

Chapter 5 TAKS–Modified (TAKS–M)



Overview

TAKS–M 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–Modified (TAKS–M) is an alternate assessment based on modified academic achievement standards designed for students receiving special education services who meet participation requirements for TAKS–M. TAKS–M has been designed to meet federal requirements mandated under the No Child Left Behind Act of 2001 (NCLB). According to federal regulations, all students, including those receiving special education services, will be assessed on grade-level curriculum. TAKS–M covers the same grade-level content as the Texas Assessment of Knowledge and Skills (TAKS), but TAKS–M tests have been changed in format (larger font, fewer items per page, etc.) and test design (fewer answer choices, simpler vocabulary and sentence structure, etc.). TAKS–M is administered in English for the same grades and subjects as TAKS, including Student Success Initiative (SSI) retest opportunities. However, successful performance on TAKS–M is not a requirement for graduation. Therefore, it is not considered an exit level test, and there are no grade 11 retest opportunities. Any student who meets participation requirements for TAKS–M may take this assessment but only two percent of the tested population can count as proficient for Adequate Yearly Progress (AYP) performance calculations.

The assessments administered for TAKS–M are illustrated in Table 16.

**Table 16.** 2008–2009 TAKS–M Assessments

2008–2009 TAKS–M Assessments	
Grade	Test Administration
Grade 3	Mathematics (April) and Reading (March, April, June)
Grade 4	Writing (March); Mathematics and Reading (April)
Grade 5	Mathematics (April, May, July); Reading (March, April, July); and Science (April)
Grade 6	Mathematics and Reading (April)
Grade 7	Writing (March); Mathematics and Reading (April)
Grade 8	Mathematics (April, May, July); Reading (March, April, July); Science and Social Studies (April)
Grade 9	Reading (March) and Mathematics (April)
Grade 10	English Language Arts (March); Mathematics, Science, and Social Studies (April)
Grade 11	English Language Arts (March); Mathematics, Science, and Social Studies (April)

Linguistically accommodated testing (LAT) administrations of TAKS–M are available for eligible immigrant ELLs taking grades 3–8 and 10 reading/ELA and mathematics tests and grades 5, 8 and 10 science tests.

TAKS–M Participation Requirements

The participation requirements were developed as a result of recommendations from the TAKS–M steering committee and educator advisory committees to assist admission, review, and dismissal (ARD) committees in determining which students should be assessed with TAKS–M.

TAKS–M has specific participation requirements that must be carefully considered when recommending this assessment for students receiving special education services. All students have the right to be instructed in grade-level Texas Essential Knowledge and Skills (TEKS) curriculum so that they can reach their academic potential. The participation requirements for TAKS–M describe the type of grade-level instruction of the TEKS that a student should be receiving in order to participate in TAKS–M. 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 program (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 a disability that significantly affects academic progress in the grade-level curriculum and precludes the achievement of grade-level proficiency within a school year are assessed with TAKS–M.

An ARD committee may decide that a student’s knowledge and skills in one or more subject areas can best be assessed with TAKS–M if the student meets all of the following participation criteria.

The student

- needs extensive modifications and/or accommodations to classroom instruction, assignments, and assessments to access and demonstrate progress in the grade-level TEKS;
- demonstrates academic progress in such a way that even if significant growth occurs during the school year, the ARD committee is reasonably certain that the student will not achieve grade-level proficiency as demonstrated by multiple valid measures of evidence;
- meets some but not all of the participation criteria of TAKS–Alternate (TAKS–Alt); and
- requires an alternate form of TAKS which is more closely aligned with instructional modifications in order to demonstrate knowledge of the grade-level TEKS.

Test Development

The test development process for TAKS–M follows as closely as possible the procedures used for development of other statewide assessments in Texas, coupled with additional requirements specific to TAKS–M. The blueprints for TAKS–M are aligned to the grade-level TEKS curriculum in the same manner as the TAKS assessments and include the same grade-level content standards as the TAKS blueprints but with fewer items.

