Chapter 6 TAKS–Alternate (TAKS–Alt)

Overview

The Texas Assessment of Knowledge and Skills–Alternate (TAKS–Alt) is an alternate assessment based on alternate academic achievement standards and is designed for students with significant cognitive disabilities receiving special education services who meet the participation requirements for TAKS–Alt. TAKS–Alt has been designed to meet federal requirements mandated under the No Child Left Behind Act of 2001 (NCLB). This assessment is not a traditional paper or multiple-choice test. Instead, it involves teachers observing students as they complete standardized state-developed assessment tasks that link to the grade-level Texas Essential Knowledge and Skills (TEKS). Teachers then evaluate student performance based on the dimensions of the TAKS–Alt rubric and submit results through an online instrument.

The assessment requirements for TAKS–Alt are illustrated in Table 19. TAKS–Alt was administered during the window of January 4 through April 16 for all subjects and grade levels.

Table 19. 2009–2010 TAKS–Alt Assessments

2009–2010 TAKS–Alt Assessments	
Grade	**Test Administration**
Grade 3	Mathematics and Reading
Grade 4	Mathematics, Reading, and Writing
Grade 5	Mathematics, Reading, and Science
Grade 6	Mathematics and Reading
Grade 7	Mathematics, Reading, and Writing
Grade 8	Mathematics, Reading, Science, and Social Studies
Grade 9	Mathematics and Reading
Grade 10	English Language Arts, Mathematics, Science, and Social Studies
Grade 11	English Language Arts, Mathematics, Science, and Social Studies

TAKS–Alt Participation Requirements

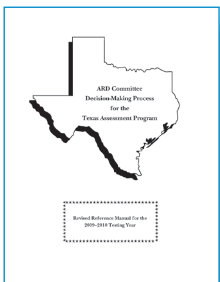
TAKS–Alt has specific participation requirements that the admission, review, and dismissal (ARD) committees must carefully consider when recommending these assessments for students receiving special education services. All students have the right to be exposed to as much of the TEKS curriculum as possible so that they can reach their academic potential. The participation requirements for TAKS–Alt describe the type of grade-level instruction of the TEKS (accessed through prerequisite skills) that a student should be receiving to participate in TAKS–Alt. The members of the ARD committee must weigh the benefits of rigorous and challenging expectations with the possibilities of success, given each student's individual strengths, needs, instruction, and accommodations. Keeping these high standards in mind, the ARD committee must choose the assessment that best matches the educational needs of each individual student. ARD committees should promote high expectations in determining the annual measurable goals documented in each student's individualized education plan (IEP). It is important to emphasize that the academic instructional decisions made by the ARD committee and documented in the IEP must always guide assessment decisions.

Students receiving special education services who have the most significant cognitive disabilities and who are unable to participate in other statewide assessments even with substantial accommodations and/or modifications can be considered for TAKS–Alt. After a cognitive disability has been established for the student, the ARD committee may decide that a student's knowledge and skills can best be assessed with TAKS–Alt if the student meets all of the following participation criteria.

The student

- requires supports to access the general curriculum that may include assistance involving communication, response style, physical access, or daily living skills;
- requires direct, intensive, individualized instruction in a variety of settings to accomplish the acquisition, maintenance, and generalization of skills;
- accesses and participates in the grade-level TEKS through activities that focus on prerequisite skills;
- demonstrates knowledge and skills routinely in class by methods other than paper-and-pencil tasks; and
- demonstrates performance objectives that may include real-life applications of the grade-level TEKS as appropriate to the student's abilities and needs.

ARD Manual



The *ARD Committee Decision-Making Process for the Texas Assessment Program: Revised Reference Manual for the 2009–2010 Testing Year* explains to ARD committees how to make assessment decisions by using the state-mandated curriculum—the TEKS—as documented in each student's IEP. Module 1 of the online TAKS–Alt training gives more detail on selecting the appropriate assessment for students with cognitive disabilities.

Test Development

The test development process for TAKS–Alt follows as closely as possible the same procedures used for other statewide assessments in Texas but with additional requirements that are specific to TAKS–Alt. During assessment task development, careful attention is given to the following three criteria:

1. Standard 3.6 in the Standards for Educational and Psychological Testing (AERA, APA, NCME 1999). The development process followed this guidance.

   > The type of items, the response formats, scoring procedures, and test administration procedures should be selected based on the purposes of the test, the domain to be measured, and the intended test takers. To the extent possible, test content should be chosen to ensure that intended inferences from test scores are equally valid for members of different groups of test takers. The test review process should include empirical analyses and, when appropriate, the use of expert judges to review items and response formats. The qualifications, relevant experiences, and demographic characteristics of expert judges should also be documented (p. 44).

2. The National Alternate Assessment Center (2005) identified three questions that must be asked when determining whether or not instruction is linked to grade-level curriculum expectations. The questions listed below were guiding principles during the development of TAKS–Alt, as well as subsequent internal Texas Education Agency (TEA)

and educator item review meetings, when determining whether or not the standardized assessment tasks for TAKS–Alt were linked to the TEKS grade-level curriculum.

- Does the assessment task cover academic content? Does the task reflect the grade-level curriculum?
- Does the assessment task access the grade-level TAKS objective and knowledge and skills statements?
- Is the assessment task meaningful to the student? Will the skill be useful to the student in the immediate future?

3. As standardized assessment tasks were developed, attention was also given to the criteria of fairness, principles of alignment, and universal design. When developing TAKS–Alt, these principles were considered from the beginning to bridge the gap between the grade-level content and the learning styles of students with significant cognitive disabilities. In incorporating universal design for TAKS–Alt, consideration was given to 1) students' response modes, allowing students to show what they know and can do; 2) differentiated supports and materials, allowing students to access the content of the assessment; and 3) multiple means of engagement to allow students more time to complete the task, meaningful activities, and context (CAST, 2002). Within the principles of universal design, each item has precisely defined constructs, has maximum legibility, has maximum readability and comprehensibility, is amenable to accommodations, is accessible and non-biased, and considers special populations.

All of the criteria listed above were specifically considered in the development of assessment tasks for each essence statement. An essence statement is a statement that summarizes the TEKS knowledge and skills statements and student expectations for each TAKS–tested objective. Standardized assessment tasks were developed at three levels of complexity (described below). Each assessment task includes three standardized predetermined criteria that specify what the student is expected to do to demonstrate the skill. The assessment tasks are based on the TEKS Vertical Alignment for TAKS–Alt, and TEKS Curriculum Framework for TAKS–Alt, which assesses prerequisite skills linked to the grade-level TEKS. The assessment tasks were developed by content specialists and special education assessment specialists. Once the assessment tasks were reviewed and approved by TEA, educator review meetings were convened. For more information about the TAKS–Alt vertical alignment and curriculum framework, refer to the TAKS–Alt resources page on TEA's Student Assessment website.

