Chapter 6 Texas English Language Proficiency Assessment System (TELPAS)

Overview

The Texas English Language Proficiency Assessment System (TELPAS) measures the progress that English language learners (ELLs) make in acquiring the English language. Title III, Part A of the Elementary and Secondary Education Act (ESEA) requires states to conduct annual statewide English language proficiency assessments for ELLs in grades K–12 in the language domains of listening, speaking, reading, and writing. Prior to No Child Left Behind, Texas developed and administered English language proficiency tests in the domain of reading, as required by Texas state law.

The TELPAS assessments are performance-based and holistically rated, with the exception of the reading assessments for grades 2–12, which are multiple-choice tests. For each language domain, TELPAS measures four levels, or stages, of increasing English language proficiency: beginning, intermediate, advanced, and advanced high.

TELPAS measures learning in alignment with the Texas English Language Proficiency Standards (ELPS) that are a part of the Texas Essential Knowledge and Skills (TEKS)

curriculum. The ELPS outline the instruction that ELLs must receive to support their ability to develop academic English language proficiency and acquire challenging academic knowledge and skills. The ELPS are composed of second language acquisition knowledge and skills that ELLs are expected to learn, as well as proficiency level descriptors characterizing the four English language proficiency levels reported in Texas.

TELPAS and the State of Texas Assessments of Academic Readiness (STAAR®) are used to show the extent to which districts and the state meet federal Annual Measurable Achievement Objective (AMAO) accountability indicators that are specific to the English language proficiency and academic achievement of ELLs. Composite performance rather than individual language domain performance is used in TELPAS AMAO indicators. For information about how TELPAS composite results are generated, refer to the TELPAS Comprehension and Composite Scores section in this chapter. More information about AMAOs is available on the Texas Education Agency (TEA) website.

The TELPAS results are also used at the student level to help teachers design instruction and plan interventions that appropriately address the student's linguistic and academic needs.

Participation Requirements

All K–12 ELLs are required to participate in TELPAS, including students classified as limited English proficient (LEP) whose parents have declined bilingual/English as a second language (ESL) program services. ELLs are required to be assessed annually until they meet bilingual/ESL program exit criteria and are reclassified as non-LEP. The rare circumstances in which a student might not be required to participate in one or more TELPAS language domains include:

Committee Decisions

In rare cases, it might be necessary for the admission, review, and dismissal (ARD) committee, in conjunction with the language proficiency assessment committee (LPAC), to determine that an ELL receiving special education services should not be assessed in reading, writing, listening, and/or speaking for reasons associated with the student's particular disability. Participation must be considered on a domain-by-domain basis. The reason for not assessing the student must be related to the student's disability and be well-supported and documented in the student's individualized education program (IEP) by the ARD committee and the student's permanent record file by the LPAC.

Newly Enrolled ELLs—Holistically Rated Domains

An ELL from another Texas school district, state, or country who enrolls on or after the first day of the TELPAS testing window will not be assessed by the receiving district in

the holistically rated domains. However, newly enrolled students in grades 2–12 are required to take the TELPAS reading assessment.

Test Development

TELPAS Reading Assessments for Grades 2–12

The TELPAS reading tests for grades 2–12 employ a multiple-choice answer format. Six grade-cluster tests are administered, as shown in Table 6.1.

Table 6.1. Grade Clusters for 2–12 Reading

Grade Clusters for 2–12 Reading
2
3
4–5
6–7
8–9
10–12

As with other components of the Texas assessment program, TEA involves educators and assessment experts in the TELPAS test development process. As part of the ongoing process to replenish the item banks, committees of Texas educators continue to review annually developed field-test items.
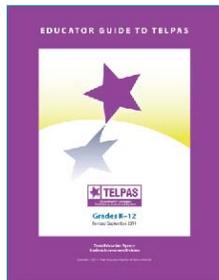
More information about the TELPAS reading tests for grades 2–12, including item specifications and samples, is available in the Educator Guide to TELPAS available on the TEA's Student Assessment Division website. This guide is provided to familiarize educators with TELPAS. It shows the integral relationship between TELPAS and the ELPS, and includes explanatory information as well as links to student video segments, authentic student writing samples, and sample test questions.

TELPAS Holistically Rated Assessments

The TELPAS holistically rated components assess reading in grades K–1 and listening, speaking, and writing in grades K–12. To conduct these assessments, teachers are specially trained to rate the English language proficiency of ELLs based on an evaluation of student writing, classroom observations in core content areas, and daily interactions with the students.

The rating process is designed to identify a student's level of English language proficiency and is holistic rather than a measure of isolated skills. Teachers are trained to use the ELPS proficiency level descriptors (PLDs) as holistic rating rubrics to assign

proficiency ratings of beginning, intermediate, advanced, or advanced high in each domain assessed. More information on the PLDs is provided in the Scores and Reports section.

TEA developed the TELPAS holistically rated components in collaboration with test development experts; bilingual/ESL consultants; and members of an ELL focus group composed of teachers, bilingual/ESL directors, assessment directors, campus administrators, and university professors. Like the TELPAS reading tests for grades 2–12, these assessments are aligned with the ELPS and designed to assess the English communication skills that ELLs need to engage meaningfully and effectively in learning the academic knowledge and skills required by the state. The holistically rated assessments draw upon second language acquisition research, research-based standards, the experience of Texas practitioners, and observational assessment practices.

Together with the TELPAS reading tests for grades 2–12, the holistically rated components of TELPAS combine classic multiple-choice testing methods with authentic, performance-based assessments to measure the construct of academic English language proficiency. The process of rating students holistically helps teachers better understand and meet the educational needs of ELLs and avoids drawbacks associated with adding assessments and field-testing activities that take students away from needed instruction.

More information about the TELPAS holistically rated assessments is available in the Educator Guide to TELPAS on TEA's Student Assessment Division website.

Training

Each year, TELPAS raters participate in holistic rating training activities in preparation for providing accurate and reliable TELPAS scores.

The TELPAS rater training activities are primarily Web-based. However, a wide variety of paper-based training materials are also prepared so that regional and district trainers can conduct face-to-face training on administration procedures and provide raters with supplemental support as needed.

