Chapter 15: Scaling

Rationale

The basic score on any test is the raw score, which is the number of items correct. However, the raw score alone does not present a broad picture of test performance because it can be interpreted only in terms of a particular set of test questions. When new test forms are administered in subsequent administrations, the questions on the new forms may be different. The set of questions on one test may be slightly easier or slightly harder than the set of questions that were on another test. Because the overall difficulty of the tests may vary, the raw scores or percentage correct cannot be directly compared to indicate differences in student performance. For example, 75% of students may earn a raw score of 34 or higher on a test in year 1, and 80% of students may earn a raw score of 34 or higher in year 2. Now assume that the questions on the test in year 2 were slightly easier than that those on the test in year 1. Has student performance improved or are the questions just easier?

Unlike raw scores, scale scores do allow direct comparisons of student performance between specific sets of test questions from different test administrations. A scale score is a conversion of the raw score onto a "scale" that is common to all test forms for that assessment. The scale score takes into account the difficulty level of the specific set of questions. It tells us about a student's performance relative to the passing standards or proficiency levels for Texas Assessment of Knowledge and Skills (TAKS), TAKS–Modified (TAKS–M), Texas English Language Proficiency Assessment System (TELPAS) reading, and Algebra I end-of-course (EOC) assessments.

Scaling is the process of creating the scale scores. This process involves using specialized statistical methods to determine the difficulty of a particular set of test questions. The specific statistical methods used for scaling of Texas assessments are described below in the Rasch Partial-Credit Model section of this chapter. The initial result of the scaling process provides a "scale" that is common to all test forms for that assessment but does not have desirable properties for reporting. To facilitate interpretation of scaled test scores by users of the test, this initial scale is transformed to a more user-friendly metric. The transformation is described for TAKS, TAKS–M, TELPAS reading, and Algebra I EOC in the sections below. For the TAKS–Alternate (TAKS–Alt) assessment, no scaling is conducted. Raw scores are reported based on the TAKS–Alt rubric. For more information on the TAKS–Alt assessment, see Chapter 3: Assessments for Students with Disabilities.

Rasch Partial-Credit Model

Test items (multiple-choice, gridded response, short-answer, and essay) for all Texas assessments are scaled and equated using the Rasch Partial-Credit Model (RPCM). The

RPCM is an extension of the Rasch one-parameter Item-Response Theory (IRT) model attributed to Georg Rasch (1980), as extended by Wright and Stone (1979), Masters (1982), Wright and Masters (1982), and Linacre (2001).

The RPCM was selected because of its flexibility in accommodating multiple-choice (correct/incorrect) data as well as multiple-response category data, and for its ability to maintain a one-to-one relationship between derived scores (i.e., scale scores) and the raw scores. It is the underlying Rasch scale that facilitates equating of multiple test forms and allows for comparisons of student performance across years. Additionally, the underlying Rasch scale facilitates the critical maintenance of equivalent performance standards across years. 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:

$$P_{xni} = \frac{\exp \sum_{j=0}^{x} (\theta_{n} - \delta_{ij})}{\sum_{k=0}^{m_{i}} \exp \sum_{j=0}^{k} (\theta_{n} - \delta_{ij})}, i = 0, 1, ..., m_{j},$$

where

$$\sum_{j=0}^{0} \left(\theta_n - \delta_{ij}\right) \equiv 0 ,$$

$$\sum_{j=0}^{k} \left(\theta_n - \delta_{ij}\right) \equiv \sum_{j=1}^{k} \left(\theta_n - \delta_{ij}\right), \text{ and}$$

$$\sum_{i=0}^{m_i} P_{xni} = 1 .$$

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*. (See 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 application of the RPCM means that all multiple-choice items and open-ended tasks will be placed on the same scale. All common item- and step-difficulty estimates will be on the same underlying logistic scale as that of the student proficiency level estimates. Estimates of items being field-tested can be obtained from a form-by-form or a concurrent calibration, with the common item set serving as an anchor. In this way, all field-test items can be placed on the same logistic scale as that of the common items.

At the conclusion of these calibrations, all item- and task-difficulty estimates as well as all student proficiency level estimates are directly comparable because they are on the same underlying logistic scale.

