Chapter 6 Texas English Language Proficiency Assessment System (TELPAS)

Overview

Participation Requirements

Test Development

Training

Test Administrations

Scores and Reports

Audits

Performance Standards

Scaling

Equating

Reliability

Validity

Sampling

Test Results

Overview

TELPAS measures the progress that emergent bilingual* (EB) students make in acquiring the English language. As required by Title III, Part A of the Elementary and Secondary Education Act (ESEA), and then reauthorized by 34 C.F.R. §200.6 of the Every Student Succeeds Act (ESSA), states must conduct annual statewide English language proficiency assessments for EB students in grades K–12 in the language domains of listening, speaking, reading, and writing. Prior to ESEA, Texas developed and administered English language proficiency tests in the domain of reading, as required by Texas state law.

The 2021–2022 TELPAS online assessments for grades 2–12 include multiple-choice items, technology enhanced items, and the automated scoring of speaking items. All four

*Note: As of the 2021-2022 school year, the term “emergent bilingual student” replaced the terms “English language learner” and “English learner” due to legislative requirements.

language domains for grades K–1 and the writing domain for grades 2–12 are performance-based and holistically rated assessments. For each language domain, TELPAS measures four levels, or stages, of English language proficiency: beginning, intermediate, advanced, and advanced high.

TELPAS assesses English language proficiency in direct alignment with the Texas English Language Proficiency Standards (ELPS) that are an integral part of each foundation and enrichment subject of the Texas Essential Knowledge and Skills (TEKS) curriculum. The ELPS outline the instruction that EB students 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 EB students are expected to learn, as well as proficiency level descriptors (PLDs) 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 ESSA requirements that are specific to the English proficiency and academic achievement of EB students. Composite performance rather than individual language domain performance is used to determine EB students' progress in achieving English language proficiency in TELPAS. For information about how TELPAS composite results are generated, refer to the TELPAS Composite Score section in this chapter.

Teachers also use TELPAS student-level results to design instruction and plan interventions that appropriately address the student's linguistic and academic needs.

Participation Requirements

All K–12 EB students are required to participate in TELPAS, including EB students whose parents have declined bilingual or English as a second language (ESL) program services. EB students are required to be assessed annually until they meet Emergent Bilingual/English Learner Reclassification Criteria and are reclassified as non-EB/English proficient.

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 EB student receiving special education services should not be assessed in listening, speaking, reading, and/or writing for reasons associated with the student's 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 in the student's permanent record file by the LPAC.

Newly Enrolled EB Students

An EB student 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 online TELPAS listening and speaking test and reading tests.

Test Development

TELPAS Assessments for Grades 2–12

TELPAS reading has been assessed online since 2008. Starting in 2018, TELPAS listening and speaking have also been assessed online. The TELPAS reading test for grades 2–12 employs an online multiple-choice answer format. For the 2022 administration, writing items were added as an embedded field test in the online reading test to prepare for a standardized writing test. These items consisted of multiple-choice, constructed response and short text entries. TELPAS also administers the listening and speaking test online but uses a variety of item formats including picture-based, drag-and-drop, passage-based, and non-passage-based types. In the rare instance that circumstances prevented students from accessing an online assessment, a special administration was available with TEA approval. TELPAS reading was available in paper form, while a special administration of TELPAS listening and speaking was holistic.

Listed below are TELPAS definitions of English language proficiency for each of the domains:

- TELPAS defines English language proficiency in listening as the ability to understand spoken language, comprehend and extract information, and follow social and instructional discourse. EB students who are English proficient in the domain of listening understand spoken English well enough to participate meaningfully, and with minimal second language acquisition support, in grade-level academic instruction.
- TELPAS defines English language proficiency in speaking as the ability to use spoken English appropriately and effectively in learning activities and in social interactions. This definition relates specifically to the communication skills that an EB student needs in order to use English as an effective medium for academic instruction.
- TELPAS defines English language proficiency in reading as the ability to comprehend and interpret written text at the grade-appropriate level. This definition is not tied specifically to the language arts discipline, but, more broadly, to the ability to read texts typically encountered during all grade-level instruction.
- TELPAS defines English language proficiency in writing as the ability to produce written text with content and format to fulfill grade-appropriate classroom assignments. This definition relates specifically to the communication skills that an EB student needs in order to use English as an effective medium for academic instruction. As previously described, the grades 2–12 TELPAS writing assessment consists of performance-based student writing collections, holistically rated by

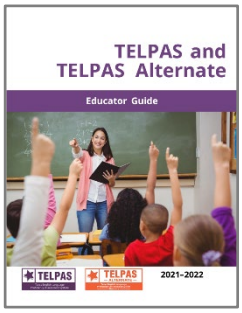
educators trained to be raters.

Six grade-cluster tests are administered for TELPAS reading, and four grade-cluster tests are administered for combined TELPAS listening and speaking, as shown in Table 6.1.