The TELPAS online basic training courses are provided to teach raters the essentials of second language acquisition theory and how to use the PLDs from the ELPS to accurately identify the English language proficiency levels of their ELLs based on how well the students understand and use English during daily academic instruction and classroom interactions. The online courses contain numerous practice rating activities that comprise student writing collections and video segments in which ELLs demonstrate their reading (K–1 only), speaking, and listening skills in authentic Texas classroom settings. The courses give teachers practice applying the scoring rubrics (PLDs) and provide teachers with detailed feedback about their rating accuracy. In the 2014–2015 school year, more than 43,000 online training courses were completed.

Each year, all raters are also required to complete online calibration activities to demonstrate their ability to apply the PLD rubrics consistently and accurately before they rate students for the operational assessment. Beginning in the 2010–2011 school year, calibration activities were provided for all holistically rated domains—listening, speaking, reading (K–1 only), and writing. There are two sets of initial calibration activities, with 10 students per set, and all applicable language domains are represented. In order to demonstrate sufficient calibration, raters are required to rate at least 70 percent of students correctly within a set. Raters finish the calibration activities when they demonstrate sufficient accuracy. If sufficient accuracy is not obtained on the first or second set, supplemental training is provided to the rater, followed by a third and final online calibration set. Individuals not successful on the final set are either not used as raters (a district decision) or are provided rater support in accordance with test administration procedures. In the 2014–2015 school year, approximately 123,000 teachers successfully calibrated within the three attempts. The percentage of successfully calibrated teachers by the end of the third set was approximately 95 percent.

Test Administrations

During the 2014–2015 school year, approximately 860,000 TELPAS assessments were administered. Districts administered the TELPAS assessments to eligible students as indicated in Table 6.2.

Table 6.2. TELPAS Assessments Administered in 2014–2015

TELPAS Assessment		Assessments Administered			
Grade/Grade Cluster	Grade	Listening	Speaking	Reading	Writing
Grades K–1	Kindergarten	108,256	108,177	108,042	108,016
	Grade 1	115,881	115,779	115,566	115,548
Grade 2	Grade 2	111,676	111,579	111,378	111,176
Grade 3	Grade 3	105,930	105,827	105,624	105,408
Grades 4–5	Grade 4	93,341	93,256	93,101	92,899
	Grade 5	78,434	78,392	78,277	78,068
Grades 6–7	Grade 6	61,311	61,272	61,228	61,016
	Grade 7	48,594	48,563	48,601	48,307
Grades 8–9	Grade 8	42,139	42,128	42,187	41,938
	Grade 9	38,954	38,940	38,876	38,714
Grades 10–12	Grade 10	25,866	25,853	25,685	25,690
	Grade 11	19,197	19,190	19,107	19,081
	Grade 12	14,667	14,658	14,464	14,499

Administration procedures that support the integrity of the assessment process are a vital part of standardized testing. For the holistically rated components of TELPAS, district personnel involved in the test administrations sign security oaths, verify the correct assembly and contents of student writing collections, and implement procedures to support the validity and reliability of the rating process.

Details about the TELPAS holistic rating training and administration procedures, including descriptions of the Web-based training components, are found in the TELPAS section of the District and Campus Coordinator Manual for the Texas assessment program and in the TELPAS Manual for Raters.
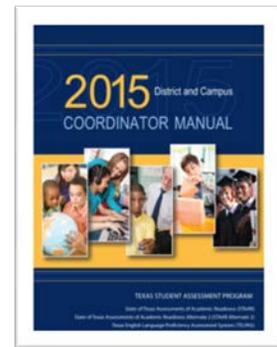
Scores and Reports

English language proficiency tests are not designed to measure mastery of learning objectives with a pass or fail score because the process of acquiring and becoming academically proficient in a second language takes longer than a school year. The TELPAS results provide an annual indicator of where each ELL is on a continuum of English language development designed for second language learners. This continuum is divided into four proficiency levels: beginning, intermediate, advanced, and advanced high. The progress of students along this continuum is the basis for the TELPAS reporting system, which enables districts and the state to evaluate whether ELLs are making steady annual growth in learning to listen, speak, read, and write in English in the context of grade-level academic instruction.

Students who take the TELPAS assessments receive proficiency ratings in each language area assessed—listening, speaking, reading, and writing—as well as a composite rating that combines the four language-area ratings into one overall English language proficiency rating. The following descriptions provide a synopsis of the abilities associated with each level of proficiency defined in the ELPS. The complete set of PLDs that are used as the TELPAS assessment rubrics are found on the TELPAS Resources page on TEA's Student Assessment Division website.

Beginning level of English language proficiency: Students who receive this rating are in the early stages of acquiring English. These students typically have a small vocabulary of high-frequency survival words in English and little or no ability to use English in academic settings.

- Beginning listeners struggle to understand simple conversations and to identify and distinguish individual words and phrases spoken in English.

- Beginning speakers mainly use single words and short phrases and lack the knowledge of English grammar necessary to connect ideas and speak in sentences.

■ Beginning readers' ability to derive meaning from English text is minimal. They rely heavily on previous knowledge of the topic, their limited vocabulary, and pictures to gain meaning from English text.

■ Beginning writers lack the English vocabulary and grasp of English language structures and grammar necessary to build writing skills in English and address grade-level appropriate writing tasks in a meaningful way.

Intermediate level of English language proficiency: Students who receive this rating are able to use common, basic English in routine academic activities but need considerable English language support to make instruction understandable. Socially, these students are able to communicate simply in English about familiar topics and are generally able to understand casual conversations but do not comprehend all the details.

■ Intermediate listeners usually understand simple or routine directions in English and short, simple conversations and discussions on familiar topics. They frequently understand only part of what they hear and seek clarification by requesting the speaker to repeat, slow down, or rephrase speech.

■ Intermediate speakers know enough English to speak in a simple manner using basic, high-frequency vocabulary. They are able to participate in short conversations and speak in sentences, though they might hesitate frequently and for long periods to think of how to communicate their intended meaning.

■ Intermediate readers are able to understand short, connected texts on familiar topics but tend to interpret English very literally and have difficulty following story lines that have a surprise twist or nonstandard format. Because their English vocabulary consists mainly of high-frequency, concrete words, they rely heavily on prior knowledge of a topic for comprehension and need the support of pictures that illustrate meaning.

■ Intermediate writers have a limited ability to use the English language to build writing skills and a limited ability to address grade-level appropriate writing tasks in English. They frequently exhibit features of their primary language when expressing themselves in English and sometimes cannot be understood by individuals not accustomed to the writing of ELLs.