For TAKS–M to be meaningfully reported in the accountability system, issues of validity, reliability, fairness, accessibility, and consistency in meaning are carefully considered as a part of the item modification and review processes. As TAKS–M items are developed and reviewed, attention is also given to the standards of fairness and the principles of alignment and universal design.

Using results from a literature review of modifications that are appropriate for students with disabilities who are eligible to be assessed with TAKS–M, the Texas Education Agency (TEA) developed modification guidelines for reading/English language arts, mathematics, science, social studies, and writing to ensure that 1) the modifications were appropriate for the students with disabilities taking TAKS–M, 2) the modifications did not affect the construct of the items, and 3) the item modifications would be consistent across development years.



TEA and Pearson special education, content, and assessment experts modified existing TAKS items. Modifications (for example, simplification of wording, removal of one of the distractors, addition of pre-reading text boxes, and simplification of figures) were made to TAKS items while preserving the construct of the original item.

TEA Review

After the items were modified, additional TEA and Pearson content and assessment experts reviewed the items and verified that they were still aligned to the grade-level TEKS curriculum.

In addition, each item was examined to ensure the adherence to the modification guidelines and absence of bias. This intense review helps ensure that the modifications made to the items make them fair and accessible to the students who take TAKS–M.

Educator Review Committee Meetings

The next step in the development of TAKS–M items is to convene educator review committees. These committees consist of special education teachers who have experience teaching students eligible to take TAKS–M and general education teachers who are knowledgeable about the TEKS curriculum.

Table 17 provides the demographic information of the teachers who participated in the 2009 TAKS–M educator review committees.

Table 17. TAKS–M Educator Review Committees’ Demographic Data

Demographic		Number	Percent
Gender	Female	177	85
	Male	32	15
	Total	209	100
Ethnicity	African American	32	15
	Hispanic	44	21
	White	132	63
	Other	1	1
	Total	209	100
Position	Special Ed	103	49
	General Ed	95	46
	ELL	11	5
	Total	209	100

The educator review committee members compare the original TAKS items and the modified TAKS–M versions of the items to verify that each TAKS–M item maintains the intent of the TEKS measured by the corresponding TAKS item.



Committee members then judge each TAKS–M item for alignment to the student expectation, appropriateness as a measure of student competency, adequacy of student preparation, and any potential bias. Committee members discuss each test item and recommend whether the item should be field-tested as written, revised, or rejected. The range of agreement shown in the compilation of the committees' judgments by grade and subject for the match to the constructs of the TEKS curriculum, appropriateness of the items for the assigned grade level, and elimination of bias is 91.84% to 99.94%.

Evaluating Potential Bias in the Assessment

As with all items, an important concern in the review of modified items is the elimination of bias toward any particular group or disability category. All modified items are reviewed to eliminate language that showed potential bias or is otherwise likely to disadvantage a particular group of students. The modified items are also reviewed for other ways in which bias might appear and unfairly inhibit the performance of any group of students.

During the February 2009 educator review meetings, each participant completed an Item Judgment Form on which individual items were evaluated in response to the question "Is this item free from bias on the basis of personal characteristics such as gender or ethnicity?"

Summaries of the committees' judgments relating each modified item to specific content standards and sub-content standards (TEKS student expectations) strongly demonstrate that the committee members believed TAKS–M items are free from bias and indicate that the TAKS–M items are judged to be fair and accessible to the students eligible to be assessed with TAKS–M.

TEA field-tested all recommended items to collect student responses from all eligible students taking TAKS–M.

Training

The TEA TAKS–M team provided ongoing training from October 2008 to April 2009 to assist regional Education Service Center (ESC) staff, district test coordinators, teachers, administrators, and other interested persons in understanding the TAKS–M assessment program. The presentations included information on federal legislation, changes to the Texas assessment program, test administration, and assessment updates. Training opportunities were conducted via the Texas Education Telecommunication Network (TETN), 2008 ESC Training, and the 2008 Texas Assessment Conference.