Table 6.1. Grade Clusters for TELPAS 2–12

Grade Clusters for TELPAS 2–12 Reading	Grade Clusters for TELPAS 2–12 Listening and Speaking
Grade 2	Grades 2–3
Grade 3	Grades 4–5
Grades 4–5	Grades 6–8
Grades 6–7	Grades 9–12
Grades 8–9	
Grades 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 tests for grades 2–12 is available in the TELPAS and TELPAS Alternate Educator Guide available on the TELPAS Resources webpage. This guide is provided to familiarize educators with TELPAS. It shows the integral relationship between TELPAS and the ELPS. It also explains the TELPAS language domains of listening, speaking, reading, and writing as well as providing sample test questions for reading and annotated test item descriptions for listening and speaking.

TELPAS Holistically Rated Assessments

The TELPAS holistically rated components assess all four domains in grades K–1 and writing in grades 2–12. To conduct these assessments, raters are specially trained to rate the English language proficiency of EB students based on an evaluation of student writing, classroom observations in core content areas, and daily interactions with the students.

The rating process identifies a student's level of English language proficiency and is holistic rather than a measure of isolated skills. Raters are trained to use the ELPS 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 and ESL consultants; and members of an EB focus group composed of teachers, bilingual and ESL directors, assessment directors, campus

administrators, and university professors. Like the TELPAS listening, speaking, and reading tests for grades 2–12, these assessments align with the ELPS assessing the English communication skills that EB students need to engage meaningfully and effectively in learning the academic knowledge and skills required by the state's curriculum, the TEKS. The holistically rated assessments draw on second language acquisition research, research-based standards, the experience of Texas practitioners, and observational assessment practices.

Together, the online TELPAS tests for grades 2–12 and the holistically rated writing K–1 and 2–12 component of TELPAS combine multiple-choice testing and modern technology-enhanced methods with an authentic, performance-based writing assessment to measure the construct of academic English language proficiency.

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 online. TEA has additional training resources to provide support as needed posted on the TELPAS Resources webpage.

The TELPAS online basic training courses are provided to teach new raters the essentials of second language acquisition theory. They also teach raters how to use the PLDs from the ELPS to accurately identify the English language proficiency levels of their EB students based on how well the students understand and use English during daily academic instruction and classroom interaction. The online basic training course for grades 2–12 contains practice rating activities that comprise student writing collections. Online courses for K–1 contain numerous practice rating activities that comprise student writing samples and video segments in which EB students demonstrate their listening, speaking, reading, and writing skills in authentic Texas classroom settings. The courses give raters practice applying the scoring rubrics (i.e., PLDs) and provide detailed feedback about their rating accuracy.

New raters were 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. Returning raters who had completed online calibration activities successfully at least once were not required to complete calibration activities. However, a campus coordinator at his or her discretion may have required additional training requirements, including calibration activities. Beginning in the 2010–2011 school year, calibration activities were provided for all holistically rated domains—listening, speaking, reading (K–1 only), and writing. Starting in 2018, calibration activities were provided for all domains for K–1 and for writing for grades 2–12. In addition, calibration activities were provided for raters of EB students approved for a special administration for listening and speaking. There are two sets of calibration activities, and all applicable language domains are represented. To demonstrate sufficient calibration, raters are required to rate at least seven out of ten students correctly within a set for grades K–1 and grades 2–12. Raters finish the calibration activities when they demonstrate sufficient accuracy. If sufficient accuracy is not obtained on the first set, the rater attempts a

second 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 2021–2022 school year, approximately 113,400 raters successfully calibrated within the two attempts.

Test Administrations

The 2022 TELPAS testing window lasted six weeks. The testing window was extended by one week for holistic tests exclusively, to allow districts more time to enter student ratings into the online data entry interface.

During the 2021–2022 school year, more than 893,000 TELPAS reading and over 892,000 TELPAS listening and speaking assessments were administered online. Holistic rating information was also collected online for writing for these students, as well as for an additional approximately 195,000 K–1 students for all domains. Districts administered the TELPAS assessments to eligible students as indicated in Table 6.2.

Table 6.2. TELPAS Assessments Administered in 2021–2022

TELPAS Grade	Assessments Administered			
	Listening	Speaking	Reading	Writing
Kindergarten	95,188	95,119	95,045	95,060
Grade 1	100,132	100,028	99,945	99,940
Grade 2	96,840	96,835	96,891	96,403
Grade 3	98,822	98,817	98,876	98,362
Grade 4	100,354	100,352	100,374	99,884
Grade 5	98,982	98,979	99,031	98,489
Grade 6	94,690	94,688	94,763	94,162
Grade 7	92,519	92,516	92,640	92,026
Grade 8	81,951	81,949	82,033	81,430
Grade 9	84,571	84,570	84,774	83,291
Grade 10	59,135	59,133	59,276	58,697
Grade 11	48,788	48,788	48,977	48,568
Grade 12	36,047	36,045	36,226	36,155
Total	1,088,019	1,087,819	1,088,851	1,082,467

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 online training components, are found in the District and Campus Coordinator Resources 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 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 EB student 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 EB students 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 webpage 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 use common, basic English in routine academic activities but need considerable English language support to make instruction understandable. Socially, these students 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 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 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 EB students.