Advanced level of English language proficiency: Students who receive this rating have an emerging academic English vocabulary, which they are able to use in classroom instruction when given second language acquisition support. In social situations, these students can understand most of what they hear but have some difficulty with unfamiliar grammar and vocabulary.

■ Advanced listeners can usually understand longer conversations and class discussions in English but occasionally depend on visuals, verbal cues, and gestures to support understanding.

- Advanced speakers are able to participate comfortably in most conversations and academic discussions in English, with occasional pauses to restate, repeat, or search for words or phrases to clarify meaning. They can narrate, describe, and explain in some detail and have an ability to speak in English using a variety of sentence patterns and basic grammar structures.

- Advanced readers have an emerging grade-level appropriate English vocabulary and are familiar with the basic structure of the English language. They use this knowledge to understand texts that introduce them to unfamiliar topics, and, with support, they can move beyond literal comprehension to begin to think critically about ideas presented in grade-level appropriate texts written in English.

- Advanced writers have enough knowledge of English to address grade-level appropriate writing tasks with second language acquisition support. They can express themselves using a variety of verb tenses and sentence patterns, and they can communicate their ideas in some detail, although they often require assistance when topics are abstract, academically challenging, or unfamiliar.

Advanced high level of English language proficiency: Students who receive this rating are able to use academic English in classroom activities with little second language acquisition support from others, even when learning about unfamiliar material. Students at this level have a large enough vocabulary in English to communicate clearly and fluently in most situations.

- Advanced high listeners can understand long conversations and class discussions in English, with little dependence on visuals, verbal cues, and gestures to support understanding. In both social and instructional interactions, they can understand main points and details at a level nearly comparable to native English-speaking peers.

- Advanced high speakers are able to use abstract and content-based vocabulary and can participate in extended discussions in English on a variety of social and grade-level appropriate academic topics with only rare disruptions or hesitations.

- Advanced high readers might have occasional difficulty with low-frequency vocabulary or new English expressions but demonstrate, at a level nearly comparable to native English-speaking peers, comprehension of both explicit and implicit information in grade-level appropriate texts.

- Advanced high writers have acquired the English vocabulary and command of English language structures to address grade-level appropriate writing tasks. They are nearly comparable to native English-speaking peers in their ability to express themselves clearly and precisely, with occasional exceptions when dealing with complex or abstract ideas or when attempting to use low-frequency words and expressions.

Language Domain Scores

Results for the multiple-choice reading tests include proficiency level ratings, the number of items answered correctly (raw scores), and scale scores. For the holistically rated domains of TELPAS, language domain scores consist of the proficiency level ratings of beginning, intermediate, advanced, and advanced high. The scores are recorded on student rating rosters, the rosters are filed at the local level, and the scores are submitted to Pearson through a secure website.

RAW SCORE

The number of multiple-choice reading items answered correctly is provided for the entire test and by groups of items associated with each proficiency level. The raw score can be interpreted only in terms of the specific set of test items on a test form because the difficulty of items might vary across different test forms over time. Thus, differences in student performance across tests or administrations cannot be compared using raw scores alone. To facilitate fair comparisons of student performance across different test forms and different administrations, raw scores are converted to scale scores.

SCALE SCORE

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

In the case of the TELPAS reading test, scale scores are reported on a vertical scale. The TELPAS reading vertical scale score can be used to evaluate a student's progress from one year to the next. For example, a student's score on the grade 2 assessment can be directly compared to the student's score on the grade 3 assessment the following year. The change in the student's vertical scale score is an indication of the progress the student has made over time in English language proficiency in the linguistic domain of reading. Vertical scaling for TELPAS reading is discussed further in the Scaling section of this chapter.

In grades 2 through 12, a student's vertical scale score on the TELPAS reading assessment determines the student's proficiency level for the domain of reading. To facilitate the monitoring of a student's progress from one year to the next, TELPAS results for individual students include the student's proficiency level rating and scale score for the previous and current year. Proficiency level cut scores are discussed in the Performance Standards section of this chapter.

TELPAS Comprehension and Composite Scores

In addition to receiving a rating of beginning, intermediate, advanced, or advanced high for each domain, students also receive a comprehension score, composite score, and composite rating.

The comprehension score is determined from the listening and reading proficiency ratings. This score ranges from 1.0 to 4.0. The listening and reading ratings are each converted to a number from 1 (beginning) to 4 (advanced high). The average of the two numbers is the comprehension score.

The TELPAS composite results indicate a student's overall level of English language proficiency and are determined from the student's listening, speaking, reading, and writing proficiency ratings. Each domain rating is weighted, as shown in Table 6.3. These weights were used for the first time in 2014.

Table 6.3. Weights of the Language Domains in TELPAS Composite Ratings

Listening	Speaking	Reading	Writing
10%	10%	50%	30%

The weights emphasize the domains of reading and writing. Listening and speaking receive less weight, so that students do not attain high composite proficiency ratings before they acquire the English reading and writing proficiency needed to support their full potential for academic success.

Figure 6.1 provides a student example to show how composite results are generated.

Figure 6.1. Sample Calculation of Composite Results

Each domain rating is converted to a domain score from 1 (beginning) to 4 (advanced high).

Domain	Proficiency Level	Domain Score
Listening	Advanced	3
Speaking	Intermediate	2
Reading	Advanced	3
Writing	Intermediate	2

Each domain score is multiplied by the appropriate weight in Table 6.3 and then summed to obtain the TELPAS composite score, as shown:

Composite Score = (Listening × .10) + (Speaking × .10) + (Reading × .50) + (Writing × .30)

Using the sample scores from the chart above, the composite score is calculated as follows:

Composite Score = (3 × .10) + (2 × .10) + (3 × .50) + (2 × .30)

Composite Score = 2.60

The TELPAS composite scores are converted to the TELPAS composite ratings as shown below. This example composite score of 2.60 would result in a composite rating of advanced.

