Chapter **7** # 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 No Child Left Behind Act of 2001 (NCLB) 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 the NCLB Act, Texas developed and administered English language proficiency tests in the domain of reading in fulfillment of Texas state law.

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), which are 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 that characterize the four English language proficiency levels reported in Texas.

TELPAS and Texas Assessment of Knowledge and Skills (TAKS) results 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 results are used in TELPAS AMAO indicators. For information about how TELPAS composite results are generated, see the "TELPAS Comprehension and Composite Scores" section in this chapter. More information about AMAO accountability indicators is available on TEA's Student Assessment Division website.

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.

Test Development and Administration

Historical Overview

The Reading Proficiency Tests in English (RPTE) implemented in 1999–2000 for ELLs in grades 3–12 were the first reading tests of English language proficiency in the Texas assessment program. In response to federal requirements for assessing additional grades and language domains, TELPAS was created, and RPTE was retained as a component of the TELPAS assessment system.

A number of changes were made to RPTE in 2004 and 2005 to meet federal requirements. One of the primary purposes of the English language proficiency assessments under NCLB is to provide annual data on how many ELLs attain English language proficiency. RPTE's purpose was somewhat different. It was intended primarily to be an interim measure of reading proficiency for ELLs who were not yet proficient enough in English to take the regular state reading assessments. Once ELLs reached the advanced level on RPTE, they stopped taking RPTE and participated in the regular state reading assessments. To meet federal requirements, RPTE participation requirements changed and an advanced high level of proficiency began to be reported so that attainment of English language proficiency in reading could be reported using RPTE.

More substantive revisions to RPTE were implemented in 2007–2008, at which time the name RPTE was discontinued. Table 29 summarizes the main revisions that were made.

Table 29. Summary of 2007–2008 TELPAS Reading Test Revisions

Grade 2 holistically rated component replaced by multiple-choice test
Six grade-cluster tests rather than four
Alignment to revised English language proficiency standards
More emphasis on reading in mathematics and science contexts
More advanced and advanced high reading material
Shift from paper-based testing to online assessment program

The holistically rated components of TELPAS were implemented in 2003–2004 to assess K–1 ELLs in listening, speaking, reading, and writing and ELLs in grades 2–12 in listening, speaking, and writing. Within the TELPAS system, these assessments were originally called Texas Observation Protocols (TOP). In 2007–2008, the component name TOP was discontinued. All English language proficiency assessment components now use just the name TELPAS.
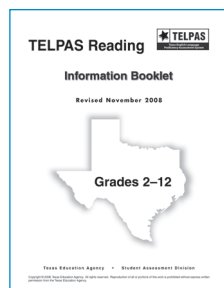
TELPAS Reading Tests for Grades 2–12

The TELPAS reading tests for grades 2–12 employ a multiple-choice test answer format. Six grade-cluster tests are administered, as shown in Table 30.

Table 30. 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, the Texas Education Agency (TEA) involves educators and assessment experts in the test development process. Committees of educators reviewed potential items both before and after the initial field testing that was conducted. As part of the ongoing process to replenish the item banks, committees of Texas educators continue to convene to review new items before and after annual field testing.

TELPAS Reading

Information Booklet

Revised November 2008

Grades 2–12

Texas Education Agency • Student Assessment Division

More information about the TELPAS reading tests for grades 2–12, including item specifications and samples, is available in the *TELPAS Reading Information Booklet.*

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 holistically rate the English language proficiency of ELLs based on an evaluation of their writing, classroom observations, and daily interactions with the students. Writing in grades 2–12 is assessed through a collection of students' classroom writing assignments.

The rating process is designed to capture an overall level of English language acquisition 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. The ratings are submitted to Pearson with the students' TELPAS reading tests, and performance reports are generated and distributed to the districts. The PLDs can be found on TEA's Student Assessment Division website.

The TELPAS holistically rated components were developed by TEA in collaboration with test development contractors, 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 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.

