAW SCORE

The raw score is the number of items answered correctly on an EOC assessment (for example, geometry). By itself the raw score has limited utility; it can be interpreted only in reference to the total number of items on a test, and raw scores should not be compared across tests or administrations.

SCALE SCORE

Because Algebra I is the only EOC assessment with established passing standards, the Algebra I Confidential Student Report (CSR) provided the student's raw score and the student's scale score, including whether the student had achieved the Met Standard or Commended Performance standard for the Algebra I EOC assessment. Refer to chapter 3 for detailed information about scaling.

Report Formats

Two types of reports are typically provided for the various testing programs: standard and optional. Standard reports are provided automatically to districts. Information contained in standard reports is sufficient to satisfy mandatory reporting requirements. Because the EOC assessment program is currently voluntary and not part of the reporting for accountability purposes, the EOC assessment program did not provide optional reports for 2009–2010.

Reports that include "Confidential" in the title contain student-level results. These reports are available for authorized users. All other reports present test results in an aggregated format and are considered public information.

Standard Reports

The standard reports available for the EOC assessment program include the CSR and Confidential List of Student Results (CLSR). These reports are available in PDF format via the online testing system within 24 hours from the close of the testing day session. In addition, Summary Reports and an EOC Data File (Confidential) are posted online for superintendent access following the close of the testing window.

Additional Reports in 2009–2010

To provide additional performance information to districts, three reports were designed for the operational EOC assessments for which performance standards have not yet been established (geometry, biology, chemistry, physics, world geography, and U.S. history). These new reports were provided at the district and campus levels and contained aggregated information about students for whom an online record was submitted. A separate report was provided for each assessment.

The reports for 2009–2010 included the Raw Score Frequency Distribution Report—All Students, Raw Score Distribution by Objective Summary Report—All Students, and Cumulative Raw Score Frequency Distribution Report—All Students.

Standard Setting

In the EOC assessment program, performance standards currently exist for the Algebra I EOC assessment only. These standards include Met Standard at a scale score of 1100 and Commended Performance at a scale score of 1400. These standards were set and approved in fall 2005. Performance standards for all 12 assessments will be set (or reset, in the case of Algebra I) prior to the first mandatory administrations for graduation purposes in the 2011–2012 school year. Planning for setting performance standards on all 12 EOC assessments continued during the 2009–2010 school year. A process and timeline for setting the performance standards was established and presented to the Texas Technical Advisory Committee (TTAC) for its feedback. According to the current plan, the performance standards will be set prior to the first high-stakes EOC administration in spring 2012. Research studies, as mandated in Texas Education Code , §39.0242, will provide information for the standard-setting process. Data collection for these studies also took place during the 2009–2010 school year. This was done through the mandatory testing of sampled campuses for the following EOC assessments: Algebra I, geometry, Algebra II, English I, biology, chemistry, and world geography.

Refer to chapter 3 for detailed information about standard setting.

Scaling

As with many other tests in the Texas assessment program, the EOC assessment program uses the Rasch Partial-Credit Model (RPCM) to place test items on the same scale across administrations for a given EOC assessment. Once performance standards have been set for an assessment, its initial 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.

Scale Score

Refer to chapter 3 for detailed information about scale scores. TEA established the performance standards for the current Algebra I EOC assessment in November 2005. Using the RPCM scaling procedures described in chapter 3, a unique scale transformation was developed for Algebra I so that the resulting set of scale scores would have the panel-recommended Commended Performance cut set at a scale score of 1400 and the Met Standard cut set at a scale score of 1100. The linear transformation of the underlying Rasch proficiency level estimate is as follows:

$$SS_j = (\theta_j \times T1) + T2$$

where SS_j is the scale score for student j, θ_j is the Rasch partial credit model proficiency level estimate for student j, and $T1$ and $T2$ are scale score transformation constants that establish the scale score system such that a scale score of 1100 is the cut score for the Met Standard performance level, and a scale score of 1400 is the cut score for the Commended Performance level. Values for $T1$ and $T2$ are provided in Table 46 for EOC assessments.

Table 46. Scaling Constant for EOC Assessments

EOC Assessment	T1	T2
Algebra I	155.0468	1009.0186

Raw Score

In 2010 the EOC assessment program also reported results for several assessments using the raw score. This score is the number of items that a student answers correctly. Because no performance standards have been set for EOC assessments in geometry, biology, chemistry, physics, world geography, and U.S. history, only raw scores for these tests were reported.