Advanced level of English language proficiency: Students who receive this rating have an emerging academic English vocabulary, which they 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 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 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 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 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 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 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 online assessments include proficiency level ratings, the number of items answered correctly (raw scores), and scale scores. For the holistically rated assessments (all domains for K-1 students; the writing domain for students in grades 2-12; and the listening and speaking domains for students in grades 2-12 eligible for special administrations), the domain score consists of the proficiency level rating of beginning, intermediate, advanced, or advanced high. The scores are recorded on student rating rosters; the rosters are filed at the local level; and the scores are submitted to Cambium Assessment through the secure Data Entry Interface (DEI).

Raw Score

The number of total points scored on the online items is provided separately for each domain. 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.

Prior to 2018, TELPAS reading scale scores were reported on a vertical scale. Starting in 2018, TELPAS listening, speaking, and reading scale scores were reported on grade-band horizontal scales. Horizontal scaling for TELPAS is discussed further in the Scaling section of this chapter.

In grades 2–12, a student's scale score on a TELPAS domain determines the student's proficiency level for that domain. To facilitate the monitoring of a student's progress from one year to the next, TELPAS results for individual students include each student's proficiency level rating and scale score for the previous and the current year.

Proficiency level cut scores are discussed in the Performance Standards section of this chapter.

TELPAS Composite Score and Composite Rating

In addition to receiving a rating of beginning, intermediate, advanced, or advanced high for each domain, a composite score and composite rating are also calculated for students. The composite rating is provided to students on the TELPAS report card.

The TELPAS composite scores and ratings 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 equally weighted, as shown in Table 6.3. These weights were used for the first time in 2018 to coincide with the new TELPAS listening, speaking, and reading assessments.

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

Listening	Speaking	Reading	Writing
25%	25%	25%	25%

After a TELPAS composite score is calculated, a TELPAS composite rating is determined according to the rules below. All the criteria listed for a particular rating must be met for a student to receive that rating.

Beginning:

- a student whose composite score fails to meet the intermediate requirements will receive a beginning proficiency rating

Intermediate:

- a TELPAS composite score of 1.5 or higher
- a minimum proficiency level of intermediate in at least half of the domains in which the student was assessed

Advanced:

- a TELPAS composite score of 2.5 or higher
- a minimum proficiency level of intermediate in all domains
- a minimum proficiency level of advanced in at least half of the domains in which the student was assessed

Advanced high:

- a TELPAS composite score of 3.5 or higher
- a minimum proficiency level of advanced in all domains

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 × 0.25) + (Speaking × 0.25) + (Reading × 0.25) + (Writing × 0.25)

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

Composite Score = (3 × 0.25) + (2 × 0.25) + (3 × 0.25) + (2 × 0.25) = 2.5

The TELPAS composite scores are converted to the TELPAS composite ratings. This example composite score of 2.5 would result in a composite rating of advanced due to the ratings profile having

- a TELPAS composite score of 2.5 or higher,
- a minimum proficiency level of intermediate in all domains, and
- a minimum proficiency level of advanced in at least half of the domains in which the student was assessed.

A small subset of EB students with disabilities who cannot be assessed in all four domains will receive a composite score if they have results for at least two domains. This is only applicable to students who have a decision from the admission, review, and dismissal (ARD) committee, in conjunction with the language proficiency assessment committee (LPAC), to not be evaluated in one or two domains.

In these instances when not all four domains are assessed, the composite score will be calculated based on the number of tests that that were assessed.

Figure 6.2 provides a student example to show how composite results missing one domain are generated.

Figure 6.2. Sample Calculation of Composite Results Missing One Domain

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

Domain	Proficiency Level	Domain Score
Listening	Intermediate	2
Speaking	Intermediate	2
Reading	Beginning	1
Writing	Not Assessed	Not Assessed

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 × $\frac{1}{3}$)) + (Speaking × $\frac{1}{3}$)) + (Reading × $\frac{1}{3}$))

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

Composite Score = (2 × $\frac{1}{3}$) + (2 × $\frac{1}{3}$) + (1 × $\frac{1}{3}$) = 1.7

Round to one decimal point with 0.05 rounding up to 0.1.