TELPAS Composite Score	TELPAS Composite Rating
1.0–1.4	Beginning
1.5–2.4	Intermediate
2.5–3.4	Advanced
3.5–4.0	Advanced High

Report Formats

Two types of reports are provided for the various testing programs: standard and optional. Standard reports are provided automatically to districts. Information contained in standard reports satisfies mandatory reporting requirements. To receive optional reports that detail student performance data in additional formats, a district must select the corresponding optional reports in the *Administration Details* screen in the Texas Assessment Management System, delivered through PearsonAccess. Generally, districts are required to pay a nominal fee for each optional report requested.
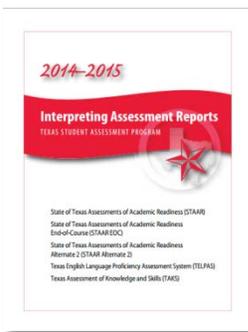
For more information about reporting of the TELPAS results, refer to the TEA publication Interpreting Assessment Reports.

Use of Test Results

The TELPAS student performance reports are used in the following ways:

- helping parents monitor the progress their child is making in acquiring English
- informing instructional planning for individual students
- reporting results to local school boards, school professionals, and the community
- evaluating programs, resources, and staffing patterns
- evaluating district effectiveness in accountability measures

Parent Brochures

To assist teachers and parents in understanding students' TELPAS results, TEA's Student Assessment Division produces brochures titled Understanding the Confidential Student Report—A Guide for Parents. The brochures provide a brief summary of the TELPAS program, explain a sample Confidential Student Report so parents can

understand their child's test report, and give a brief summary of the meaning of each proficiency level. The brochures, available in both English and Spanish, are provided to districts each spring for distribution with individual student TELPAS results.

Audits

Since the 2004–2005 school year, TEA has conducted periodic audits of the TELPAS assessment processes as a means of collecting reliability and validity evidence for the assessment program. Audits allow for the collection of information from school districts that can be used to evaluate the training, administration, and scoring of the holistically rated assessments. Information collected during TELPAS audits has been useful in the refinement of TELPAS holistic rating training and administration procedures. For the listening and speaking domains, an audit process is used in which documentation is collected from teachers at selected sites to evaluate the accuracy of holistic ratings. The most recent TELPAS listening and speaking audit occurred in spring 2011. More recently, a TELPAS writing audit was conducted in spring 2013. During the TELPAS writing audits, expert raters provide second ratings of writing samples of students in the state, and testing personnel at the sampled sites complete questionnaires that allow the state to evaluate conformity with training and administration procedures. See the Interrater Reliability section of this chapter for more details.

Performance Standards

Performance standards relate levels of test performance directly to what students are expected to learn, as defined in the statewide curriculum. This is done by establishing cut scores that distinguish between performance levels or categories. Standard setting is the process of establishing these cut scores that define the performance levels for an assessment.

For holistically rated assessments, standards are established through descriptions of student performance in the scoring rubrics and student exemplars used in scorer training. For the TELPAS holistically rated assessments, the scoring rubrics are the PLDs in the ELPS. The student exemplars are the student writing collections and student videos used in rater training.

For multiple-choice tests, standards are established by determining the score students need to obtain to be classified into specified performance categories. For the TELPAS multiple-choice reading tests, the performance categories are the proficiency levels described in the ELPS.

The original TELPAS reading proficiency level standards were established in 2008 when the Texas Assessment of Knowledge and Skills (TAKS) was the academic assessment in Texas. Detailed information about the standard-setting activities for the TELPAS reading tests that were revised in 2007–2008 is available in the TELPAS Proficiency Level Setting Report on TEA's Student Assessment Division website.

The move from TAKS to STAAR in 2011–2012 made it necessary to review the original TELPAS reading proficiency level standards so that performance on TELPAS could be a meaningful indicator of the level of English language proficiency required to access the language in STAAR assessments. In August 2013, a standards review was conducted with committees of educators, and the commissioner of education approved the new standards. The new standards were first implemented with the 2014 spring administration of TELPAS reading. Table 6.4 shows the scale score ranges and corresponding raw score cuts from the proficiency level standards review conducted in August 2013. While the scale score ranges remain constant from year to year, slight fluctuations in raw score cut scores might occur. For more information about scale scores and the potential for raw score fluctuations in standardized assessments, refer to the Equating section in chapter 3, "Standard Technical Processes." More detailed information about the standards review process is available in the TELPAS Standards Review Technical Report on TEA's Student Assessment Division website.

Table 6.4. Approved Raw and Scale Score Cut Scores from 2013 TELPAS Reading Proficiency Level Standards Review

TELPAS Reading Assessment	Raw Score Cut Scores	Total Number of Questions on Test	Beginning Level	Intermediate Level	Advanced Level	Advanced High Level
Grade 2	20, 31, 39	49	≤ 578	579–644	645–700	≥ 701
Grade 3	26, 37, 47	58	≤ 619	620–673	674–731	≥ 732
Grades 4–5	25, 37, 47	61	≤ 642	643–697	698–765	≥ 766
Grades 6–7	25, 38, 50	63	≤ 651	652–711	712–782	≥ 783
Grades 8–9	27, 37, 49	63	≤ 660	661–719	720–795	≥ 796
Grades 10–12	24, 37, 48	64	≤ 679	680–736	737–814	≥ 815

Scaling

Scaling is a statistical procedure that places raw scores on a common scoring metric in order to make test scores easier to interpret and compare across test administrations. As with many of the other programs in the Texas assessment program, the TELPAS reading tests for grades 2–12 use the Rasch partial-credit model (RPCM) to place test items for a given TELPAS assessment on the same scale across administrations. Once performance standards have been set for an assessment, its Rasch scale is then transformed to a more user-friendly metric to facilitate interpretation of the test scores. Details of the RPCM scaling method used in Texas are provided in chapter 3, "Standard Technical Processes."

Reporting Scales

VERTICAL REPORTING SCALES

Scale scores are reported on a vertical scale for TELPAS reading tests. A vertical scale allows for the direct comparison of student scores across grade levels in a particular subject. Student increases in vertical scale scores provide information about the student's year-to-year growth. A vertical scale system was developed for the revised grades 2–12 TELPAS reading tests using a spring 2008 vertical scaling study, and the proficiency level standards established in summer 2008 and revised in summer 2013 were mapped onto the vertical scale score system.

The vertical scale scores can be computed through a linear transformation of the underlying Rasch proficiency level estimate (θ) as follows:

$$SS_\theta = A \times (\theta - V_g) + B \qquad\qquad (1)$$

where SS_θ is the scale score for a Rasch proficiency level estimate (θ). A and B are vertical scale score transformation constants, and V_g is the vertical scaling constant for each grade for Equation (1). The values of A, B, and V_g for the TELPAS assessments are provided in Table 6.5. Once established, these same transformations are applied each year to the proficiency level estimates for that year's set of test questions.

