Chapter **5** # TAKS–Modified (TAKS–M)

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 2 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. 2009–2010 TAKS–M Assessments

2009–2010 TAKS–M Assessments	
Grade	**Test Administration**
Grade 3	Mathematics and Reading (April)
Grade 4	Writing (March); Mathematics and Reading (April)
Grade 5	Mathematics and Reading (April, May, June); and Science (April)
Grade 6	Mathematics and Reading (April)
Grade 7	Writing (March); Mathematics and Reading (April)
Grade 8	Mathematics and Reading (April, May, June); 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 students receiving special education services who are 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 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 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. Within the principles of universal design, each item has precisely defined constructs, has maximum legibility, has maximum readability and comprehensibility, is amendable to accommodations, is accessible and non-biased, and considers special populations.

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) modified existing TAKS items and 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.

Because the items for the TAKS–M assessments are developed on a three-year cycle, no item development occurred during the 2009–2010 year. The TAKS items that were released in 2010 were modified for the TAKS–M release.

Training

The TEA TAKS–M team provided ongoing training from August 2009 to April 2010 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), 2009 ESC Training, and the 2009 Texas Assessment Conference.

The presentations addressed the following topics:

- how to choose the appropriate assessment
- accommodation policy 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 2010, all TAKS–M grades and subjects had an operational administration. Approximately 372,000 TAKS–M assessments were administered to approximately 153,000 students who met participation requirements. Districts administered the TAKS–M assessments to eligible students as indicated below.

Spring/Summer 2010

March:
- Writing at grades 4 and 7
- Reading at grade 9
- ELA at grades 10 and 11

April:

- Mathematics at grades 3–11
- Mathematics at grades 3, 4, 6, 7, and 10 (LAT)
- Reading at grades 3–8
- Reading at grades 3, 4, 6, and 7 (LAT)
- ELA at grade 10 (LAT)
- Science at grades 5, 8, 10, and 11
- Science at grades 5, 8, and 10 (LAT)
- Social Studies at grades 8, 10, and 11

May:

- Mathematics at grades 5 and 8 (LAT)
- Mathematics retests at grades 5 and 8
- Reading at grades 5 and 8 (LAT)
- Reading retests at grades 5 and 8

June:

- Mathematics retests at grades 5 and 8
- Reading retests at grades 5 and 8

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.
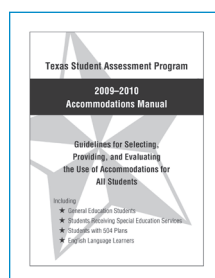
Information about testing accommodations appropriate for TAKS–M administered in 2009–2010 is available in the *2009–2010 Accommodations Manual*.

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, refer to 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.
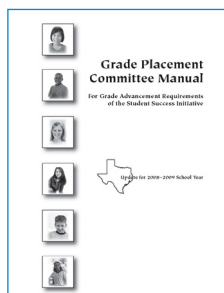
More information about SSI requirements for TAKS–M assessments administered in 2009–2010 is available in the *2009–2010 Grade Placement Committee Manual*.

Scores and Reports

Description of Scores

For a detailed description of how test scores are derived, refer to 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, refer to chapter 3.
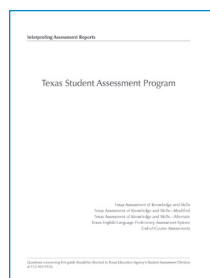
Report Formats

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.

Standard and optional reports were provided in spring 2010 for all grades and subjects.

Standard and Optional Reports for TAKS–M

The standard reports available for the 2009–2010 TAKS–M program include the Confidential Student Report (CSR), 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.
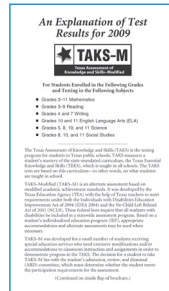
The optional reports available include the Confidential Electronic Individual Student Record File and Optional Confidential Student Item Analysis Report.

More information about scoring and reporting for TAKS–M is available in the *2009–2010 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 again in fall 2010.

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.

The standards used to define student performance for TAKS–M assessments in 2009–2010 were set in August 2008 and August 2009. The standards set in August 2008 were for the AYP grades and subjects, and the standards set in August 2009 were for the non-AYP grades and subjects. In August 2008, panels of educators were convened to recommend cut scores which were then reviewed by TEA and later approved by the Commissioner of Education. The same process occurred for the non-AYP grades and subjects in August 2009. A description of the standard setting process and the approved cut scores for the AYP grades and subjects is available in chapter 9 of the *2007–2008 TAKS–M Technical Report*. A description of the standard-setting process and the approved cut scores for the non-AYP grades and subjects is available in chapter 5 of the *2008–2009 Technical Digest*.

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 and 2009 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 (refer to 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 17 for TAKS–M.

