State of Texas Assessments of Academic Readiness Bridge Study for AYP

Table of Contents

Executive Summary

During the transition to a new assessment system, the State of Texas Assessments of Academic Readiness (STAAR), performance standards will not be available for most STAAR assessments in time to assign Adequate Yearly Progress (AYP) statuses in summer 2012. The Texas Education Agency (TEA) proposes a bridge study to facilitate evaluating AYP in 2012.

- A bridge study will identify the *Met Standard* performance standard for the previous assessment system, the Texas Assessment of Knowledge and Skills (TAKS), on the STAAR assessments in 2012.

- Students who meet the bridged TAKS *Met Standard* performance standard on STAAR assessments will be counted as proficient for AYP purposes in 2012.

- The bridge study methods, as supported by the Texas Technical Advisory Committee (TTAC), vary by test type and content area (in some cases) due to different testing conditions for the assessments and the availability of student data.

- Content alignment, or overlap, analyses will determine whether there is sufficient shared content between the two assessments so that mapping the TAKS *Met Standard* performance standards onto the STAAR assessments will result in a meaningful interpretation.

- The empirical analyses will statistically map the TAKS *Met Standard* performance standard to the STAAR 2012 assessments using student performance data. The impact data analyses can provide supplemental data to support the results from the empirical analyses.

- The bridge study will only be used for AYP evaluations in spring 2012 while Texas transitions from TAKS to STAAR. In spring 2013, AYP will use the performance standards for STAAR for AYP evaluations.

State of Texas Assessments of Academic Readiness Bridge Study for AYP

Overview

The Texas Education Agency (TEA), in collaboration with the Texas Higher Education Coordinating Board (THECB) and Texas educators, is developing a new assessment system, the State of Texas Assessments of Academic Readiness (STAAR), in response to requirements set forth by the 80th and 81st Texas legislatures. This new system will focus on increasing postsecondary readiness of graduating high school students and helping to ensure that Texas students are competitive with other students nationally and internationally. The STAAR program, similar to the Texas Assessment of Knowledge and Skills (TAKS), includes general education, modified, alternate, and linguistically accommodated assessments.

STAAR will replace the Texas Assessment of Knowledge and Skills program [TAKS, TAKS (Accommodated), TAKS–Modified (TAKS–M), TAKS–Alternate (TAKS–Alt), and TAKS Linguistically Accommodated Testing (LAT)] beginning in spring 2012. To facilitate the evaluation of Adequate Yearly Progress (AYP) in 2012, Texas plans to conduct a bridge study that will identify the existing TAKS *Met Standard* performance standard used for AYP evaluations on the STAAR assessments. Since STAAR performance standards will not yet be available for the majority of the tests until late fall 2012, performance standards used with the TAKS assessments will be carried over to the STAAR program for the 2012 AYP evaluations that will be released in early August 2012.

AYP evaluations include students taking assessments in reading/ELA and mathematics in grades 3–8 and 10. In 2011–2012, students taking general, linguistically accommodated, modified, or alternate assessments in grades 3–8 or alternate assessments in grade 10 will take STAAR 3–8, STAAR L, STAAR end-of-course (EOC), STAAR Modified, or STAAR Alternate assessments. Some students enrolled in grade 8 or lower will be taking advanced courses in which the STAAR EOC assessments are required for graduation. Therefore, some of the STAAR EOC assessments will be included in the bridge study in order to have continuity in the AYP calculations based on TAKS performance standards. For STAAR assessments, the TAKS performance standards will be bridged to the STAAR assessments. Students will be considered passing on the STAAR assessments for AYP purposes only if those students meet the TAKS performance standard mapped onto the STAAR assessment.

Students taking general or modified assessments in grade 10 in 2011–2012 will be administered the TAKS or TAKS–M assessments (including accommodated and linguistically accommodated forms) which have defined performance standards. Therefore, a bridge study is not needed for grade 10 for these assessments. Table 1 lists the 2011–2012 assessments and corresponding performance standards proposed for 2012 AYP evaluations.

Table 1. Assessments and Performance Standards for 2011–2012 AYP Evaluations

Enrolled Grades	Assessments in 2011–2012	Performance Standard Planned for AYP Calculations
Grade 3–8	STAAR reading and mathematics*	Bridged to TAKS *Met Standard*
	STAAR EOC English I reading and Algebra I for enrolled grades 8 and lower	Bridged to TAKS *Met Standard* for grade 9 reading and mathematics
	STAAR Modified reading and mathematics	Bridged to TAKS–M *Met Standard*
	STAAR Alternate reading and mathematics	Bridged to TAKS–Alt *Met Standard*
Grade 10	TAKS ELA and mathematics*	TAKS *Met Standard*
	TAKS–M ELA and mathematics*	TAKS–M *Met Standard*
	STAAR Alternate English I, Algebra I, English II, and Geometry	Bridged to TAKS–Alt *Met Standard* for grades 9 and 10 reading/ELA and mathematics

*Includes English, Spanish, accommodated and linguistically accommodated assessments, where applicable

Bridge Study Process

The STAAR bridge study consists of three stages: content overlap analysis, empirical analysis, and impact data analysis. Other states, such as Georgia, New Jersey, and New Hampshire, have implemented bridge studies while transitioning to a new assessment or calculating safe harbor (Erpenbach, 2008; Erpenbach 2011; Forte & Erpenbach, 2006). The methods proposed for the Texas bridge study are supported by the Texas Technical Advisory Committee (TTAC) and are consistent with other approved procedures and methodologies (Erpenbach, 2008; Erpenbach 2011; Forte & Erpenbach, 2006). Appendix 1 provides a timeline for completing the three stages and implementing the bridge study results in 2012 AYP evaluations.

Stage 1. Content Overlap Analysis

The STAAR program at grades 3–8 will assess the same grades and subjects as are assessed on TAKS. For high school, grade-level subject-area TAKS tests will be replaced with twelve STAAR end-of-course (EOC) assessments. Content alignment, or overlap, analyses are needed to determine whether there is sufficient shared content between the two assessments so that mapping the TAKS *Met Standard* performance standards onto the STAAR assessments would support a meaningful interpretation. The content overlap studies have two major components:

1. Content standard-level comparisons (curriculum overlap, as represented by student expectation [SE] overlap); and
2. Test-level comparisons (purpose, assessment type, administration, item formats, test blueprints, number of items, and performance levels).

The first component is a detailed review of the content standards (as defined by the Student Expectations or SEs from the Texas Essential Knowledge and Skills curriculum standards) for the course measured by STAAR. The SEs represented on the STAAR assessments are listed along with those that overlap with the SEs assessed on TAKS. An overall percentage of overlap across the two assessments is also shown. As an indication of non-overlap, those SEs represented on TAKS and not overlapped on STAAR are delineated. In addition to this quantitative component, a second component is included that shows an overall qualitative summary of the two tests. Detailed analysis

sheets containing information for both of these components for each set of compared assessments will be prepared using templates such as the one included in Appendix 2. An overall evaluation of the content similarity between the TAKS and STAAR assessments will be captured in a summary rating of good, moderate, or weak. The content overlap analyses are expected to confirm that the tests evaluated for AYP bridge study purposes have either good or moderate content similarity such that bridging will be appropriate.

Stage 2. Empirical Analysis

The empirical analysis stage statistically maps the TAKS *Met Standard* performance standard to the STAAR 2012 assessments using student performance data. Different empirical methods, as supported by the Texas Technical Advisory Committee (TTAC), will be conducted based on the availability of student data. For STAAR 3–8, student data were collected in spring 2011 by embedding STAAR field-test items in the TAKS operational assessments (see Appendix 3 for technical details). For STAAR EOC, student data were collected in spring 2011 by administering both TAKS and STAAR assessments to the same students (see Appendix 3 for technical details). For STAAR Modified and STAAR Alternate, the first administration of STAAR items will not occur until spring 2012. For the modified assessments, the empirical methods will match 2011 TAKS–M student data to 2012 STAAR Modified student data (see Appendix 4 for technical details). For the alternate assessments, the scoring rubrics for TAKS–Alt and STAAR Alternate assessments will be applied to STAAR Alternate 2011–2012 student data to be evaluated for classification accuracy (see Appendix 5 for technical details). The empirical methods will be used to identify the raw score on each of the STAAR Alternate assessments that best represents the TAKS–Alt *Met Standard* performance standards.

Stage 3. Impact Data Analysis

The third stage involves evaluating the percent of students attaining the TAKS *Met Standard* on the TAKS 2011 assessments (referred to as impact data) in relation to student performance on STAAR assessments in 2012. The impact data analysis can provide supplemental data to support the results from the empirical analyses. As supported by the TTAC, the impact data analysis will identify the percentage of students at and above the *Met Standard* on each TAKS 2011 assessment and find the raw score on the STAAR assessment that corresponds to the TAKS passing percentage in 2011 (see Appendix 6).

Summary

The bridge study will facilitate AYP evaluations for 2012 during the transition from TAKS to STAAR especially for STAAR assessments that will not yet have performance standards established. The three stages of the bridge study will be completed by summer 2012 for inclusion in AYP calculations. The bridge study results will also be used as one piece of information in the STAAR standard-setting process (to help determine a lower bound for STAAR performance standards). However, students will not be held accountable for the bridge study results. After the completion of all standard-setting activities and approval of the performance standards by the Texas commissioner of education, students will receive score reports based on the STAAR performance standards in January 2013 for performance on STAAR assessments in spring 2012. In spring 2013, AYP will use the performance standards for STAAR for AYP evaluations. Texas will begin submissions of the STAAR assessment system for federal peer review in spring 2012.