Table 6.5. Vertical Scale Score Linking Constants for TELPAS Reading Tests

Grade/Grade Cluster	V_g	A	B
Grade 2	0		
Grade 3	0.73978		
Grades 4–5	1.16885	48	575
Grades 6–7	1.24793		
Grades 8–9	1.61612		
Grades 10–12	1.90165		

Further information about vertical scaling appears in chapter 3, "Standard Technical Processes." Additional information specific to the generation of the TELPAS reading vertical scale for grades 2–12 is available in the 2008 TELPAS Reading Vertical Scaling Study on TEA's Student Assessment Division website.

SCALE FOR HOLISTICALLY RATED ASSESSMENTS

The scale for the TELPAS holistically rated assessments (grades K–1 reading and grades K–12 listening, speaking, and writing) ranges from 1 to 4 and is defined by the four proficiency levels: beginning, intermediate, advanced, and advanced high.

SCALE FOR COMPOSITE SCORES

The TELPAS composite rating uses a scale from 1.0 to 4.0. More information about the calculation of the composite rating is available in the TELPAS Comprehension and Composite Scores section of this chapter.

Equating

Used in conjunction with the scaling process, equating is the statistical process that takes into account the slight differences in difficulty across test forms and administrations and allows the scores to be placed onto a common scale. TEA statistically equates the results of different tests, enabling the comparison of scale scores across test forms and testing administrations. Equating for the TELPAS reading assessment is done using the Rasch measurement model. In the 2014–2015 school year, equating activities for the TELPAS reading assessments included pre-equating, post-equating, and field-test equating. Refer to chapter 3, "Standard Technical Processes," for detailed information about equating.

Pre-Equating

The pre-equating process takes place prior to test administration. It links a newly developed test form onto the scale of the item bank through the use of a set of items that appeared previously on one or more test forms. This permits the difficulty level of the newly developed form to be closely determined even prior to its administration, and thus the anticipated raw scores that correspond to scale scores at performance standards can be identified. Pre-equating was conducted for all spring 2015 TELPAS reading test forms during the test construction process.

Post-Equating

Post-equating is conducted for the online forms of TELPAS reading assessments in the spring. The post-equating process uses data from the operational test administration to re-estimate item difficulties and place them onto the scale of the item bank. For the TELPAS reading assessments, post-equating uses conventional common-item/non-equivalent groups equating procedures as described in the technical details and procedures in chapter 3, "Standard Technical Processes."

Field-Test Equating

To replenish the item bank as new tests are created each year, newly developed items must be field-tested and equated to the item bank scale as described in the technical details and procedures in chapter 3, "Standard Technical Processes." Whenever possible, embedded designs are used to field-test new items so that test takers will be unable to distinguish between field-test items and operational items on each test form. This results in student response data that are more stable. In the 2014–2015 school year, field-test equating was conducted for all TELPAS reading assessments through an embedded field-test design.

Equating of the TELPAS holistically rated assessments is not necessary. The difficulty level of holistically rated assessments is maintained through the use of consistent rating rubrics developed to define the proficiency levels. The training activities completed by raters before administering the assessment provide consistency in the way the rubrics are applied each year. The training maintains the difficulty of the assessment across administrations by calibrating the teachers to the assessment rubric every time they administer the holistically rated portions of TELPAS.

Reliability

During the 2014–2015 school year, reliability estimates for the TELPAS scores were obtained mainly through analyses of internal consistency, classical standard error of measurement, conditional standard error of measurement, classification accuracy, and interrater reliability. Refer to chapter 3, "Standard Technical Processes," for detailed information about reliability.

Internal Consistency

The Kuder–Richardson Formula 20 (KR20) was used to calculate the reliability estimates for TELPAS reading scores. As a general rule, reliability coefficients from 0.70 to 0.79 are considered adequate, those from 0.80 to 0.89 are considered good, and those greater than 0.90 are considered excellent. However, what is considered appropriate might vary depending on how assessment results are used. For the spring 2015 TELPAS reading assessments, internal consistency estimates ranged from 0.92 to 0.94. This indicates that the reliability estimates were in the highest range in terms of appropriateness for student-level interpretations. In addition to the overall test reliability, Appendix D presents reliability estimates by proficiency level (beginning, intermediate, advanced, and advanced high) and by gender.

Classical Standard Error of Measurement

Classical standard error of measurement (SEM) represents the amount of variance in a score that results from factors other than what the assessment is intended to measure. The SEM is helpful for quantifying the margin of uncertainty that occurs on every test. Refer to chapter 3, "Standard Technical Processes," for detailed information about SEM. The SEM values (shown in Appendix D) for TELPAS reading tests are between 2.8 and 3.1 raw score points across grades/grade clusters.

Conditional Standard Error of Measurement

The SEM index provides only an estimate of the average test score error for all students regardless of their individual levels of proficiency. By comparison, conditional standard error of measurement (CSEM) provides an estimate of test score error at each score point on a test. More specifically, CSEM is an estimate of the average test score measurement error that is conditional on the proficiency or scale score estimate. Appendix D provides CSEM values for all primary administrations of TELPAS.

Classification Accuracy

Classification accuracy provides an estimate of the accuracy of student classifications into performance categories based on current test results. Appendix D provides classification accuracy rates for each grade-level/grade cluster of the TELPAS reading tests.

Interrater Reliability

Evidence that the holistically rated components of TELPAS result in reliable observation and rating of student performance is collected through periodic interrater reliability studies. Evidence of interrater reliability is collected through the audit process by having a second rater provide independent ratings for a sample of students. No interrater reliability studies were conducted in 2014–2015.

In addition, the composite reliability estimates of TELPAS are analyzed annually to evaluate the impact of the reliability of the listening, speaking, and writing domains on the TELPAS composite reliability estimates. The results of these analyses show that the weighted TELPAS composite ratings have reliability estimates that are at least 0.91. The high internal consistency reliability of the TELPAS reading scores and the high interrater reliability of the TELPAS writing ratings, combined with the heavy weighting of these domains, produce highly reliable TELPAS composite ratings. Additional information is available in the TELPAS Composite Reliability Estimates 2015 Report on the TEA's Student Assessment Division website.