Table 17. 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. In the 2009–2010 school year, field-test equating and pre-equating activities were conducted for TAKS–M for all grades and subjects.

Field-Test Equating

To replenish the item bank in order to create new tests, newly developed items must be field-tested and equated to the scale of the original form as discussed in the technical details and procedures in chapter 3. This process is called field-test equating. TAKS–M uses a stand-alone census field test design for all its assessments. This type of design is used to obtain a sufficient number of students' data to conduct field-test analyses and to adhere to the assessment's reduced test length design.

TAKS–M field tests contain common linking items across all forms. This is called a common-items/nonequivalent groups design. With this design, linking items only appear on the field test and are not present on operational tests. It is the operational tests that are used to set the base scale for TAKS–M. 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. Field tests for the AYP grades and subjects took place in fall 2007, and field tests for the non-AYP grades and subjects took place in spring 2008.

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 onto the base test scale. These linking items from the 2007 and 2008 field-test administrations were then put onto the fall 2009 field-test with the newly developed items in order to place the item statistics of the new items onto the base scale. The newly equated item statistics (p-values, point biserial correlations, and Rasch item difficulties) were then used to inform decisions about which items to use on the operational tests and the sequence of those items. Once the new tests were created, those test forms were then pre-equated.

Pre-Equating

The pre-equating process is one in which a newly developed test form is linked, before it is administered, to a set of items that appeared previously on one or more test forms. In essence, the raw score to scale score (RSSS) table for each test is established by using field-test item statistics equated to the base test scale before the operational administration of the test. The pre-equated RSSS table that is created is then applied to the operational test administration results and no calibration of the operational test is necessary. By using this process, the difficulty level of the newly developed test form is known prior to

its administration, and the anticipated raw scores that correspond to scale scores at performance standards can be identified. The pre-equating process is used for TAKS–M in order to facilitate reporting on the same schedule as TAKS.

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 2009–2010 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 0.76 to 0.90. This year's range of reliability coefficients increased from last year's range of 0.71 to 0.88, indicating greater reliability. TAKS–M will be used in conjunction with other criteria to make student-level decisions; therefore, reliabilities of 0.70 and above are an acceptable starting point for TAKS–M reliability estimates. The internal consistency values for each TAKS–M assessment is available 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 1.948 to 3.201 (in raw score points). 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 37 to 237 scale score points across all TAKS–M grades and subjects. In general, the CSEM values occurring in the middle of the score range for mathematics ranged from 54 to 69, for reading the CSEM values ranged from 37 to 52, and for science the CSEM values ranged from 58 to 72. The CSEM values in the middle of the score range for writing were 37 and 39 scale score points, while the CSEM values in the middle of the score range for social studies were 47 and 48 scale score points. The CSEM values for TAKS–M is available 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 77.7 to 89.9. 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 2009–2010.

Evidence Based on Test Content

Validity evidence based on test content refers to evidence of the relationship between tested content and the construct the test is intended to measure. 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 students' knowledge and understanding of the materials in 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 to demonstrate 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 essay 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 is available 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.

Correlations among the scale scores of TAKS–M subjects were calculated. Correlation between the TAKS–M reading and mathematics scale scores were calculated resulting in a moderate correlation of 0.506. 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 other subject scores had similarly moderate correlations. Science and math had a correlation of 0.587, writing and reading had a correlation of 0.612, the science and reading correlation was 0.608, social studies and reading was 0.597, and social studies and mathematics resulted in a correlation of 0.533. Social studies and science had the strongest correlation at 0.706, while the correlation between mathematics and writing scores was the weakest among the subjects at 0.457. This overall range of moderate correlations suggests that all the scores among subjects are related and neither redundant nor irrelevant.

The correlations between the total test score and the TAKS–M objective scores were also calculated within grade and subject. Across all subjects and grades, the correlations between each objective and test score ranged from 0.410 to 0.929. The magnitudes of these correlations were found to support theoretical relations between objectives and the overall test. More specifically, the range of correlations within reading across all grades was 0.717 to 0.925. For mathematics, the range was 0.410 to 0.824. Science had a correlation range of

0.553 to 0.879, while social studies correlations ranged from 0.623 to 0.850. Lastly, the correlations of objective scores to total test scores for writing ranged from 0.644 to 0.834.

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 students' performance on the assessment; therefore, the lack of meaningful empirical relationships between these measures is expected and is reflected in the overall results of the correlation calculations between total score and gender as well as between total score and ethnicity. The correlations are 0.005 regarding gender and 0.068 regarding ethnicity.

Evidence Based on Consequences of Testing

Another way to provide validity evidence is by documenting the intended and unintended consequences of administering an assessment. Validity evidence showing the impact of administering the TAKS–M assessment was collected through educator surveys during the 2008–2009 school year.