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 in order to effectively 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 enables the state to avoid drawbacks related to adding assessments and field-testing activities that take students away from needed instruction.

Holistic Rating Training

TELPAS raters receive annual holistic rating training from individuals trained directly by TEA. Comprehensive, standardized online training courses are used as an integral part of the training of new raters and annual refresher training of returning raters. Raters learn the essentials of second language acquisition theory and how to use the PLDs from the ELPS to officially identify the English language proficiency levels of their ELLs based on how well the students are able to understand and use English during daily academic instruction and classroom interactions. The courses contain numerous rating activities that comprise student writing collections and video segments in which ELLs demonstrate their reading, speaking, and listening skills in authentic Texas classroom settings. The courses give teachers ample practice in applying the scoring rubrics (PLDs) and provide teachers with detailed feedback before they rate students for the live assessment. The online courses are described below.

KINDERGARTEN AND GRADE 1

- *Level 1: Basic Training Course for Kindergarten and Grade 1*
- *Level 2: Refresher Course for Kindergarten and Grade 1*

GRADES 2–12

- *Level 1: Basic Training Course for Grades 2–12*
- *Level 2: Refresher Course for Grades 2–12*

OTHER TELPAS TRAINING

- *Assembling and Verifying Grades 2–12 Writing Collections*
  This course is used not only in the training of raters, but also in the training of testing coordinators, campus administrators, and writing collection verifiers.

QUALIFICATION COMPONENT OF TRAINING

The TELPAS training model for grades 2–12 includes an online qualifying component. The qualifying component of TELPAS rater training requires teachers to demonstrate their ability to independently rate student writing collections in accordance with the PLDs. The qualification process provides districts and the state evidence that raters are prepared to rate students in the live assessment in a consistent and accurate manner. Planning is underway to expand the qualification component to include all language domains of the K–12 training system.

TRACKING ONLINE TRAINING AND QUALIFICATION

To monitor the adequacy of training in support of the assessment's validity and reliability, school districts and TEA track qualification as well as course participation and performance electronically. During the 2008–2009 school year, more than 37,000 qualification certificates were issued to Texas teachers as part of the training system for new raters. Over the course of the TELPAS program, more than 150,000 qualification certificates have been issued. Approximately 90% of the teachers trained met the

qualification requirements of their training. If a teacher is not successful in qualifying but is assigned to be a rater, a qualified rater must provide rating support during the live assessment.

Administration Procedures

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 oaths of test integrity, 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 these administration procedures are found in the TELPAS section of the *2009 District and Campus Coordinator Manual* for the Texas assessment program.

Scores and Reports

English language proficiency tests are not designed to measure mastery of learning objectives with a pass or fail score. This is because becoming academically proficient in a second language takes longer than a school year for students who enter U.S. schools without knowledge of the English language. TELPAS test results provide a measure of progress, indicating annually 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 and the key to helping districts monitor whether their 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 summarize information contained in the ELPS PLDs in order to provide a synopsis of the abilities associated with each level of proficiency. The complete set of ELPS proficiency level descriptors, which are used as TELPAS assessment rubrics, are found on TEA's Student Assessment Division website.

Beginning level of English language proficiency: Students who receive this rating are in the early stages of learning 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 use mainly 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-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 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 as well as 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 may 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-appropriate writing tasks in English. They frequently exhibit features of their primary language when expressing themselves in English and are sometimes unable to be understood by individuals not accustomed to English language learners.

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 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, 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-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-appropriate texts written in English.

- Advanced writers have enough knowledge of English to address grade-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, with little dependence on visuals, verbal cues, and gestures to support understanding. In both social and instructional interactions, they are able to 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 on a variety of social and grade-appropriate academic topics with only rare disruptions or hesitations.

- Advanced high readers may 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-appropriate texts.

- Advanced high writers have acquired the English vocabulary and command of English language structures to address grade-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

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 rosters, the rosters are filed at the local level, and the scores are submitted to Pearson through a secure website.