Equating

During the 2009–2010 school year, two types of equating were conducted in the EOC assessment program: pre-equating and field-test equating. In addition, comparability analyses were conducted for the five EOC assessments that were administered to students both online and on paper in spring 2010: geometry, Algebra II, biology, chemistry, and English I. Algebra II and English I were first-time stand-alone field tests in spring 2010. Refer to chapter 3 for detailed information about equating and comparability analyses.

Pre-Equating

In 2009–2010 pre-equating was conducted for Algebra I, the only EOC assessment that currently has scale scores and performance standards. For Algebra I, the 2010 test form was constructed from a bank of items that had been equated to the base-test scale that was established in 2005. Through this pre-equating process, a raw score to scale score conversion table for Algebra I was generated prior to the operational administration of the test. The pre-equating design allowed for student test results on Algebra I to be reported to districts and campuses as soon as possible after the test was administered. Refer to chapter 3 for a detailed description of the pre-equating process.

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 scale of the original form. During 2009–2010, field-test equating was conducted for nine EOC assessments: Algebra I, geometry, Algebra II, biology, chemistry, physics, English I, world geography, and U.S. history. The field-test equating process for each test depended on the model in which field-test items were placed on the test form—either through embedded field testing or stand-alone field testing.

Because seven of the EOC assessments (Algebra I, geometry, biology, chemistry, physics, world geography, and U.S. history) were operational assessments with embedded field-test items, live test items common to each form of the test were used to place the embedded field-test items onto the baseline scale of each EOC assessment. Chapter 3 provides more details about the procedures for equating embedded field-test items.

For Algebra II and English I, where no operational test forms existed, newly constructed items were placed in stand-alone field-test forms. For these two assessments, a set of linking items common across all field-test forms was used to equate the field-test items to each other. The field-test form taken by the most students (that is, with the largest n-count) became the baseline scale and the items from the other field-test forms were moved onto that scale, using the linking items. The specific procedure for equating stand-alone field-test items is described in further details in chapter 3.

Comparability Analyses

In spring 2010, five EOC assessments were administered in both online and paper modes. Of these assessments, Algebra II and English I were stand-alone field tests, whereas biology, chemistry, and geometry were operational tests. In order to evaluate whether the mode of presentation impacted item difficulty, comparability analyses were conducted at the item level for all of the items on these assessments. For the English I EOC assessment, the comparability analysis was conducted separately for the reading and writing components of the assessment.

Three types of item-level analyses were conducted across testing modes: (a) comparison of item p-values, (b) comparison of Rasch item difficulties, and (c) Mantel-Haenszel differential item functioning analysis. The results of these comparability studies indicated that, for all of the assessments except English I, very few items showed differences between the online and paper versions of the tests. The percentage of items that demonstrated a mode effect for biology, chemistry, geometry, and Algebra II were: 0.8%, 3.3%, 5.8%, and 5.3%, respectively. For English I, more reading items were identified as showing a mode effect than writing items. The percentages of items that demonstrated a mode effect for English I reading was 33.7% as compared to 21.3% for English I writing.

The specific procedures used to evaluate comparability as well as the detailed results of the analyses are available in the "2010 End-of-Course Comparability Study Report" in the TEA technical report series. The results from the comparability study were used in field-test equating and will be used in future test construction. Refer to chapter 3 for detailed information about comparability analyses.

Reliability

During the 2009–2010 school year, reliability for the seven operational EOC assessments was estimated through several reliability indices, including internal consistency, classical standard error of measurement, conditional standard error of measurement, and classification accuracy. Refer to chapter 3 for detailed information about reliability.

Internal Consistency

The Kuder-Richardson Formula 20 (KR20) was used to calculate the reliability estimates for all EOC assessments. As a general rule, reliability coefficients from 0.70 to 0.79 are considered adequate, 0.80 to 0.89 are considered good, and above 0.90 are considered excellent. However, appropriate levels of reliability depend on how an assessment is being used. For the EOC assessment program,

the internal consistency estimates range from 0.88 to 0.94. The internal consistency estimates for the EOC assessments are available in Appendix F. Refer to chapter 3 for detailed information about internal consistency.

Classical Standard Error of Measurement

For the EOC assessments, Standard Error of Measurement (SEM) values range from 2.96 to 3.70. The SEM values for the EOC assessments are provided in Appendix F. Refer to chapter 3 for detailed information about standard error of measurement.