Validity

Validity refers to the extent a test measures what it is intended to measure. The results of the TELPAS assessments are used to guide instructional planning related to the progress that ELLs make in acquiring English. Validity evidence for an assessment can come from a variety of sources, including test content, response processes, internal structure, relationships with other variables, and analysis of the consequences of testing.

The sections that follow describe how these types of validity evidence were collected for the TELPAS assessments in 2014–2015.

Evidence of the validity of the reading, writing, listening, and speaking domains of TELPAS has been continually collected since the first administration in 2003–2004. In addition to the studies described in this year's technical digest, a wide range of validity studies and analyses has been conducted and documented in the Technical Report Series and Technical Digests for previous years. The Technical Report Series and Technical Digests are available on TEA's Student Assessment Division website.

Evidence Based on Test Content

Validity evidence based on test content refers to evidence of the relationship between tested content and the construct the test is intended to measure. TELPAS measures

student performance in direct alignment with the English language acquisition skills and PLDs defined by the Texas ELPS that are part of the TEKS curriculum. The ELPS outline the instruction that ELLs must receive to support their ability to develop academic English language proficiency. TELPAS assesses the ELPS for listening, speaking, reading, and writing.

TELPAS MULTIPLE-CHOICE COMPONENT

Test Design and Alignment with Standards. The multiple-choice TELPAS reading tests for grades 2–12 are designed to assess English language reading proficiency in a manner that provides information about how well ELLs read and understand the English they need for academic success in Texas schools, as well as the types of language supports they require to independently comprehend written text.

The test is built using four levels, or degrees, of built-in linguistic support, addressing the gradually reduced degree of linguistic accommodation that ELLs need as they progress from knowing little or no English to becoming fluent English readers. The levels of linguistic support are integrally related to the four proficiency levels assessed, as each proficiency level described in the ELPS is characterized by the degree of linguistic accommodation that students at that level need to read and understand English.

Each reading selection and test question is written to reflect a particular proficiency level associated with a particular degree of linguistic accommodation. The test blueprints require a specified number of items per proficiency level and per reporting category (reading skill category). Score reports inform teachers about how successfully students demonstrate the comprehension and analytical reading skills of the ELPS at the four proficiency levels. The content validity of the TELPAS reading assessment is supported by this test design, in that it provides built-in, staged linguistic accommodations validated by second language acquisition theory and empirical data as it measures the ELPS-aligned reading skills that students need for academic success in all subject areas. The staged linguistic accommodation test design is shown in Table 6.6.

Table 6.6. Staged Linguistic Accommodation Test Design

TELPAS Reading Levels	Degree of Linguistic Accommodation Applied to Passage and Item Development	
Advanced High	Minimal	Minimal linguistic accommodation; texts highly comparable to those written for native English speakers
Advanced	Moderate	Occasional picture support; contextual aids and organizational features support comprehension of longer texts on both familiar and unfamiliar language arts and content area topics
Intermediate	Substantial	Frequent picture support; short texts written primarily on familiar topics; commonly used, everyday English and routine academic English
Beginning	Extensive	Maximum picture support; short texts that require comprehension of words, phrases, and short sentences that use the type of high-frequency, concrete vocabulary first acquired by learners of a second language

The TELPAS reading material requires students to comprehend the types of written English they encounter in everyday life as well as grade-level core content instruction. Most of the topics and contexts come from the content areas of language arts, mathematics, and science, although other subjects are eligible as well. Items that assess the higher proficiency levels challenge students' ability to think critically and conceptually while reading complex English and academic content. The construct measured is the ability to read the English required for meaningful engagement in the learning of the state's grade-level academic content standards.

Test Development and Construction. Although their test designs differ, the quality assurance steps used to develop the multiple-choice TELPAS reading assessments and the STAAR assessments are the same. This process adheres to the Standards for Educational and Psychological Testing (AERA/APA/NCME, 2014), is grounded in the state's standards, and is guided by assessment experts and educators who have first-hand knowledge of the standards and the students. As with STAAR, the TELPAS reading test construction process involves multiple reviews by both content and psychometric experts. The fact that the state follows the same thorough development processes for the STAAR and TELPAS reading tests—and includes the STAAR assessment and content area experts throughout the development process—further supports the content validity of TELPAS and its link to the state's academic content standards.

TELPAS HOLISTICALLY RATED COMPONENTS

Test Design and Alignment with Standards. Like the reading tests, the TELPAS holistically rated components are aligned with the ELPS and are designed to assess the English communication skills that ELLs need in order to engage meaningfully and successfully in learning the academic knowledge and skills required by the state. The holistically rated assessments draw upon second language acquisition research, research-based standards, the experience of Texas practitioners, and observational assessment practices.

The TELPAS holistically rated components are based on ongoing observations of the ability of ELLs to understand and use English during the grade-level core content area instruction that is required by the state-mandated curriculum and assessed on the state-mandated assessments. The TELPAS holistically rated assessments measure the ELPS student expectations from the cross-curricular second language acquisition knowledge and skills and use the ELPS PLDs as assessment rubrics. Rater training and administration procedures require these ratings to be based on the ability of the students to use English in a variety of core content areas.

The state's decision to implement holistically rated TELPAS assessments was made in order to attain the following goals:

■ to develop the ability of all Texas teachers to meet the instructional needs of the state's growing ELL population;

■ to minimize the number of additional instructional days devoted to standardized testing; and

■ to avoid logistically impractical speaking and listening assessments, given the state's large ELL population.

The TELPAS holistically rated assessments support all these goals. Due to the direct involvement teachers across the state have in the assessment process, the holistically rated assessments have an immediate and substantial positive effect on classroom instruction.

Evidence Based on Response Processes

An additional source of validity evidence is whether the way students respond to test questions on the TELPAS assessments supports the accurate measurement of the construct.

TELPAS MULTIPLE-CHOICE COMPONENT

Theoretical and empirical evidence was used to pilot test multiple-choice reading assessments and determine the appropriateness of each item type used on the assessments. A variety of question-and-answer and cloze (i.e., fill-in-the-blank) response formats are used. The items are written in alignment with the second language acquisition characteristics of students at each of the four proficiency levels assessed.

Validity evidence of the appropriateness of the item types and each item's conformity to the proficiency-level and item specifications is gathered annually through educator and expert review and through analyses of student responses to the items during field testing. Educators evaluate whether the content assessed by the item in its format is appropriate and whether students are able to accurately demonstrate the knowledge being assessed by the construct. When items are field-tested, statistical data such as item difficulty for students at each proficiency level, item point-biserial correlations, and differential item functioning can be gathered and evaluated.