Scores for the multiple-choice reading tests consist of proficiency level ratings, the number of items answered correctly (raw scores), and scale scores. The number of items answered correctly is provided for the test overall and by proficiency level. The raw scores are converted to scale scores that 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. Vertical scaling for TELPAS reading is discussed further in the "Scaling" section of this chapter.

A student's vertical scale score on the TELPAS reading assessment determines the student's proficiency level for the domain of reading. Proficiency level cut scores are discussed in the "Standard Setting" section of this chapter. So that progress from one year to the next can be monitored easily, TELPAS results for individual students include the student's proficiency level rating and scale score for the previous and current year.

For a detailed description of raw scores and scale scores, see chapter 3.

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 31.

Table 31. Weights of the Language Domains in TELPAS Composite Ratings

Year	Listening	Speaking	Reading	Writing
2009	5%	5%	75%	15%

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

Table 32 uses a student example to show how composite results are generated.

Table 32. 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 31 and then summed to obtain the TELPAS composite score, as shown:

Composite Score = (Listening x .05) + (Speaking x .05) + (Reading x .75) + (Writing x .15)

Composite Score = (3 x .05) + (2 x .05) + (3 x .75) + (2 x .15)

Composite Score = 2.80

TELPAS composite scores are converted to TELPAS composite ratings as shown below. This example composite score of 2.80 would result in a composite rating of advanced.

TELPAS Composite Score	TELPAS Composite Rating
1.0–1.5	Beginning
1.6–2.5	Intermediate
2.6–3.5	Advanced
3.6–4.0	Advanced High

Use of Results

Reports of TELPAS student performance are used in the following ways.

- Helping parents monitor the progress their children make in learning 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

Standard and Optional Reports

The standard reports available for the 2008–2009 TELPAS program include the Confidential Student Report, Confidential Student Label, Confidential Campus Roster, Campus and District Summary Reports.

The optional reports available include the Electronic Individual Student Record File (Confidential), additional copies of the Confidential Campus Roster, Confidential Student Report, Confidential Student Label, and the Campus and District Summary Reports. See chapter 4 for detailed information about standard and optional reports.

Parent Brochures

To assist teachers and parents in understanding students' TELPAS results, TEA's Student Assessment Division produces a series of brochures titled *Understanding the Confidential Student Report—A Guide for Parents*. Each grade-level brochure provides a brief summary of the TELPAS program, explains a sample Confidential Student Report so parents can understand their child's test report, and gives a brief summary of the meaning of each proficiency level. The brochures, developed in both English and Spanish, are provided to districts each spring for distribution with individual student TELPAS results.

Interpreting TELPAS Reports
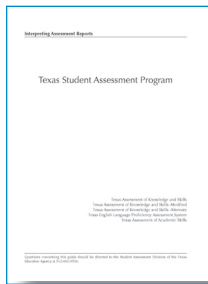
For more information about the reporting of TELPAS results, see the TEA publication *Interpreting Assessment Reports.*

Audits

Since the 2004–2005 school year, TEA has conducted periodic audits of the TELPAS assessment processes as one of the means of collecting reliability and validity evidence for the assessment program. Audits allow for the collection of information from school districts which can be used to evaluate the training, administration, and scoring of the holistically rated assessments. During the audits, expert raters provide second ratings of 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. Information about audits that have been conducted is provided in the "Reliability" and "Validity" sections of this chapter. Additional information can be found in technical digests from previous years.

Standard Setting

Standard setting is the process of relating levels of test performance directly to what students are expected to learn as expressed in the statewide curriculum standards.

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 number of questions students need to answer correctly 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.

Table 33 shows the scale score ranges and corresponding raw score cuts from the proficiency level setting activities that were conducted in 2008 when revisions to the TELPAS reading tests were implemented. While the scale score ranges remain constant from year to year, slight fluctuations in raw score cut points may occur. For more information about scale scores and the potential for raw score fluctuations in standardized assessments, refer to chapter 3.