Complexity Level of Assessment Tasks

Three assessment tasks of varying complexity levels were developed for each essence statement. To establish the verbs that define the complexity levels for the assessment tasks, Bloom's work on learning taxonomies (Bloom, Englehart, Furst, Hill, & Krathwohl, 1956) was reviewed. Webb's depth of knowledge (Webb, 1997), Cook's extended depth of knowledge (Cook, 2008), and Browder and Flowers' depth of knowledge scales (Flowers, Browder, Wakeman, & Karvonen, 2007) were considered when developing the three complexity levels of the assessment tasks. Using a combination of cognitive scales, the verbs were selected that define each complexity level and show how the student demonstrates knowledge. Each verb is specifically defined, and suggestions for possible ways the student can respond are provided to further standardize task implementation. The complexity levels are described below.

Level 1: Beginning Awareness

Level 1 assessment tasks are the least complex and involve responding with knowledge at the beginning awareness level. Some of the skills students at this level are expected to demonstrate may include acknowledging features, indicating preferences, responding to stimuli, participating in processes, exploring materials, or anticipating outcomes.

Level 2: Basic Recall

Level 2 assessment tasks are moderately complex and involve recalling or reciting information at a basic level. Some of the skills students at this level are expected to demonstrate may include identifying or sorting elements, assisting in procedures, choosing options, examining features, or matching or replicating components.

Level 3: Application

Level 3 assessment tasks are the most complex and involve applying knowledge beyond basic recall. Some of the skills students at this level are expected to demonstrate may include determining distinguishing features, organizing information, comparing components, generating ideas, making inferences, or justifying answers.

Educator Review Committee Meetings

Once TEA approved the assessment tasks, committees composed of Texas educators reviewed the TAKS–Alt assessment tasks to ensure the alignment with grade-level standards using prerequisite skills linked to the grade-level, to eliminate potential bias, and to judge the appropriateness of task content and complexity level. The committees included special education classroom teachers who had experience teaching students with significant cognitive disabilities, and general education teachers who were knowledgeable of the TEKS curriculum. The educator review committees were convened in March 2009.

Evaluating Bias in the Assessment

An important concern in the development and review of the standardized assessment tasks was the elimination of bias toward any particular group of students with significant cognitive disabilities. All assessment tasks are expected to be fair and free from bias. The assessment tasks were reviewed for ways in which bias might appear and unfairly inhibit the performance of any student. An Assessment Task Judgment Form was completed by every member of the educator review committee during the meeting for each grade and subject area convened for the TAKS–Alt educator review meetings. Judgments were collected for every assessment task related to its potential bias in response to the question, "Are these assessment tasks free from bias on the basis of personal characteristics such as gender or ethnicity?" The range of agreement shown in the compilation of the committees' judgments by grade and subject for the match of the assessment tasks to the appropriateness of the tasks for the assigned grade level, link to the grade-level TEKS curriculum, and elimination of bias is 89.81% to 100.00%.

Training

Because the TAKS–Alt assessment features specifically developed materials, unique administration requirements, and an online reporting system, teacher training on TAKS–Alt is extremely important. For test results to be comparable across students, classrooms, campuses, and school districts, TEA developed four web-based training modules that standardize teacher training across the state and assist teachers in effectively implementing TAKS–Alt.

To further standardize the statewide training, TEA offers additional modes of training via the Texas Education Telecommunication Network (TETN), on-site training as requested by school districts and regional Education Service Centers (ESCs), and PowerPoint presentations on TEA's website that can be downloaded and used for individual or group training sessions.

TAKS–Alt Online Training and Qualification

TAKS–Alt online training is mandatory for all test administrators. This training is provided through online modules and qualification activities which are accessed through the Texas Training Center. Teachers administering the TAKS–Alt assessment for the first time are required to complete Modules 1–3 and pass the qualification activity with a score of 80% or above for each of these three modules prior to conducting any assessment observations. Upon successfully passing the qualification activity for each of Modules 1–3, the test administrator can print a qualification certificate. Teachers who qualified previously on Modules 1–3 are only required to complete Module 4. It is also recommended that new teachers and those who have not qualified previously complete all four modules. There is no qualification activity for Module 4, but test administrators completing the module receive a certificate of completion.

On the training center website, users have secure access to the online training modules, and from this site they can access training certificates, track their training status, and access online resources. In addition, for administrative monitoring, the training center tracks and reports the completion and qualification status for each user on the three TAKS–Alt training modules. The electronic monitoring of each test administrator's status on the training modules and performance on the qualification activities allows TEA and districts to ensure that all test administrators are trained and adequately prepared to assess students with TAKS–Alt.

The training modules that can be accessed from the Texas Training Center are described below.

Module 1: Overview of the TAKS–Alt Assessment

Module 1 explains the features of TAKS–Alt, complexity levels, steps for administering TAKS–Alt, role of the ARD committee, and the TAKS–Alt participation requirements.

Module 2: Implementing the TAKS–Alt Assessment

Module 2 explains how to select and provide access to an assessment task through the addition of presentation supports and response modes, and how to conduct and record observations and document when cues and prompts are given.

Module 3: The TAKS–Alt Online Instrument

Module 3 provides instructions for evaluating student performance and maintaining documentation, as well as information about the scoring rubric, generalization process, automated scoring feature, and submission process.

Module 4: Beyond the Basics

Module 4 clarifies and refines how to choose the appropriate complexity level, provide access to the tasks, plan supports, document the observation, and evaluate student performance. A template to help organize and plan an assessment task for a specific student is presented and discussed.

Test Administrations

The TAKS–Alt assessment process is designed to mirror the instructional process for a student with a significant cognitive disability. The assessment was provided during an assessment window that was open from January 4 to April 16, 2010. That window provided the time teachers needed for selecting appropriate assessment tasks, determining appropriate implementation of the tasks, evaluating and documenting student performance, and entering results in the online assessment system. Teachers were able to submit completed student assessment results at any time throughout the TAKS–Alt assessment window.

District and campus testing coordinators were given access to the TAKS–Alt online assessment system beginning on November 16, 2009. Approximately 77,400 assessments were administered to approximately 26,700 students who met participation requirements in spring 2010.

Testing Accommodations

Students being assessed with TAKS–Alt may use whatever accommodations and supports are routinely and successfully used as instructional accommodations. Accommodations may include cues and prompts. The difference between a cue and a prompt is related to the degree to which the student is assisted. A cue is a hint and does not lead the student to a direct answer. A prompt is much more invasive as it takes the student step-by-step through the task leading to a direct answer.

Because of the design of TAKS–Alt, linguistic accommodations are not necessary for limited English proficient (LEP)-exempt immigrants receiving special education services. The TAKS–Alt assessment can be administered using any language or other communication method routinely used by the student.

A chart suggesting accommodations, supports, and materials that can be used for TAKS–Alt is available on the TAKS–Alt Resources page of TEA's Student Assessment Division website and in the *Accommodations Manual*.

Student Success Initiative

Students receiving special education services who take TAKS–Alt are not subject to Student Success initiative (SSI) requirements because multiple testing opportunities are included in the TAKS–Alt process.

Scores and Reports

Scoring TAKS–Alt Assessments

In order to incorporate varied skill-level performance into the scoring of TAKS–Alt, Complexity Level was added as an additional component to the scoring rubric. The scoring rubric has three dimensions: 1) Demonstration of Skill; 2) Level of Support; and 3) Generalization of Skill.