Appendix 1. Timeline for Completion of the STAAR Bridge Study

Table 2. Timeline for Completion of the STAAR Bridge Study

Task*	Type	Date
STAAR EOC English I reading and Algebra I content overlap analyses with TAKS reading and mathematics for grade 9, respectively	Content Analysis	December 1, 2011–February 1, 2012
STAAR 3–8 content overlap analyses with TAKS	Content Analysis	December 1, 2011–February 1, 2012
STAAR Alternate window	Test Administration	January 9–April 20, 2012
STAAR Modified content overlap analyses with TAKS–M	Content Analysis	February 15–March 15, 2012
Review analyses plans with Texas Technical Advisory Committee	Empirical and Impact Analyses	March 22–March 23, 2012
STAAR Alternate content overlap analyses with TAKS–Alt	Content Analysis	March 16–April 16, 2012
STAAR and STAAR Modified reading and mathematics for grades 5 and 8	Test Administration	March 27–March 28, 2012
STAAR and STAAR Modified reading and mathematics for grades 5 and 8 results bridged to the TAKS and TAKS–M passing standard	Empirical and Impact Analyses	April 16–April 30, 2012
STAAR and STAAR Modified reading and mathematics for grades 3, 4, 6, and 7	Test Administration	April 24–April 25, 2012
STAAR Alternate results bridged to the TAKS–Alt passing standard	Empirical and Impact Analyses	April 26–May 15, 2012
STAAR Algebra I and English I window	Test Administration	May 7–May 19, 2012
STAAR and STAAR Modified reading and mathematics for grades 3, 4, 6, and 7 results bridged to the TAKS and TAKS–M passing standard. STAAR Algebra I and English I results bridged to TAKS passing standard.	Empirical and Impact Analyses	May 14–May 31, 2012
Apply bridge study results to student data	Reporting Process	June 1–June 15, 2012

*Includes English, Spanish, and linguistically accommodated assessments, where applicable

Appendix 2. Example Content Overlap Template

Table 3. Component 1: Content Standard-Level Comparisons

		Grade 3 Mathematics Assessed TEKS		TAKS Grade 3 Math	STAAR Grade 3 Math	Percentage of SEs Assessed on TAKS Also Assessed on STAAR
STAAR Reporting Category	**SE**	**Knowledge and Skills Statement**	**Student expectation**	**SE Assessed**	**SE Assessed**	
1	1A	**Number, operation, and quantitative reasoning.** The student uses place value to communicate about increasingly large whole numbers in verbal and written form, including money. The student is expected to:	(A) use place value to read, write (in symbols and words), and describe the value of whole numbers through 999,999;	yes	yes	91%
	1B		(B) use place value to compare and order whole numbers through 9,999; and	yes	yes	
	1C		(C) determine the value of a collection of coins and bills.	yes	yes	
	2C	**Number, operation, and quantitative reasoning.** The student uses fraction names and symbols (with denominators of 12 or less) to describe fractional parts of whole objects or sets of objects. The student is expected to:	(C) use fraction names and symbols to describe fractional parts of whole objects or sets of objects.	yes	yes	

		Grade 3 Mathematics Assessed TEKS		TAKS Grade 3 Math	STAAR Grade 3 Math	Percentage of SEs Assessed on TAKS Also Assessed on STAAR
STAAR Reporting Category	**SE**	**Knowledge and Skills Statement**	**Student expectation**	**SE Assessed**	**SE Assessed**	
1 (continued)	3A	**Number, operation, and quantitative reasoning.** The student adds and subtracts to solve meaningful problems involving whole numbers. The student is expected to:	(A) model addition and subtraction using pictures, words, and numbers; and	yes	yes	91%
	3B		(B) select addition or subtraction and use the operation to solve problems involving whole numbers through 999.	yes	yes	
	4A	**Number, operation, and quantitative reasoning.** The student recognizes and solves problems in multiplication and division situations. The student is expected to	(A) learn and apply multiplication facts through 12 by 12 using [concrete] models [and objects];	no	yes	
	4B		(B) solve and record multiplication problems (up to two digits times one digit); and	yes	yes	
	4C		(C) use models to solve division problems and use number sentences to record the solutions.	yes	yes	
	5A	**Number, operation, and quantitative reasoning.** The student estimates to determine reasonable results. The student is expected to:	(A) round whole numbers to the nearest ten or hundred to approximate reasonable results in problem situations; and	yes	yes	
	5B		(B) use strategies including rounding and compatible numbers to estimate solutions to addition and subtraction problems.	yes	yes	

		Grade 3 Mathematics Assessed TEKS		TAKS Grade 3 Math	STAAR Grade 3 Math	Percentage of SEs Assessed on TAKS Also Assessed on STAAR
STAAR Reporting Category	**SE**	**Knowledge and Skills Statement**	**Student expectation**	**SE Assessed**	**SE Assessed**	
2	6A	**Patterns, relationships, and algebraic thinking**. The student uses patterns to solve problems. The student is expected to:	(A) identify and extend whole-number and geometric patterns to make predictions and solve problems;	yes	yes	100%
	6B		(B) identify patterns in multiplication facts using concrete objects, pictorial models, or technology; and	yes	yes	
	6C		(C) identify patterns in related multiplication and division sentences (fact families) such as 2 x 3 = 6, 3 x 2 = 6, 6 ÷ 2 = 3, 6 ÷ 3 = 2.	yes	yes	
	7A	**Patterns, relationships, and algebraic thinking.** The student uses lists, tables, and charts to express patterns and relationships. The student is expected to:	(A) generate a table of paired numbers based on a real-life situation such as insects and legs; and	yes	yes	
	7B		(B) identify and describe patterns in a table of related number pairs based on a meaningful problem and extend the table.	yes	yes	

		Grade 3 Mathematics Assessed TEKS		TAKS Grade 3 Math	STAAR Grade 3 Math	Percentage of SEs Assessed on TAKS Also Assessed on STAAR
STAAR Reporting Category	**SE**	**Knowledge and Skills Statement**	**Student expectation**	**SE Assessed**	**SE Assessed**	
3	8A	**Geometry and spatial reasoning.** The student uses formal geometric vocabulary. The student is expected to:	(A) identify, classify, and describe two- and three-dimensional geometric figures by their attributes. The student compares two- dimensional figures, three-dimensional figures, or both by their attributes using formal geometry vocabulary.	yes	yes	100%
	9A	**Geometry and spatial reasoning.** The student recognizes congruence and symmetry. The student is expected to:	(A) identify congruent two-dimensional figures; and	yes	yes	
	9C		(C) identify lines of symmetry in two-dimensional geometric figures.	yes	yes	
	10A	**Geometry and spatial reasoning.** The student recognizes that a line can be used to represent numbers and fractions and their properties and relationships. The student is expected to:	(A) locate and name points on a number line using whole numbers and fractions, including halves and fourths.	yes	yes	

		Grade 3 Mathematics Assessed TEKS		TAKS Grade 3 Math	STAAR Grade 3 Math	Percentage of SEs Assessed on TAKS Also Assessed on STAAR
STAAR Reporting Category	**SE**	**Knowledge and Skills Statement**	**Student expectation**	**SE Assessed**	**SE Assessed**	
4	11A	**Measurement.** The student directly compares the attributes of length, area, weight/mass, and capacity, and uses comparative language to solve problems and answer questions. The student selects and uses standard units to describe length, area, capacity/volume, and weight/mass. The student is expected to:	(A) use linear measurement tools to estimate and measure lengths using standard units;	yes	yes	100%
	11B		(B) use standard units to find the perimeter of a shape; and	yes	yes	
	11C		(C) use [concrete and] pictorial models of square units to determine the area of two-dimensional surfaces.	yes	yes	
	12A	**Measurement.** The student reads and writes time and measures temperature in degrees Fahrenheit to solve problems. The student is expected to:	(A) use a thermometer to measure temperature; and	yes	yes	
	12B		(B) tell and write time shown on analog and digital clocks.	yes	yes	

		Grade 3 Mathematics Assessed TEKS		TAKS Grade 3 Math	STAAR Grade 3 Math	Percentage of SEs Assessed on TAKS Also Assessed on STAAR
STAAR Reporting Category	**SE**	**Knowledge and Skills Statement**	**Student expectation**	**SE Assessed**	**SE Assessed**	
5	13A	**Probability and statistics.** The student solves problems by collecting, organizing, displaying, and interpreting sets of data. The student is expected to:	(A) collect, organize, record, and display data in pictographs and bar graphs where each picture or cell might represent more than one piece of data;	yes	yes	100%
	13B		(B) interpret information from pictographs and bar graphs; and	yes	yes	
	13C		(C) use data to describe events as more likely than, less likely than, or equally likely as.	yes	yes	
Underlying Processes and Mathematical Tools	14A	**Underlying processes and mathematical tools.** The student applies grade 3 mathematics to solve problems connected to everyday experiences and activities in and outside of school. The student is expected to:	(A) identify the mathematics in everyday situations;	yes	yes	62.5%
	14B		(B) solve problems that incorporate understanding the problem, making a plan, carrying out the plan, and evaluating the solution for reasonableness;	yes	yes	
	14C		(C) select or develop an appropriate problem-solving plan or strategy, including drawing a picture, looking for a pattern, systematic guessing and checking, acting it out, making a table, working a simpler problem, or working backwards to solve a problem; and	yes	yes	
	14D		(D) use tools such as real objects, manipulatives, and technology to solve problems.	no	yes	

		Grade 3 Mathematics Assessed TEKS		TAKS Grade 3 Math	STAAR Grade 3 Math	Percentage of SEs Assessed on TAKS Also Assessed on STAAR
STAAR Reporting Category	**SE**	**Knowledge and Skills Statement**	**Student expectation**	**SE Assessed**	**SE Assessed**	
Underlying Processes and Mathematical Tools (continued)	15A	**Underlying processes and mathematical tools.** The student communicates about grade 3 mathematics using informal language. The student is expected to:	(A) explain and record observations using objects, words, pictures, numbers, and technology; and	no	yes	62.5%
	15B		(B) relate informal language to mathematical language and symbols.	yes	yes	
	16A	**Underlying processes and mathematical tools.** The student uses logical reasoning. The student is expected to:	(A) make generalizations from patterns or sets of examples and non-examples; and	yes	yes	
	16B		(B) justify why an answer is reasonable and explain the solution process.	no	yes	
Number/Percentage of SEs represented on TAKS that are also on STAAR				**32**	**36**	**89%**

		Grade 3 Mathematics Assessed TEKS		TAKS Grade 3 Math	STAAR Grade 3 Math
	SE	**Knowledge and Skills Statement**	**Student expectation**	**SE Assessed**	**SE Assessed**
SEs on TAKS That Are Not Assessed on STAAR	2B	**Number, operation, and quantitative reasoning.** The student uses fraction names and symbols (with denominators of 12 or less) to describe fractional parts of whole objects or sets of objects. The student is expected to:	(B) compare fractional parts of whole objects or sets of objects in a problem situation using concrete models	yes	no
	11D	**Measurement.** The student directly compares the attributes of length, area, weight/mass, and capacity, and uses comparative language to solve problems and answer questions. The student selects and uses standard units to describe length, area, capacity/volume, and weight/mass. The student is expected to:	(D) identify concrete models that approximate standard units of weight/mass and use them to measure weight/mass	yes	no
	11E		(E) identify concrete models that approximate standard units for capacity and use them to measure capacity	yes	no
	11F		(F) use concrete models that approximate cubic units to determine the volume of a given container or other three-dimensional geometric figure	yes	no