The advantages of an RPCM scaling include the following:

- All items, independent of type, are placed on the same common score scale.
- The RPCM provides the same score scale onto which students' achievement results are placed. Hence, direct comparisons can be made regarding the kinds of items students with various achievement levels can answer. This is very helpful in describing the results of the assessments to students, parents, and teachers.
- All field-test items can be placed on the same scale as those of the live, or operational, part of the assessment. This is invaluable in linking student performance back to all banked items and useful in the construction of multiple future forms that are psychometrically balanced.
- This design allows for the pre-equating of future test forms, which is a valuable component of the complex test construction process.
- Such an approach supports post-equating of the test. In this way, a link is established between previous forms and the current administration. This current form is on the same scale as the previous forms so that comparisons in form difficulties and passing rates can be ascertained. Because both pre-equated and post-equated item difficulty estimates are available, any difficulty drift or scale drift can be quantified.
- Establishing a common scaling allows for the direct comparison of performance-level standards established by the State Board of Education (SBOE) for future test forms.

TAKS

Scale Scores

The RPCM model provides an underlying "ability" scale for the TAKS tests, but because it is centered on zero and has both negative and positive values it is not a metric that is useful for reporting purposes. For the TAKS tests, scale scores have been developed through a linear transformation of the underlying Rasch "ability" scale to one that is easier to understand because it does not have negative numbers. Use of this scale not only is easier to understand but it can ensure that the performance standards are maintained at the same level of difficulty across administrations. The TAKS scale scores, which are known as "derived" scores, are useful in describing different aspects of student performance and maintaining performance standards over test administrations.

Derived scores are computed using the one-parameter IRT or RPCM. The advantage of using IRT models in scaling is that all the items measuring performance in a particular content area can be placed on the same scale of difficulty. The further value of the Rasch model over more complex IRT models is that the Rasch model assumes that for each total score point, there is only one student proficiency (or ability) estimate. This relationship allows the Rasch difficulty values for individual items to be used in computing a Rasch ability level for any total score point on any test constructed from these items.

The SBOE established the performance standards for most TAKS tests independently at each grade level and test content area in November 2002. During the spring 2003 operational test administration, tests were initially calibrated onto a Rasch partial-credit model scale. For TAKS developed since 2002, the SBOE established performance standards and the initial calibrations onto a Rasch partial-credit model scale have been conducted. Calibration of the TAKS operational test data was accomplished by Pearson with independent verification of the analyses performed by the Texas Education Agency (TEA) and an external psychometric consultant. The extensive verification procedure was part of a TEA quality assurance plan that was implemented to ensure the accuracy of the results of the Rasch partial-credit scaling of TAKS.

A unique scale transformation was then developed in each grade and content area so that the resulting set of scale scores would have the panel-recommended Met Standard performance level cut set at a scale score of 2100 and the panel-recommended Commended Performance level cut set at a scale score of 2400. It was felt that establishing the recommended cut scale scores to have the same value, regardless of test or grade level, would aid in the interpretability of the scale scores. This linear transformation of the underlying Rasch proficiency level estimate is as follows:

$$SS_j = (\theta_j \times T\mathbf{1}) + T\mathbf{2},$$

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 *T*1 and *T*2 are scale score transformation constants that

establish the scale score system so 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 *T*1 and *T*2 are provided in Tables 8 and 9.

These linear transformations established the original scale score system based on the Rasch partial credit scaling of the spring 2003 test results. Statistical equating has been applied to maintain the same level of difficulty for newly developed forms.

The resulting TAKS scale score system has a range of approximately 1000 to 3200. For tests containing constructed-response items (open-ended or essay questions), it is important to note that the total score is a combination of the number-correct score on the multiple-choice questions and the number of points achieved on the constructed-response questions.

For the grade 10 and exit level English language arts (ELA) tests, the total score to Rasch proficiency level estimate (and the subsequent scale score) table incorporates a weighted essay score (essay score \times 4). This helps ensure that the appropriate value is placed on the direct writing sample given the amount of time and effort put into writing to the essay prompt. Thus, for ELA, the total number of attainable score points after weighting is greater than would be the case if each question was worth one point.