The TELPAS composite scores are converted to the TELPAS composite ratings. This example composite score of 1.7 would result in a composite rating of intermediate due to the ratings profile having

- a TELPAS composite score of 1.5 or higher, and
- a minimum proficiency level of intermediate in at least half of the domains in which the student was assessed.

Assessment Reports

Standard and additional reports are provided for the various testing programs. Standard reports are provided automatically to districts. Information contained in standard reports satisfies mandatory reporting requirements. Districts had the option to request additional

reports through an online form. They were required to pay a nominal fee for each additional report requested.

YEARLY PROGRESS INDICATOR

The student's yearly progress indicator provides information about the yearly proficiency level progress that an EB student makes in acquiring the English language. This measure is based on a comparison of a student's composite rating in the previous year with his or her composite rating in the current year. The yearly statewide summary reports provide the number and percentage of students who progressed one, two, or three proficiency levels. The yearly statewide summary reports also provide the number and percentage of students who progressed at least one proficiency level. The yearly progress indicator is set as follows:

- If a student received a composite rating one level higher than the previous year, the student's yearly progress indicator is 1. Additionally, if a student received an advanced high composite rating (4) in the current year and an advanced high composite rating (4) in the previous year, the student's yearly progress indicator is also 1.
- If a student received a composite rating two levels higher than the previous year, the student's yearly progress indicator is 2.
- If a student received a composite rating three levels higher than the previous year, the student's yearly progress indicator is 3.
- If a student received a current year composite rating that is the same as the previous year's composite rating (excluding an advanced high composite rating of 4) or lower than the previous year's rating, the yearly progress indicator is 0.

For more information about reporting 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

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 was used in which documentation was 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. Starting in 2018, with the replacement of holistically scored assessments with an online assessment, no further audits are needed for TELPAS listening and speaking.

A TELPAS writing audit was conducted in spring 2019. During the TELPAS writing audit, expert raters provided second ratings of writing samples of students in the state, and testing personnel at the sampled sites completed questionnaires that allowed the state to evaluate conformity with training and administration procedures. Refer to 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 accomplished 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 online tests, standards are established by determining the score students need to obtain to be classified into specified performance categories. For the TELPAS online listening, speaking, and 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.

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 change to the TELPAS reading test design in spring 2018, in addition to the development of an online test for the listening and speaking domains, required establishing cut scores for the four TELPAS proficiency levels. The new standards were first implemented with the spring 2018 administration. Table 6.4 shows the scale score ranges from the proficiency level standard setting meetings conducted in summer 2018. The scale score ranges remain constant from year to year, even though 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 standard setting process is available in the TELPAS Standard Setting Technical Report on TEA's Student Assessment Division website.

Table 6.4. Approved Scale Score Cut Scores from 2018 TELPAS Standard Setting

Domain	TELPAS Assessment Grade Bands	Beginning Level	Intermediate Level	Advanced Level	Advanced High Level
Reading	Grade 2	1000 to 1439	1440 to 1524	1525 to 1599	1600 to 2000
	Grade 3	1000 to 1434	1435 to 1524	1525 to 1599	1600 to 2000
	Grades 4–5	1000 to 1430	1431 to 1524	1525 to 1599	1600 to 2000
	Grades 6–7	1000 to 1446	1447 to 1524	1525 to 1599	1600 to 2000
	Grades 8–9	1000 to 1437	1438 to 1524	1525 to 1599	1600 to 2000
	Grades 10–12	1000 to 1426	1427 to 1524	1525 to 1599	1600 to 2000
Listening	Grades 2–3	1000 to 1441	1442 to 1524	1525 to 1599	1600 to 2000
	Grades 4–5	1000 to 1455	1456 to 1524	1525 to 1599	1600 to 2000
	Grades 6–8	1000 to 1430	1431 to 1524	1525 to 1599	1600 to 2000
	Grades 9–12	1000 to 1447	1448 to 1524	1525 to 1599	1600 to 2000
Speaking	Grades 2–3	1000 to 1410	1411 to 1524	1525 to 1599	1600 to 2000
	Grades 4–5	1000 to 1466	1467 to 1524	1525 to 1599	1600 to 2000
	Grades 6–8	1000 to 1459	1460 to 1524	1525 to 1599	1600 to 2000
	Grades 9–12	1000 to 1484	1485 to 1524	1525 to 1599	1600 to 2000

Scaling

Scaling is a statistical procedure that places raw scores on a common scoring metric 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 listening, speaking, and 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

Scale scores for TELPAS assessments are reported on a horizontal scale. Horizontal scale scores allow for direct comparisons of student performance between specific sets of test items from different test administrations for a specific grade/grade band and subject. Refer to Chapter 3, "Standard Technical Processes," for detailed information about the scaling process for the different types of reporting scales.