Table 33. Approved Raw and Scale Score Cut Points from 2008 TELPAS Reading Proficiency Level Setting (Online Versions)

TELPAS Reading Assessment	Raw Score Cut Points	Total Number of Questions on Test	Beginning Level	Intermediate Level	Advanced Level	Advanced High Level
Grade 2	19, 30, 39	48	< = 550	551–614	615–669	> = 670
Grade 3	25, 36, 46	58	< = 596	597–647	648–698	> = 699
Grades 4–5	25, 37, 47	61	< = 609	610–667	668–717	> = 718
Grades 6–7	26, 39, 50	63	< = 612	613–673	674–730	> = 731
Grades 8–9	26, 37, 49	63	< = 632	633–680	681–737	> = 738
Grades 10–12	26, 39, 50	64	< = 643	644–703	704–756	> = 757

A summary of the standard setting activities for the TELPAS reading tests that were revised in 2007–2008 may be found in chapter 14 of the *2007–2008 Technical Digest*. More detailed information is available in the "TELPAS Reading Proficiency Level Setting" report on TEA's Student Assessment Division website.

Scaling

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 on the same scale across administrations for a given TELPAS assessment. Once performance standards have been set for an assessment, its initial scale is then transformed to a more user-friendly metric to facilitate interpretation of the test scores. Details of the RPCM scaling method used in Texas are provided in chapter 3.

Vertical Scale Score

A vertical scale allows for the direct comparison of student scores across years. Student increases in vertical scale scores provide information about the year-to-year growth of students. A vertical scale system was developed for the revised grades 2–12 TELPAS reading tests via a spring 2008 vertical scaling study, and the proficiency-level standards established in the summer of 2008 were mapped onto the vertical scale score system.

The vertical scale scores can be computed through a linear transformation of the student ability estimates using commercial software. For TELPAS, the Winsteps software is used to obtain student ability estimates.

The linear transformation is as follows:

$$VS_j = \{\theta_j - LC_V\}*A1 + A2$$

where VS_j is the vertical scale score for student j, θ_j is the estimated student ability for student j, LC_V refers to the vertical scale linking constant, and $A1$ and $A2$ refer to the vertical scale transformation constants.

Table 34. Vertical Scale Score Linking Constants for TELPAS Reading tests

Grade/Grade Cluster	LC_V	A1	A2
2	0		
3	0.73978		
4–5	1.16885		
6–7	1.24793	48	575
8–9	1.61612		
10–12	1.90165		

Appendix E provides frequency distributions and summary statistics for TELPAS assessments based on the vertical scale score as well as mean p-values by grade/grade cluster.

Further information about vertical scaling can be found in chapter 3. 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 Report" 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 calculation of the composite rating can be found in the "TELPAS Comprehension and Composite Scores" section of this chapter.

Equating

Used in conjunction with the scaling process, equating is the process that "balances" the slight difficulty differences across test forms and administrations to place the scores onto a common scale. Statistical methods are used to equate the results of different tests, enabling the comparison of scale scores across test forms and testing administrations.

Equating for grades 2–12 TELPAS reading is done using the Rasch IRT model. Pre-equating is done first (before the test form is administered) to link the newly built test form to the TELPAS reading base scale. This is done using item information collected during TELPAS field testing. (Field-test items are

embedded within each live test form so students are unable to tell which items are live test items and which items are field-test items.) After the test form is administered, post-equating is done using the item information obtained from the live test administration to move the test form onto the base scale. This allows the most recent information about how students responded to each item to be used when controlling for differences in the difficulty of test forms. Additional information about equating can be found in chapter 3.