Once a score has been determined for these three dimensions, Complexity Level is incorporated into the scoring by weighting the Demonstration of Skill dimension depending on the level of task the student completes. Through weighting, students successfully completing more complex tasks receive higher scores than students successfully completing less complex tasks.

Three standardized assessment tasks have been developed for each essence statement. Each task varies in terms of the skill-level performance required to complete the task. Teachers will determine the task that is most appropriate for their student. A Level 3 task is most complex because it involves applying knowledge; a Level 2 task is moderately complex because it involves recalling information; and a Level 1 task is least complex because it involves responding with knowledge at the beginning awareness level. The highest possible score obtainable for a student is directly related to the Complexity Level of the tasks he or she completes.

Scoring Steps for TAKS–Alt

The steps necessary to complete the scoring of TAKS–Alt are presented below. The rubric can be used by teachers as a tool to help them understand the scoring process.

1) Test administrator enters primary information into system.
   - Determine which assessment tasks (select the appropriate Complexity Level [CL] or Levels) to administer.
   - Respond to the evaluation questions for student performance for each assessment task.
   - The TAKS–Alt system assigns student scores based on the responses to the evaluation questions.

2) Test administrator scores Demonstration of Skill (Initial DS).

   Initial Score
   - Based on the response to "Did the student demonstrate the skill?" with the possible outcomes being Yes/No.
   - A student will receive 2 points for each predetermined criteria he or she completes (for each response of Yes, the student receives 2 points for a maximum of 6 points per assessment task).

   Weighted Score
   - The initial score for DS is weighted by the assessment task CL.
   - Initial DS x CL weighting = Demonstration of Skill (DS)
   - The weighting for each complexity level are:
     - CL 3 is weighted by 1.5
     - CL 2 is weighted by 1.2
     - CL 1 is weighted by 1.0
   - A student will receive a maximum of 9 points per assessment task.

3) Test administrator scores Level of Support (LS)
   - Based on the response to "How did the student perform the task?" with the possible outcomes being independently, needed cueing, needed prompting, or N/A if the response to Initial DS was No.

- A student will receive a maximum of 6 points per essence statement:
    - 2 points for each predetermined criterion completed independently,
    - 1 point for each predetermined criterion completed with cueing,
    - 0 points for each predetermined criterion completed with prompting, or
    - 0 points for N/A.

4) Test administrator determines if student is eligible for Generalization of Skill (GS).
    - Generalization questions will only appear if the student is eligible based on his or her score on DS and LS.
    - A student is eligible for Generalization of Skill if
        - a Complexity Level 2 or 3 assessment task was completed,
        - the skill was demonstrated for all three predetermined criteria, and
        - there was no prompting.

5) Test administrator enters generalization information into system.
    - Respond to the evaluation questions for student performance regarding generalization for each assessment task.

6) The TAKS–Alt System scores Generalization of Skill (GS).
    - Based on the response to "Did the student generalize the predetermined criteria in a different context?" with the possible outcomes being Yes/No.
    - A student will receive 1 point for each predetermined criterion he or she completes without prompting (for each response of Yes, the student receives 1 point if there was no prompting for a maximum of 3 points for generalization).

7) Essence Score
    - Each essence score will be calculated by adding together the following:

	Demonstration of Skill (0–9 points)
	Level of Support (0–6 points)
+	Generalization of Skill (0–3 points)
	Essence Score (0–18 points)

8) Total Score

- The total score will be calculated by adding together each essence score.
- The total score will be rounded to the nearest whole number.

	Essence Score A (0–18 points)
	Essence Score B (0–18 points)
	Essence Score C (0–18 points)
+	Essence Score D (0–18 points)
	Total Score (0–72 points)

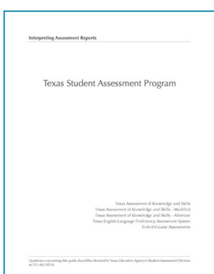
Report Formats

Two types of reports are provided for TAKS–Alt, standard and optional reports. Standard reports are provided automatically to districts. Information contained in standard reports is sufficient to satisfy mandatory reporting requirements. To receive optional reports, a district must have completed the Optional Reports Order Form and returned it with the scorable materials. Generally districts are required to pay a nominal fee for each optional report requested.

Standard and Optional Reports

The standard reports available for the 2009–2010 TAKS–Alt program include the Confidential Student Report (CSR), Confidential Student Label, Confidential List of Students' Results, and Summary Report.

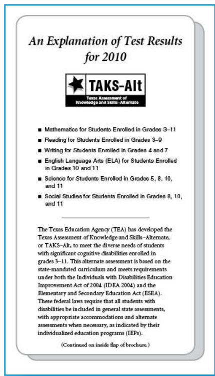
The Confidential Electronic Individual Student Record File was available as an optional report.


Interpreting Assessment Reports

Texas Student Assessment Program

More information about scoring and reporting for TAKS–Alt is available in the *2009–2010 Interpreting Assessment Reports*.

Parent Brochure


An Explanation of Test Results for 2010

TAKS–Alt
Texas Assessment of Knowledge and Skills–Alternate

- Mathematics for Students Enrolled in Grades 3–11
- Reading for Students Enrolled in Grades 3–9
- Writing for Students Enrolled in Grades 4 and 7
- English Language Arts (ELA) for Students Enrolled in Grades 10 and 11
- Science for Students Enrolled in Grades 5, 8, 10, and 11
- Social Studies for Students Enrolled in Grades 8, 10, and 11

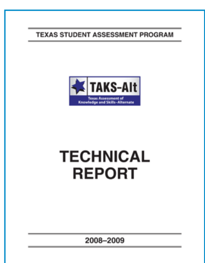
The Texas Education Agency (TEA) has developed the Texas Assessment of Knowledge and Skills–Alternate, or TAKS–Alt, to meet the diverse needs of students with significant cognitive disabilities enrolled in grades 3–11. This alternate assessment is based on the state-mandated curriculum and meets requirements under both the Individuals with Disabilities Education Improvement Act of 2004 (IDEA 2004) and the Elementary and Secondary Education Act (ESEA). These federal laws require that all students with disabilities be included in general state assessments, with appropriate accommodations and alternate assessments when necessary, as indicated by their individualized education programs (IEPs).

(Continued on inside flap of brochure.)

TEA developed a TAKS–Alt parent brochure that summarizes the assessment for students receiving special education services. The brochure includes a sample CSR with explanations of each element of the report to help parents better understand their child's score report. Test objectives for each subject area assessed with TAKS–Alt are summarized. The brochure, developed in both English and Spanish, was distributed with individual student results in spring 2010.

Standard Setting

The standards used to define student performance for TAKS–Alt assessments in 2009–2010 were set in April 2009. In April 2009, panels of educators were convened to recommend cut scores which were then reviewed by TEA and later approved by the Commissioner of Education. The approved cut score for Commended Performance was 68 out of a possible 72 points, and the approved cut score for Met Standard was 44 out of a possible 72 points.


TEXAS STUDENT ASSESSMENT PROGRAM

TAKS–Alt

TECHNICAL REPORT

2008–2009

A description of the standard setting process for TAKS–Alt is available in chapter 9 of the *2008–2009 TAKS–Alt Technical Report*.