Horizontal Reporting Scales

Prior to 2018, TELPAS reading was reported on a vertical scale. The reporting scales for each assessment for the three domains (listening, speaking, and reading) are independent horizontal scales with lowest obtainable scale scores of 1000 and highest obtainable scale scores of 2000. The cut scores on the reporting scale for the advanced and advanced high proficiency levels are 1525 and 1600, respectively, to create common points of reference across the assessments for each grade and domain. It is important to note that although the advanced high and advanced scale score values are fixed across horizontally scaled assessments, the intermediate scale score values vary across TELPAS. For a given assessment, the intermediate, advanced, and advanced high scale score values remain constant over time.

The TELPAS scale scores represent linear transformations of Rasch proficiency level estimates (θ). Specifically, the transformation is made by first multiplying θ by a slope constant (A) and then adding an intercept constant (B). This operation is described by the equation below:

$$SS_\theta = A * \theta + B$$

where SS_θ is the scale score for a Rasch proficiency level estimate (θ). A and B are referred to as the horizontal scaling constants. The values of A and B 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. Horizontal Scaling Constants for TELPAS Tests

	Grade Cluster	*A*	*B*
Reading	Grade 2	66.7438	1423.0422
	Grade 3	88.0488	1396.6160
	Grades 4–5	86.5951	1391.3838
	Grades 6–7	79.5756	1380.2599
	Grades 8–9	68.8452	1408.3486
	Grades 10–12	64.4607	1389.4972

	Grade Cluster	*A*	*B*
Listening	Grades 2–3	67.4946	1497.4015
	Grades 4–5	64.5661	1482.9804
	Grades 6–8	67.6285	1486.0798
	Grades 9–12	53.7172	1497.3517
Speaking	Grades 2–3	35.0533	1511.4519
	Grades 4–5	24.6208	1522.0652
	Grades 6–8	19.5008	1530.4446
	Grades 9–12	21.0574	1545.1456

Further information about scaling appears in Chapter 3, “Standard Technical Processes."

SCALE FOR HOLISTICALLY RATED ASSESSMENTS

The scale for the TELPAS holistically rated assessments (all domains for grades K–1, and writing for grades 2–12) ranges from 1 to 4 and is defined by the four proficiency levels: beginning, intermediate, advanced, and advanced high.

SCALE FOR COMPOSITE SCORE

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 Composite Score and Composite Rating section of this chapter.

Equating

Used in conjunction with the scaling process, equating is the statistical process that considers 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 online TELPAS assessments is done using the RPCM. In the 2021–2022 school year, equating activities for all online TELPAS 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 before test administration. It links a newly developed test form onto the scale of the item bank using 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 before its administration, and thus, the anticipated raw scores that correspond to scale scores at performance standards can be identified. Pre-equating is conducted for all TELPAS test forms during the test construction process.

Post-Equating

Post-equating is conducted for the online forms of TELPAS assessments after the test administration. 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 online assessments, post-equating uses conventional common-item/non-equivalent groups equating procedures. Post-equating is conducted on all online test forms.

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. 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. Field-test equating is conducted for all online TELPAS 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 using 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 raters to the assessment rubric every time they administer the holistically rated portions of TELPAS.

Reliability

Reliability refers to the expectation that repeated administrations of the same test should generate consistent results. Reliability is a critical technical characteristic of any measurement instrument because unreliable scores cannot be interpreted as valid indicators of students' knowledge and skills.

Reliability estimates for the TELPAS scores are obtained mainly through analyses of internal consistency, classical standard error of measurement, conditional standard error of measurement, classification consistency and accuracy, and interrater reliability. Refer to Chapter 3, "Standard Technical Processes," for detailed information about reliability.

Internal Consistency

Internal consistency is a measure of the consistency with which students respond to the items in a test. The Kuder–Richardson Formula 20 (KR20) was used to calculate the reliability estimates for TELPAS reading scores. For the TELPAS listening and speaking scores, an extension of KR20 for polytomous items, coefficient alpha is reported. Generally, reliability coefficients from 0.70 to 0.79 are considered adequate, those from 0.80 to 0.89 are considered good, and those equal to or greater than 0.90 are considered excellent. However, what is considered appropriate might vary depending on how assessment results are used.

Internal consistency estimates for the spring 2022 TELPAS assessments indicate that the reliability estimates were all in the good to excellent ranges in terms of appropriateness for student-level interpretations. In addition to the overall test reliability, Appendix D presents reliability estimates by reporting category and gender as well.

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.

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 administrations of TELPAS.

Classification Consistency and Accuracy

Classification consistency provides an estimate of the consistency of student classifications into proficiency levels for parallel test forms, while classification accuracy provides an estimate of the accuracy of student classifications into proficiency levels based on current test results. Appendix D provides classification consistency and accuracy rates for each grade cluster of the TELPAS listening, speaking, and 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.

