

TAKS–Alternate (TAKS–Alt)

[Overview](#)[TAKS–Alt Participation Requirements](#)[Test Development](#)[Training](#)[Test Administrations](#)[Testing Accommodations](#)[Student Success Initiative](#)[Scores and Reports](#)[Standard Setting](#)[Scaling](#)[Equating](#)[Reliability](#)[Validity](#)[Sampling](#)

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 20. TAKS–Alt was administered during the window of January 4 through April 16 for all subjects and grade levels.

Table 20. 2008–2009 TAKS–Alt Assessments

2008–2009 TAKS–Alt Assessments	
Grade	Test Administration
Grade 3	Mathematics and Reading
Grade 4	Writing, Mathematics, and Reading
Grade 5	Mathematics, Reading, and Science
Grade 6	Mathematics and Reading
Grade 7	Writing, Mathematics, and Reading
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 must be carefully considered 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 in order to participate in TAKS–Alt. The members of the admission, review, and dismissal (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.

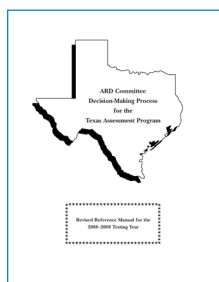
Students receiving special education services who have the most significant cognitive disabilities and are unable to participate in other statewide assessments even with substantial accommodations and/or modifications can be considered for TAKS–Alt. After a cognitive disability has been established for 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 2008–2009 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 three 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 in order 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).

All of the criteria listed above were specifically considered in the development of assessment tasks for each essence statement. 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.

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 Flower’s 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 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.

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 1: BEGINNING AWARENESS

Level 1 assessment tasks are the least complex and involve responding to 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.

Educator Review Committee Meetings

Once TEA approved the assessment tasks, committees composed of Texas educators reviewed the TAKS–Alt assessment tasks to eliminate potential bias and to judge the appropriateness of task content and complexity level. The committees included special education specialists, 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 to review the standardized assessment tasks. These committees reviewed the alignment of the assessment tasks with grade level standards using prerequisite skills linked to the grade-level. The composition of the TAKS–Alt educator review committees is described in Table 21.

Table 21. TAKS–Alt Educator Review Committees’ Demographic Data

Demographic		Number	Percent
Gender	Female	84	84
	Male	16	16
	Total	100	100
Ethnicity	African American	14	14
	Hispanic	21	21
	White	63	63
	Other	2	2
	Total	100	100
Position	Special Ed	44	44
	General Ed	45	45
	ELL	11	11
	Total	100	100

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 three 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 the TEA 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.

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

Topics covered in this training module explain the features of TAKS–Alt including the complexity level of the assessment tasks and the steps for administering TAKS–Alt. Module 1 also reviews the role of the ARD committee and the TAKS–Alt participation requirements.

Module 2: Implementing the TAKS–Alt Assessment

Topics covered in this training module explain how to select and implement an assessment task through the addition of supports and response modes. Module 2 describes how to conduct the observation, provide appropriate cues and prompts, and record the results of the observation.

Module 3: The TAKS–Alt Online Instrument

Topics covered in this training module include information about the TAKS–Alt rubric, the generalization process, maintaining documentation, the automated scoring feature, and the submission process. Module 3 also provides instructions on using the online instrument for evaluating student performance.



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 ran from January 5 to April 10, 2009. The assessment 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 17, 2008. The assessment was administered to approximately 24,900 students who met participation requirements in spring 2009.

Testing Accommodations

Students being assessed with TAKS–Alt may use whatever accommodations and supports are routinely and successfully used as instructional accommodations. 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 provided 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.

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 not included in the TAKS–Alt process.

Scores and Reports

Scoring TAKS–Alt Assessments

In order to incorporate 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: Level 3—most complex, Level 2—moderately complex, or Level 1—least complex. 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 essence statement).

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
- 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 3 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 they complete 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)

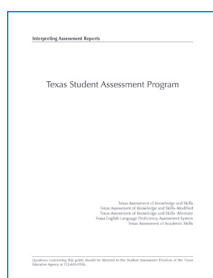
Report Formats

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

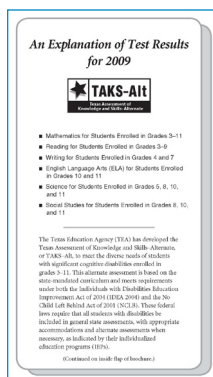
The standard reports available for the 2008–2009 TAKS–Alt program include the Confidential Student Report, Confidential Student Label, Confidential List of Students' Results, and Summary Report.