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 and qualification 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

The concept of reliability is based on the idea that repeated administrations of the same test should generate consistent results about student performance. Reliability is a critical technical characteristic of any measurement instrument, because unreliable instruments cannot be interpreted in a valid way. Reliability estimates for TELPAS are mainly obtained through analyses of internal consistency, classical standard error of measurement, conditional standard error of measurement, classification accuracy, and interrater reliability.

Internal Consistency

Internal consistency is a measure of the consistency with which students respond to the items within a test. The Kuder-Richardson Formula 20 (KR20) was used to calculate the reliability estimates for TELPAS. As a general rule, reliability coefficients from 0.70 to 0.79 are considered adequate, 0.80 to 0.89 are considered good, and above 0.90 are considered excellent. However, what is considered appropriate may vary in accordance with how assessment results are used. For the spring 2009 TELPAS reading tests, internal consistency estimates were in the excellent range, with reliabilities for both the online version and paper version ranging from 0.93 to 0.96. 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 E presents reliability estimates for each subgroup of items (beginning, intermediate, advanced, and advanced high). As expected, reliabilities for each item subgroup are lower (ranging from 0.70 to 0.90) than the overall test reliability because the reliability estimate is based on fewer items. However, values are still considered adequate to good and no student-level interpretations are based on responses to only one subgroup of items. All reliability estimates are calculated for all students and for gender groups.

Classical Standard Error of Measurement

Classical standard error of measurement (SEM) provides a reliability estimate for a test score by representing the amount of variance in a score resulting from factors other than language proficiency. The SEM provides an estimate of the average test score error for all students regardless of their proficiency level. See chapter 3 for detailed information about SEM. The SEM values (shown in Appendix E) for TELPAS reading tests are between 2 and 3 raw score points across grades/grade clusters.

Conditional Standard Error of Measurement

Conditional standard error of measurement (CSEM) provides a reliability estimate at each score point on a test. See chapter 3 for detailed information about CSEM. In general, the CSEM values for TELPAS reading tests are between 15 and 25 scale score points, with a CSEM value of 15 occurring in the middle of the score range. These values should be interpreted based on the TELPAS vertical scale range of approximately 200 to 1000 scale score points. Appendix E 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. Classification accuracy rates for the 2008–2009 school year ranged from 81% to 85%, with highest estimates in grades 8–9 and lowest estimates in grade 2. Appendix E provides classification accuracy rates for each grade/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 in two general ways. First, information about the consistency with which raters adhere to the strict administration protocol is provided through voluntary surveys that are periodically conducted and through mandatory questionnaires that a sample of campus and district personnel are required to complete during audits of the rating process. The information gathered provides evidence of the efficacy of the training and administration procedures used for TELPAS. Second, evidence of interrater reliability is collected through the audit process by having a second rater provide independent ratings for a sample of audited students. For writing audits, the second rater provides second ratings based on the same collection of student work used by the first rater. For listening and speaking audits, the second rater provides a second rating based on independent observations of the student during classroom instruction.

In addition, an analysis of the composite reliability estimates of TELPAS has been conducted to evaluate the impact of various potential reliabilities of the listening, speaking, and writing domains on TELPAS composite reliability estimates. The results of this analysis indicate that the weighted TELPAS composite ratings have reliability estimates that exceed 0.89 even with conservative (lower-bound) estimates of TELPAS listening, speaking, and writing reliabilities. The high internal consistency reliability of TELPAS reading scores and high interrater reliability of TELPAS writing ratings combined with the heavy weighting of these domains produce TELPAS composite ratings with high internal consistency. Additional information is available in the "TELPAS Composite Reliability Estimates" report on TEA's Student Division website.

Validity

Validity refers to the extent to which the test measures what it is intended to measure. Validity evidence for an assessment can come from a variety of sources including test content, the response process, the internal structure, relationships with other variables, and the consequences of testing. The sections that follow describe how these types validity of evidence were collected for the TELPAS assessments in 2008–2009.

The results of TELPAS assessments are used to guide educational planning related to the progress that ELLs make in acquiring the English language. Refer to the "Use of Results" section of this chapter for a description of the specific purposes for which TELPAS results are to be used.