In 2019, a writing audit was conducted for grades 2–12. For the audit, districts were required to submit writing collections for EB students selected for the pure random sample. The collections included writing from classroom instruction in a variety of core content areas. About 2,200 students were selected for the 2019 writing audit, spread across grade levels and stratified across proficiency levels. The Pearson Performance Scoring Center rescored the writing collections after the original scores were collected from the raters. Rescoring was completed in September 2019. The results of this audit process can be found in the TELPAS Writing Audit Report and add to the body of validity and reliability evidence collected to support the assessment system. This process enables the state to evaluate the classroom activities on which the assessments are based and the way raters statewide interpret the PLD rubrics. The same information

collected during TELPAS audits has been useful in the refinement of TELPAS holistic rating training and administration procedures.

For TELPAS speaking, an additional validity check is performed on the automated scoring of the responses to check interrater reliability between automated and human scoring. A random sample of approximately 5,000 students per grade band is selected for human scoring. The grade band correlations between the total raw scores on the human scored and automated scored samples are presented in Table 6.6.

Table 6.6. 2022 TELPAS Speaking Validity Correlations of Total Raw Scores

Grade	N	Interrater Correlation
Grades 2–3	5678	0.86
Grades 4–5	5387	0.84
Grades 6–8	5183	0.83
Grades 9–12	5049	0.89

When students are ready to respond to a prompt, they use a speech capture tool in the online testing interface to record their responses. They have up to 45 seconds for some items, and 90 seconds for other items, for a response. Students are allowed two recording attempts per item. After the first attempt, students are allowed to listen to their recording. If desired, they may delete the first recording and record a second response.

Table 6.7 shows the median speech time per item that the randomly sampled students recorded on the 2021–2022 administration by proficiency level.

Table 6.7. TELPAS Speaking Median Speech Time per Item in Seconds by Proficiency Level

TELPAS Proficiency Level	Median Speech Time per Item in Seconds			
	Grades 2–3	Grades 4–5	Grades 6–8	Grades 9–10
Beginning	0.8	2.2	0.0	0.0
Intermediate	13.6	13.1	9.0	12.7
Advanced	26.5	24.4	20.5	25.9
Advanced High	43.3	43.0	41.0	42.7

In addition, the composite score reliability estimates of TELPAS are analyzed annually to evaluate the impact of the reliability of the listening, speaking, reading, and writing domains on the TELPAS composite reliability estimates. The composite score reliability estimates were calculated using a stratified alpha approach. The equation of stratified alpha can be seen as below:

$$\alpha_{Strat} = 1 - \frac{\sum_{i=1}^{k} w_i^2 \sigma_{X_i}^2 (1 - \rho_{X_i X_i'})}{\sigma_Z^2}$$

where k is the number of the components/domains, w_i is the weight of each domain, X_i represents the domain score of each domain, $\rho_{X_i X_i'}$ is the internal consistency of each domain and Z is the composite score. The internal consistency of writing was

constrained to the interrater reliability (perfect agreement) based on the 2019 TELPAS writing audit. The internal consistency values of listening, speaking, and reading on the categorical scale were estimated based on their internal consistency values on the continuous scale. The results of these analyses, presented in Table 6.8, show that the weighted TELPAS composite scores have reliability estimates of at least 0.906.

Table 6.8. Estimated Reliability of the TELPAS Composite Score

Grade	Subject	Mean	Standard Deviation	Internal Consistency	Composite Reliability
Grade 2 (*n*= 95930)	Listening	2.864	0.903	0.734	0.918
	Speaking	2.036	0.716	0.710	
	Writing	2.112	0.976	0.910	
	Reading	2.011	0.940	0.789	
Grade 3 (*n*= 97935)	Listening	3.338	0.824	0.734	0.927
	Speaking	2.310	0.780	0.724	
	Writing	2.441	0.982	0.930	
	Reading	2.520	1.067	0.821	
Grade 4 (*n*=99519)	Listening	2.636	0.964	0.761	0.929
	Speaking	2.395	0.798	0.729	
	Writing	2.778	0.972	0.930	
	Reading	2.645	1.015	0.822	
Grade 5 (*n*=98122)	Listening	2.900	0.962	0.758	0.929
	Speaking	2.431	0.817	0.738	
	Writing	3.011	0.942	0.930	
	Reading	2.958	1.002	0.821	
Grade 6 (*n*=93614)	Listening	3.038	0.905	0.745	0.917
	Speaking	2.372	0.712	0.723	
	Writing	3.028	0.921	0.890	
	Reading	2.652	1.029	0.806	
Grade 7 (*n*=91316)	Listening	3.158	0.906	0.768	0.924
	Speaking	2.299	0.727	0.750	
	Writing	3.077	0.910	0.890	
	Reading	2.782	1.046	0.820	
Grade 8 (*n*=80748)	Listening	3.239	0.903	0.787	0.923
	Speaking	2.284	0.747	0.756	
	Writing	3.154	0.895	0.890	
	Reading	2.680	0.935	0.794	