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



More information about scoring and reporting for TAKS–Alt can be found in the [2008–2009 Interpreting Assessment Reports](#).

Parent 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 2009.



Standard Setting

Standard setting is the process of relating levels of test performance directly to what students are expected to learn as expressed in the statewide curriculum by establishing cut scores that define performance categories like “Met Standard” and “Commended Performance.” Through the process of standard setting, cut scores (or student performance on the assessment tasks in terms of the Complexity Level, Demonstration of Skill, Level of Support, and Generalization) are determined to reflect the level of performance a student must demonstrate to match the TAKS–Alt performance level descriptors.

Standards were initially set for TAKS–Alt in the summer of 2007. However, due to significant changes to the assessment during 2008 and the beginning of 2009, new standards were needed. Therefore, in April 2009, TEA and Pearson convened five standard-setting panels to recommend cut scores for all TAKS–Alt assessments. The panels were grouped by the following subject areas:

- reading grades 3–9, English language arts (ELA) grades 10 and 11
- mathematics grades 3–11
- science grades 5, 8, 10, and 11
- writing grades 4 and 7
- social studies grades 8, 10, and 11

Panelists made judgments about the number of score points a student is required to obtain in order to achieve Met Standard or Commended Performance, and also reviewed student score patterns to set the recommended cut points. During the meetings, panelists reviewed the content measured by the assessment tasks, engaged in table and whole group discussions, and considered the impact on students when making cut-score recommendations.

The standard-setting meeting included sessions in which panelists 1) received training on the assessment and the standard-setting process, 2) became familiar with the TAKS–Alt performance level descriptors for each grade and subject, 3) reviewed assessment tasks, 4) applied a modified extended Angoff procedure (Angoff, 1971; Hambleton & Plake, 1995) to make judgments about the number of score points a borderline student would need to obtain on each assessment task in order to achieve Met Standard or Commended Performance, and 5) applied a modified performance profile procedure (Morgan, 2003) to set recommended cut points by reviewing student score patterns.

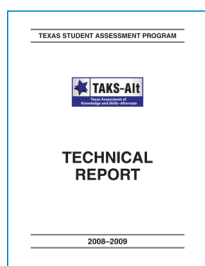
After the standard-setting meetings were completed, the panel recommendations from the meetings were reviewed as part of a vertically moderated standard-setting (VMSS) procedure (Cizek & Bunch, 2007) to determine the final cut score recommendation. VMSS is a method to adjust individual grade or content area standards recommended by discrete standard-



setting panels. TEA implemented a 3-step process to determine the final recommended cut scores. The first step was reviewing the recommendations from standard-setting panels. The second step was considering common cut scores across grades and subjects. The third step was making recommendations to the commissioner of education after reviewing how well student performance at the cuts reflected the performance level descriptors, the panelists' comments from the standard-setting evaluation forms, and the impact data.

Approved Cut Scores

The final cut score recommendations were approved by the Commissioner of Education for use during the spring 2009 administration. The approved cut score for Commended Performance was 68 out of a possible 72 points. The approved cut score for Met Standard was 44 out of a possible 72 points.



For additional information regarding the 2008–2009 administrations of TAKS–Alt, see the [TAKS–Alt Technical Report 2008–2009](#).

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. 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](#).

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 can be found in [Appendix D](#).

Equating

See [chapter 3](#) for detailed information about the equating process. The distribution of raw scores for all TAKS–Alt assessments can be found 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 2008–2009 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 ratings took place during the 2008–2009 assessment window. The following grades and subjects were included in the study:

- reading grades 5 and 8
- ELA grade 10
- mathematics grades 5, 8, and 10
- science grades 5, 8, and 10

A sample size of approximately 400 students per test was selected to obtain the desired statistics for the study, for a total of approximately 4,300 students. Although approximately 4,300 students were sampled for the study, not all students who were sampled were able to participate in the study. The final sample size was 3,084 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. The second rater 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 on the assessment task.