Additionally, scale scores for writing and ELA are impacted by the essay score requirement of the standards. For writing and ELA, a student is required to attain a score of 2 or higher on the essay prompt to achieve Met Standard. For writing, a student is required to attain an essay score of 3 or higher on the essay prompt to achieve Commended Performance. Essay score requirements of the standards can be found online at http://www.tea.state .tx.us/student.assessment/taks/standards/scalescorecuts0305.pdf.

If a student receives a score of "0" or "1" on the essay prompt, the highest scale score he or she can receive is one scale score point less than Met Standard. For example, at the panel-recommended standard of 2100, the highest scale score a student can receive if he or she scores below a 2 on the essay prompt is 2099. All students receiving a "0" or "1" on the essay prompt with scale scores higher than this value as obtained through the Rasch calibration have their scores artificially "re-mapped" to this value to reflect the essay score requirement of the passing standard. Similarly, for writing, the highest scale score a student can receive if he or she scores below a 3 on the essay prompt is 2399 (one scale score point less than the Commended Performance standard of 2400). Students with scale scores above this value based on the Rasch calibration will have their scores re-mapped. This can be observed as a spike (large number of students) at the re-mapped value in the scale score distribution in Appendix B.

English	<i>T</i> 1	T2
Gr 3 Reading	125.89173	1992.23668
Gr 3 Mathematics	146.69927	1967.23716
Gr 4 Reading	129.42192	1996.07420
Gr 4 Mathematics	142.51781	1976.29454
Gr 4 Writing	110.88114	1981.33501
Gr 5 Reading	155.92516	1954.52183
Gr 5 Mathematics	170.35775	1939.18228
Gr 5 Science	187.96992	1832.51880
Gr 6 Reading	166.38935	1988.85191
Gr 6 Mathematics	174.31726	1987.91400
Gr 7 Reading	139.08206	1964.53408
Gr 7 Mathematics	133.98839	2046.53863
Gr 7 Writing	135.59322	2002.82034
Gr 8 Reading	153.76730	1948.53921
Gr 8 Mathematics	153.68852	2025.61475
Gr 8 Social Studies	145.41929	2085.16723
Gr 9 Reading	123.21847	1944.27650
Gr 9 Mathematics	184.61538	2009.90769
Gr 10 ELA	97.06539	1983.74478
Gr 10 Mathematics	141.04372	2038.64598
Gr 10 Science	160.42781	1996.84492
Gr 10 Social Studies	145.20813	2046.85382
Gr 11 ELA	113.48162	2017.62369
Gr 11 Mathematics	140.58107	2064.71415
Gr 11 Science	129.47777	2070.86750
Gr 11 Social Studies	126.47555	2093.29680

Table 8. Scale Score Transformation Constants for the TAKS Tests(English versions)

Table 9. Scale Score Transformation Constants for the TAKS Tests(Spanish versions)

Spanish	<i>T</i> 1	<i>T</i> 2
Grade 3 Reading	148.66204	1995.19326
Grade 3 Mathematics	146.69927	1968.26406
Grade 4 Reading	165.10732	2006.21904
Grade 4 Mathematics	198.15059	1923.64597
Grade 4 Writing	151.04980	1998.90237
Grade 5 Reading	190.23462	1967.02600
Grade 5 Mathematics	190.71837	1915.95677
Grade 5 Science	189.27455	1841.07256
Grade 6 Reading	187.96992	2057.89474
Grade 6 Mathematics	202.56583	1970.76300

Following the spring 2003 operational test calibration analyses, these linear transformations were applied to the resulting Rasch student proficiency (ability) estimates at each total

score point, yielding the final raw score to scale score conversion tables. These tables can be found online at http://www.tea.state.tx.us/student.assessment/scoring/convtables/ index.html.

Vertical Linking

TAKS is a standards-referenced assessment reflecting the curriculum as specified in the TEKS at each grade level. As part of the process for setting standards for student performance, groups of Texas educators participated in advising the SBOE on a recommended score point for each subject area at which students are assumed to have sufficient mastery of the TEKS student expectations at that grade level. The TAKS scale score system was set so that a scale score of 2100 is the minimum attainable panel-recommended Met Standard score and a scale score of 2400 is the minimum attainable Commended Performance score at each grade level and for each content area (though there are additional requirements for ELA). It was argued that such a scoring system, once the standards phase-in was completed, would be easier for students, parents, schools, and the public to understand since the meaning of a scale score of 2100 and 2400 would remain the same regardless of grade and subject.