Scaling

Scaling is the statistical procedure used to make test scores easier to interpret and compare across test administrations by placing raw scores on a common scoring metric. Most programs in the Texas assessment program use the Rasch Partial-Credit Model (RPCM) to place test items on the same scale across administrations. Once performance standards have been set for an assessment, the 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.

Unlike the other programs in the Texas assessment program, scaling is not done for the TAKS–Alt assessments. All results are reported using the raw score scale which is the number of items a student answers correctly. This is because all TAKS–Alt assessments consist of four assessment tasks based on teacher observation. Thus, there is not the same type of variation across tests forms as is typically observed in a multiple-choice assessment. Through training, teachers are able to consistently apply the TAKS–Alt rubric and maintain the integrity of the TAKS–Alt raw score scale across assessment tasks and administrations. The distribution of raw scores for all TAKS–Alt assessments is available in Appendix D.

During the administration of the 2009–2010 TAKS–Alt assessments, there were 77,450 TAKS–Alt assessments completed for all grades and subjects. Overall descriptive statistics were calculated including the mean and standard deviation of the total score for the assessments (refer to Table 20). The distribution of raw scores for each TAKS–Alt assessment is available in Appendix D.

Table 20. Descriptive Statistics for Operational Data—Overall

Number of Assessments	Total Score Mean	Standard Deviation	Median	Mode	Minimum	Maximum
77,450	57.81	10.37	61	65	0	72

In order to incorporate skill-level performance into the scoring of TAKS–Alt, Complexity Level was added as a component to the scoring rubric. The number of assessment tasks completed at each complexity level is shown in Table 21. Across all assessments, a total of 309,788 individual tasks were completed. The results show that approximately 50% of the assessment tasks were completed at Complexity Level 2. Complexity Level 3 had the lowest number of assessment tasks completed.

Table 21. Number of Assessment Tasks by Complexity Level

	Complexity Level 3	Complexity Level 2	Complexity Level 1	Overall
Number of Assessment Tasks	61,426	151,523	96,839	309,788
Percentage of Total	*20%*	*49%*	*31%*	*100%*

The TAKS–Alt scoring rubric is comprised of three dimensions: 1) Demonstration of Skill, 2) Level of Support, and 3) Generalization of Skill. The Demonstration of Skill scores are weighted based on the complexity level of the task. The essence statement score for a complexity level assessment task ranges from 0 to 18 for Complexity Level 3 tasks, 0 to 16.2 for Complexity Level 2 tasks, and 0 to 12 points for Complexity Level 1 tasks. The most frequently occurring essence score was at the maximum score point of each complexity level, indicating that students were doing very well at the complexity levels determined. Additional teacher training will be provided to ensure teachers are selecting tasks that are challenging for students. The frequency distribution of scores on an individual essence statement (one assessment task with three predetermined criteria is completed for each essence statement) was calculated for each complexity level (refer to Tables 22 to 24).

Table 22. Frequency Distribution for the Complexity Level 3 Essence Scores

Score	0	3	4	5	6	7
Number of Essences	55	22	56	109	44	53
Percentage of Total	*<1%*	*<1%*	*<1%*	*<1%*	*<1%*	*<1%*
Score	**8**	**9**	**10**	**11**	**12**	**13**
Number of Essences	144	402	613	746	833	1,110
Percentage of Total	*<1%*	*1%*	*1%*	*1%*	*1%*	*2%*
Score	**14**	**15**	**16**	**17**	**18**	**Overall**
Number of Essences	145	1,549	3,527	8,553	43,465	61,426
Percentage of Total	*<1%*	*3%*	*6%*	*14%*	*71%*	*100%*

Table 23. Frequency Distribution for the Complexity Level 2 Essence Scores

Score	0	2	2.4	3.4	4	4.4	4.8	5.8
Number of Essences	235	0	74	180	0	378	97	140
Percentage of Total	*<1%*	-	*<1%*	*<1%*	-	*<1%*	*<1%*	*<1%*
Score	**6**	**6.8**	**7.2**	**7.8**	**8.2**	**8.8**	**9.2**	**10.2**
Number of Essences	0	456	771	480	635	1,325	2,187	2,206
Percentage of Total	—	*<1%*	*1%*	*<1%*	*<1%*	*1%*	*1%*	*2%*
Score	**11.2**	**12.2**	**13.2**	**14.2**	**15.2**	**16.2**	**Overall**	
Number of Essences	3,234	369	5,337	8,096	18,797	106,528	151,523	
Percentage of Total	*2%*	*<1%*	*4%*	*5%*	*12%*	*70%*	*100%*	

Table 24. Frequency Distribution for the Complexity Level 1 Essence Scores

Score	0	2	3	4	5	6	
Number of Essences	375	160	107	517	111	1,538	
Percentage of Total	*<1%*	*<1%*	*<1%*	*1%*	*<1%*	*2%*	
Score	**7**	**8**	**9**	**10**	**11**	**12**	**Overall**
Number of Essences	602	1,914	2,836	4,295	6,704	77,680	96,839
Percentage of Total	*1%*	*2%*	*3%*	*4%*	*7%*	*80%*	*100%*

To complete the assessment, a student was evaluated on four assessment tasks that could vary in terms of complexity level for each task. The frequency distribution of students for each possible complexity level combination was calculated to evaluate the combinations of complexity levels being selected for a single assessment (refer to Table 25). The majority of students (33%) completed four Complexity Level 2 assessment tasks. The second most common combination (25%) was students completing four Complexity Level 1 assessment tasks followed by students completing four Complexity Level 3 tasks (11%).

Table 25. Frequency Distribution of Overall Complexity Level Combination

CL Combination	**1111**	**1112**	**1113**	**1122**	**1123**	
Number of Selections	19,403	2,789	48	2,971	132	
Percentage of Total	*25%*	*4%*	*<1%*	*4%*	*<1%*	
CL Combination	**1133**	**1222**	**1223**	**1233**	**1333**	
Number of Selections	22	3,785	447	193	39	
Percentage of Total	*<1%*	*5%*	*1%*	*<1%*	*<1%*	
CL Combination	**2222**	**2223**	**2233**	**2333**	**3333**	**Overall**
Number of Selections	25,445	5,185	4,447	3,978	8,559	77,443
Percentage of Total	*33%*	*7%*	*6%*	*5%*	*11%*	*100%*

In addition to the previous tables, correlation coefficients were calculated between the different dimensions of the scoring rubric and the complexity levels. The overall correlations were significant at the $p \leq 0.01$ level (refer to Table 26). The correlations between Generalization of Skill and the dimensions of Level of Support and Demonstration of Skill were moderate. However, the correlation between Demonstration of Skill and Level of Support was lower. Correlations between complexity level and Demonstration of Skill were strong, which is to be expected since the Demonstration of Skill score is weighted by the complexity level of the task. The relationship between complexity level and generalization was of moderate strength, which is due in part to the fact that students completing Complexity Level 1 assessment tasks are not eligible for generalization. The correlation between complexity level and Level of Support was quite low, indicating that these two components of TAKS–Alt are assessing different constructs of the assessment.