Table 4. Component 2: Test-Level Comparisons

Comparing STAAR Grade 3 Mathematics and TAKS Grade 3 Mathematics

Assessment Features	TAKS Grade 3 Mathematics	STAAR Grade 3 Mathematics
Purpose	Developed to accurately measure student learning of the grade 3 mathematics curriculum, the Texas Essential Knowledge and Skills (TEKS student expectations), which were adopted to be effective September 1, 1996, and revised in 2005	Developed to accurately measure student learning of the grade 3 mathematics curriculum, the Texas Essential Knowledge and Skills (TEKS student expectations), which were adopted to be effective September 1, 1996, and revised in 2005
Assessment Type	A criterion-referenced test measuring student performance on the grade 3 mathematics Texas Essential Knowledge and Skills	A criterion-referenced test measuring student performance on the grade 3 mathematics Texas Essential Knowledge and Skills
Administration	Administered in April Administered by school personnel Paper version Untimed	Administered in April Administered by school personnel Paper version Four-hour time limit
Item formats	Multiple-choice & constructed response (gridded numerical answers); presented in a machine-scorable test booklet to allow students to mark answers directly in the booklet	Multiple-choice & constructed response (gridded numerical answers); students mark answers on separate answer documents
Blueprint Categories/ Number of Items	Number, operation, and quantitative reasoning - 10 Patterns, relationships, and algebraic thinking - 6 Geometry and spatial reasoning - 6 Measurement - 6 Probability and statistics - 4 Mathematical processes and tools – 8 Total number of items - 40	Number, operation, and quantitative reasoning - 15 Patterns, relationships, and algebraic thinking - 8 Geometry and spatial reasoning - 9 Measurement - 8 Probability and statistics – 6 Total number of items – 46* * Mathematics process skills are assessed in context in 75% of the items
Performance Levels	Commended Met Standard Did Not Meet Standard	Advanced Academic Performance Satisfactory Academic Performance Unsatisfactory Academic Performance

Appendix 3. STAAR 3–8 Empirical Analysis

The STAAR program at grades 3–8 will assess the same grades and subjects as are assessed on TAKS. Similarly to TAKS, STAAR will assess the Texas Essential Knowledge and Skills (TEKS) curriculum standards. The Texas State Board of Education (SBOE) periodically updates the state's curriculum standards by content area on a rotating basis. Since the new curriculum comes into effect for different content areas at different times the methodology for the empirical analyses for reading and mathematics are different as discussed below.

The method proposed for identifying the TAKS *Met Standard* performance standard on the STAAR 3–8 assessments consists of a single group design in which students are administered both a TAKS base test, or operational test, and STAAR field-test items within the same test form. Because of the similarity in content and item format between STAAR and TAKS items, it was possible to embed STAAR field-test items in the TAKS assessments in 2011. This is especially true for mathematics where the curriculum standards have not been revised recently (last adopted in 2005 and amended in 2009). Therefore, it is likely that some of the TAKS mathematics items will transition to STAAR mathematics assessments. However, TAKS reading items will not be used on STAAR reading assessments because of recent revisions to the English language arts and reading curriculum standards, which were last adopted in 2008 and amended in 2010.

For STAAR 3–8, the statistical methods for establishing the bridge between TAKS and STAAR 3–8 assessments were presented and discussed at the February 2010 Texas Technical Advisory Committee (TTAC) meeting and the August 2011 TTAC meeting (see Appendices 7 and 8 for the TTAC comments). The following sections provide the technical details of the bridge study design.

Data Collection Design

For STAAR 3–8, the data collection consists of student performance data from the TAKS 3–8 assessments in spring 2011, including field-test data for embedded STAAR items in the TAKS base tests. Figure 1 illustrates the embedded field-test design for STAAR items in a TAKS base test across multiple forms (form 1 to form n). The base test items are common across the forms. The number of embedded STAAR field-test items ranges from eight to ten items per test form. The number of test forms ranges from 25 to 61 test forms across the grades and subjects. The test forms are spiraled at the student-level resulting in approximately randomly equivalent groups.

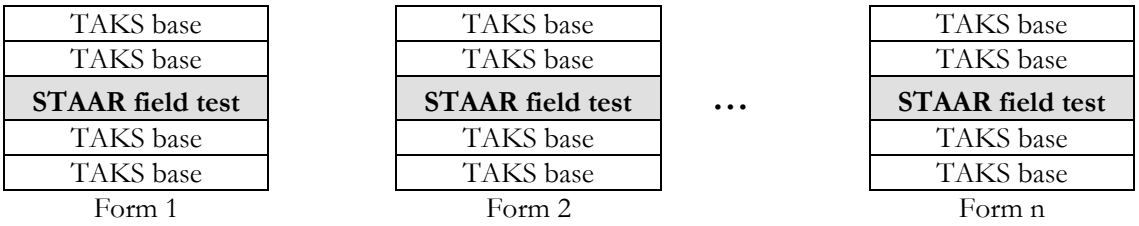

Figure 1. 2011 STAAR 3–8 Embedded Field Test Design

Empirical Methods

STAAR 3–8 reading and mathematics assessments are scaled and equated using a statistical model known as the Rasch Partial-Credit Model (RPCM) to place test items and measures of student

proficiency on the same scale across assessments. The RPCM is an extension of the Rasch one-parameter Item-Response Theory (IRT) model attributed to Georg Rasch (1966), as extended by Wright and Stone (1979), Masters (1982), Wright and Masters (1982), and Linacre (2001). The RPCM maintains a one-to-one relationship between scale scores and raw scores, meaning each raw score is associated with a unique scale score.

The RPCM is defined by the following mathematical measurement model where, for a given item involving $m + 1$ score categories, the probability of person n scoring x on prompt i is given by:

$$P_{xni} = \frac{\exp \sum_{j=0}^{x} (\theta_n - \delta_{ij})}{\sum_{k=0}^{m_i} \exp \sum_{j=0}^{x} (\theta_n - \delta_{ij})} \quad x = 0, 1, \ldots, m_i \qquad (1)$$

The RPCM provides the probability of a student scoring x on the m steps of question/prompt i as a function of the student's proficiency level, θ_n (sometimes referred to as "ability"), and the step difficulties, δ_{ij}, of the m steps in prompt i. (Refer to Masters, 1982, for an example.) Note that for multiple-choice and gridded-response questions, there are only two score categories: (a) 0 for an incorrect response and (b) 1 for a correct response, in which case the RPCM reduces to the standard Rasch one-parameter IRT model, and the resulting single-step difficulty is more properly referred to as an item difficulty. The underlying Rasch scale enables the maintenance of equivalent performance standards across test forms.

Since the same underlying constructs were being assessed for both TAKS and STAAR assessments, and the same students were administered both TAKS and STAAR items, the bridge study for STAAR 3–8 reading and mathematics consisted of calibrating the STAAR items with the TAKS items, and establishing a bridge using the linear relationship between the Rasch scales. The method of calibrating the embedded field-test items differed for reading and mathematics because of the changes in the reading curriculum and the decision to include TAKS mathematics items in future STAAR mathematics assessments.

STAAR 3–8 Mathematics. The embedded field-test items were calibrated separately by form in Winsteps (Linacre, 2001). This process is similar to field-test equating for TAKS as previously approved through the peer review process. Each form was then linearly transformed to be on the same Rasch scale as the base-test items, using the base-test items as a common item set. For each form, a field-test equating constant was calculated as the difference between the mean post-equated Rasch difficulty value for the base-test items and the mean Rasch difficulty for the base-test items which were calibrated with the field-test items. The difference in the means was used as the equating adjustment constant (i.e., scale linking) to place the field-test items for a particular form onto the current Rasch scale. A stability check was not conducted during field-test equating because all base-test items were previously evaluated during post-equating. For each form, the derivation of the field-test constant (C_{FT}) can be represented as follows:

$$C_{FT} = \bar{b}_{base} - \bar{b}_{2011unscaled} \qquad (2)$$

The $\bar{b}_{base}$ is the mean for the post-equated Rasch item difficulty of the base-test items. The $\bar{b}_{2011unscaled}$ is the mean of the unscaled Rasch item difficulty from the Winsteps calibration. Figure 2 illustrates the form by form field-test equating process and the source of the mean Rasch item difficulty values for calculating each form's equating constant.

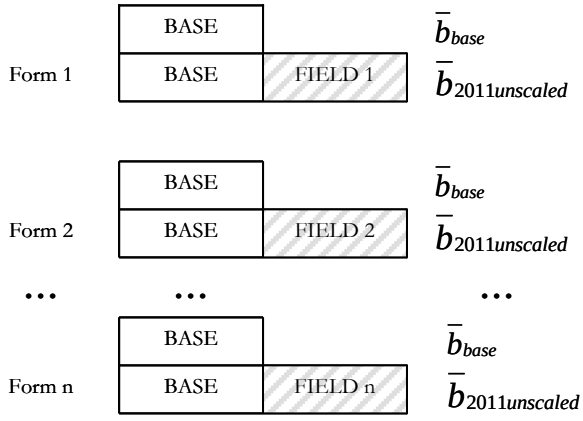

Figure 2. STAAR Mathematics Form by Form Field-Test Equating

For each form, the field-test constant is added to the unscaled 2011 calibrated Rasch item difficulty for all field-test items to place the items onto the current Rasch scale. The field-test constant is added to field-test items only, as depicted in Equation 3. The $b_{2011equated}$ is the Rasch item difficulty for the STAAR field-test items on the current Rasch scale.

$$b_{2011equated} = b_{2011unscaled} + C_{FT} \qquad (3)$$

In spring 2012, the STAAR operational test forms will be post-equated using the field-test statistics of the base-test items to link to the current Rasch scale. Once the STAAR base-test items' Rasch item difficulty values are estimated and transformed to the current Rasch scale, then the raw score associated with the TAKS *Met Standard* ability level, θ_{ms}, will be identified through the creation of a raw-score-to-ability-level table. A separate table will be developed for the STAAR 2012 operational test forms at each grade and subject that designates the TAKS *Met Standard* ability level, θ_{ms}.

STAAR 3–8 Reading. Unlike STAAR 3–8 mathematics, the current TAKS reading items will not be used for STAAR 3–8 reading due to extensive revisions to the Texas Essential Knowledge and Skills curriculum standards. Therefore, the field-test items did not need to be placed on the current Rasch scale even though the STAAR field-test items were embedded in the 2011 TAKS base test forms. A single concurrent Winsteps calibration of both TAKS base-test items and STAAR field-test items across all test forms using an incomplete data matrix was conducted for the STAAR 3–8 reading embedded field-test equating. The concurrent calibration allowed for a single calibration in which all field-test items were calibrated together. (Linacre, 2001; Skaggs & Wolfe, in press). The base-test items were included in the concurrent calibration in order to provide field-test item Rasch item difficulty

estimates on the same scale, and to compute a scaling constant for the bridge to TAKS. Figure 3 illustrates the incomplete data matrix for the STAAR field-test equating forms (FIELD). The concurrent calibration resulted in all items having a Rasch item difficulty, denoted b_{STAAR}.