Conditional Standard Error of Measurement

During the 2009–2010 school year, only the Algebra I EOC assessment had Conditional Standard Error of Measurement (CSEM) estimates. The Algebra I EOC assessment has CSEM values because it is the only EOC assessment that is reported using scale scores. CSEM is typically only computed for assessments that have a scale score reporting system. The CSEM estimate for Algebra I at the scale score of 1100 (Met Standard cut) is 47, and the CSEM estimate for Algebra I at the scale score of 1400 (Commended Performance cut) is 74. Appendix F contains the Algebra I CSEM estimates.

Classification Accuracy

As with CSEM, classification accuracy was only calculated for Algebra I because it is the only EOC assessment with performance standards. In 2009–2010 the Algebra I assessment had an 87.6% agreement between expected and observed classifications for students who were in the two levels of performance (Met Standard and Commended Performance). Appendix F contains the classification accuracy table for Algebra I. Refer to chapter 3 for detailed information about classification accuracy.

Validity

The sections that follow describe how validity evidence was collected for the EOC assessments in 2009–2010. Refer to chapter 3 for detailed information about validity.

Evidence Based on Test Content

Evidence based on test content is information that shows the relationship between content of the test and the test constructs that are intended to be measured by the test. The EOC assessments have been developed to align with the content defined by the TEKS. Content validity evidence has been collected at all stages of the test development process.

Established test development processes for the Texas assessment program were followed in developing the EOC assessments to support the use of EOC scores in making inferences about students' knowledge and understanding of the TEKS.

The following activities took place during the 2009–2010 school year to support the content validity of EOC assessments.

RELATIONSHIP TO THE STATEWIDE CURRICULUM

As part of the transition to a high-stakes graduation program in 2012, teachers, curriculum specialists, test development specialists, college educators, and TEA staff members worked together in advisory committees to identify appropriate assessment objectives for new EOC assessments in Algebra I, geometry, Algebra II, biology, chemistry, physics, English I, English II, and English III. The input of the advisory committees is reflected in draft assessed curriculum documents and item specifications.

Prototype items were developed for the new assessments in English II, English III, and world history. As part of the item development process, these prototypes were reviewed by advisory committees and TEA staff to identify how well they measured the student expectations to which they were aligned and to provide information for item-development guidelines and test-item types.

EDUCATOR INPUT

Following item development for each EOC assessment, committees of Texas educators met to review test items and confirm that each item appropriately measured the TEKS to which it is aligned. The committees also reviewed and edited the items for content and bias. These reviews occurred for Algebra I, geometry, Algebra II, biology, chemistry, physics, English I, world geography, and U.S. history. There were two rounds of educator input per course: item review, in which the items were revised and edited, as appropriate, before field testing occurred; and data review, in which the educators reviewed the field-test performance data of each item and made a judgment about whether the item appropriately measured the construct and was eligible for placement on an operational, or live, test.

As EOC assessments transition from the low-stakes environment in which they were developed and administered beginning in 2005 to the high-stakes graduation measure under which they are now developed and will be administered in 2012, evidence of curricular and instructional validity is being gathered. The evidence will determine 1) if there is a match between the tested content drawn from the state curriculum standards and classroom curricular materials, and 2) if there is a match between the tested content and what is being taught in the classrooms. This evidence has been collected as part of the item-by-item judgments made by educators during item review.

TEST DEVELOPER INPUT

Item writers and reviewers followed test development guidelines that informed how the content of the assessed TEKS should be measured. At each stage of development, writers and reviewers verified the alignment of the test items with the assessed objectives.

Test Expert Input

TEA, in conjunction with Pearson, receives ongoing input from a panel of national testing experts related to plans for collecting validity evidence for Texas testing programs, including EOC assessments.

In February 2009, the TTAC provided input on validity studies designed to examine the relationship between students' end-of-course scores with performance on college placement tests such as the ACT and SAT, and longitudinal studies that examine the relationship between the EOC college-readiness performance standard and performance in college courses in the same content area. In February 2010, the TTAC provided input on conducting studies that link performance on EOC assessments with commonly taken college placement tests, such as ACCUPLACER, COMPASS and THEA.

Evidence Based on the Response Process

Response processes refer to the cognitive behaviors required to respond to a test item. For the EOC assessments, Texas collects validity evidence based on response process through the different item types on each test and the mode in which tests were administered.