Table 26. Overall Correlations among Complexity Level, Demonstration of Skill, Level of Support, and Generalization of Skill

	Complexity	Demonstration	Support	Generalization
Complexity	1.000			
Demonstration	0.920*	1.000		
Support	0.012*	0.182*	1.000	
Generalization	0.477*	0.492*	0.403*	1.000

n=77,450 (56,413 were eligible for generalization of skill), * $p \leq .01$

Equating

Refer to chapter 3 for detailed information about the equating process. The distribution of raw scores for all TAKS–Alt assessments available in Appendix D.

Equating is not done for TAKS–Alt. The difficulty level of the assessments and assessment tasks is taken into consideration through the differential weighting of the Complexity Level of the task. In addition, consistency across administrations is maintained through the training and qualification done by teachers before administering the assessment.

Reliability

During the 2009–2010 school year, reliability estimates for TAKS–Alt were obtained from an interrater reliability study.

The Interrater Reliability Study

Assessments that are not traditional paper-and-pencil or multiple-choice tests may require a different approach to gather reliability evidence. Interrater reliability is an alternate method often used to provide reliability evidence. Interrater reliability for TAKS–Alt was evaluated by having two raters observe the same student performing a specific assessment task at the same time. Both raters evaluated the student's performance using the assessment's performance evaluation questions. The two independent ratings were then compared to determine the reliability for TAKS–Alt.

The first and second ratings for this study took place during the 2009–2010 TAKS–Alt assessment window, between January 4, 2010 and April 16, 2010. The following grades and subjects were included in the study:

- grades 6 and 9 mathematics
- grades 6 and 9 reading
- grade 11 science

A target sample size of approximately 200 students per test (1000 students total) was selected to obtain the desired statistics for the study, for a total of approximately 1,500 students. In an attempt to obtain the target sample size of 1,000 students, approximately 1,500 students were selected by campuses to take part in the study. Not all students who were selected were able to participate in the study, resulting in a final sample size of 796 students.

When a student was selected for the interrater study, the student's primary teacher was notified by TEA of the subject area and essence statement to be observed by the second rater. To standardize the qualifications for selection of the second rater, TEA developed and published the following guidelines.

TAKS–Alt second raters should be professionals or under the supervision of professionals who hold valid education credentials such as Texas teacher certificates or permits. Those selected may include the following:

- teachers (including general, special education, and teachers for the visually and auditorily impaired)
- paraprofessionals (when appropriate)
- assessment specialists
- speech therapists
- occupational therapists
- physical therapists

The second rater for the study should be someone who knows the student well enough to provide a knowledgeable rating and is available to observe the assessment task. The second rater is required to complete all the TAKS–Alt training modules and successfully complete the online training qualification activities before taking part in the study.

Once the second rater qualified for the study, the primary teacher was able to plan a task and schedule when the task would be administered to the student.

The primary teacher implemented the assessment task while the second rater observed and both raters documented their own observations on a data observation sheet, which was used to complete the student performance evaluation questions on the TAKS–Alt online system by the primary teacher. Second raters entered their data into the TAKS–Alt interrater reliability online tool by answering similar student performance evaluation questions. Once the observation was complete, the raters gave their separate, independent ratings.

The second rater was asked to respond to the performance evaluation questions for demonstration of skill and level of support. The generalization of skill dimension was not included in the study because the student may demonstrate this skill at a different time and the second rater may not be present when generalization occurs.

The results for the interrater reliability study are provided in the following tables. There were several statistics calculated for the study's results. The results are shown for each dimension of the rubric (except for generalization of skill) and the score combining Demonstration of Skill and Level of Support.

Correlation coefficients were calculated between the two sets of ratings as shown in Table 27. Complexity Level correlations showed the strongest relationship between rater 1 and rater 2 (0.95 to 0.99). The correlation between the raters for Demonstration of Skill ranged from 0.88 to 0.99 and the correlation for Level of Support ranged from 0.73 to 0.82. The patterns of all correlations across subject areas were quite similar. The correlation for mathematics ranged from 0.73 to 0.99; for reading ranged from 0.72 to 0.97; and the correlation for science ranged from 0.81 to 0.95.

Table 27. Correlations between First and Second Ratings

Subject Area	Grade	Complexity Level	Demonstration of Skill	Level of Support	Combined Score	Sample Size
Mathematics	6	0.99	0.95	0.80	0.88	186
	9	0.99	0.99	0.73	0.89	146
Reading	6	0.98	0.92	0.82	0.89	205
	9	0.97	0.90	0.82	0.86	130
Science	11	0.95	0.89	0.81	0.74	127

NOTE: All correlations significant at $p<.01$

Perfect agreement rates between rater 1 and rater 2 were calculated for Complexity Level, Demonstration of Skill, and Level of Support, as shown in Table 28. Perfect agreement rates were high for Complexity Level (98% to 99%) and Demonstration of Skill (97 % to 100%). Perfect agreement rates were slightly lower for Level of Support (83% to 91%) and the Combined Score (83% to 91%). Results were similar across subject areas.

Table 28. Percentages of Perfect Agreement between First and Second Ratings

Subject Area	Grade	Complexity Level	Demonstration of Skill*	Level of Support	Combined Score
Mathematics	6	99	97	83	83
	9	99	100	86	86
Reading/ELA	6	99	97	84	84
	9	98	98	83	83
Science	11	98	98	91	91

Conclusions from the Interrater Reliability Study

The range of the correlation coefficients and the strength of agreement level of the kappa coefficients indicate that the relationships between the first and second ratings are high. This trend occurred for all three grade levels and across all three subjects for Complexity Level, Demonstration of Skill, Level of Support, and for the score combining Demonstration of Skill and Level of Support—supporting the reliability of the TAKS–Alt scores. The agreement rate results indicate that first and second raters had high levels of agreement. Interrater reliability analyses is available in the *2008–2009 TAKS–Alt Technical Report*.

Validity

Refer to chapter 3 for detailed information about validity evidence. Texas collects validity evidence annually to support the various uses of TAKS–Alt scores. The sections that follow describe how these types of validity evidence were collected for the TAKS–Alt assessments in 2009–2010.

Evidence Based on Test Content

Validity evidence based on test content refers to evidence of the relationship between tested content and the construct the test is intended to measure. The TAKS–Alt test development process, which was conducted in 2008–2009, played an intricate role in providing content validity evidence for the assessment in 2009–2010. The assessment tasks developed during that process were administered in the 2009–2010 school year. To link the TAKS–Alt assessment with the Texas grade-level content standards assessed on TAKS, a vertical alignment was conducted on the reading/English language arts, writing, mathematics, science, and social studies TEKS curriculum from the tested grades and subjects.

A task force of content experts, curriculum specialists, and assessment specialists, including TEA Student Assessment Division and Curriculum Division experts, aligned the TEKS knowledge and skills statements and their accompanying student expectations. The task force developed the vertical alignment and curriculum framework documents to help teachers access the grade-level TEKS curriculum. Curriculum and content specialists who did not participate in the alignment process verified the alignment as well as educator advisory committees, the steering committee, and educator item-review committees.