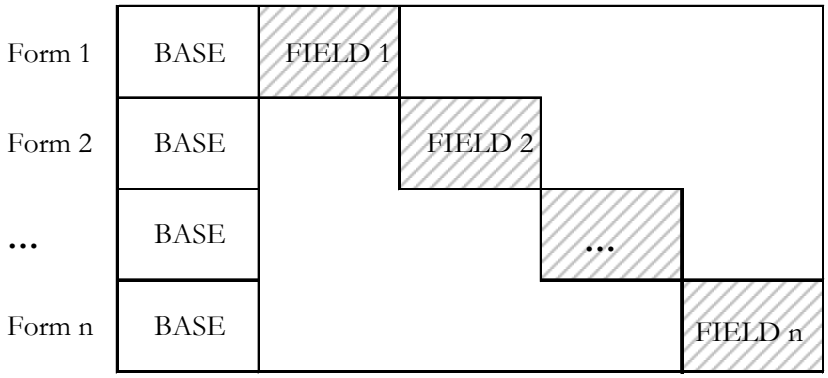

Figure 3. STAAR Reading Incomplete Data Matrix for Field-Test Equating

Since the base-test items for TAKS were included in the concurrent calibration, an equating constant could be computed for the TAKS base-test items. Similar to the process for STAAR 3–8 mathematics, a bridge equating constant (C_{BR}) was calculated as the difference between the mean post-equated Rasch difficulty value for the post-equated base-test items and the mean Rasch difficulty for the base-test items when calibrated with the field-test items. A stability check was not conducted since all base-test items were previously evaluated during post-equating. The derivation of the bridge equating constant (C_{BR}) can be represented as follows:

$$C_{BR} = \bar{b}_{base} - \bar{b}_{STAAR} \qquad (4)$$

The $\bar{b}_{base}$ is the mean for the post-equated Rasch item difficulty of the base test items. The $\bar{b}_{STAAR}$ is the mean of the base-test items' Rasch item estimated during the concurrent field-test item calibration. The equating constant was used to identify the TAKS *Met Standard* on the current Rasch scale.

Since RPCM places test items and measures of student proficiency on the same scale, the Rasch item difficulty (b_{STAAR}) is on the same scale as the student ability (θ_{STAAR}) scale. The bridge equating constant (C_{BR}) can be used to linearly transform the TAKS *Met Standard* ability level, denoted θ_{MS}, to the STAAR Rasch scale. This is the STAAR ability value, denoted θ_{STAAR_MS}, that corresponds to the TAKS *Met Standard* ability level, denoted θ_{MS}, and is represented as follows:

$$\theta_{STAAR\_MS} = \theta_{MS} - C_{BR} \qquad (5)$$

The TAKS *Met Standard* ability level, denoted θ_{MS}, is known for each grade and subject because these values were established in 2002 during the TAKS standard-setting committee meetings.

In spring 2012, the STAAR operational test forms will be post-equated through the field-test statistics to the STAAR Rasch scale. The bridge study results will be applied to the 2012 post-equated tests through the Rasch scale. The Rasch item difficulty values for the STAAR 3–8 reading base-test items will be used to create a raw score to ability table that provides a one-to-one correspondence between the possible raw score values and the STAAR Rasch ability values, θ_{STAAR}. The raw score associated with the STAAR Rasch ability level, denoted θ_{STAAR_MS}, that equals the TAKS *Met Standard* derived in equation 5, will indicate the bridge study result for STAAR 3–8 reading assessments.

Table 5 provides an example of how the bridge study results for grade 7 mathematics might be displayed. In this table, the TAKS *Met Standard* for grade 7 mathematics is indicated on a raw score table based on the STAAR grade 7 mathematics 2012 base test. In this hypothetical example, a raw score of 33 corresponds to the TAKS *Met Standard* performance standard and students obtaining a raw score of 33 or greater will be evaluated as proficient for AYP purposes in 2012.

Table 5. Example Bridge Study Result for STAAR Grade 7 Math Raw Score Scale

STAAR Grade 7 Math Raw Score	TAKS Grade 7 Math Performance Standard
…	…
30	
31	
32	
33	**TAKS Met Standard**
34	
35	
36	
37	
38	
39	
40	
…	…

STAAR English I Reading and Algebra I. In the 2011–2012 school year, students in grades 3–8 who are also enrolled in English I and/or Algebra I will take the STAAR EOC assessment associated with that course, as required for graduation. In order to provide a bridge between the TAKS *Met Standard* performance standards and the STAAR assessments for AYP, additional analyses to bridge TAKS to STAAR EOC are proposed.

Students that take both TAKS and STAAR EOC assessments in spring 2011 will allow for a single-group design to be created for the bridge study analyses. Students taking STAAR English I Reading will be bridged to TAKS grade 9 reading. Students taking STAAR Algebra I will be bridged to TAKS grade 9 mathematics. English I reading and Algebra I examinees in 2011 represent a sample, rather than a census. As such, descriptive analyses will focus on the extent to which the STAAR examinees are representative of the Texas test-taking population. Two sets of descriptive analyses are planned. First, demographic characteristics of the STAAR English I Reading and Algebra I samples will be compared to demographic characteristics of grade 9 students in Texas who did *not*

take the STAAR assessments in 2011. This will provide insight regarding the demographic similarity of STAAR-tested and untested populations. Second, the TAKS grade 9 reading and mathematics scores for those Texas students who also took STAAR English I reading and Algebra I tests in 2011 will be compared to the TAKS scores of students who did *not* take the STAAR assessments in 2011. This analysis evaluates the comparability of the 2011 STAAR sample to the full population of Texas students.

In addition, content overlap analyses will be conducted to determine the similarity of assessed material across the STAAR and TAKS tests to be linked. The content overlap analyses will focus not only on the subject matter covered in each test, but also the extent to which the depth and complexity of assessed material is similar across tests. Each link (e.g., Algebra I – TAKS grade 9 mathematics) will be classified according to the strength of the content overlap observed.

Following the descriptive analyses outlined above, the single-group bridge studies will utilize the equipercentile method described in Pommerich et al. (2004) and detailed in Kolen and Brennan (2004). The specific steps for the equipercentile equating between a TAKS assessment (*X*) and a STAAR EOC assessment (*Y*) include:

1. Let $T_X \in (1000, 3000)$ (note: the upper and lower boundary will be determined by the observed TAKS scores, or, if applied, pre-smoothing) with an interval of 1 between each T_X on the TAKS assessment (*X*). For each T_X, use the following equipercentile function to find the corresponding raw score on the given STAAR EOC assessment (*Y*):

$$\frac{\Pr(X < T_X) + 0.5 \times \Pr(X = T_X) - \Pr(Y < u^*(T_X))}{\Pr(Y = u^*(T_X))} + u^*(T_X) - 0.5 \qquad (6)$$

   Where $u^*(T_X)$ is smallest raw score on the STAAR EOC assessment such that, $\Pr(X < T_X) + .5 \times \Pr(X = T_X) < \Pr(Y \le u^*(T_X))$.

2. The resulting table will list the STAAR EOC raw score that has the same percentile rank (Pr) as each TAKS score. Under this approach, many consecutive TAKS scale scores may be expected to correspond with the same STAAR EOC raw score (e.g., the TAKS scale scores 2073 through 2103 may all correspond with a STAAR EOC raw score of 29). As such, any given STAAR EOC raw score may be associated with a band of TAKS scale scores (e.g., 29 = 2073-2103). The equipercentile equating results are summarized using this STAAR raw score-to-TAKS score band approach.

3. The STAAR EOC raw score associated with the TAKS score band that includes the *Met Standard* scale score cut (i.e., 2100) will serve as the bridge study result for STAAR EOC assessments.

Note: When a single STAAR EOC form is administered to a population of examinees, a single table is produced providing STAAR EOC scores on the Rasch scale (θ) that correspond to each obtainable STAAR EOC raw score. This "raw-score-to-Theta" table facilitates the interpretation of bridge study results using either raw scores or θ estimates. Because a one-to-one

correspondence exists between raw scores and θ estimates, the equipercentile equating function presented in Equation 6 can utilize either measure; bridge study results will be equivalent.

Table 6 provides an example of the bridge study results for STAAR Algebra I to grade 9 TAKS mathematics. In Table 6, the TAKS score bands for grade 9 mathematics are mapped to obtainable raw scores on the STAAR Algebra I 2012 base test. In this hypothetical example, the TAKS grade 9 mathematics scale score needed to attain *Met Standard* was a value of 2100. An equipercentile linking procedure between STAAR Algebra I raw scores, and TAKS scale scores indicated that a TAKS score band of 2073–2103 (which includes 2100) corresponded to an Algebra I raw score of 29. This raw score corresponded to an Algebra I ability estimate (on the Rasch scale) of roughly -0.15. Therefore, based on this bridge study hypothetical example, a score of 29 on the Algebra I test (or, equivalently, a θ estimate of -0.15) is associated with the TAKS *Met Standard*.

Table 6. Example Bridge Study for STAAR Algebra I to Grade 9 TAKS Mathematics

STAAR Algebra I Raw Score	TAKS Grade 9 Mathematics Scale Score	STAAR Algebra I Ability
…	…	…
27	2026–2056	-0.33
28	2057–2072	-0.24
29	**2073–2103**	**-0.15**
30	2104–2118	-0.07
31	2119–2135	0.02
…	…	…

Appendix 4. STAAR Modified 3–8 Empirical Analysis

STAAR Modified is an alternate assessment based on modified academic achievement standards. The assessment is for students receiving special education services that need extensive modifications and accommodations to classroom instruction, assignments, and assessments to access and demonstrate progress in the grade-level curriculum. These are students receiving special education services who can make academic progress even though they may not reach grade-level achievement standards in the same time frame as their non-disabled peers.

Only those students who meet the established participation requirements for STAAR Modified are allowed to take the assessment. As with the previous modified assessment, TAKS–M, a student's admission, review, and dismissal (ARD) committee will determine whether the student meets the participation requirements for STAAR Modified. The STAAR Modified assessment will have specific participation requirements. The participation requirements specifically describe student needs and learning behaviors and appear as a worksheet to guide campuses through a series of steps in order to determine appropriate placement in the modified assessment. The requirements are intended to help the ARD committees better identify the students who are eligible for the assessment and to encourage active, thoughtful, and collaborative discussions by the members of the ARD committee when making these decisions.

To allow this unique population of students the opportunity to demonstrate their knowledge of the grade-level curriculum, the test design and format of the general STAAR assessments were modified to create the STAAR Modified assessments for grades 3–8.

STAAR Modified assessments are defined by

- a shorter test blueprint that reflects the STAAR blueprint,
- test development based on STAAR content and item statistics,
- modification of STAAR items to be accessible to students receiving special education services who are eligible for STAAR Modified (i.e., fewer answer choices, simplification of wording, chunking of passages, inclusion of pre-reading text boxes, simplification of figures),
- embedded field-test items administered on live-test forms,
- pre-equated live-test forms so reporting can occur at the same time as STAAR,
- common-item equating from year to year, and
- modified academic achievement standards which are not linked to the general academic achievement standards.