These scales are grade and subject specific; however, they cannot be compared from one year to another. Unlike TAAS, TAKS currently does not have a vertical scale score system, but one will be put in place for spring 2009 in grades 3–8 reading and math. TAKS also did not have a measure of student-level growth from grade to grade, like the Texas Learning Index (TLI), in 2008. The Texas Growth Index (TGI) provided a measure of growth; however, the growth is intended only for interpretation at aggregate educational units, such as campuses and districts. The TGI is calculated at the student level, but the reliability of student-level growth is not strong enough for interpretation at the student level. The TGI may add value to campuses struggling to show the movement of students within and across the proficiency levels. For more information about the TGI, see Chapter 12: Texas Growth Index. Texas will implement a new student-level growth measure in spring 2009.

TAKS-M

The first operational administration of TAKS–M occurred in spring 2008. Performance standards for TAKS–M were set in August 2008. The tests were scaled so that, like TAKS, cuts are at 2400 and 2100 for Commended Performance and Met Standard, respectively. The scale transformation values *T*1 and *T*2 were determined after standard setting using the same approach used for TAKS. A vertical scale for TAKS–M has not been established. Additional information about TAKS–M can be found in Chapter 3: Assessments for Students with Disabilities.

TELPAS Reading Tests for Grades 2–12

Rather than linking performance to grade-level expectations in the traditional sense, Texas English Language Proficiency Assessment System (TELPAS) measures performance in terms of language proficiency levels that describe what second language learners can read and understand at various stages of English acquisition. TELPAS reports performance at four language proficiency levels—beginning, intermediate, advanced, and advanced high. Students who enter U.S. schools knowing no English, regardless of their grade level at the time of entry, progress from one proficiency level to the next as they become fluent in English.

Both TAKS and TELPAS reading tests measure many of the same reading skills. However, the manner in which TELPAS assesses these skills reflects the stages of second language acquisition which occur on a continuum spanning from little or no knowledge of English to English fluency. TELPAS reading tests are constructed as four mini-tests within a test. Each English language learner takes the entire assessment. The English used on each mini-test is appropriate for students at that stage of English acquisition. The stages of English acquisition are defined in proficiency level descriptors within the English language proficiency standards (ELPS) of the TEKS curriculum.

The TELPAS reading test is based on an underlying Rasch vertical scale that allows the test results to track individual student progress in English proficiency over time. In conjunction with the development of the second version of this assessment, vertical scaling was obtained via a spring 2008 TELPAS reading scaling study. A sample of items, called anchor items, were selected from the TELPAS reading item bank to be administered within in the spring 2008 live administration in place of field-test item locations. The selected items were administered on both their intended grade level/clusters and adjacent grade level/clusters. When identifying those items, Pearson and TEA content experts verified the appropriateness of the items to be administered at the adjacent off-grade level/clusters. This type of design is very common in vertical linking studies in educational testing.

Using grade 2 as the base form (for vertical linking purposes), the vertical linking constants are cumulative across the TELPAS reading grade clusters with respect to scale distance from the base form scale (in this case grade 2). These cumulative vertical linking constants (referred to as the scaling constants) are equal to the vertically scaled mean item difficulty of the test items at each grade cluster. Since grade 2 was used as the base (and a scaling constant of zero), it would have a scaled mean item difficulty value of zero (see Table 10 for vertical scaling constants for TELPAS reading).

Grade Cluster	Constant		
2	0.000		
3	0.740		
4–5	1.169		
6–7	1.248		
8–9	1.616		
10–12	1.902		

Table 10. Vertical Scaling Constants for TELPAS Reading

As with TAKS, the underlying Rasch scale does not have desirable properties for reporting purposes, so the final student proficiency (ability) levels are subjected to a linear transformation in order to derive the TELPAS reading scale scores. Detailed information about the scale transformation process and scale score cuts can be found in the "TELPAS Reading Vertical Scale" report in the 2008 Texas Education Agency Technical Report Series. As can be seen from the vertical scaling constants, the TELPAS reading vertical scale is centered on the grade 2 cluster as the zero point. The TELPAS reading scale score transformation is as follows:

$$SS_i = (\theta_i \times \mathbf{48}) + \mathbf{575},$$

where θ_i is the vertically scaled Rasch student proficiency level for student *j*.