Multiple-choice was the primary item type to which students were asked to respond on all EOC assessments. This type of item was used because it most closely resembles what students typically experience in classroom testing. The multiple-choice items were developed so that students were required to recall and apply what they had learned about the course, thereby supporting an accurate measurement of the construct being assessed. For the Algebra I, geometry, Algebra II, chemistry, and physics assessments, griddable items, which required the students to determine a numerical answer and then grid in (or type in, for online tests) their answers, were also administered. These item types facilitated the assessment of the students' knowledge and skills at an even deeper level by requiring students to generate answers independently without being influenced by answer choices provided with questions. These skills were essential in assessing the construct in the mathematics and science content areas. In addition, cluster items were used for the biology, English I, world geography, and U.S. history EOC assessments. Cluster items are a group of multiple-choice items associated with a common stimulus, such as a reading passage, a graph, or a map. The use of cluster items typically required students to apply their knowledge and skills within the context introduced by the stimulus in order to respond correctly to the items. This skill was also important in assessing the construct in content areas such as science, social studies, and English language arts.

Evidence Based on Internal Structure

Texas collects evidence that shows the relationship among test questions and test objectives to demonstrate that the parts of a test conform to the test construct. When tests are designed to measure a single construct, the internal components of the test should exhibit a high level of homogeneity, which can be evaluated in terms of the

internal consistency estimates of reliability. Refer to the "Reliability" section for descriptions and estimates of internal consistency in the EOC assessments during 2009–2010.

Evidence Based on Relationship to Other Variables

Another source of validity evidence is the relationship between test performance and performance on some other measure, sometimes called criterion-related validity. Planning for validity studies that will correlate student performance between EOC assessments in the mathematics and English content areas occurred during the 2009–2010 school year. In addition, future studies that evaluate the relationship between the EOC assessments and other external measures such as the high school TAKS assessments, course performance, and AP, IB, SAT, PSAT, and PLAN assessments are being planned. Results for these validity studies can be used to provide criterion-related validity evidence for the EOC assessments and may also be used in the standard-setting process to support the validity of the performance standards.

Evidence Based on the Consequences of Testing

Another way to provide validity evidence is by documenting the intended and unintended consequences of administering an assessment. Validity evidence indicating the impact of EOC testing on students was collected through a question on the online survey at the end of each operational EOC assessment (Algebra I, geometry, biology, chemistry, and U.S. history) during the 2008–2009 school year. The survey was voluntary for students and asked them the question "For what percent of your course grade does this test count?" The survey results showed that, across the five EOC assessments, most students (71–75%) who responded to the question did not know whether their EOC test scores would count toward their final grade in the course, but some students (about 12–21% across the five tests) responded that the EOC test scores did not affect their course grades. This provided evidence that the results from the EOC assessments had academic consequences for some students in 2008–2009. TEA plans to collect additional validity evidence related to the impact of EOC assessments on student performance.

Student Growth and Projection Measures

Student growth and projection measures track a student's performance across time. Improvement measures track student performance from year to year, whereas projection measures use current student performance to predict future performance. No student growth or projection measures are currently used with the EOC assessment program. During the 2009–2010 school year, TEA and Pearson began initial planning for the development of student growth and projection measures in the EOC assessments to meet federal and state accountability requirements.

Sampling

During the 2009–2010 school year, sampling was used to select campuses to participate in an assessment's first-time stand-alone field test or to collect data for research studies that will provide information for the EOC standard-setting process in February 2012. Stand-alone field tests were conducted for the two new EOC assessments: Algebra II and English I. Data for standard-setting research studies were collected for five operational EOC assessments: Algebra I, geometry, biology, chemistry, and world geography. The sampling for geometry, Algebra II, biology, chemistry, and English I, was conducted in both testing modes (i.e., on paper and online), while the sampling for Algebra I and world geography was conducted in the online mode only. Campuses were specifically assigned to be part of each of the EOC samples. Campuses selected for one of the dual-mode EOC assessments were specifically assigned to test in one of the two modes: paper or online. Participation in the EOC assessments by all selected campuses was mandatory. Campuses not selected to participate in one of the EOC assessments could volunteer to test in one of the available modes.

EOC Sampling Process

A stratified sampling design was used for the EOC assessment program in which the campus was the sampling unit, but the student was the observation unit. Each campus was classified into one of five strata based on its campus size, or estimated student count for each EOC assessment. Because the campus was the sampling unit, it was necessary to obtain the student course enrollment from each campus as an estimate of the number of students that would participate in the corresponding EOC assessment. The estimated student counts for each campus were based on the number of students who were enrolled in each of the sampled courses in fall 2008.