For STAAR Modified 3–8, the statistical methods for establishing the bridge with TAKS–M were presented and discussed at the August 2011 TTAC meeting (see Appendix 8 for the TTAC comments). The following sections provide the technical details of the bridge study design.

Data Collection

The first administration of STAAR Modified assessments will be in spring 2012. As a result, a single group design is not possible for the STAAR Modified bridge studies since there will not be the same students taking both STAAR Modified and the TAKS–M assessments for the same grade and subject test. For many of the grades and subjects, there is a common TAKS–M assessment that both groups of students have taken in an earlier grade. The data from the earlier assessment can be used

along with demographic variables to create matched samples for the analyses. The participation requirements for STAAR Modified may result in a student population for STAAR Modified that has different characteristics than the TAKS–M student population. The matched sample process will use the entire STAAR Modified and TAKS–M student populations to create matched samples that are representative of the STAAR Modified student population.

Figure 4 illustrates the data collection design for identifying the TAKS–M *Met Standard* performance standard on the STAAR Modified assessments. In this example, the grade 6 reading test for TAKS–M is bridged to the grade 6 reading test for STAAR Modified. The students taking the TAKS–M test are different students than the students taking the STAAR Modified assessment (denoted by the dashed line in Figure 4). Both student groups previously took the TAKS–M grade 5 reading test (denoted A in Figure 4). The students' performance on the TAKS–M grade 5 reading test can be used to create matched samples (denoted B in Figure 4). Once the matched samples are identified, empirical analyses for a single group design can be applied (denoted C in Figure 4). The matching procedure can use student performance from other subjects and/or other variables to create the matched samples.

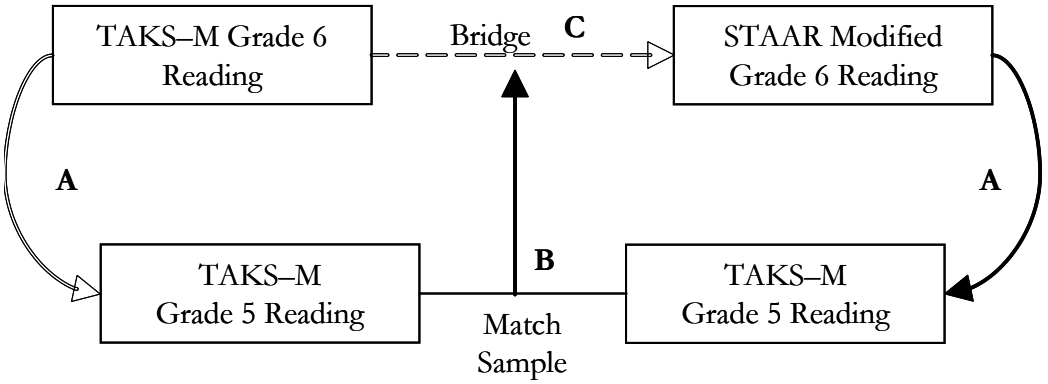

Figure 4. Example Matched Samples Data Collection Design

The STAAR Modified operational test forms will be calibrated through Winsteps and a STAAR Modified Rasch scale will be established. The Rasch item difficulty values for the STAAR Modified base-test items will be used to create a raw score to ability table that provides a one-to-one correspondence between the possible raw score values and the STAAR Modified Rasch ability values.

Matched Samples

A sample that emulates a single group will be created from the two cohorts by matching the two groups on attributes related to each of the assessments being linked. Attributes that could be used to create a matched sample include demographic variables such as gender, ethnicity, and social economic status, as well as previous TAKS–M scores in the same content area. The matching procedure involves creating a composite score based on previous years' TAKS–M scores in the same content area through linear regression (Way, Davis, & Fitzpatrick, 2006). In order to promote similarity in the two samples being matched on demographic variables, all models include demographic variables such as gender and ethnicity. Once the regression results are generated, an

evaluation of the statistical contribution of the independent variables in relation to the dependent variable across regression models is conducted.

A generic model is denoted in Equation 7:

$$\hat{Y}_{predicted} = \beta_0 + \beta_1 X_{1(MatchingVar1)} + \beta_2 X_{2(MatchingVar2)} + \ldots \quad (7)$$

Empirical Methods

Once the matched samples are created, then an analysis method for a single group design can be used to establish the relationship between the TAKS–M and the STAAR Modified grade 6 reading test. The proposed method is the equipercentile method described in Pommerich et al. (2004) and detailed in Kolen and Brennan (2004).

The specific steps for creating equipercentile equating between a STAAR Modified assessment (*X*) and a TAKS–M assessment (*Y*) include:

1. Let $\theta_X \in (-3, +3)$ (note: the upper and lower boundary will be determined by the observed ability scores, or, if applied, pre-smoothing) with an interval of 0.1 between each θ_X on the STAAR Modified assessment (*X*). For each θ_X, use the following equipercentile function to find the corresponding scale score on the TAKS–M scale scores (*Y*):

$$\frac{\Pr(\theta_X < 0.5) + \Pr(\theta_X = u^*(\theta_X)) \times \Pr(\theta_X = u^*(\theta_X))}{\Pr(Y = u^*(\theta_X))} + u^*(\theta_X) - 0.5 \qquad (8)$$

   Where $u^*(\theta_X)$ is smallest scale score on the TAKS–M scale score such that, $\Pr(\theta_X < .5) + \Pr(\theta_X = u^*(\theta_X)) \times \Pr(\theta_X = u^*(\theta_X))$.

2. Round each *Y* obtained in Step 1 to the nearest whole number. The resulting table will list the corresponding TAKS–M scale score that has the same percentile rank (Pr) as the ability value. The only TAKS–M scale score of interest is the scale score associated with the TAKS–M *Met Standard* performance standard.

3. The raw score associated with the STAAR Modified Rasch ability level that equals the TAKS–M *Met Standard* derived in equation 8 will indicate the bridge study result for STAAR Modified assessments.

Table 7 provides a hypothetical example of the bridge study results for STAAR Modified grade 8 reading to grade 8 TAKS–M reading. In this table, the TAKS–M *Met Standard* for grade 8 reading is indicated on a raw score table based on the STAAR Modified grade 8 reading base test. In this hypothetical example, a raw score of 32 corresponds to the TAKS–M *Met Standard* performance standard for grade 8 TAKS–M reading.

Table 7. Example Bridge Study for STAAR Modified Grade 8 Reading and TAKS–M Grade 8 Reading

STAAR Modified Grade 8 Reading Raw Score	TAKS–M Grade 8 Reading Scale Score (Y)	STAAR Modified Grade 8 Reading Ability (θ_X)
…	…	…
30		-0.1
31		0.0
32	**2100**	**0.1**
33		0.2
34		0.3
35		0.4
36		0.5
37		0.6
38		0.7
…	…	…

Appendix 5. STAAR Alternate Empirical Analysis

STAAR Alternate

STAAR Alternate is an assessment based on alternate academic achievement standards that is designed for students with significant cognitive disabilities receiving special education services. Students must meet established participation requirements to be eligible for this assessment. For example, students must have a significant cognitive disability but also require individualized instruction and specialized supports to access the content standards, and they must also access the content standards through prerequisite skills rather than actual grade-level content.

An Admission, Review, and Dismissal (ARD) committee determines whether participation in this assessment is appropriate for students. STAAR Alternate is not a traditional paper or multiple-choice assessment. The assessment involves teachers observing students as they complete four standardized state-developed assessment tasks for each subject area that link to the grade-level Texas Essential Knowledge and Skills (TEKS) through prerequisite skills.

For STAAR Alternate, teacher observations of the assessment tasks will be conducted during an assessment window that will begin in January 2012 and close in April 2012. To conduct the assessment, teachers:

- select state-developed assessment tasks of the appropriate complexity level for each student,
- administer the three standardized predetermined criteria (that specify what the student is expected to do to demonstrate the skill) associated with each of the four assessment task (primary observation),
- record observations of the assessment tasks on state-developed documentation forms,
- evaluate student performance on each dimension (Demonstration of Skill, Level of Support, and Generalization of Skill) of the STAAR Alternate scoring rubric (see Appendix 9), and
- enter student performance results into the STAAR Alternate online testing interface.

Students are evaluated during the primary observation for their Demonstration of Skill and Level of Support. The Demonstration of Skill portion of the scoring rubric is weighted according to the complexity level of the task. Demonstration of Skill refers to whether or not the student demonstrated the skill outlined in the state-developed assessment task. Teachers indicate students' performance as yes, they demonstrated the skill, or no, they did not demonstrate the skill. Level of Support refers to the amount of independence with which a student demonstrated the skill. Students can demonstrate skills independently, with cueing, or with prompting.

Complexity Level will be incorporated into the scoring by weighting the primary observation Demonstration of Skill dimension by the level of complexity of the task the student completes. Through weighting, students receive more credit for successfully completing more complex tasks. For each of the assessment tasks, teachers determine the complexity level of the task that is most appropriate for their student. The assessment tasks are available at three complexity levels: Level 3/most complex, Level 2/moderately complex, or Level 1/least complex.

If a student demonstrates the skill independently or with cueing, he/she is eligible to repeat the task for a second observation, called the generalization observation, in which the task is repeated using different materials. The generalization observation is evaluated for both Demonstration of Skill and Level of Support but is not weighted by complexity level.

Each assessment task will be scored from 0–21 points based on the teacher's response to a set of performance evaluation questions about how the student performed the task. Each of the four assessment task scores will be added together for a total score ranging from 0–84 points (reporting is done on the raw score scale).

Students will be allowed to use accommodations and supports that are routinely and successfully used as instructional accommodations. Teachers will enter information about the student's performance on each assessment task in the STAAR Alternate online testing interface. Student scores are calculated within the online system based on the information entered by the teacher. The STAAR Alternate scoring rubric will be available to teachers to provide an overview of how scores are applied by the online system.

Students in grades 3–11 who are eligible for an alternate assessment based on alternate academic achievement standards will take STAAR Alternate beginning in the 2011–2012 school year. STAAR Alternate is replacing the previous TAKS–Alt assessment. It is expected that the content being assessed and the STAAR Alternate assessment tasks will be more rigorous than what was used with TAKS–Alt. New performance standards, cut scores, will be set using recommendations from a panel of stakeholders in September 2012. Otherwise, the administration procedures and policies are similar between TAKS–Alt and STAAR Alternate.