In addition to the vertical alignment conducted by the state, TEA contracted with Dr. Norman Webb to conduct an independent vertical alignment study. In Dr. Webb's alignment study, four indicators of Categorical Concurrence—Extended Depth of Knowledge, Range of Knowledge, and Balance of Representation—were used to determine the degree of alignment.

The study aligned the TEKS (the objectives, the knowledge and skills statements, and the essence statements) to the state-standardized assessment tasks for TAKS–Alt for the following grades and subjects:

- grades 3–8 and 10 mathematics
- grades 3–8 and 10 reading/English language arts
- grades 5, 8, and 10 science

In general, the alignment studies showed full or reasonable degrees of alignment between the assessments and the curriculum standards in all subject areas, providing support for the content validity of TAKS–Alt.

Educator judgments from educator review meetings provided content-validity evidence. Educators from across the state reviewed the content of every item to ensure that each item matched to the appropriate content standard. Committees included special education specialists, special education classroom teachers, teachers of English language learners, and general education teachers.

An Assessment Task Judgment Form was completed by every member of the educator review committees. Judgments were collected for every assessment task in response to the question, "Does the assessment task measure the objective statement/knowledge and skills statement/essence statement it was designed to measure?" Teachers used the TEKS curriculum documents to verify the match of the objective/essence statement to each assessment task. Summaries of the committees' judgment related to assessment tasks' alignment to specific content standards clearly demonstrated that the TAKS–Alt is an appropriate and reliable measure of the state's content standards.

The committees also evaluated the assessment tasks for bias to ensure that the tasks were measuring the content instead of another construct. Judgments were collected for every assessment task on the Assessment Task Judgment Form related to its potential bias in response to the question, "Are these assessment tasks free from bias on the basis of personal characteristics such as gender or ethnicity?"

Summaries of the committees' judgments relating each essence statement to specific content standards and sub-content standards (TEKS student expectations) clearly demonstrate that the committee members believed TAKS–Alt is free from bias.

After the assessment tasks were reviewed by educator committees, two consultants reviewed the assessment tasks for any bias toward disability groups. One consultant was from the Texas School for the Blind and Visually Impaired, and one consultant was from the Texas School for the Deaf. The assessment tasks were revised to reflect suggestions from the educator review committees and the consultants.

Evidence Based on Response Processes

Texas collects evidence to suggest that the way in which students respond to test questions on the TAKS–Alt assessments supports the accurate measurement of the construct. The TAKS–Alt audit provided additional validity evidence for the TAKS–Alt assessment by showing that the assessment is being administered and implemented correctly, as well as scored accurately. Because the response processes used for this assessment are the teacher observations of student performance, the audit involved a review of teacher-submitted documentation about the teachers' observations taken while the students completed their assessment tasks.

Approximately 4% of TAKS–Alt students (a target sample of 1,000 students) were asked to participate in the audit. The sample was selected to be representative of the population in terms of gender, ethnicity, district size, and region of the state. TAKS–Alt assessment documentation was requested for each sampled student. The submitted documentation was then placed in student folders that were reviewed during the audit meetings. The meetings were held over a two-week period in July 2010 and included the following grades and subjects:

- grades 6 and 9 mathematics
- grades 6 and 9 reading
- grade 11 science

Across all meetings, a total of 839 student folders were viewed by 27 auditors. The reduction of the sample typically resulted from student movement across campuses and the submission of incorrect documentation.

To ensure that all student folders were reviewed by at least two educators, auditors were assigned to review specific student folders in a specific order. Folders were organized into sets, with each auditor reviewing at least six sets of folders. Each folder was assigned to be viewed by at least two auditors, with some folders viewed by as many as four auditors. However, there were several mathematics and science folders reviewed by only one auditor. This occurred because fewer panelists than expected attended this meeting; as a result, there was not adequate time for all folders to be reviewed by at least two auditors.

Within each student folder was information about the four essence statements that were assessed in spring 2010. This information included the essence statement being assessed, the assessment task and the three predetermined criteria for each essence statement, the documentation of the student's responses during the task, and the responses to the student performance evaluation questions entered into the online system.

While the auditors were making their judgments, they were asked to evaluate the following: how well the documentation supported the responses to the student performance evaluation questions (for Demonstration of Skill, Level of Support, and Generalization of Skill) from the online assessment system, and whether the complexity level of the assessment task was maintained during the administration of the assessment task. In addition to the rating options of how well the documentation supported a student's performance evaluation, a "Not Documented" response category was added to the 2010 audit judgment form for those instances where there was no documentation to review or the documentation was not adequate to make a judgment. A third question, whether the complexity level of the assessment task was maintained when the student was given the opportunity to generalize the skill, was also added to the 2010 audit judgment form. After reviewing the folders, the auditors entered their judgments on an audit judgment form.

Table 29 shows the frequency of responses to all five items on the audit judgment form. The highest rates of agreement were for the statement regarding how well the documentation supported the student's performance evaluation for Level of Support, while the agreement rates were lower for the statements pertaining to how well the documentation supported the Demonstration of Skill performance evaluation and Generalization of Skill. The highest percentage of auditor agreement occurred when auditors were asked whether the complexity level of the assessment task was maintained during the implementation of the assessment task as well as Generalization of Skill. These high results indicate that the teachers are selecting the appropriate complexity level for their students.

Table 29. Frequency of Responses to Audit Judgment Form Items

Audit Item	Mean Audit Rating	Frequency	Percent	Total
The documentation supports the student's performance evaluation for demonstration of skill. (16)	Strongly Agree	469	55.90	839
	Agree	308	36.71	
	Disagree	57	6.79	
	Strongly Disagree	5	0.60	
The documentation supports the student's performance evaluation for level of support. (23)	Strongly Agree	497	59.31	838
	Agree	287	34.25	
	Disagree	52	6.21	
	Strongly Disagree	2	0.24	
The documentation supports the student's performance evaluation for generalization of skill. (48)	Strongly Agree	265	47.32	560
	Agree	238	42.50	
	Disagree	52	9.29	
	Strongly Disagree	5	0.89	
The complexity level of the assessment task. (46)	Increased	16	1.91	836
	Maintained	819	97.97	
	Decreased	1	0.12	
The complexity level of the generalization of skill. (109)	Increased	3	0.56	537
	Maintained	530	98.70	
	Decreased	4	0.74	

NOTE: Mean Audit Rating is a one-step calculation. 1) Get the mean rating of each folder across panelists. Frequency of folders with "Not Documented" is listed in parentheses below each audit item.

Further teacher training on student documentation may result in higher agreement rates in future audits. Future studies will allow audit results to be monitored over time. Previous audit analyses available in the *2008–2009 TAKS–Alt Technical Report*.

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. A measure of internal consistency is used to provide evidence of the internal structure of a test. A measure of internal consistency is currently not available for TAKS–Alt. Various methods for determining internal consistency for TAKS–Alt are currently being investigated.

Evidence Based on Relationships 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. Several analyses were conducted to show that TAKS–Alt scores are related to other variables as expected and related weakly, if at all, with irrelevant characteristics.