Evidence of the validity of the reading, writing, listening, and speaking domains of TELPAS has been collected since the first administration in 2003–2004 and continues to be collected. In addition to the studies described in this year's digest, a wide range of validity studies and analyses have been conducted and documented in the Technical Report Series and Technical Digests for previous years. These documents 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 proficiency level descriptors defined by the Texas ELPS, which 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 Components

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 U.S. schools and 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 test questions per proficiency level and per test objective (reading skill category). The 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 its test design in that it provides built-in, staged linguistic accommodations validated by second language acquisition theory and empirical data as it measures ELPS-aligned reading skills students need for academic success in all subject areas. The staged linguistic accommodation test design is shown in Table 35.

Table 35. 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

TELPAS reading material requires students to comprehend the type of written English they encounter in everyday life and grade-level core content instruction. Most of the topics and contexts come from the content areas of language arts, mathematics, and science, although other subject matter is eligible as well. Questions that assess the higher proficiency levels tap the ability of students to think critically and conceptually when 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 the test designs differ, the quality assurance steps used to develop the multiple-choice TELPAS reading tests and TAKS tests are the same. The process adheres to the Standards for Educational and Psychological Testing (AERA/APA/NCME, 1999), is grounded in the state's standards, and is guided by assessment experts and educators who have first-hand knowledge of the standards and students. As with TAKS, 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 TAKS and TELPAS reading tests—and includes TAKS assessment and content area experts throughout the development process—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 also aligned with the ELPS and designed to assess the English communication skills that ELLs need 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 very 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 proficiency level descriptors as assessment rubrics. Rater training and administration procedures require 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 stems from the goal 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 high ELL population. The TELPAS holistically rated assessments address these priorities. Because of the direct involvement teachers across the state have in the assessment process, the holistically rated assessments have a direct and significant positive effect on classroom instruction.

Evidence Based on Response Processes

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

TELPAS MULTIPLE-CHOICE COMPONENTS

Theoretical and empirical evidence was used to pilot and determine the appropriateness of each item type used on the multiple-choice reading tests. A variety of question-and-answer and cloze (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 the item's format is appropriate and whether students are able to accurately demonstrate the knowledge of the construct assessed. When items are field-tested, data are gathered about students' responses to items, and statistical information such as item difficulty for students at each proficiency level, item point-biserial correlations, and differential item functioning is evaluated.

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 and attaining academic language proficiency so that 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 to engage in state-required academic instruction, which provides strong validity evidence related to the response process.

Evidence Based on Internal Structure

Texas collects evidence that shows the relationship among test questions and test objectives to demonstrate that the parts of a test conform to the test construct.

TELPAS MULTIPLE-CHOICE COMPONENTS

Internal structure is evaluated annually through estimation of internal consistency reliability for the TELPAS multiple-choice components. As stated in the "Reliability" section of this chapter, internal consistency reliability estimates provide a measure of the consistency with which students respond to the items within a test.

The internal consistency of TELPAS reading tests is evaluated each year using KR20 statistics. As noted in the "Reliability" section of this chapter, the 2008–2009 internal consistency estimates were in the excellent range for both the online and paper version of the tests. Across grade clusters, the internal consistency estimates were 0.90 or higher (with 1.0 being perfectly reliable), indicating that it is appropriate to use TELPAS reading test scores for student-level interpretations.

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 district administrators to follow procedures to ensure the integrity of the test administration. Stringent training requirements are used to maximize rating accuracy associated with

the holistically assessed components of TELPAS. Refer to the "Holistic Rating Training" and "Administration Procedures" sections of this chapter for detailed information.

In addition, the TELPAS 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 collected 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 as intended.

The TELPAS holistically rated assessments squarely address the state's goal of having both a valid and authentic assessment and a critical ongoing professional development tool that supports effective instruction so that teachers 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 assessments is by analyzing the relationship between test performance and performance on external measures.