Grade	Subject	Mean	Standard Deviation	Internal Consistency	Composite Reliability
Grade 9 (n=80464)	Listening	2.693	0.951	0.800	0.927
	Speaking	2.081	0.896	0.782	
	Writing	2.920	0.968	0.860	
	Reading	2.590	0.971	0.809	
Grade 10 (n=56546)	Listening	2.803	0.915	0.783	0.921
	Speaking	2.180	0.920	0.785	
	Writing	3.057	0.894	0.860	
	Reading	2.560	0.933	0.812	
Grade 11 (n=46757)	Listening	2.837	0.895	0.780	0.916
	Speaking	2.216	0.930	0.786	
	Writing	3.145	0.849	0.860	
	Reading	2.603	0.925	0.811	
Grade 12 (n=34559)	Listening	2.820	0.871	0.759	0.906
	Speaking	2.193	0.931	0.777	
	Writing	3.229	0.804	0.860	
	Reading	2.585	0.903	0.798	

For TELPAS speaking items, field-test items are examined for human–human and human–machine agreement. Evidence of interrater reliability is gathered by examining the perfect agreement rates and the Pearson correlations.

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 EB students 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 2021–2022.

Evidence of the validity of the listening, speaking, reading, and writing 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, either separately, or as part of the Technical Digests for previous years.

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 EB students must receive to support their ability to develop academic English language proficiency. TELPAS assesses the ELPS for listening, speaking, reading, and writing.

TELPAS Online Component

Test design and alignment with standards: The online TELPAS listening, speaking, and reading tests for grades 2–12 are designed to assess English language proficiency in a manner that provides information about how well EB students understand and produce the English they need for academic success in Texas schools, as well as the types of language supports they require to independently comprehend written or spoken English.

The tests are built using four levels, or degrees, of built-in linguistic support, addressing the gradually reduced degree of linguistic accommodation that EB students need as they progress from knowing little or no English to becoming fluent in English. 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 understand and speak English.

Each passage and test question combination 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 reporting category (reading skill category) or a specified number of points per reporting category (listening or speaking skill category).

Score reports inform teachers about how successfully students demonstrate the comprehension and analytical listening, speaking, or reading skills of the ELPS at the four proficiency levels. The content validity of the TELPAS online assessments 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 skills that students need for academic success in all subject areas. The staged linguistic accommodation test design is shown in Table 6.9.

Table 6.9. Staged Linguistic Accommodation Test Design

TELPAS Levels	Degree of Linguistic Accommodation Applied to Stimulus and Item Development	
Beginning	Extensive	Maximum picture support; short stimuli 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
Intermediate	Substantial	Frequent picture support; short stimuli written primarily on familiar topics; commonly used, everyday English and routine academic English
Advanced	Moderate	Occasional picture support; contextual aids and organizational features support comprehension of longer stimuli on both familiar and unfamiliar social and content-area topics
Advanced High	Minimal	Minimal linguistic accommodation; stimuli highly comparable to those intended for native English speakers

The online TELPAS material requires students to comprehend the types of written or spoken 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 mathematics, language arts, and science, although other subjects are eligible as well. Items that assess the higher listening, speaking, and reading levels challenge students' ability to think critically and conceptually while listening, reading, or responding to complex English and academic content. The constructs measured are the ability to listen, speak, or 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 online TELPAS 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 online TELPAS 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 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 online tests, the TELPAS holistically rated components are aligned with the ELPS and are designed to assess the

English communication skills that EB students need to engage meaningfully and successfully in learning the academic knowledge and skills required by the state. The holistically rated assessments draw on 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 EB students 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.

Evidence Based on Response Processes

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

TELPAS Online Component

Theoretical and empirical evidence was used to pilot test the online assessments and determine the appropriateness of each item type used on the assessments. A variety of question-and-answer, cloze (i.e., fill-in-the-blank), drag-and-drop functionality, click on it (i.e., click on the correct answer), picture-based prompt, and open-ended style prompt response formats were used. The items were written and developed 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 can 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 administration. The new interface was designed to enhance the students' testing experience and provide improved testing conditions for students to demonstrate what they had 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. In 2021–2022, the TELPAS test was administered on the Cambium Test Delivery System (TDS). The same online administration platform was used for STAAR, providing students with the same familiar system, functionality, and

supports across all assessments.

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 EB students 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. This provides strong validity evidence related to the response process.

Evidence Based on Internal Structure

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

TELPAS Online 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 online TELPAS tests is evaluated each year using KR20 and coefficient alpha 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 raters to perform their duties and prepare district administrators to follow procedures 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 the TELPAS holistically rated domains they will assess, whether only writing or all domains, depending on grade level and accommodations. Additional support is provided to raters who cannot calibrate on their first two attempts in order to help them assess assigned students in a manner consistent with the PLDs. 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 the 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 EB students.