TEA initiated a sampling model in spring 2006 that provided a "relief year" to campuses so that each campus would have a minimum of one in every five years during which it would not be asked to participate in TAKS stand-alone field testing. This "relief year" process implemented originally for TAKS was extended to include the EOC field tests so that campuses that had participated in either the TAKS or EOC stand-alone field testing for the previous four years were exempted from selection in the 2010 EOC samples. Stand-alone field-test samples for TAKS were being selected at the same time as the EOC samples. Sampling for TAKS and EOC was coordinated across programs in order to reduce campus testing burdens.

In addition to the "relief year" process, the following factors were considered in determining each EOC sample for 2010:

- The sample was chosen to be representative of the overall population of Texas high school students taking the course in terms of ethnic composition and campus size.

- The sample was selected to include a minimum of 280 students per form from each major ethnic and gender subgroup (i.e., African American, Hispanic, white, male, and female groups).

  o Campuses were not assigned to more than four total EOC and TAKS assessments.

  o For any online EOC assessments, each sampled campus was required to test all of its enrolled students in the course, regardless of grade, up to a maximum of 300 students during the window.

  o For any paper EOC assessments, each selected campus was required to test all enrolled students in the course, regardless of grade, during the window.

  o Campuses were not assigned to more than two online EOC assessments.

  o Campuses selected for testing as part of the National Assessment of Educational Progress (NAEP) in 2010 were considered as participating in one test and were therefore not assigned to more than three total EOC and TAKS assessments.

  o Campuses selected for the Algebra I sample in fall 2009 were also expected to test the same students in geometry during spring 2010. This was done to empirically link the performance on these two mathematics EOC assessments for the purpose of informing the EOC standard-setting process.

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

  o Campuses with fewer than 15 students enrolled in the course for the EOC assessment.

  o Campuses who had been academically unacceptable (i.e., failed to meet the state accountability standards) for 3 or more years.

  o Campuses who had failed in meeting Adequate Yearly Progress (AYP) for 5 or more years.

  o Campuses that are part of the Juvenile Justice Alternative Education Program (JJAEP), Disciplinary Alternative Education Programs (DAEP), or Texas Youth Commission (TYC).

The process for selecting campuses to be included in each EOC sample was as follows:

1. All eligible campuses were divided into five even-sized strata based on campus size (i.e., strata 1, 2, 3, 4, and 5). If a number of campuses of equal size appeared around the threshold between strata, the placement in the upper or lower stratum was done randomly.

2. Campuses were sorted randomly within each stratum.

3. One campus was randomly selected from each stratum. Each campus was chosen into the sample in ascending and descending order of strata (e.g., 5-4-3-2-1-1-2-3-4-5-5-4-3-2-1-...). For example, the first campus was chosen from stratum 5, the second campus from stratum 4, the third campus from stratum 3, the fourth campus from stratum 2, the fifth campus from stratum 1, etc. Thus, one campus was selected in one stratum first before moving to the next stratum.

4. The number of students in the sample was evaluated relative to the target total number of students after the campus had been selected. Step 3 was repeated until the target number of students was reached.

5. A "fit index" was calculated for the resulting sample of campuses. This index indicated how well the selected campuses reflected the demographic breakdown of the students enrolled in the course statewide.

6. Steps 1 to 5 (from dividing campuses into five strata to calculating the fit index) were repeated up to 1,000 times. Any sample for which the fit index indicated that the sample was within a reasonable percentage of the target demographic breakdown was reviewed by a psychometrician, who selected a final sample using professional judgment.

7. Once the final sample was determined, it was regenerated using the appropriate random number seed so that additional detailed output descriptive statistics for this sample could be generated.

The final sample was determined after evaluating four key elements: fit to statewide ethnic percentages, number of campuses, number of students, and distribution of campus size strata within the sample. A summary of the number of campuses and students selected for the 2009–2010 EOC samples is provided in Table 47.

Table 47. Sampling Summary for 2010 EOC Assessments

EOC Assessment	Number of Sampled Campuses	Number of Expected Students at the Time of Sampling
Algebra I Online (Fall 2009)	197	40,365
Algebra II Online	294	40,728
Algebra II Paper	213	67,951
Geometry Online	455	67,175
Geometry Paper	266	24,152
Biology Online	450	67,771
Biology Paper	279	54,931
Chemistry Online	395	54,960
Chemistry Paper	244	32,533
English I Writing Online	241	32,533
English I Writing Paper	143	32,214
English I Reading Online	241	32,214
English I Reading Paper	143	67,013
World Geography Online	169	40,365