The presentations addressed the following topics:

- how to choose the appropriate assessment
- new accommodation policy changes and clarifications
- documenting accommodation use for alternate assessments

- written composition modification guidelines, explanation of TAKS–M rubric, and discussion of TAKS–M sample student papers
- modification guidelines and processes
- sample TAKS–M items

Test Administrations

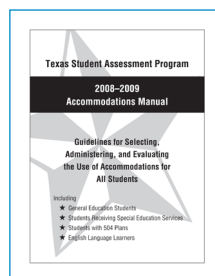
In spring 2009, approximately 398,000 TAKS–M assessments were administered to approximately 147,000 students who met participation requirements.

Spring 2009 represented the first operational administration for grades and subjects that do not fall under federal accountability requirements to demonstrate adequate yearly progress (AYP). These non-AYP grades and subjects include writing (grades 4 and 7), mathematics (grades 9 and 11), reading/English language arts (ELA) (grades 9 and 11), science (grade 11), and social studies (grades 8, 10, and 11). Data from this administration were used to set performance standards for the non-AYP grades and subjects in August 2009. The performance standards were applied retroactively to the spring 2009 test scores.

Testing Accommodations

Accommodations are practices and procedures that provide equitable access to grade-level curriculum during instruction and assessment. The decision to use a particular accommodation with a student eligible to be assessed with TAKS–M is made on an individual basis and takes into consideration both the needs of the student and whether the student routinely receives the accommodation in classroom instruction and testing.

TEA's Accommodations Manual provides guidance to district and campus personnel in selecting, providing, and evaluating the use of accommodations in instruction and assessment. Students eligible for TAKS–M may receive accommodations specified in the Accommodations Manual when certain conditions and eligibility criteria are met.



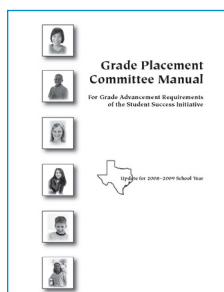
Information about testing accommodations appropriate for TAKS–M administered in 2008–2009 can be found in the [2008–2009 Accommodations Manual](#).



Student Success Initiative

In 1999 the Texas Legislature enacted the SSI, which originally tied grade-level promotion to satisfactory performance on state-mandated assessments in reading at grade 3, and reading and mathematics at grades 5 and 8. (In 2009, the Texas Legislature eliminated SSI grade advancement requirements for grade 3 students. For more information, see [chapter 1](#).)

Students have up to three opportunities to meet the passing standard on the required TAKS tests. Beginning with the 2008–2009 school year, the SSI grade advancement requirements were extended to students taking TAKS–M tests. Students are allowed three testing opportunities to meet the standard. If a student does not meet the standard, a grade placement committee (GPC) is formed to develop an accelerated instruction plan (AIP) and make promotion decisions for the student. For students receiving special education services, the ARD committee functions as the GPC.



More information about SSI requirements for TAKS–M assessments administered in 2008–2009 can be found in the [2008–2009 Grade Placement Committee Manual](#).

Scores and Reports

Description of Scores

For a detailed description of how test scores are derived, see [chapter 2](#).

RAW SCORE

The number of items that a student answers correctly on the TAKS–M assessment is the student's raw score. The raw score can be interpreted only in terms of a specific set of test questions. The difficulty of items may vary between test forms over time. Therefore, differences in student performance across test scores cannot be compared using raw scores. To compare student scores across different test forms and different administrations, raw scores are converted to scale scores.

SCALE SCORE

Unlike raw scores, scale scores allow direct comparisons of student performance across separate test forms and different administrations. TAKS–M raw scores are transformed into a scale that is common to all test forms. This score accounts for differences in the difficulty of the test forms used for each administration.