For STAAR Alternate, the statistical methods for establishing the bridge between TAKS–Alt and STAAR Alternate assessments were presented and discussed at a November 2011 conference call with specific TTAC members (see Appendix 11 for the TTAC comments). The following sections provide the technical details of the bridge study design.

Empirical Methods

The goal of the empirical method is to identify the raw score on the STAAR Alternate scale that corresponds to the cut score for *Met Standard* on the TAKS–Alt scale. The empirical method will involve comparing TAKS–Alt and STAAR Alternate by using both the TAKS–Alt and the STAAR Alternate scoring rubrics to derive scores for students in the 2011–2012 school year such that students receive an actual score using the STAAR Alternate rubric and a theoretical score using the TAKS–Alt rubric.

STAAR Alternate rubric allows for more differentiation in scores. Ultimately the changes result in an increase to the STAAR Alternate raw score range (0–84) compared to the TAKS–Alt raw score range (0–72). Previously, the TAKS–Alt cut score for *Met Standard* was set at a raw score of 44. The STAAR Alternate performance standards will be set in fall 2012 (see Appendices 9 and 10 for the STAAR Alternate and TAKS–Alt scoring rubrics, respectively).

The empirical approach to identifying the TAKS–Alt cut score on the STAAR Alternate scale can be conceptualized as an attempt to maximize the classification accuracy of students' performance into two categories (e.g., *Met Standard* using theoretical TAKS–Alt score and *Met Standard* using actual STAAR Alternate score). Each student will have a TAKS–Alt theoretical score and a STAAR Alternate actual score based on their 2011–2012 STAAR Alternate assessment results. Deriving a score based on the TAKS–Alt rubric allows for categorization of students as *Met Standard* or *Did Not Meet Standard* using the established TAKS–Alt cut score. This is a known, unchanging value. Therefore, the number of students who achieve Met Standard using the TAKS-Alt cut score is also a

known value. Figure 5 shows a conceptual example of classification accuracy. The graphic on the left side of the figure illustrates a scatter plot of students' theoretical TAKS–Alt scores (vertical axis) and actual STAAR Alternate scores (horizontal axis). The TAKS–Alt cut score for *Met Standard* (raw score = 44) is indicated by the solid horizontal line that bisects the graph. The goal in maximizing classification accuracy is to place the vertical dashed line on the STAAR Alternate raw score scale that results in consistently classifying students in the same category (e.g., *Met Standard*) as derived using TAKS–Alt. In other words, the goal is to locate the STAAR Alternate raw score that results in the most data points in the top right and bottom left quadrants of the graphic.

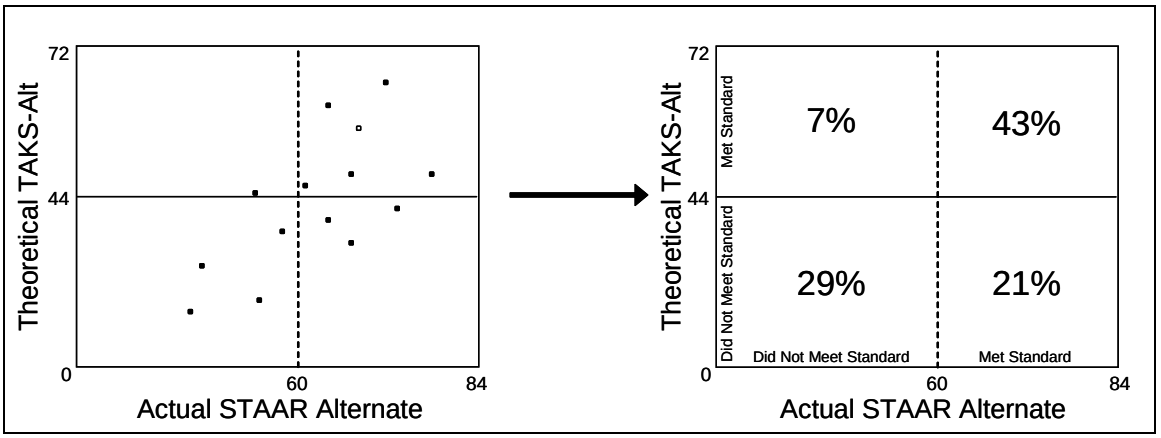

Figure 5. Example of Classification Accuracy Estimates Between Actual STAAR Alternate Scores and Theoretical TAKS–Alt Scores

The graphic on the right side of Figure 5 illustrates the percent of students' in each quadrant, if the cut score for actual STAAR Alternate scores that aligned most closely to the TAKS–Alt score was located at 60. More explicitly, the top right quadrant of the graphic shows the percent of students who would attain *Met Standard* using their theoretical TAKS–Alt score and *Met Standard* with their actual STAAR Alternate scores (43%). The bottom left quadrant shows the percent of students who *Did Not Meet Standard* with their theoretical TAKS–Alt scores and who *Did Not Meet Standard* with their actual STAAR Alternate scores (29%).

Computationally, the frequencies in each of the four quadrants will be derived through various iterations, or placements, of the cut score on the STAAR Alternate scale. Calculating the frequencies, given the STAAR Alternate cut score, will allow us to identify the maximum classification accuracy value as defined by Equation 9. Classification accuracy refers to the level of certainty that a student's classification into a performance category accurately reflects that student's true performance.

$$Accuracy = \frac{N_{TP} + N_{TN}}{N_{TP} + N_{TN} + N_{FP} + N_{FN}} \qquad (9)$$

where N_{TP} is the number of true positives, N_{TN} is the number of true negatives, N_{FP} is the number of false positives, and N_{FN} is the number of false negatives. In this context, a student's theoretical TAKS–Alt score is considered the true score because the goal of the bridge study is to identify a

STAAR Alternate raw score that aligns as closely as possible with the performance category a student would have received if the TAKS–Alt rubric had been used. Therefore, classification accuracy is maximized at the STAAR Alternate raw score cut that: (1) classifies students who *Met Standard* with their theoretical TAKS–Alt score as *Met Standard* with their actual STAAR Alternate score and (2) classifies students who *Did Not Meet Standard* with their theoretical TAKS–Alt score as *Did Not Meet Standard* with their actual STAAR Alternate score.

To align with the conceptual framework, applying Equation 9 gives the graphic displayed in Figure 6. Conceptually, a 2×2 table like Figure 6 will be produced for every combination of actual STAAR Alternate raw score points and theoretical TAKS–Alt raw score points to ascertain classification accuracy for each STAAR Alternate raw score point. The STAAR Alternate raw score that produces the highest classification accuracy will be identified as that which aligns most closely to the TAKS–Alt cut score of 44.

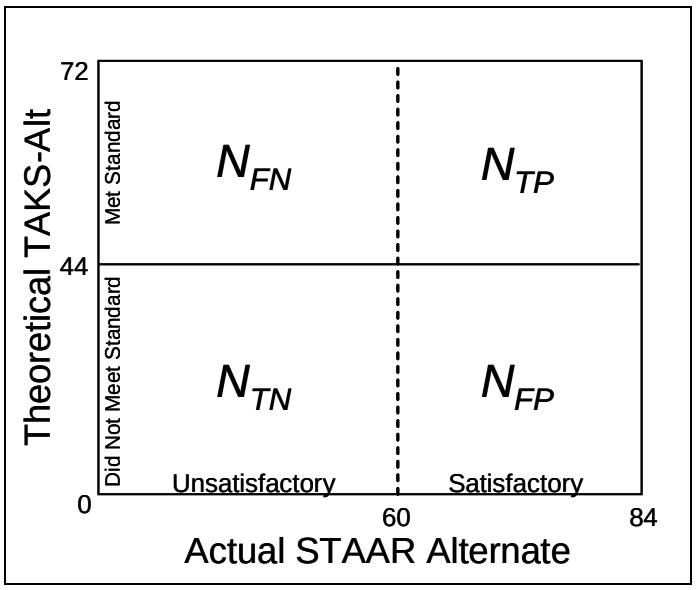

Figure 6. Example of Classification Accuracy Equation Applied to Figure 5

For verification of the classification accuracy analysis, logistic regression can be used to evaluate the relationship between STAAR Alternate performance and a student's probability of achieving *Met Standard* on TAKS–Alt. The specific steps for establishing this relationship are:

1. For each attainable raw score point using the TAKS–Alt scoring rubrics, create the dichotomous variable, Y_i, for each student (i) such that
   - $Y_i = 0$, if the student's proficiency using the TAKS–Alt scoring rubric is below the *Met Standard* cut (raw score = 44);
   - $Y_i = 1$, if the student's proficiency using the TAKS–Alt scoring rubric is at or above the *Met Standard* cut (raw score =44).

2. Let X_i represent the level of performance (STAAR Alternate raw score) each student (i) achieved using the STAAR Alternate scoring rubric and p be the probability that $Y_i = 1$

given X_i; that is, $p = \text{Prob}(Y_i = 1 | X = X_i)$, then using logistic regression, derived the regression coefficients (*a* and *b*) in the equation:

$$\text{logit}(\hat{p}) = aX + b \qquad (10)$$

where $\hat{p}$ is the expected probability that $Y = 1$ and $\text{logit}(\hat{p}) = \ln(\frac{\hat{p}}{1-\hat{p}})$.

3. A probability for each raw score on the STAAR Alternate scale will result. The minimum STAAR Alternate raw score with a probability greater than 0.50 will serve as the STAAR Alternate score that aligns most closely with the TAKS–Alt *Met Standard* performance standard.

The raw score derived through the logistic regression analysis is expected to be similar to the raw score resulting from the classification accuracy analyses. Students attaining a raw score associated with the TAKS–Alt *Met Standard* bridged to the STAAR Alternate assessments will be considered proficient for AYP evaluations in 2012.

Appendix 6. Impact Data Analyses

The purpose of the impact data analysis stage is to validate the empirical method results and, if needed, make an adjustment to the identified TAKS performance standards on the STAAR assessments. The method for the impact data analysis involves identifying the percentage of students at and above the *Met Standard* on the TAKS 2011 assessment and finding the raw score on the STAAR assessment that corresponds to the TAKS passing percentage in 2011 given the percent of students at each raw score on the STAAR 2012 assessment. It is expected that the STAAR raw scores identified by using the same percentages will be equal or close to the TAKS cut points on the STAAR test identified through empirical analyses in the previous stage.

Figure 7 illustrates a hypothetical example of the results of the impact data analysis. The TAKS 2011 grade 8 reading raw score table identifies the TAKS *Met Standard* at a raw score equal to 33. In 2011, 80% of grade 8 students achieved a raw score equal to or greater than 33, indicating the percent passing in 2011 was 80%. The STAAR 2012 grade 8 reading raw score table indicates that a raw score equal to 29 is associated with 80% of students having a total raw score equal to or greater than 29. A STAAR raw score equal to 29 represents the same passing percentage as the raw score of 33 on the TAKS 2011 assessment.