In 2008–2009 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.

Measures of Student Progress

In 2009–2010 the Texas Projection Measure (TPM) for TAKS–M was implemented. The TPM uses a multilevel, regression-based projection model that estimates whether the student is likely to meet the standard (pass) and/or achieve commended performance (obtain the highest performance level) on TAKS–M tests at a future grade (grade 5, 7 [writing only], 8, or 11). This measure is based on (1) a student's current performance on TAKS–M and (2) the TAKS–M scores of other students in the same enrolled grade in the student's school district. As with TAKS and TAKS (Accommodated), the TPM was used as a measure to assist campuses in meeting Adequate Yearly Progress (AYP) for federal reporting purposes and as a criterion for evaluating campuses in the state accountability system.

The TPM implemented for TAKS–M is designed to be as similar as possible to the model used with TAKS. Both measures require at least two years of data to develop equations and require the equations to be developed the year prior to implementing them. Projection accuracy will also be evaluated in the same manner for both measures. The TPM for TAKS–M, however, does differ from the TPM for TAKS in that a TPM is reported only for students who are assessed with TAKS–M in all the subject areas needed to implement the TPM. Also, for TAKS–M, a district mean predictor is used when developing the TPM equations rather than a campus mean predictor. For a full description of the process used to develop the TPM for TAKS–M, refer to "Procedures for Developing the Texas Projection Measure Equations for TAKS–M" on the Texas Projection Measure for TAKS–M page on TEA's Student Assessment Division website.

For the 2009–2010 school year, the TPM was reported for the first time for TAKS–M grades 4, 7, and 10 reading/English language arts (ELA) and mathematics, and grade 10 science. These were the only subjects and grades for which two years of data were available in 2009, which is the year before the TPM was implemented. To provide sufficient time for data collection, the remaining grades and subjects would be phased in over time. Because the TPM for those seven grades and subjects was first reported and used in 2010, the accuracy of the 2010 projections cannot be evaluated until 2011 performance data become available. There are data available, however, indicating the number of students who were projected to Meet the Standard or achieve Commended Performance. The numbers of students assessed with TAKS–M who were projected to Meet the Standard or achieve Commended Performance are shown on the next page, in Table 18. The results show that

between 72% and 89% of all students assessed with TAKS–M in the grades for which there was a TPM received a projection. Of those students with a projection, between 37% and 89% were projected to Meet the Standard, and between 0.1% and 15% were projected to achieve Commended Performance. Once performance data become available, TEA will evaluate the accuracy of these projections.

Table 18. TAKS–M Projection Classifications for 2009–2010, All Students

Grade/ Subject	Total Tested (Pct. Relative to All Students Tested)	Students With/Without Projections (Pct. Relative to All Students Tested)		Projection Classifications (Pct. Relative to Students with Projections Group)		Projection Classifications (Pct. Relative to Students with Projections Group)	
		Students With Projections	Students Without Projections	Projected to Meet Standard	Projected Not to Meet Standard	Projected to Achieve Commended Performance	Projected Not to Achieve Commended Performance
Grade 4 to Grade 5 Reading	14134 (100.00)	11849 (83.83)	2285 (16.17)	10601 (89.47)	1248 (10.53)	830 (7.00)	11019 (93.00)
Grade 4 to Grade 5 Mathematics	13577 (100.00)	11849 (87.27)	1728 (12.73)	10443 (88.13)	1406 (11.87)	1732 (14.62)	10117 (85.38)
Grade 7 to Grade 8 Reading	13650 (100.00)	12183 (89.25)	1467 (10.75)	9925 (81.47)	2258 (18.53)	649 (5.33)	11534 (94.67)
Grade 7 to Grade 8 Mathematics	15198 (100.00)	12183 (80.16)	3015 (19.84)	8202 (67.32)	3981 (32.68)	284 (2.33)	11899 (97.67)
Grade 10 to Grade 11 ELA	11011 (100.00)	9739 (88.45)	1272 (11.55)	7101 (72.91)	2638 (27.09)	620 (6.37)	9119 (93.63)
Grade 10 to Grade 11 Mathematics	13574 (100.00)	9739 (71.75)	3835 (28.25)	4966 (50.99)	4773 (49.01)	13 (0.13)	9726 (99.87)
Grade 10 to Grade 11 Science	11571 (100.00)	8891 (76.84)	2680 (23.16)	3321 (37.35)	5570 (62.65)	19 (0.21)	8872 (99.79)

NOTE: Numbers in parentheses indicate percentages.

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 2009–2010 stand-alone census field testing was used for TAKS–M. TAKS–M uses a stand-alone census field-test design to obtain a sufficient number of students to conduct field-test analyses and to adhere to the assessment's reduced test-length design. However, there was no sampling for special studies because no special studies were conducted for TAKS–M this year.