The scale score can be used to determine whether a student attained Met Standard or Commended Performance. Performance level cut scores are discussed in the “[Standard Setting](#)” section of this chapter.

For a detailed description of raw scores and scale scores, see [chapter 3](#).

Report Formats

Two types of reports are provided for TAKS–M, 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.

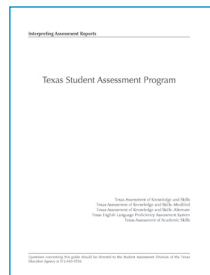
TAKS–M standards were approved for mathematics (grades 3–8 and 10), reading/ELA (grades 3–8 and 10), and science (grades 8 and 10) in August 2008. Standard and optional reports were provided in spring 2009 for these AYP grades and subjects.

Results for non-AYP TAKS–M grades and subjects reported during spring 2009 consisted of raw scores at the objective and subject level. These include writing (grades 4 and 7), mathematics (grades 9 and 11), reading/ELA (grades 9 and 11), science (grade 11), and social studies (grades 8, 10, and 11). Standards for the non-AYP TAKS–M grades and subjects were determined during summer 2009, and Met standard and Commended Performance results were reported in fall 2009.

Standard and Optional Reports for TAKS–M

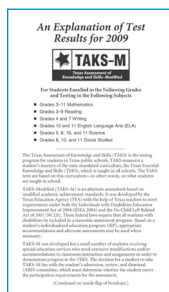
The standard reports available for the 2008–2009 TAKS–M program include the Confidential Student Report, Confidential Student Label, Confidential List of Students’ Results, Confidential Campus Roster—Students Not Meeting Standard, Statewide Summary Report, Demographic Performance Summary, Written Performance Summary Report, and TAKS–M Item Analysis Summary Report.

The optional reports available include the Confidential Electronic Individual Student Record File and Optional Confidential Student Item Analysis Report.



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

Parent Brochure



TEA developed a parent brochure that provides a brief summary of the TAKS–M program and includes a sample CSR with explanations of each element of the report to help parents better understand their child’s score report. The brochure, developed in both English and Spanish, was distributed with individual student results in spring 2009, and in again in fall 2009 when non-AYP grades and subjects were reported.

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 standard-setting process, cut scores (or the number of questions a student must answer correctly) are determined to reflect the level of performance a student must demonstrate to match the performance level descriptors for TAKS–M.

In August 2009, TEA convened two standard-setting panels to recommend cut scores that would define performance standards for TAKS–M. The panels recommended cut scores for the assessment in the following grades and subjects: social studies at grades 8, 10, and 11; and writing at grades 4 and 7. The TAKS–M performance standards were recommended and approved for mathematics, reading, and science in August 2008.

The approach for setting standards was an item-mapping method (Lewis, Mitzel, & Green, 1996), the same approach that was originally used to set performance standards for TAKS. The three-day standard-setting meetings included sessions in which panelists 1) reviewed the test items, 2) became familiar with the performance level descriptors for each assessment, and 3) applied an item-mapping procedure to set recommended cut points. During the item-mapping procedure, panelists reviewed the content assessed by the test items, engaged in table and whole group discussions, and considered the impact on students when making their cut-score recommendations.

Recommended cut scores resulting from the standard-setting meetings were reviewed by TEA and then approved by the Commissioner of Education. Table 18 reports the two cut scores, the total number of points possible, and the percent correct at the cuts. Once approved, these raw score cuts were mapped onto the TAKS–M scale score system with the Met Standard cut point at 2100 and the Commended Performance cut point at 2400.



Immediately following the spring 2009 TAKS–M administration, raw scores were reported to districts for students taking a writing or social studies assessment. After standards were set, a second report was sent out with the performance standards applied to student scores.