Correlation coefficients were calculated between the two sets of ratings as shown in Table 22. Complexity Level correlations showed the strongest relationship between rater 1 and rater 2 (0.93 to 0.98). Correlations between the raters were less strong for Demonstration of Skill (ranging from 0.81 to 0.91) and Level of Support (ranging from 0.64 to 0.84) but still indicated a high level of agreement between the two raters. The patterns of correlations across subject areas were quite similar, with the highest correlations for mathematics (0.74 to 0.98) and slightly lower correlations for reading/ELA (0.72 to 0.97) and science (0.72 to 0.97).



**Table 22.** Correlations Between First and Second Ratings

Subject Area	Grade	Complexity Level	Demonstration of Skill	Level of Support	Combined Score	Sample Size
Mathematics	5	0.98	0.89	0.84	0.84	349
	8	0.96	0.91	0.84	0.86	342
	10	0.98	0.83	0.80	0.74	317
Reading/ELA	5	0.97	0.86	0.77	0.82	346
	8	0.96	0.90	0.83	0.86	341
	10	0.94	0.81	0.72	0.72	299
Science	5	0.93	0.86	0.64	0.75	335
	8	0.97	0.89	0.73	0.82	331
	10	0.95	0.83	0.72	0.75	318

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 23. Perfect agreement rates were highest for Complexity Level (95% to 98%) and Demonstration of Skill (94 % to 97%). Perfect agreement rates were slightly lower for Level of Support (73% to 82%) and the Combined Score (73% to 80%). Results were similar across subject areas.

Table 23. Percent Perfect Agreement Between First and Second Ratings

Subject Area	Grade	Complexity Level	Demonstration of Skill*	Level of Support	Combined Score
Mathematics	5	97	95	80	79
	8	96	96	76	75
	10	98	96	76	74
Reading/ELA	5	97	95	77	77
	8	97	97	77	76
	10	95	97	82	80
Science	5	95	96	73	73
	8	96	96	76	75
	10	97	94	78	77

*Demonstration of Skill score component is unweighted

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.



Future interrater reliability studies will allow agreement rates to be monitored over time. Additional interrater reliability analyses can be found in the *2008–2009 TAKS–Alt Technical Report*.

Validity

See [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 validity of evidence were collected for the TAKS–Alt assessments in 2008–2009.

Evidence Based on Test Content

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:

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

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



Educator judgments from educator review meetings also 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 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 13% of TAKS–Alt students (a target sample of 3,000 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 2009 and included the following grades and subjects:

- science grades 5, 8, and 10
- mathematics grades 5, 8, and 10
- writing grades 4 and 7
- reading grades 5 and 8
- ELA grade 10
- social studies grades 8 and 10

Across all meetings, a total of 2,648 student folders were viewed by 91 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 viewed by at least two auditors, with some folders viewed by as many as five auditors.

Within each student folder was information about the four essence statements that were assessed in spring 2009. 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 things: 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. Auditors entered their judgments on an audit judgment form.

Table 24 shows the frequency of responses to all four 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 Demonstration of Skill, while the agreement rates were lower for the statements pertaining to how well the documentation supported the Level of Support 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. These high results indicate that the teachers are selecting the appropriate complexity level for their students.

Table 24. Frequency of Responses to Audit Judgment Form Items

Audit Item	Mean Audit Rating	Frequency	Percent	Total
<i>The documentation supports the student's performance evaluation for demonstration of skill.</i>	Strongly Agree	1277	48.23	2648
	Agree	1118	42.22	
	Disagree	208	7.85	
	Strongly Disagree	45	1.70	
<i>The documentation supports the student's performance evaluation for level of support.</i>	Strongly Agree	1166	44.03	2648
	Agree	1145	43.24	
	Disagree	286	10.80	
	Strongly Disagree	51	1.93	
<i>The documentation supports the student's performance evaluation for generalization of skill.</i>	Strongly Agree	645	37.83	1705
	Agree	777	45.57	
	Disagree	216	12.67	
	Strongly Disagree	67	3.93	
<i>The complexity level of the assessment task either increased, was maintained, or decreased.</i>	Increased	191	7.30	2616
	Maintained	2408	92.05	
	Decreased	17	0.65	

Further training on student documentation may result in higher agreement rates in future audits. Additional training with this focus will be provided for the 2009–2010 school year. Additional audit analyses can be found 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. There are several analyses performed to show that TAKS–Alt scores are related to other variables as expected and are related weakly, if at all, with irrelevant characteristics.



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

Table 25. Overall TAKS–Alt Correlation Between Reading Total Score and Mathematics Total Score

	Mathematics Total	Reading Total
Mathematics Total	1.000	
Reading Total	.828*	1.000

$n = 23,937$

* The above correlation was significant at the $p \leq .05$ level.