This scale score system results in scale scores in the range of 200 to 999 and is flexible enough for development of future test forms. TELPAS reading student proficiency levels based on this scale score metric were then determined based on the spring 2008 test, using the final raw score proficiency-level cuts as shown in Table 11.

TELPAS Grade Clusters	Beginning SS Range	Intermediate SS Range	Advanced SS Range	Advanced High SS Range
2	550 and below	551–614	615–669	670 and above
3	596 and below	597–647	648–698	699 and above
4–5	609 and below	610–667	668–717	718 and above
6–7	612 and below	613–673	674–730	731 and above
8–9	632 and below	633–680	681–737	738 and above
10–12	643 and below	644–703	704–756	757 and above

Table 11. Scale Score Ranges Associated withTELPAS Reading Proficiency Levels

Descriptors of the four TELPAS reading proficiency levels are found in Chapter 10: Scores and Reports. Additional information about TELPAS reading can be found in Chapter 4: Texas English Language Proficiency Assessment System (TELPAS).

Algebra I End-of-Course Assessment

Scale Scores

TEA established the new performance standards for the Algebra I EOC assessment in November 2005, replacing the old performance standards used in 2003 and 2004, which only had one cut scale score (at 1500) and two performance levels. Using a procedure similar to TAKS, a unique scale transformation was developed so that the resulting set of scale scores would have the panel-recommended Commended Performance cut set at a scale score of 1400 and the Met Standard cut set at a scale score of 1100. This 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 model proficiency level estimate for student j, and T1 and T2 are scale score transformation constants that establish the scale score system so that a scale score of 1400 is the cut score for Commended Performance and a scale score of 1100 is the cut score for Met Standard. Values for T1 and T2 are 155.0468 and 1009.0186, respectively.

This linear transformation established the original scale score system based on the Rasch dichotomous scaling of the spring 2005 test results. Because the new standards were set after the reporting of spring 2005 administration results, the performance levels for students were not provided for the students tested in that administration; instead, the test report displayed their raw scores. Fall 2005 was the first time the performance standards were used with the new Algebra I EOC assessment scores.

Quality Control

The scaling process for TAKS, TAKS–M, TELPAS reading, and the Algebra I EOC assessment is independently conducted by at least two different psychometricians at Pearson. Once each party completes the Rasch calibrations and applies the scaling transformation, the separate results are compiled. These compiled results are reviewed for differences. If any differences are detected, the results and procedures are reviewed until consensus is reached.

TAAS Exit Level

The Texas Assessment of Academic Skills (TAAS) testing program is still ongoing, in retest form, for students for whom TAAS is their high school graduation requirement. TAAS provided two derived scores that described different aspects of student performance: scale scores and Texas Learning Index (TLI) scores.

Scale Scores

For the TAAS tests, scale scores were developed to maintain performance standards at the same level of difficulty across administrations. The original test form, on which the 70% correct standard was established (see Chapter 14: Standards), was calibrated using the Rasch model. This calibration produced a relationship between the raw score and the Rasch achievement score, which was then transformed so that 1500 represented the passing standard. This transformation is reproduced in the formula

TAAS Scale Score =
$$\left(\frac{\theta - \theta_{at Standard}}{\sigma_{\theta}}\right) \times 200 + 1500$$
,

where θ is the Rasch student proficiency level estimate, $\theta_{at \ Standard}$ is the Rasch student proficiency level associated with the 70% raw score, σ_{θ} is the standard deviation of the Rasch student proficiency levels, and 200 and 1500 are the spread and centering constants, respectively. This transformation established the original scale score system, and statistical equating was required to maintain the same level of difficulty for newly developed forms. The resulting TAAS scale has a range of approximately 400 to 2400, with 1500 corresponding to 70% of the items correct on the first administration of the test, when the passing standards were set.