Table 18. TAKS–M Cut Scores by Grade and Subject

Grade		Mathematics			Reading/ELA			Science			Social Studies			Writing		
		Raw Score Cut*	Total Points	Percent Correct	Raw Score Cut*	Total Points	Percent Correct	Raw Score Cut*	Total Points	Percent Correct	Raw Score Cut*	Total Points	Percent Correct	Raw Score Cut*	Total Points	Percent Correct
3	Met	15	32	47	14	30	47									
	Comm	27		84	26		87									
4	Met	15	34	44	14	32	44							15	33§	45
	Comm	28		82	28		88							29		88
5	Met	17	35	49	18	34	53	24	32	75						
	Comm	29		83	31		91	30		94						
6	Met	18	37	49	17	34	50									
	Comm	29		78	31		91									
7	Met	17	38	45	19	38	50							24	44±	55
	Comm	29		76	34		89							39		89
8	Met	19	40	48	20	38	53	24	40	60	20	38	53			
	Comm	31		78	35		92	35		88	34		89			
9	Met	19	42	45	18	30	60									
	Comm	34		81	28		93									
10	Met	20	45	44	23	45§	51	25	44	57	23	40	58			
	Comm	36		80	39		87	36		82	36		90			
11	Met	23	48	48	29	45§	64	26	44	59	24	44	55			
	Comm	38		79	42		93	38		86	39		89			

*The cut scores shown are those approved in 2008 for AYP grades/subjects and those approved in 2009 for non-AYP grades/subjects. For subsequent administrations, shifts may occur in the number of items (raw score) needed to achieve Met Standard and Commended Performance.

§For the grade 4 writing test and the grades 10 and 11 English language arts tests, the written composition prompt is rated on a scale of 1–3 and is equal to a maximum of nine score points (written composition rating x 3).

±For the grade 7 writing test, the written composition prompt is rated on a scale of 1–3 and is equal to a maximum of twelve score points (written composition rating x 4).



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. As with many of the other programs in the Texas assessment program, the TAKS–M assessment program uses the Rasch Partial-Credit Model (RPCM) to place test items on the same scale across administrations for a given TAKS–M assessment. Once performance standards have been set for an assessment, its initial scale is then transformed to a more user-friendly metric to facilitate interpretation of the test scores. Details of the RPCM scaling method used in Texas are provided in [chapter 3](#).


Scale Score

A scale score is a conversion of the raw score onto a “scale” that is common to all test forms for that assessment. Scale scores allow direct comparisons of student performance between specific sets of test questions from different test administrations.

After the August 2008 standard-setting recommendations, a unique scale transformation was applied such that the resulting scale scores have the recommended cut score for the Met Standard performance level at a scale score of 2100 and the recommended cut score for the Commended Performance level at a scale score of 2400 (see the “[Standard Setting](#)” section of this chapter). The linear transformation of the underlying Rasch proficiency level estimate is as follows:

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

where SS_j is the scale score for student j , θ_j is the Rasch partial credit model proficiency level estimate for student j , and $T1$ and $T2$ are scale score transformation constants that establish the scale score system such that a scale score of 2100 is the cut score for the Met Standard performance level and a scale score of 2400 is the cut score for the Commended Performance level. Values for $T1$ and $T2$ are provided in Table 19 for TAKS–M.

Table 19. Scale Score Transformation Constants for TAKS–M


Grade	Mathematics		Reading/ELA		Science		Social Studies		Writing	
	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2
3	149.3503	2121.4766	135.1047	2120.5765	—	—	—	—	—	—
4	160.3592	2139.7691	132.0306	2133.7866	—	—	—	—	94.6522	2108.8027
5	167.6821	2111.8048	125.1617	2084.5551	178.6033	1885.0509	—	—	—	—
6	201.3558	2111.6585	120.7730	2100.5193	—	—	—	—	—	—
7	203.7767	2147.3169	127.9591	2100.3581	—	—	—	—	114.7491	2062.1558
8	211.6104	2123.1713	119.6554	2086.1439	184.7404	2018.9544	140.2262	2083.9441	—	—
9	173.7116	2134.4470	128.2709	2043.1888	—	—	—	—	—	—
10	179.3508	2141.1430	134.7285	2077.4330	234.1555	2032.6803	150.7083	2048.5934	—	—
11	200.2269	2117.0794	117.6194	2004.0461	188.1822	2021.1705	153.4841	2069.9939	—	—

Equating

Used in conjunction with the scaling process, equating is the process that “balances” the slight difficulty differences across test forms and administrations to place the scores onto a common scale. Through the use of sophisticated statistical methods, the results of different tests are equated to enable the comparison of scale scores across test forms and test administrations.