Correlations among TAKS–Alt subject scores were calculated. The results in the tables below indicated that the correlations were high but not so high as to indicate redundancy. The correlations among subjects ranged from 0.850 to 0.889. This finding provides strong validity evidence, as empirical results match the theory underlying the relationship between these constructs.

Table 30. Overall TAKS–Alt Correlation between Subject Scores

	Correlation	*n*-Count
Mathematics Total & Reading Total	0.866*	26,330
Science Total & Reading Total	0.850*	10,901
Social Studies Total & Reading Total	0.868*	7,797
Writing Total & Reading Total	0.879*	6,057
Science Total & Math Total	0.863*	10,907
Social Studies Total & Math Total	0.871*	7,803
Writing Total & Math Total	0.850*	6,057
Social Studies Total & Science Total	0.889*	7,800

* The above correlations were significant at the $p \leq .05$ level.

NOTE: The correlation between any subject variable and itself is 1.

Additional validity evidence was gathered in the form of discriminant validity analyses, demonstrating that the TAKS–Alt test scores were unrelated to demographic variables (e.g., gender and ethnicity). Theoretically, student characteristics should not relate to their performance on the assessment; therefore, the lack of meaningful empirical relationships between these measures is expected and is shown in the tables below.

Table 31. Overall TAKS–Alt Correlation between Total Score and Gender

	Total Score	Gender
Total Score	1.000	
Gender	-0.026	1.000

n = 77,450

* The above correlation was significant at the p ≤ .05 level.

Table 32. Overall TAKS–Alt Correlation between Total Score and Ethnicity

	Total Score	Gender
Total Score	1.000	
Ethnicity	0.011	1.000

n = 75,932

* The above correlation was significant at the p≤.05 level.

Evidence Based on Consequences of Testing

Another way to provide validity evidence is by documenting the intended and unintended consequences of administering an assessment.

The intended consequences of TAKS–Alt are directly tied to its primary purpose: to measure and improve student achievement based on alternate achievement standards. Validity evidence that shows that TAKS–Alt has a positive impact on student learning and instruction has been collected through educator review meetings and focus group meetings during the 2008–2009 test development process as well as through teacher and district test coordinator surveys collected during the 2009–2010 school year.

Survey results showed that teachers generally agreed that students would be adequately prepared for the 2009–2010 administration of the assessment.

Educator committee meetings to review the assessment tasks provided documentation showing that the behaviors being measured are relevant and important for the special education population. The teachers indicated that the assessment tasks were designed to be age-appropriate and linked to the TEKS.

Most teachers clearly saw the link between the assessment task and the associated essence statement. These consequences of the assessment are intentional and lead to the improvement of student achievement.

According to educators, unintended consequences include increased collaboration of special education teachers and general education teachers. The alignment of TAKS–Alt with the TEKS curriculum requires special education teachers to be more focused on grade-level content than they have been in the past. Though not used as much as some of the other resources available to special education teachers, the general education teacher has started to emerge as a resource for grade-level content and instructional activities that can be adapted for use by special education teachers. This increased collaboration allows for sharing of ideas that will increase special education teachers' familiarity with the TEKS.

Additionally, during educator committee meetings and focus groups, teachers expressed surprise at the amount of interest their students demonstrated during the assessment tasks and how much the students were capable of learning. Teachers indicated that they need to allow students to be more independent and only use the least invasive cues and prompts during classroom instruction so that students have the opportunity to demonstrate what they know during the assessment without unnecessary support from the teacher. These consequences of the assessment should lead to the improvement of student achievement. TEA will continue to gather validity evidence regarding the consequences of administering TAKS–Alt.

Measures of Student Progress

In 2009–2010 Texas implemented the TAKS–Alt growth measure to assist campuses in meeting Adequate Yearly Progress (AYP) for federal reporting purposes and as a means for evaluating campuses in the state accountability system. The TAKS–Alt growth measure describes a student's score changes from one year to the next and is used to determine whether a student is on track to meet the standard and/or achieve commended performance on TAKS–Alt at a future grade. The model for the TAKS–Alt growth measure is based on a student continuing to make progress at the same rate from the current year to future grades. The measure consists of two parts: 1) a student's stage change from the prior year to the current year, which is determined by representing the student's scores from a previous school year and the current school year in terms of the stage of performance achieved each year; and 2) a determination of whether the progress made is sufficient to designate the student as on track to meet the state's performance standards in the next growth target grade (grade 5, 8, or 11).

The TAKS–Alt growth measure uses a transition table approach to growth. To implement this approach, the performance categories, which divide the assessment's raw score scale into three sections, are divided into sublevels, or stages. Dividing the performance categories into stages provides a mechanism through which parents and teachers can observe meaningful changes in student performance along the score scale and allows them to track student performance from year to year by reviewing a student's stage changes, or stage transitions. The number of stage changes a student demonstrates from year to year can then be used to determine whether a student is making progress each year and to determine whether a student on track to Meet the Standard or achieve Commended Performance in a future grade. For a full description of the process used to develop the TAKS–Alt growth measure, refer to "Procedures for Developing the TAKS–Alt Growth Measure" on TEA's Student Assessment Division website.

In spring 2010, stage change information was reported in grades 4–11 mathematics and reading/English Language Arts (ELA) and in grade 11 science and social studies. Beginning in 2011, stage change information will also be reported for grade 10 science and social studies. To determine stage change, two years of student data are needed for tests administered no more than two years apart; as a result, stage change can only be determined if a student has a score for the current grade and a previous grade (either one or two grades away) in the same subject.

Additionally, on track information was reported for grades 4–10 mathematics and reading/ELA. On track information for grade 10 science and social studies will be reported in spring 2011. For the on track component, two years of data (for assessment administered no more than two years apart) prior to the growth target grade are needed. The growth target grades are grades 5, 8, and 11, which also are the target grades used for the Texas Projection Measure. Like the evaluation of the projections for the TPM, TEA will evaluate the accuracy of the on track designations for the TAKS–Alt growth measure.

Because the on track designations were first reported and used in 2010, the accuracy of the designations cannot be evaluated until 2011 performance data become available. There are data available, however, indicating the number of students who were designated as on track to Meet the Standard or achieve Commended Performance (refer to Table 33). The results on the next page show that between 74% and 84% of all students in the grades and subjects in which the growth measure was used received an on track designation. Of those students, between 92% and 96% were on track to Meet the Standard, and between 31% and 50% were on track to achieve Commended Performance. Once performance data become available, TEA will evaluate the accuracy of the on track designations.