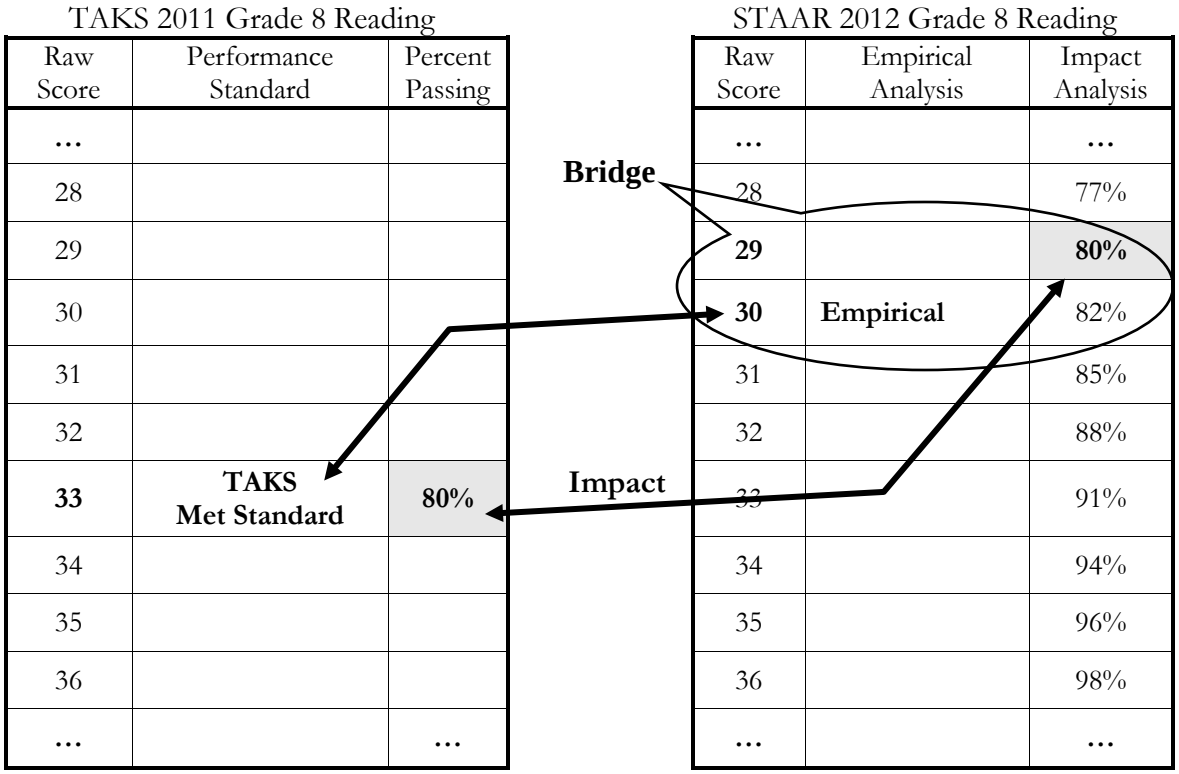

Figure 7. Example Raw Score Tables for TAKS 2011 and STAAR 2012 Grade 8 Reading

At the conclusion of the analyses presented for each method of the bridge study, there will be two scores identified for each grade and subject—one by the empirical method and one by the impact data analysis (as illustrated in Figure 7). The raw score from the impact data analysis is anticipated to be similar to the raw score from the empirical method for a grade or subject. TEA proposes implementing the result from the empirical method. However, in the case that the scores identified by the two methods differ, the student populations will be evaluated for systematic differences that may account for the disparate results. The impact data analyses include the student populations testing in spring 2011 and spring 2012; whereas, some of the empirical analyses are based on samples of the student populations. The samples will be evaluated with respect to the student populations. Additionally, TAKS impact data from previous years may serve as a reference when evaluating differences between the empirical method and the impact data from 2011. The result from the bridge study may be applied by selecting the lower of the two scores or averaging the scores, as suggested by TTAC members (see Appendices 8 and 11). TEA will evaluate averaging the scores to allow for consideration of both the impact data and the empirical analyses. Students attaining a raw score associated with the TAKS *Met Standard* bridged to the STAAR assessments will be considered proficient for AYP evaluations in 2012.

Appendix 7. TTAC Meeting Notes from February 2010

Attendees
TTAC: Carol Allman, Gregory Cizek, Barbara Dodd, Richard Duran, Michael Kolen, William Mehrens, Susan Phillips, Barbara Plake, Joseph Ryan, Stephen Sireci, Roger Trent
TEA: Criss Cloudt, Julie Guthrie, Cathy Kline, Lizettte Reynolds, Mi-Suk Shim, Tomoko Traphagan, Marianna Vassileva, Cari Wieland, Gloria Zyskowski
Pearson: Michael Bay-Borelli, Aimee Boyd, Ticia Carter, Sandi Cowes, Laurie Davis, Tish Denny, Stacy Duke, Phyllis Garrett, Elizabeth Hanna, Leslie Keng, Jadie Kong, Amy LaSalle, Chow-Hong Lin, Paul Matzen, Dan Murphy, Kimberly O'Malley, Ha Phan, Barbara Poynter, Amy Reilly, Martha Scarborough, Walter Sherwood, Natasha Williams, Wenyi You, Malena Zou

Bridge Study Between TAKS and the New 3–8 Program

General:

a) We should carefully consider how to add cognitive complexity to items without adding construct irrelevant variance.
b) Typically a bridging study is used to maintain the same performance expectations for inclusion in AYP. We should consider how we will use information from student performance in AYP if the performance expectations increase.
c) We might consider reporting student performance on STAAR in 2012 using TAKS standards to mitigate motivation effects and help schools transition between assessment programs.
d) We should consider the student implications separately from school/AYP implications as we transition from TAKS to STAAR.

Q1: Does the TTAC know if and how other states have linked new assessments to previous assessments?

a) The transition from TAAS to TAKS is one example to which we might look.
b) Many programs (South Dakota, Ontario, CA) are trying to maintain the performance standards between assessment programs when smaller changes to curriculum are made or when transitions are made between vendors.
c) A revisiting of standards may occur when the constructs are expected to be similar.
d) GA was able to present items from the old and new test together.

Q2: Given the purpose of the linking studies, does the TTAC have any feedback on the proposed analysis method when items for the new assessment are embedded as field-test items and the constructs are similar?

a) We can check to see how well the embedded STAAR items are functioning relative to the TAKS items with which they appear.

Q3: Does the TTAC have recommendations for how high the correlations should be to move forward with concordance analyses or recommendations of other analyses to conduct if the correlations are too low?

a) We could consider invariance analyses to look at differences across subgroups.

Q4: Given the purpose of the linking studies, does the TTAC have any feedback on the proposed analysis method when the new assessment is administered as a stand-alone field test?

a) We could consider embedding TAKS items into the STAAR stand-alone field test.
b) The characteristics of the grade 8 STAAR stand-alone sample and whether they take the STAAR field test before or after the grade 8 TAKS test should be considered.
c) We could look at the change in construct by evaluating the underlying structure of the STAAR and TAKS tests. Is the structure more related to strands than to old/new?
d) We are likely to have a significant motivation effect in the stand-alone field test.
e) We should consider embedding the reading items even though the reading construct assessed in STAAR may be somewhat different from the reading construct assessed in TAKS.
f) We should consider a separate calibration of the STAAR reading items in addition to linking the STAAR reading items onto the same scale as TAKS reading to evaluate whether they are measuring something differently.
g) Support for embedding reading is also given by the early implementation of the new reading curriculum in classrooms.
h) We could look at the STAAR items in CFA framework to evaluate dimensionality. We should take random sets of items from old tests to use as a baseline comparison.
i) We could evaluate differences in item parameters between 2011 field test and 2012 live test.
j) Because of the spiraling of field-test forms, we will have a random groups design which we might fall back upon if we see differences in dimensionality.
k) We could consider a social moderation approach to compare TAKS writing rubric and STAAR writing rubric.
l) We should look at benchmark papers from TAKS when selecting benchmark papers for STAAR.
m) We could consider ways to motivate student performance in the STAAR writing stand-alone field test. For example, we could give them the higher score of their STAAR writing field test and TAKS writing live test.
n) We should consider whether embedding STAAR reading items for Spanish will be feasible given the number of available forms.

Q5: Given the purpose of the linking studies, does the TTAC have any suggestions for linking the current TAKS to the New 3–8 assessments if a new assessment has a different construct and the new assessment items are embedded in the 2011 administration of the current TAKS assessments?

a) No discussion for this question.

Q6: Does the TTAC have any suggestions for addressing the potential issues or reactions to the possible solutions to the potential issues in our current linking approach?

a) No discussion for this question.

Appendix 8. TTAC Meeting Notes from August 2011

Attendees
TTAC: Wayne Camara, Gregory Cizek, Barbara Dodd, Robert Linquanti, Susan Phillips, Rachel Quenemoen, Charlene Rivera, Stephen Sireci, Michael Kolen, Joanne Lenke and Suzanne Lane
TEA: Laura Ayala, Criss Cloudt, Julie Guthrie, Mi-Suk Shim, Tomoko Traphagan, Marianna Vassileva, Victoria Young, Gloria Zyskowski, Ester Regalado, Shannon Housson, Glenn Kirchner, Tong Zhang, Nancy Stevens, Linda Roska, and Pat Sullivan
THECB: Lynnette Heckmann
Pearson: Aimee Boyd, John Cernohous, Sandi Cowes, Laurie Davis, Stacy Duke, Estella Frie, Ezra Hodge, Leslie Keng, Malena McBride, Katie McClarty, Eric Moyer, Dan Murphy, Kimberly O'Malley, Ha Phan, Sonya Powers, Walter Sherwood, Denny Way, Jon Twing, Natasha Williams, Wenyi You, Melinda Taylors, Matthew Gaertner, Phyllis Garrett, Wanchen Chang, Tish Denny, Ian Hembry, Amy LaSalle, Sara Tucker, and Darrel Baker

STAAR Bridge Study
General:

a) The TTAC believes stage 1 (empirical) and stage 2 (impact data) of the proposed STAAR bridge study process provide sufficient support for the state's purposes.
b) The qualitative data in Stage 3 (content judgments) is not likely to capture information that is useful for the bridge study purposes.

1. **What guidance does the TTAC have for communicating about the differences in TAKS and STAAR student performance during the transition to STAAR?**
   a) The state should plan to manage the perception that a drop in campuses meeting AYP is due to the change in the assessment program; other factors such as the increase in AYP targets should be clearly communicated.

2. **Does the TTAC have any feedback regarding the use of impact data?**
   a) Consider the representativeness of the sample, the degree of content overlap, the similarity of item format, and the construct measured by the STAAR and TAKS tests being linked.