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 resulting from 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

- learn how developing academic language proficiency in English relates to and supports academic achievement in English;
- learn how to adjust content instruction for EB students to make it more comprehensible and how to target steady progress in English acquisition; and
- practice observing student behaviors in the instructional environment to make better instructional decisions about students.

Evidence based on the consequences of testing can be found by comparing performance from past administrations, which is represented in Appendix D. If long-term trends show that student performance increased after the introduction of testing, such improvement may have resulted, in part, from the use of test data to inform instruction. The proficiency level classifications of students for the listening, speaking, reading, and writing domains of TELPAS have been continually collected since the first administration.

Historically, the results have shown a gradual increase in EB student performance in all TELPAS domains over time. Due to the change in the blueprint and new standards for TELPAS reading and the complete redesign of TELPAS listening and speaking in 2018, the percentage of students at given proficiency levels cannot be directly compared to student performance from 2017 and earlier. If historical trends hold, however, over time the percentages of students across proficiency levels are expected to remain relatively stable, with the possibility of a gradual increase in performance.

Specifically, the TELPAS Proficiency Levels Trends tables in Appendix D present the results from the 2018–2022 TELPAS administrations. In 2018, TELPAS reading was shortened and redesigned and new proficiency level performance standards were set for

grades 2–12. This made comparisons to previous administrations inappropriate. TELPAS grades 2–12 listening and speaking tests also changed from a holistic rating to online item-based tests in 2018. Proficiency levels on the new listening and speaking tests, while assessed using the same TELPAS proficiency level descriptors, are not directly comparable to the previous holistically rated assessments. In 2018, calculation of the TELPAS composite rating was also changed to use equal weights instead of having higher weights for reading and writing.

Due to the COVID-19 pandemic, the 2020 TELPAS assessments were optional, and the testing window was extended, resulting in fewer students testing overall. In 2021, the TELPAS testing window was again extended, but assessments were not optional. The 2021 TELPAS reading and writing results appeared consistent with pre-pandemic results for the upper grades, while most of the elementary grade levels showed higher percentages of students in the lower two proficiency levels compared with prior years. TELPAS speaking also saw decreased scores for elementary students compared with 2018 and 2019 results, while TELPAS listening results improved for several grade levels compared with pre-pandemic results. Although 2021 was the fourth year of the online listening and speaking assessments, the interruption to instruction and the optional nature of the assessments in 2020 make it difficult to interpret trends. The 2022 administration was the fifth year of the online listening and speaking assessments.

The TELPAS Proficiency Levels Trends tables in Appendix D also present all the historical TELPAS results since 2005. In addition to the changes described above in 2018, TELPAS reading and the composite rating were also changed in 2014. In 2014, new proficiency level performance standards were applied for TELPAS reading in grades 2–12, and the domain weights for the composite rating were changed. Table 2 represents the TELPAS changes enacted in 2014 and 2018. Comparisons in performance are only appropriate across certain years. For example, TELPAS writing results for all grades, can be compared across all years going back to 2005, while TELPAS grades 2–12 listening and speaking results from 2005 until 2017, can only be compared against other results including reading and composite results from 2014-2017.

These results generally show incremental increases in English learner performance in all TELPAS domains from spring 2005 to spring 2013, and then the performance stabilizes around 2014 and afterward. The 2021 results show a general decrease in performance for the lower grades as compared with results from 2018 and 2019. The 2022 results show that performance stabilized.

Sampling

No sampling was conducted for TELPAS during the 2021–2022 school year.

Test Results

Appendix D provides frequency distributions and summary statistics for the TELPAS online assessments administered in 2021–2022, based on scale scores as well as mean p-values and reliability estimates by grade level. The percentage of students in each of

the TELPAS composite proficiency levels is provided in Table 6.10. The percentages are available by domain in the 2022 TELPAS Statewide Summary Reports on TEA's Student Assessment Division website.

Table 6.10. Percentages* of Students in Each of the TELPAS Composite Proficiency Levels in 2022

Grade	Composite Proficiency Levels Across All Domains			
	Beginning	Intermediate	Advanced	Advanced High
Kindergarten	48%	32%	13%	7%
Grade 1	26%	38%	22%	14%
Grade 2	13%	51%	31%	5%
Grade 3	6%	37%	42%	15%
Grade 4	7%	34%	42%	16%
Grade 5	5%	27%	44%	24%
Grade 6	4%	31%	46%	19%
Grade 7	4%	29%	46%	21%
Grade 8	4%	28%	47%	21%
Grade 9	9%	39%	38%	14%
Grade 10	6%	39%	40%	15%
Grade 11	5%	38%	41%	16%
Grade 12	3%	40%	41%	15%

*Percentages may not sum to 100 due to rounding.