Table 33. TAKS–Alt On Track Classification Results for 2009–2010, All Students

Grade/ Subject	Total Assessed (Pct. Relative to All Students Assessed)	On Track Designation Status (Pct. Relative to All Students Assessed Group)		On Track Classifications (Pct. Relative to the Students with an On Track Designation Group)		On Track Classifications (Pct. Relative to the Students with an On Track Designation Group)	
		Students With an On Track Designation	Students without an On Track Designation	On Track to Meet Standard	Not On Track to Meet Standard	On Track to Achieve Commended Performance	Not On Track to Achieve Commended Performance
Grade 4 Reading	3230 (100.00)	2724 (84.33)	506 (15.67)	2546 (93.47)	178 (6.53)	986 (36.20)	1738 (63.80)
Grade 4 Mathematics	3231 (100.00)	2727 (84.40)	504(15.60)	2584 (94.76)	143 (5.24)	1254 (45.98)	1473 (54.02)
Grade 5 Reading	3110 (100.00)	2469 (79.39)	641(20.61)	2335 (94.57)	134 (5.43)	1143 (46.29)	1326 (53.71)
Grade 5 Mathematics	3108 (100.00)	2474 (79.60)	634(20.40)	2368 (95.72)	106 (4.28)	1238 (50.04)	1236 (49.96)
Grade 6 Reading	3007 (100.00)	2273 (75.59)	734 (24.41)	2145 (94.37)	128 (5.63)	1023 (45.01)	1250 (54.99)
Grade 6 Mathematics	3011 (100.00)	2278 (75.66)	733 (24.34)	2168 (95.17)	110 (4.83)	1056 (46.36)	1222 (53.64)
Grade 7 Reading	2831 (100.00)	2280 (80.54)	551 (19.46)	2158 (94.65)	122 (5.35)	846 (37.11)	1434 (62.89)
Grade 7 Mathematics	2832 (100.00)	2283 (80.61)	549 (19.39)	2180 (95.49)	103 (4.51)	868 (38.02)	1415 (61.98)
Grade 8 Reading	2984 (100.00)	2275 (76.24)	709 (23.76)	2132 (93.71)	143 (6.29)	1057 (46.46)	1218 (53.54)
Grade 8 Mathematics	2986 (100.00)	2281 (76.39)	705 (23.61)	2158 (94.61)	123 (5.39)	1136 (49.80)	1145 (50.20)
Grade 9 Reading	2739 (100.00)	2026 (73.97)	713 (26.03)	1896 (93.58)	130 (6.42)	857 (42.30)	1169 (57.70)
Grade 9 Mathematics	2739 (100.00)	2029 (74.08)	710 (25.92)	1901 (93.69)	128 (6.31)	921 (45.39)	1108 (54.61)
Grade 10 ELA	2492 (100.00)	2057 (82.54)	435 (17.46)	1906 (92.66)	151 (7.34)	850 (41.32)	1207 (56.68)
Grade 10 Mathematics	2494 (100.00)	2056 (82.44)	438 (17.56)	1899 (92.36)	157 (7.64)	636 (30.93)	1420 (69.07)

Sampling

Refer to chapter 3 for detailed information about the sampling process. Two samples were selected for research study purposes for TAKS–Alt during the 2009–2010 school year: the interrater reliability sample and the validity audit sample.

Interrater Reliability Sampling

The sampling plan for the TAKS–Alt interrater reliability study for grades 6 and 9 mathematics and reading and grade 11 science was to sample approximately 4% of students (n = 1,000) taking TAKS–Alt. (Sample numbers were based on an estimate of 24,000 students participating in the assessment in 2008–2009.) When designing the sampling strategy of the interrater reliability study, the sample size for the interrater study was selected so that it

- took into account that the interrater study is only one phase of a larger set of studies,
- took into account that reliability evidence is gathered over multiple years,
- took into account the campus and district burden for all planned studies across all the testing programs,
- considered the desired outcomes of the study (correlations, agreement rates, and kappa statistics), and
- accommodated the planned level of aggregation (test level).

A data extract from the 2008–2009 school year was used to obtain the interrater reliability sample. A sample of approximately 200 students per aggregation level was selected to meet statistical requirements while taking practical and logistical considerations into account:

- Students who were in grades 5, 8, and 10 for the 2008–2009 TAKS–Alt administration were eligible to be selected for the sample. If students were sampled, districts and campuses would have to be notified that these students would participate in the study. However, notifying the appropriate district and campus was a concern because students in grades 5 and 8 typically enroll at a new campus when they are promoted to grades 6 and 9. Due to the campus movement and subsequent difficulty tracking these students, campuses were sampled instead of students. The campuses sampled selected the students to participate in the study.
- Elementary, middle, and high school campuses from the 2008–2009 TAKS–Alt administration were eligible to be selected for the sample. The number of students in grades 6, 9, and 11 was checked across campuses to ensure that selected campuses had students available in these grades.
- Elementary and middle school campuses were eligible to be sampled for grade 6 reading and mathematics.

- High school campuses were eligible to be sampled for grade 9 reading and mathematics and grade 11 science.
- To minimize campus and district burden, no district had more than 10 campuses sampled within a test, and no district had more than 20 campuses sampled across all five tests.
- Students who were in grades 6, 9, and 11 at the time of the 2009–2010 TAKS–Alt administration were eligible to be selected by the campuses.
- A campus selected for grade 6 reading or mathematics would select two students.
- A campus selected for grade 9 reading or mathematics or grade 11 science would select three students.
- Overall, a group of approximately 300 students per test would be selected across all sampled campuses. Requesting campuses to select more students than the target of 200 students per test would allow for attrition.
- Random assignment to subject and essence statement was done at the campus level, so all students at a campus completed the same essence statement within the same subject area.

Across the three grade levels, 377 districts and 600 campuses were selected to participate in the study, for a total of approximately 1,500 students. Sampling more students than the target was designed to produce a sample that would allow for attrition. Although approximately 1,500 students were selected to participate, the final number of students participating in the Interrater Reliability Study was 796. By subject, there were 333 students for mathematics, 335 students for reading and 128 students for science.

Validity Audit Sampling

For the audit, a sample of 4% (n=1000) of the TAKS–Alt students was targeted. The audit sample consisted of students who participated in the 2008–2009 audit in grades 5 and 8 mathematics and reading and grade 10 science. The audit data was collected on the same students in order to conduct the TAKS–Alt longitudinal study. Based on an early extract from the TAKS–Alt online system in March 2010, 972 of the 1,040 audit students who participated in the 2008–2009 audit could be matched from 2009 to 2010. These 972 students were eligible to be sampled for the 2009–2010 audit.

After taking into account several practical and logistical considerations, the 2010 audit sample consisted of 853 students, or approximately 3% of the students assessed with TAKS–Alt. This percentage was slightly lower than the sample target of 4% due to sampling only those students who had participated in the 2008–2009 audit and could be matched from 2009 to 2010, and to the increase in the number of students who were assessed with TAKS–Alt between 2009 and 2010 (approximately 1,800 more students). Once the students were selected to participate in the audit study, campuses were notified to submit the documentation forms for selected students for the specified subject area. The tests included in the audit are listed in Table 34.

Table 34. Grades and Subject Areas Included in the Audit

Mathematics	Reading	Science
6	6	11
9	9	—

Although approximately 853 students were sampled for the audit, the final number of students who participated in the audit consisted of 839 students. The reduction of the original sample typically resulted from student movement across districts. First, grade 176 mathematics and 181 reading students were selected from grade 6. Second, 157 mathematics and 160 reading students were selected from grade 9. Lastly,179 science students were selected from grade 11. The audit sample was selected to be representative of the population in terms of gender, ethnicity, district size, and region of the state.