A common-items/nonequivalent groups design is used for all TAKS–M grades and subjects. All students meeting the TAKS–M participation requirements are asked to participate in the field test that contains common linking items across all forms. These linking items only appear on the field test and are not present on operational tests. The field-test administrations are used to 1) establish linking items for future field tests, 2) provide item statistics (p-values, point biserial correlations, and Rasch item difficulties) to inform decisions about which items to use on the operational tests and the sequence of those items, and 3) to pre-equate the operational tests forms for all administrations occurring after the initial operational administration.

The TAKS–M base scale was established using data from the first operational administration of the assessment, which was spring 2008 for the AYP grades and subjects with the remaining non-AYP grades and subjects being administered operationally in spring 2009. Once the base scale was set, the item statistics of the linking items appearing on the field test (and therefore on the field-test scale) were moved over to the base test scale. Then these linking items will appear on the next TAKS–M field test (fall 2009) and, beginning with



the spring 2010 administration, will be used to pre-equate all future operational TAKS–M test forms to the base scale. Post-equating is not done for TAKS–M in order to facilitate reporting on the same schedule as TAKS.

In spring 2009, calibrations were performed for some of the TAKS–M assessments. These were the assessments that were being administered operationally for the first time. However, future equating for TAKS–M will follow a pre-equated model in which raw score to scale score (RSSS) tables are established before the operational administration using field-test item statistics. That table is then applied to the operational test administration results and no calibration of the operational test is necessary.

Reliability

The concept of reliability is based on the idea that repeated administrations of the same test should generate consistent results about student performance. Reliability is a critical technical characteristic of any measurement instrument, because unreliable instruments cannot be interpreted in a valid way. During the 2008–2009 school year, reliability estimates for TAKS–M assessments were conducted through internal consistency, classical standard error of measurement, conditional standard error of measurement, and classification accuracy.

Internal Consistency

Internal consistency is a measure of the consistency with which students respond to the items within a test. For tests involving dichotomously scored (multiple-choice) items, the Kuder-Richardson Formula 20 (KR20) was used to calculate the reliability estimates; and for tests involving a combination of dichotomous and a polytomous (extended response) items, the stratified coefficient alpha was used to calculate the reliability estimates. As a general rule, reliability coefficients from 0.70 to 0.79 are considered adequate, 0.80 to 0.89 are considered good, and above 0.90 are considered excellent. However, appropriate levels of reliability depend on how an assessment is being used. Reliability values for TAKS–M ranged from 0.71 to 0.88. Nunnally & Bernstein (1994) recommend reliabilities of 0.70 and above for assessments in the early stages of development. Since TAKS–M is a new assessment and will be used in conjunction with other criteria to make student-level decisions, reliabilities of 0.70 and above are an acceptable starting point for TAKS–M field-test reliability estimates. The internal consistency values for each TAKS–M assessment can be found in [Appendix C](#).

Classical Standard Error of Measurement

Classical standard error of measurement (SEM) provides a reliability estimate for a test score. The SEM represents the amount of variance in a test score resulting from factors other than achievement. The SEM is helpful for quantifying the margin of uncertainty that occurs on every test. For example, factors such as chance error, differential testing conditions, and imperfect test reliability can cause a student's observed score (the



score achieved on a test) to fluctuate above or below his or her true score (the true proficiency of the student). SEM values for TAKS–M ranged from 2.099 to 3.268. The SEM values for TAKS–M are provided in [Appendix C](#).