By examining the relationship between TELPAS test performance and performance on external measures, evidence can be collected supporting that the empirical relationships found from the comparisons are consistent with the relationships expected at the level of the construct underlying the proposed test score interpretations. Texas conducted two studies in 2008–2009 (one reading study and one cross-domain study) to investigate student performance on TELPAS compared with student performance on external measures.

RELATIONSHIP OF TAKS PASSING STANDARD AND TELPAS PROFICIENCY LEVELS

This study was conducted to show the relationship between TAKS reading and English language arts (ELA) scale scores and TELPAS reading ratings. Table 36 shows that as a student's TELPAS proficiency rating increases so does the TAKS scale score. The strong relationship between acquiring English language reading proficiency as measured on TELPAS and success on grade-level reading achievement tests provides validity evidence for TELPAS.

Table 36. 2009 English TAKS Scale Score Performance by TELPAS Reading Proficiency Rating for Students Who Participated in Both Assessments

Grade Level	TELPAS Reading Proficiency Rating	Number of Students Participating in Both Assessments	Average TAKS Reading/ELA Scale Score
3	Beginning	1,262	1957.45
	Intermediate	6,816	2059.41
	Advanced	15,119	2183.25
	Advanced High	30,092	2346.75
4	Beginning	336	1908.70
	Intermediate	4,334	1993.81
	Advanced	12,601	2089.77
	Advanced High	31,305	2250.35
5	Beginning	165	1862.79
	Intermediate	1,839	1913.43
	Advanced	6,281	1988.96
	Advanced High	28,332	2159.92
6	Beginning	216	1908.22
	Intermediate	2,260	1981.15
	Advanced	8,758	2082.49
	Advanced High	18,682	2244.07
7	Beginning	165	1883.96
	Intermediate	1,208	1927.98
	Advanced	4,853	1991.03
	Advanced High	16,699	2140.50
8	Beginning	234	1914.03
	Intermediate	1,176	1956.35
	Advanced	4,346	2047.57
	Advanced High	11,112	2205.68
9	Beginning	442	1895.87
	Intermediate	1,528	1941.38
	Advanced	4,327	2017.07
	Advanced High	10,172	2144.38
10	Beginning	250	1919.09
	Intermediate	1,354	1968.08
	Advanced	3,944	2037.88
	Advanced High	7,855	2136.54
11	Beginning	256	1943.42
	Intermediate	1,152	1981.85
	Advanced	2,825	2046.18
	Advanced High	6,348	2153.77

NOTE: The scale scores necessary for the TAKS Met Standard and Commended Performance levels are 2100 and 2400 respectively.

Evidence Based on Consequences of Testing

Another way to provide validity evidence is by 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). Others might call it "efficacy of actions following assessment" (Mehrens, 1997). The administration of TELPAS holistically rated assessments leads to improvements in students' academic language acquisition because of what educators learn during the rater training process and through direct application of the assessment process for both formative and summative purposes. Shepard (1997) claimed that, "a test carefully tied by logical and empirical evidence to the intended content domain is valid for reporting on the status or level of student achievement." 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 interact more with students, adjust content area instruction to make it comprehensible, and 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. These results show incremental increases in ELL performance in all TELPAS domains from spring 2005 to spring 2009. These increases in student performance provide additional evidence of the consequential validity of TELPAS.

Sampling

Sampling plays a critical role in the research and annual development activities necessary to support the Texas assessment program. See chapter 3 for detailed information about sampling.

A sample is a group of students smaller than the population in question that can be used to represent the overall population. Through the careful selection of student samples, TEA is able to gather reliable information about student performance on its tests while minimizing campus and district participation. In particular, sampling is used in the Texas assessment program for: 1) testing that is part of a research study, and 2) stand-alone field tests. Research studies and stand-alone field tests requiring sampling were not necessary for TELPAS during the 2008–2009 school year.