3. **Does the TTAC have any feedback regarding the use of content judgments?**
   a) The empirical study results are the more objective data. We could consider combining the results from the linking studies and impact data (e.g., averaging the two resulting cut scores).
   b) Could consider as collateral evidence a comparative analysis of the PLDs for the tests, paying specific attention to the differences at the corresponding performance categories.
   c) This is somewhat similar to what is done with the ID mapping standard setting approach.

4. **What guidance does the TTAC have regarding how other states determine federal accountability when transitioning to new assessment programs?**
   a) To our knowledge, other states have not conducted the content judgments proposed in Stage 3 when transitioning to a new assessment.
   b) Other states generally do conduct bridge studies when transitioning to a new program, and so the idea is well accepted.

STAAR Modified Bridge Study

General:

a) Consider conducting some weighting analyses if the sample in the study is different from the target (STAAR Modified) population.
b) Consider conducting the same type of content overlap analyses that has been done for the general assessments to help support the use of bridge study results for STAAR Modified.

1. **Does the TTAC have any feedback on the proposed approaches for conducting the empirical analyses for STAAR Modified?**
   a) No comments from the TTAC

2. **What guidance does the TTAC have for considering changes in participation requirements and their impact on the data?**
   a) Consider including only students who meet the new participation requirements in the sample.

3. **What guidance does the TTAC have for evaluating the results of the empirical analyses?**
   a) No comments from the TTAC

4. **What guidance does the TTAC have for selecting an analysis method when two are possible?**
   a) Judgments should be made based on the quality of matching variables and/or the common anchor test.

Appendix 9. STAAR Alternate Scoring Rubric

Scoring the Primary Observation

For each Essence Statement, the student's score on **Demonstration of Skill** and **Level of Support** is determined by teacher responses to a series of evaluation questions in the STAAR Alternate online system about student performance for the Primary Observation.

Predetermined Criteria	**Demonstration of Skill***	**Level of Support**
	Did the student demonstrate the skill?	How did the student perform the skill?
1	Yes – 2 points No – 0 points Yes but Needed Prompting – 0 points+	Independently – 2 points Needed Cueing – 1 point Needed Prompting – 0 points N/A – 0 points
2	Yes – 2 points No – 0 points Yes but Needed Prompting – 0 points+	Independently – 2 points Needed Cueing – 1 point Needed Prompting – 0 points N/A – 0 points
3	Yes – 2 points No – 0 points Yes but Needed Prompting – 0 points+	Independently – 2 points Needed Cueing – 1 point Needed Prompting – 0 points N/A – 0 points
	Level 3 task weighted by 1.5 Level 2 task weighted by 1.2 Level 1 task weighted by 1.0	No weighting
Total Points Possible	**9 points**	**6 points**

*Complexity Level (3–most complex, 2–moderately complex, or 1–least complex) is determined by the assessment task the teacher selected and observed the student complete.

+ If a student needs prompting, he or she will not receive any points for Demonstration of Skill (0 points). Prompting guides the student through each step to the end of the assessment task and leads directly to the answer.

After weighting **Demonstration of Skill**, there are a total of **9 points** possible.

There are a total of **6 points** possible for **Level of Support**, which is not weighted. A student who does not demonstrate the skill does not receive any points for Level of Support (see "N/A – 0 points" under Level of Support column in the table above).

Scoring the Generalization Observation

The student is eligible for Generalization if:

- The student is assessed with a Complexity Level 2 or 3 assessment task.
- The skill was successfully demonstrated for all three predetermined criteria.
- There was no prompting on any of the three predetermined criteria.
- The student is assessed using different materials for the Generalization Observation.

Students accessing Complexity Level 1 assessment tasks are not eligible for Generalization of Skill since their performance is being measured at a beginning awareness level.

The student's score on Generalization of Skill is determined by teacher input in response to a series of evaluation questions in the STAAR Alternate online system about student performance for the Generalization Observation.

There are a total of **6 points** possible for **Generalization of Skill**.

- The student will receive **2 points** for each predetermined criterion completed **independently**.
- The student will receive **1 point** for each predetermined criterion completed **with cueing**.
- Any predetermined criteria completed with Prompting will receive 0 points.
- Any predetermined criteria not completed will receive 0 points.

Calculating the Essence Score

Each essence score will be calculated by adding together:

Demonstration of Skill
Level of Support
\+ Generalization of Skill
Essence Score (**21 points** possible)

Calculating the Total Score

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

Essence A Score
Essence B Score
Essence C Score
\+ Essence D Score
Total Score (**84 points** possible)

Appendix 10. TAKS–Alt Scoring Rubric

Scoring the Primary Observation

For each Essence Statement, the student's score on Demonstration of Skill and Level of Support is determined by teacher input in response to a series of evaluation questions in the TAKS–Alt online instrument about student performance for the Primary Observation.

Predetermined Criteria	**Demonstration of Skill**	**Level of Support**
	Did the student demonstrate the skill?	How did the student perform the task?
1	Yes – 2 points No – 0 points	Independently – 2 points Needed Cueing – 1 point Needed Prompting – 0 points N/A – 0 points
2	Yes – 2 points No – 0 points	Independently – 2 points Needed Cueing – 1 point Needed Prompting – 0 points N/A – 0 points
3	Yes – 2 points No – 0 points	Independently – 2 points Needed Cueing – 1 point Needed Prompting – 0 points N/A – 0 points
	Level 3 weighted by 1.5 Level 2 weighted by 1.2 Level 1 weighted by 1.0	No weighting
Total Possible	9 points	6 points

*Complexity Level (3–most complex, 2–moderately complex, or 1–least complex) is determined by the assessment task the teacher selected and observed the student complete.

After weighting **Demonstration of Skill**, there are a total of **9 points** possible.

There are a total of **6 points** possible for **Level of Support**, which is not weighted.

Scoring the Generalization Observation

The student is eligible for Generalization if:

- The student is assessed with a Level 2 or 3 assessment task.
- The skill was successfully demonstrated for all three predetermined criteria.
- There was no prompting on any of the three predetermined criteria

Students accessing the Level 1 assessment task are not eligible for Generalization of Skill since student performance is being measured at a beginning awareness level.

The student's score on Generalization of Skill is determined by teacher input in response to a series of evaluation questions in the TAKS–Alt online instrument about student performance for the Generalization Observation.

There are a total of **3 points** possible for **Generalization of Skill**.

- The student will receive 1 point for each predetermined criteria completed either independently or with cueing.
- Any predetermined criteria completed with Prompting will receive 0 points.
- Any predetermined criteria not completed will receive 0 points.

Calculating the Essence Score

Each essence score will be calculated by adding together:

	Demonstration of Skill
	Level of Support
+	Generalization of Skill
	Essence Score (**18 points** possible)

Calculating the Total Score

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

	Essence A Score
	Essence B Score
	Essence C Score
+	Essence D Score
	Total Score (**72 points** possible)

Appendix 11. TTAC Meeting Notes from November 2011

Attendees
TTAC: Robert Linquanti and Stephen Sireci
TEA: Cari Wieland, Pat Otto, Debbie Owens, and Janet Borel
Pearson: Aimee Boyd, Barbara Poynter, Melinda Taylor, and Natasha Williams

STAAR Alternate Bridge Study
General:

a) Due diligence has been taken in determining the weights used to incorporate complexity level in the STAAR Alternate scoring rubric.
b) The proposed bridge study is a logical approach.

1. Does the TTAC have feedback regarding the STAAR Alternate Bridge Study process?

a) Could use logistic regression as a more straightforward approach to classification accuracy; results will likely be similar to the proposed classification accuracy approach (Livingston & Zieky, 1989).
b) Suggest using three methods: (1) classification accuracy, as proposed; (2) logistic regression, as verification of classification accuracy; and (3) impact data, as proposed.

2. Does the TTAC have suggestions for resolving differences, if needed?

a) Consider evaluating the impact of results from the three methods on districts' and schools' AYP decisions.
b) The decision on which cut score to use for the bridge study is a policy decision which would be informed by the study results and an evaluation of the consequences associated with those results.
c) Could consider combining the results (e.g., averaging the resulting cut scores).
d) Consider cross validating the logistic regression analyses, as an evaluation of the method.

3. What guidance does the TTAC have regarding how other states determine federal accountability when transitioning to new assessment programs?

a) In transitioning to new assessment programs (e.g. ELP assessment), some states "hold harmless" LEAs for increased rigor of new assessment during transition year by adjusting performance targets downward (i.e., percent of students expected to meet criteria) to hold constant the percent of LEAs that would have met performance target using old criteria – i.e., percentile rank of LEA performance distribution) -- particularly if this difference is small.

Appendix 12. References

Erpenbach, W. J. (2008). *Statewide educational accountability systems under the NCLB Act – A report on 2008 amendments to state plans.* Council of Chief State School Officers, Washington, DC.

Erpenbach, W. J. (2011). *Statewide educational accountability systems under the NCLB Act – A report on 2009 and 2010 amendments to state plans.* Council of Chief State School Officers, Washington, DC.

Forte, E. and Erpenbach, W. J. (2006). *Statewide educational accountability systems under the NCLB Act – A report on 2006 amendments to state plans.* Council of Chief State School Officers, Washington, DC.

Kolen, M. J., & Brennan, R. L. (2004). *Test equating: Methods and practices* (2nd ed.). New York: Springer.

Linacre, J. M. (2001). *WINSTEPS Rasch Measurement Program, Version 3.32*. Chicago: John M. Linacre.

Livingston, S. A., & Zieky, M. J. (1989). A comparative study of standard-setting methods. *Applied Measurement in Education, 2 (2),* 121–141.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*, 149-174.

Pommerich, M., Hanson, B. A., Harris, D. J., & Sconing, J. A. (2004). Issues in conducting linkages between distinct tests. *Applied Psychological Measurement*, 28(4), 247-273.

Rasch, G. (1966). An Individualistic Approach to Item Analysis. In *Readings in Mathematical Social Science*, edited by Paul F. Lazarfeld and Neil W. Henry. Chicago, IL: Science Research Associates.

Skaggs, G., & Wolfe, E. W. (in press). Equating designs and procedures used in Rasch scaling. In Smith, Jr., E.V., & Stone, G.E. (Eds.). *Criterion-referenced Testing: Practice analysis to score reporting using Rasch measurement models*. Maple Grove, MN: JAM Press.

Way, W. D., Davis, L. L., & Fitzpatrick, S. (2006). *Score comparability of online and paper administrations of the Texas Assessment of Knowledge and Skills*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, CA.

Wright, B. D. & Masters, G. N. (1982). *Rating Scale Analysis*. Chicago, IL: Mesa Press.

Wright, B. D. & Stone, M. H. (1979). *Best test design: Rasch measurement*. Chicago, IL: Mesa Press.