Conditional Standard Error of Measurement

Conditional standard error of measurement (CSEM) provides a reliability estimate at each score point on a test. CSEM provides an estimate of the average test score measurement error conditional on the proficiency estimate or scale score estimate. CSEM values for TAKS–M ranged from 36 to 237. The CSEM values for TAKS–M can be found in [Appendix C](#).

Classification Accuracy

Classification accuracy provides an estimate of the accuracy of student classifications into performance categories based on current test results. Classification accuracy rates for TAKS–M ranged from 80.8 to 93.2. The classification accuracy rates for TAKS–M are provided in [Appendix C](#).

Validity

Validity refers to the extent to which the test measures what it is intended to measure. Validity in the Texas assessment program is concerned with the general question of whether or not test scores will help educators to make appropriate judgments about student performance. Validity evidence for an assessment can come from a variety of sources including test content, the response process, the internal structure, relationships with other variables, and the consequences of testing. Texas collects validity evidence annually to support the various uses of TAKS–M scores. The sections that follow describe how these types of validity evidence were collected for the TAKS–M assessments in 2008–2009.

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. For TAKS–M, test results are used to make inferences about students' knowledge and understanding of the TEKS. Standards-referenced assessments, such as TAKS and TAKS–M, are based on an extensive definition of the content they assess. Test validity is therefore content based and tied directly to the statewide curriculum. Because TAKS–M is a modified version of TAKS, the test development processes for both assessments play an intricate role in building validity evidence. To achieve the highest level of content validity, the process of aligning both TAKS and TAKS–M to the curriculum was carefully approached and included review by numerous committees of Texas educators.



When TAKS was designed as the standards-referenced general assessment to measure the TEKS, advisory committees consisting of educators from school districts across the state were formed for each subject area at each grade level. Teachers, test development specialists, and TEA staff members worked together in these committees to identify the TEKS student expectations that were important to assess and to develop test objectives, item development guidelines, and test-item types. In addition, committees met to review and edit TAKS items for content and bias and to review field-test data.

The item writers as well as reviewers for each stage of development verified the alignment of test items with the objectives to ensure that the items measure appropriate content. The sequential stages of item development and item review provide many opportunities for Texas educators to offer suggestions for improving or eliminating items and to offer insights into the interpretation of the statewide curriculum.

When TAKS–M was designed as the alternate assessment based on modified achievement standards, special education content specialists developed detailed modification guidelines so that the modifications made to the TAKS items were consistent. After the items were modified, educator committees for each subject area at each grade level reviewed the original TAKS item and the modified TAKS–M version of the item to make sure that the modified item still measured the same underlying skill as the original item. In this way, the alignment between the TEKS curriculum and the TAKS items carries through to the TAKS–M items.

Evidence Based on Response Processes

Response processes refer to the cognitive behaviors required to respond to a test item. Texas collects evidence that the way in which students respond to test questions on the TAKS–M assessments supports the accurate measurement of the construct.

TAKS–M includes item (or question) types that require students to respond in various ways. These item types include: selected-response items (both stand-alone items and passage-related items) and constructed-response (or short-answer) items. Theoretical and empirical evidence has been gathered to suggest that the way in which students respond to these types of questions does not add construct-irrelevant variance.

The evidence the Texas assessment program gathers comes from several sources. When item types were initially modified for TAKS–M, the items were reviewed by educator committees to make sure that the modifications made the items accessible to the TAKS–M student population. In addition, educator review of the items is done to gather evidence that the response processes do not advantage or disadvantage certain student groups (for example: males or females, different ethnic groups, and different disability groups). The process for the review of item content involves: 1) an evaluation by educators that the content assessed by the item is appropriately assessed with the planned item type; and 2) a judgment by educators that students will be able to accurately demonstrate their knowledge of the content by responding



to each item in its planned format. When items are field-tested, data are gathered about students' responses to items, and statistical information—such as item difficulty and item point-biserial correlations—is evaluated taking item type into consideration.