In 2010, TEA began using an enhanced online interface as part of the TELPAS reading administration. The new interface was designed to enhance the students' testing experience and provide improved testing conditions for students to demonstrate what they have learned. A usability study was conducted as part of the design process, and final decisions on the components of the new interface were made based on the results of this usability study.

TELPAS HOLISTICALLY RATED COMPONENTS

The TELPAS holistically rated components are assessed through a collection of students' writing samples, classroom observations, and daily interactions with the students. As is typical of holistically scored assessments, students are evaluated on their overall performance in a global and direct way. The goal of English language proficiency assessments is to effectively assess the extent to which ELLs are making progress in attaining academic language proficiency so they can achieve their full academic potential. The TELPAS holistically rated assessments are direct measures of the ability of students to understand and use English while engaging in state-required academic instruction, which provides strong validity evidence related to the response process.

Evidence Based on Internal Structure

Texas collects evidence that reflects the relationship between test item performance and proficiency levels in order to verify that patterns of item performance are consistent with the constructs the test is intended to measure.

TELPAS MULTIPLE-CHOICE COMPONENT

Internal structure is evaluated annually by estimating the internal consistency reliability for the TELPAS multiple-choice component. Internal consistency reliability estimates provide a measure of the consistency with which students respond to the items in an assessment. The internal consistency of the TELPAS reading tests is evaluated each year using KR20 statistics that can be found in Appendix D.

TELPAS HOLISTICALLY RATED COMPONENTS

Evidence of the validity of TELPAS is supported by comprehensive training and administration procedures that prepare teachers to perform their duties and prepare district administrators to follow procedures in order to maintain the integrity of the test administration. In addition to holistic rating training opportunities, raters must perform

calibration activities to demonstrate high accuracy in rating student activities across all TELPAS holistically rated domains. Additional supplemental support training is provided to raters who are unable to calibrate on their first two attempts in order to help them identify misunderstandings they might have about the TELPAS rating rubrics and thereby improve their rating accuracy. Refer to the Training section of this chapter for detailed information about this calibration process.

The TELPAS holistic rating audits provide both validity and reliability evidence based on internal structure for the holistically rated components of the assessment by examining the extent to which raters follow the defined protocol for rating these TELPAS components. As part of the audit, reports of rater adherence to the assessment protocol are made and used to provide evidence that the internal structure of the assessment is intact and that teachers are administering the assessment and applying the scoring rubrics appropriately. Additional information can be found in the Audits section of this chapter.

The TELPAS holistically rated assessments directly support the state's goal of having a valid and authentic assessment. These holistically rated assessments also serve an ongoing and critical role as a professional development tool that supports effective instruction, enabling teachers to better understand and meet the educational needs of ELLs.

Evidence Based on Relationships to Other Variables

Another way that Texas provides validity evidence for TELPAS is by analyzing the relationship between test performance and performance on external measures. By examining this relationship, evidence can be collected to show that the relationships are consistent with those expected at the level of the construct underlying the proposed score interpretations.

TELPAS READING PROFICIENCY LEVELS AND STAAR READING PERFORMANCE STANDARDS

To examine validity evidence based on external measures, TEA conducted an analysis of the relationship between 2015 TELPAS reading performance and 2015 STAAR reading (English-version tests only) or STAAR EOC English performance. Note that, in spring 2014, the redesigned EOC English assessments were administered for the first time. Scores from these assessments reflected both reading and writing performance, so it was not possible to examine EOC reading performance alone.

As indicated earlier in this chapter, TELPAS reading measures English language proficiency in reading, that is, how well ELLs are learning to understand written English and apply reading skills for meaningful engagement in content area instruction. STAAR, on the other hand, assumes that students already understand the English language and so it focuses on assessing the degree to which students can apply literary and analytical reading skills as required by the language arts TEKS for their grade level. Because of the differences in the designs and purposes of these two assessments, one would not expect ELLs to perform at the same level of proficiency

on the two assessments. One would, however, expect ELLs who have comparatively little difficulty understanding and reading English to score higher on the STAAR reading tests when compared with ELLs who are in earlier stages of English fluency.

To examine the relationship between performance on the two tests, average performance for STAAR grades 3–8 reading, English I, and English II was calculated separately for the students classified in each of the TELPAS reading English language proficiency levels (beginning, intermediate, advanced, or advanced high). For each grade level and TELPAS proficiency level breakout group, two types of performance data were examined:

- average STAAR scale scores

- STAAR passing rates (Level II: Satisfactory Academic Performance)

The STAAR 3–8 reading Level II performance levels were set in 2012, and performance standards for the redesigned English I and II assessments were set in January 2014. STAAR performance standards are being phased in incrementally. The STAAR passing rates provided are based on the phase-in 1 standard, which was the standard students were held to in 2015. For more information about the phase-in of the STAAR standards, refer to the Standards section in chapter 4, "STAAR."

On the following page, data are presented for STAAR grades 3–8 and English I and II. The scale scores for STAAR grades 3–8 reading are on a vertical scale, and the scale scores for STAAR English I and II are on horizontal scales. Table 6.7 shows that within a grade level, the average STAAR scale score increases as a student's TELPAS proficiency level increases. In addition, the passing rate on STAAR increases as the TELPAS proficiency level increases within each grade level. State passing rates for all STAAR students can be found in the Test Results section of chapter 4, "STAAR."