Additional validity evidence was also gathered in the form of discriminant validity evidence in 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 shown in the tables below.

Table 26. Overall TAKS–Alt Correlation Between Total Score And Gender

	Total Score	Gender
Total Score	1.000	
Gender	-.024*	1.000

$n = 70,428$

* The above correlation was significant at the $p \leq .05$ level.

Table 27. Overall TAKS–Alt Correlation Between Total Score And Ethnicity

	Total Score	Gender
Total Score	1.000	
Ethnicity	.022*	1.000

$n = 68,492$

* The above correlation was significant at the $p \leq .05$ 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 is having a positive impact on student learning and instruction has been collected through teacher and district test coordinator surveys, educator review meetings, and focus group meetings.

Survey results showed that teachers generally agreed that students would be adequately prepared for the 2008–2009 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.

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 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.



Sampling

See [chapter 3](#) for detailed information about the sampling process. Two samples were selected for research study purposes for TAKS–Alt during the 2008–2009 school year: the interrater reliability sample and the validity audit sample.

Interrater Reliability Sampling

The sampling plan for the TAKS–Alt interrater reliability study was to sample approximately 17% of students ($n=3,600$) taking TAKS–Alt. (Sample numbers were based on an estimate of 21,000 students participating in the assessment in 2007–2008.) 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,
- considered the desired outcomes of the study (correlations, agreement rates, and kappa statistics), and
- accommodated the planned level of aggregation (test level).

The interrater reliability sample of approximately 400 students per aggregation level was selected to meet statistical requirements while taking practical and logistical considerations into account:

- Students who were in grades 4, 7, and 9 for the 2007–2008 TAKS–Alt administration were eligible to be selected for the sample.
- At the elementary and middle school levels, no more than 2 students per campus were selected. This requirement minimized teacher burden at any one campus.
- For elementary and middle schools, campuses for which 2 students would be sampled were placed in the sample first, and then campuses with 1 student were added to the sample to reach the desired sample size. This requirement minimized the number of campuses selected to participate in the study.
- Because there are fewer students assessed at the high school level, up to 3 students per campus were selected to minimize teacher burden at any one campus.
- For high school, campuses for which 2 students would be sampled were added to the sample first, then campuses for which 1 student would be sampled were added to the sample, and finally campuses for which 3 students would be sampled were added to the sample to reach the desired sample size. This requirement minimized the number of campuses selected to participate in the study.

Based on a target of 400 students per test, a random sample of approximately 480 students per test was selected. Sampling more students than the target was designed to produce a sample that would allow for attrition. Sample demographics were



compared to the population demographics for each grade level to verify that the samples were comparable to the population in terms of gender, ethnicity, economically disadvantaged status, and regional representation.

Validity Audit Sampling

For the audit, a sample of 15% of the TAKS–Alt students was targeted. The sample was selected using an early extract (March 2009) from the TAKS–Alt online system. The 15% sample was based on the number of students who took the assessment in 2007–2008. However, the number of TAKS–Alt students increased from 2008 to 2009, resulting in approximately 13% of the students participating in the 2009 TAKS–Alt audit. Thirteen percent of TAKS–Alt students tested in spring 2009 resulted in a sample of approximately 3,000 students. This final sample size was larger than the 10% recommended by the Texas Technical Advisory Committee (TTAC). With 13 tests in the audit, approximately 230 student folders were prepared for each tested grade/subject. The tests included in the audit are listed in Table 28.

Table 28. Grades and Subject Areas Included in the Audit

Mathematics	Reading/ELA	Science	Social Studies	Writing
Grade 5	Grade 5	Grade 5	Grade 8	Grade 4
Grade 8	Grade 8	Grade 8	Grade 10	Grade 7
Grade 10	Grade 10	Grade 10	—	—

For social studies and writing, students were randomly selected from the data extract until the desired number of folders (230 folders x 4 tests = 920 students) was reached. For mathematics, reading/ELA, and science, the audit sample was selected from those students already participating in the TAKS–Alt interrater reliability study. The interrater reliability sample consisted of 4,316 students; approximately half of these students were selected to be included as part of the audit sample (230 folders x 9 tests = 2,070 students). The audit sample was selected to be representative of the population in terms of gender, ethnicity, district size, and region of the state.