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. Measures of internal consistency are used to provide validity evidence based on internal structure. Internal consistency measures show to what degree responses to items measuring the same or a similar content are related. Two measures of internal consistency, the Kuder-Richardson Formula 20 and the stratified coefficient alpha, were used for TAKS–M. These two consistency measures also provide reliability evidence for the TAKS–M tests. As a result, the internal consistency evidence for TAKS–M can be found in the ["Reliability"](#) section of this chapter under the Internal Consistency results.

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 are done to support that TAKS–M assessments and item scores are related to outside variables as intended and are weakly related, if at all, to irrelevant characteristics.

The correlations among TAKS–M reading and mathematics scores were calculated resulting in a moderate correlation. This indicates that the scores are related but not redundant, which is to be expected since the two constructs are both academic subjects but assess different skills.

The correlations among total test score and the TAKS–M objective scores were also calculated. The magnitudes of these correlations were found to support theoretical relations between objectives and the overall test.

Additional validity evidence was collected in the form of discriminant validity evidence in analyses demonstrating that the TAKS–M 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.

Evidence Based on Consequences of Testing

Another way to provide validity evidence is by documenting the intended and unintended consequences of administering an assessment. In the 2008–2009 school year, TEA formally captured and documented the intended and unintended consequences of the TAKS–M assessment program by



administering a consequential validity survey to educators. The consequential validity survey allowed educators to document the extent to which they believed the administration of TAKS–M has led to changes in certain areas, such as student achievement, impact on teachers, and curricular and instructional reform. In addition to documenting educators’ beliefs about the extent of changes due to TAKS–M, the survey also allowed educators to document their extent of agreement with statements about the consequences of TAKS–M.

A guided discussion approach was used to generate a list of the possible intended and unintended consequences of administering TAKS–M from special education experts at internal meetings with TEA. The notes taken during these discussions were analyzed to identify recurring themes, and these themes were used to generate survey questions. The consequential validity survey was administered to both general and special education teachers during meetings for which they were selected to represent the diversity of Texas teachers.

Overall, the educators reported more positive changes due to the administration of TAKS–M. The majority of educators responded that the emphasis placed on the curriculum had increased due to the administration of TAKS–M, the teaching of higher-order thinking skills had increased, and that attention given to the lowest performing students and students performing in the middle ranges had also increased. Educators also reported that the inclusion of TAKS–M students in general education classrooms had increased, there was more cooperation between the general and special educators, professional development opportunities were increasing, and campuses and districts provided teachers with the resources they needed to teach effectively.

Although educators felt that TAKS–M was helpful for students and teachers, they also felt that it was more burdensome to prepare students to take the test and that the public did not perceive TAKS–M as being helpful to the students. Despite these responses, almost half of the educators agreed that students are responding positively to taking TAKS–M.

Sampling

Sampling is a procedure to select a smaller number of observations (in this case, Texas students) that are representative of the entire body of Texas students. The results from well drawn samples allow TEA to estimate characteristics of the larger population of Texas.

Sampling plays a critical role in the research and annual development activities necessary to support the Texas assessment program. The assessment program affects all students (or the population of students) in Texas. A sample is a group of students smaller than the population that can be used to represent the overall population. Through the careful selection of student samples, TEA is able to gather reliable information about student performance on its tests while minimizing campus and district participation. In particular, sampling is used in the Texas assessment program for: 1) testing that is part of a research study, and 2) stand-alone field tests.



In 2008–2009, there was no stand-alone field testing for TAKS–M. However, when stand-alone field testing is done for TAKS–M, census field testing is used. This method tests the entire student population instead of testing a sample of the population. Census field testing is used for this population of students because data are needed from all students to obtain a large enough number of students to conduct field-test analyses. In addition to no field-test sampling, there was also no sampling for special studies because no special studies were conducted for TAKS–M this year.