Texas Learning Index

The Texas Learning Index (TLI) was first developed in spring 1994 to better meet the needs of districts and students for longitudinal comparability. A metric with two essential characteristics was sought. First, such a metric should provide an index of student achievement toward the goal of passing the TAAS exit level test. Second, the metric should permit comparisons between administrations and between grades for use in the accountability system. The TLI provides a means for schools to be able to demonstrate improvements in their instructional programs, even in cases where the passing standard has not yet been met or the passing standard has been exceeded. Likewise, with a derived score such as the TLI, individual students are able to demonstrate improvement regardless of their current achievement relative to the passing standard.

T-Score Type Transformation

The requirements listed above led to the consideration of a vertical scale score system for the TAAS examinations that would place results for grades 3–8 and the exit level test all on the same scale. TEA convened a panel of measurement experts from across the nation to advise the agency regarding such a scaling. This panel was composed of educators, test publishers, and educational consultants. The committee expressed two main concerns regarding a vertical scaling system. First, placing both grade 3 and exit level students on the same scale could lead to misinterpretations because of the large difference in the content of the test items at these grades. Second, a vertical scale implies a linear and well-defined curriculum from grades 3 through exit level when such a wellordered curriculum may not be in place. The committee concluded that a vertical scale would not meet the needs of TEA and offered an alternative proposal of using a transformed "T-score" type of scale. A transformed T-score is expressed in terms of standard deviation units away from the mean. For example, if a student earns a raw score of 50 on a test with a mean of 40 and a standard deviation of 10, this student's score is one standard deviation unit above the mean. Traditionally, such a score is referred to as a standard score, or z-score, and can be reproduced with the following formula:

$$z = \frac{(Observed Score - Mean)}{Standard Deviation} \cdot$$

The student in the example above has an observed score of 50 and a z-score of 1.00. Because standard scores have decimals and typically range from -3.00 to 3.00, an additional transformation is usually made to simplify the reporting. A common transformation in which the scale in this example is "re-anchored" to have a mean of 50 and a standard deviation of 10 is often referred to as a T-score transformation. The following formula provides the T-score transformation:

 $T = (z \times 10) + 50.$

Such a transformation simply renames the z-score. After the transformation, the student in the example would have the following scores: a raw score of 50, a z-score of 1.00, and a T-score of 60.

TLI Established at Exit Level

The TLI is very much like the T-score described above. Unlike the T-score, however, the TLI is anchored at the exit level passing standard rather than at the mean of the distribution. To distinguish between the scale score system and the TLI, TEA chose a two-digit metric for the TLI so that it is anchored at the exit level passing standard with a value of 70 and a standard deviation of 15. The TLI is derived by the following formula:

$$TLI = \left[\left(\left(\frac{(Observed Score - Mean)}{Standard Deviation} \right) - z - Score at Passing Standard \right) \times 15 \right] + 70.$$

Because a TLI of 70 represents the passing standard, there is no difference in interpretation between a student who scores 1500 under the scale score system and a student who scores 70 on the TLI scale.

Assumptions

An assumption of the TLI is that the reference distributions on which the TLI scale was constructed should be used for all future TAAS scoring. Because the TLI is a distributionalbased metric relying on a z-score transformation, it is normative in nature. Because of this norm-referenced component, recalculating the TLI each year would make year-toyear comparisons impossible. For this reason, all TAAS administrations describe student performance in terms of the population tested in spring 1994. Additional information about the TAAS scaling, including the TLI derived score, can be found in the *2001–2002 Technical Digest*, which is located at http://www.tea.state.tx.us/ student.assessment/resources/techdigest/.

Frequency Distributions and Descriptive Statistics

Appendix B provides frequency distributions and summary statistics for TAKS, TAKS–M, TAKS–Alt, TELPAS reading, and Algebra I EOC scale scores and the EOC biology and geometry raw scores. Appendix C provides mean p-values by objective and subject area and internal consistency estimates for TAKS, TELPAS reading, and EOC tests taken. Descriptive statistics for TAKS–Alt are reported in the "2007–2008 TAKS–Alt Technical Report" in the 2007 TEA Technical Reports Series which can be found at http://www.tea.state.tx.us/student.assessment/resources/techdigest/.