Table 6.7. 2015 STAAR Grades 3–8 Reading*, English I, and English II Performance by TELPAS Reading Proficiency Level for Students Who Participated in Both Assessments

Grade Level	TELPAS Reading Proficiency Level	N	2015 Average STAAR Reading Scale Score**	STAAR Passing Rate (%), Phase-in 1 Standard
3	Beginning	2,861	1214	3
	Intermediate	13,493	1269	19
	Advanced	25,332	1359	66
	Advanced High	26,843	1477	96
4	Beginning	1,692	1284	2
	Intermediate	11,440	1327	9
	Advanced	29,209	1417	49
	Advanced High	24,851	1534	89
5	Beginning	1,123	1306	2
	Intermediate	7,734	1342	6
	Advanced	25,344	1426	38
	Advanced High	30,791	1536	83
6	Beginning	2,207	1350	1
	Intermediate	9,970	1397	8
	Advanced	27,487	1468	36
	Advanced High	20,563	1559	75
7	Beginning	2,152	1407	1
	Intermediate	7,708	1443	5
	Advanced	20,952	1507	25
	Advanced High	16,587	1585	63
8	Beginning	2,027	1428	2
	Intermediate	6,498	1464	7
	Advanced	18,436	1537	31
	Advanced High	14,216	1606	64
English I (reading and writing)	Beginning	3,467	2947	0
	Intermediate	10,701	3180	1
	Advanced	23,259	3457	13
	Advanced High	11,784	3716	43
English II (reading and writing)	Beginning	1,180	2934	0
	Intermediate	7,606	3141	1
	Advanced	18,034	3422	13
	Advanced High	9,592	3690	42

*English versions only
**Scale scores necessary to meet each STAAR performance level:
  • For grade 3, Level II is 1331 at Phase-in 1.
  • For grade 4, Level II is 1422 at Phase-in 1.
  • For grade 5, Level II is 1458 at Phase-in 1.
  • For grade 6, Level II is 1504 at Phase-in 1.
  • For grade 7, Level II is 1556 at Phase-in 1.
  • For grade 8, Level II is 1575 at Phase-in 1.
  • For English I, Level II is 3750 at Phase-in 1.
  • For English II, Level II is 3750 at Phase-in 1.

TELPAS WRITING PROFICIENCY LEVELS AND STAAR WRITING PERFORMANCE STANDARDS

TELPAS performance was also compared to STAAR writing performance by looking at average STAAR grade 4 English writing, grade 7 writing, and STAAR EOC English scale scores and passing rates for each of the four TELPAS writing proficiency levels. Table 6.8 shows that within a grade level, the average STAAR scale score increases as a student's TELPAS writing proficiency level increases. In addition, the passing rate on STAAR tends to increase as a student's TELPAS proficiency level increases within a grade level.

Table 6.8. 2015 STAAR Grades 4 and 7 Writing*, English I, and English II Performance by TELPAS Writing Proficiency Rating for Students Who Participated in Both Assessments

Grade Level	TELPAS Writing Proficiency Level	N	2015 Average STAAR Writing Scale Score**	STAAR Passing Rate (%), Phase-in 1 Standard
4	Beginning	2,413	2802	3
	Intermediate	12,972	3168	20
	Advanced	24,482	3470	52
	Advanced High	26,461	3781	81
7	Beginning	2,931	2689	1
	Intermediate	9,024	3042	9
	Advanced	17,979	3309	30
	Advanced High	17,663	3526	53
English I Writing	Beginning	4,420	2998	0
	Intermediate	12,320	3275	6
	Advanced	18,538	3470	16
	Advanced High	14,287	3617	31
English II Writing	Beginning	1,431	3002	1
	Intermediate	9,035	3230	5
	Advanced	14,468	3433	16
	Advanced High	11,765	3593	31

*English version only
**Scale scores necessary to meet each STAAR performance level:
  • For grade 4 (English), Level II is 3500 at Phase-in 1.
  • For grade 7, Level II is 3500 at Phase-in 1.
  • For English I writing, Level II is 3750 at Phase-in 1.
  • For English II writing, Level II is 3750 at Phase-in 1.

In summary, the positive empirical relationship between TELPAS reading proficiency levels and success rates on grade-level STAAR 3–8 reading and EOC English assessments adds to the body of the TELPAS validity evidence, as does the positive empirical relationship between TELPAS writing proficiency levels and success rates on grade-level STAAR 4 and 7 writing and EOC English assessments.

Evidence Based on Consequences of Testing

Another source of validity evidence comes from documenting the intended and unintended consequences of administering an assessment. The effect an assessment has on the instructional environment after the assessment is given is referred to by some researchers as consequential validity (Kane, 1992; Messick, 1989; Shepard, 1997). The administration of the TELPAS holistically rated assessments leads to improvements in students' academic language acquisition as a result of what educators learn during the rater training process and through direct application of the assessment process for both formative and summative purposes. Logical consequences of administering TELPAS are that educators (1) learn how developing academic language proficiency in English relates to and supports academic achievement in English, (2) learn how to adjust content instruction for ELLs to make it more comprehensible and how to target steady progress in English acquisition, and (3) practice observing student behaviors in the instructional environment for the purpose of making better instructional decisions about students.

Evidence of consequential validity can be found by comparing performance from past administrations in the TELPAS Statewide Summary Reports on TEA's Student Assessment Division website. These results show incremental increases in ELL performance in all TELPAS domains from spring 2005 to spring 2015. These increases in student performance provide additional evidence of the consequential validity of TELPAS. The percentage of students at higher TELPAS proficiency levels decreased in 2014 for the reading domain in grades 2–12 because the new reading performance standards were first applied in spring 2014. Over time, the percent of students in the higher proficiency levels is expected to rise, like it has in the past.

Sampling

Sampling was not conducted for the TELPAS assessments during the 2014–2015 school year.

Test Results

Appendix D provides frequency distributions and summary statistics for the TELPAS reading assessments administered in 2014–2015 based on the vertical scale score system, as well as mean p-values and reliability estimates by grade cluster. The percentage of students in each of the TELPAS composite proficiency levels is provided in Table 6.9. The percentages are available by domain in the 2015 TELPAS Statewide Summary Reports on TEA's Student Assessment Division website.

Table 6.9. Percentages* of Students in Each of the TELPAS Composite Proficiency Levels in 2015

TELPAS Assessment		Composite Proficiency Levels			
Grade Cluster	**Grade**	**Beginning**	**Intermediate**	**Advanced**	**Advanced High**
Grades K–1	Kindergarten	59%	23%	12%	6%
	Grade 1	29%	34%	22%	15%
Grade 2	Grade 2	11%	36%	33%	20%
Grade 3	Grade 3	8%	25%	35%	31%
Grades 4–5	Grade 4	5%	21%	42%	31%
	Grade 5	4%	14%	39%	43%
Grades 6–7	Grade 6	4%	17%	45%	34%
	Grade 7	5%	16%	44%	35%
Grades 8–9	Grade 8	5%	16%	45%	34%
	Grade 9	11%	22%	42%	26%
Grades 10–12	Grade 10	5%	21%	45%	29%
	Grade 11	3%	16%	46%	35%
	Grade 12	2%	13%	45%	40%

*Percentages may not sum to 100 due to rounding.