

CHAPTER 4: TEXAS ENGLISH LANGUAGE PROFICIENCY ASSESSMENT SYSTEM (TELPAS)

Overview

Title III, Part A of the No Child Left Behind Act of 2001 (NCLB) requires states to conduct annual statewide English language proficiency assessments to measure the progress of K–12 limited English proficient (LEP) students in learning the English language. In accordance with NCLB, these assessments must assess four language domains: listening, speaking, reading, and writing.

To meet these federal requirements, the Texas English Language Proficiency Assessment System (TELPAS) was developed. TELPAS consists of multiple-choice Reading Proficiency Tests in English (RPTE) and holistically rated Texas Observation Protocols (TOP). RPTE has been administered to LEP students in grades 3–12 since the 1999–2000 school year. TOP was benchmarked in spring 2004 and implemented fully in spring 2005. TOP is administered in reading in K–2 and in listening speaking, and writing in grades K–12. The component names RPTE and TOP will no longer be used beginning with the 2007–2008 school year, the year in which the second edition of RPTE will be implemented. The assessment components will simply be referred to as TELPAS reading, writing, listening, and speaking.

TELPAS is used to show the extent to which districts and the state meet federal Annual Measurable Achievement Objective (AMAO) accountability indicators. Composite ratings rather than individual language domain ratings are used in the AMAO accountability measures. For information about how TELPAS composite ratings and composite scores are generated, see Appendix 7 in the *2005–2006 Technical Digest*. This digest can be found online at <http://www.tea.state.tx.us/student.assessment/resources/techdig06/index.html>.

Historical Overview of First Edition of RPTE

The impetus for the development of the first edition of RPTE came from state legislation enacted in 1995, which required the commissioner of education to propose a way to evaluate the progress of LEP students eligible by state law for exemption from the state-mandated assessments on the basis of limited English proficiency. After seeking input from national experts and state stakeholders, the commissioner proposed to develop a test that would measure the annual progress of LEP students in learning to read in English. Such a test would not only be useful for measuring student progress in the state reading curriculum but would also provide educators with information about how well English learners were able to understand the written English used in other content areas and on the state assessments.

In May 1998 TEA convened an advisory committee of Texas educators, assessment specialists, and state and national experts in second language acquisition to provide direction regarding the design of the test. In subsequent steps of test development, groups of educators helped

identify which TEKS student expectations to assess. In addition, national consultants as well as Texas bilingual and English as a second language (ESL) teachers articulated what second language learners can and cannot read and understand as they progress from one stage of second language development, or proficiency level, to the next. Afterward, the testing contractor, in collaboration with TEA staff, developed prototype test items. Texas educator committees reviewed these prototypes and provided the necessary input to allow the test developer to draft a large number of items which were field-tested in spring 1999. Additional committees of educators reviewed the field-test items before and after the field test and determined which items were appropriate for inclusion on tests. Finally, before implementing RPTE in spring 2000, TEA conducted a field study to determine the appropriate length of the tests and configuration of test items. Numerous districts assisted with this study by administering pilot test forms and providing input.

An RPTE test was developed for each of the following grade clusters: grade 3, grades 4–5, grades 6–8, and grades 9–12. Each RPTE test was originally constructed with items assessing the characteristics and abilities associated with three levels of second language reading proficiency—beginning, intermediate, and advanced. (Recent changes to RPTE grade clusters and the addition of an advanced high proficiency rating are discussed below.) Following the spring 2000 baseline administration, statistical analyses were performed to determine the range of raw scores representing each proficiency level. Subsequent reviews by educators and psychometricians concluded that these ranges appropriately identified the stages of language development assessed. Figure 6 below shows the raw score ranges from the spring 2000 baseline administration. An equating process is used to ensure that subsequent tests maintain the same levels of difficulty as the baseline tests.

Figure 6. Spring 2000 RPTE Raw Score Ranges and Proficiency Levels

Grade Levels	Score Ranges		
	Beginning	Intermediate	Advanced
3	0–37	38–47	48–56
4–5	0–37	38–47	48–56
6–8	0–39	40–48	49–57
9–12	0–39	40–49	50–60

The next step in developing the RPTE assessments was implementation of a scaling study conducted in fall 2000. This study generated data used to develop a cohesive scaled score system for documenting students’ annual progress.

Prior to the 2004–2005 school year, English language learners who reached the advanced RPTE proficiency level in a given year did not take RPTE in subsequent years. The TAKS tests were used to determine whether LEP students had acquired enough English to attain the full range of English reading skills expected at their grade level. To fulfill NCLB requirements for annually assessing the English language proficiency in reading of all LEP students, two RPTE policy changes were implemented in the 2004–2005 school year. First, an advanced high RPTE proficiency rating was provided with RPTE test results. Second, LEP students were required to

take RPTE annually, until they met the state criteria for reclassification as non-LEP. These policy changes would enable RPTE to show whether LEP students who demonstrated performance at the advanced level in a previous year were continuing to make the progress in second language acquisition associated with the advanced high level of English language proficiency.

Several options were considered in proposing cut scores for the advanced high proficiency category. It was determined that the advanced high rating should identify English language learners (ELLs) who needed minimal second language acquisition support when reading grade-appropriate texts and should strongly correlate with the Met Standard level of performance on the TAKS reading and English language arts tests. Options included setting the cut scores

- halfway between the cut score for the advanced proficiency level and the top of the scale (maximum scale score attainable on the 2005 test),
- three-fourths of the way between the cut score for the advanced proficiency level and the top of the scale, or
- at the scale score for which the predicted TAKS score is just above 2100 (that is, predicting passing TAKS).

The impact of setting different scale score cuts was investigated. Specifically, the percentage of students who would have been placed in the advanced high proficiency category had it existed in previous years and how those students performed on TAKS the following year was studied. An ELL focus group that assists the Texas Education Agency with LEP student assessment matters reviewed the data and collaborated with the test contractor and agency staff in establishing the final cut scores. Appendix 8 of the *2005–2006 Technical Digest* provides the technical information associated with establishing the advanced high RPTE cut scores. This digest can be found online at <http://www.tea.state.tx.us/student.assessment/resources/techdig06/index.html>.

Relationship Between Performance on RPTE and TAKS in Spring 2007

Every spring most ELLs take both the RPTE and English TAKS reading or English language arts tests. Table 2 shows the relationship between the performance of ELLs on both assessments. Some ELLs take TAKS in Spanish, and a small number of recent immigrant ELLs qualify to take English TAKS with linguistic accommodations, or in the case of eligible first-year immigrants, to take only RPTE. Only ELLs who took both RPTE and English versions of TAKS without linguistic accommodations are included in Table 2.

Table 2. 2007 TAKS Scale Score Performance by RPTE Proficiency Rating for Students Who Participated in Both Assessments

Grade Level	RPTE Proficiency Rating	N	Average TAKS Reading/ELA Scale Score
3	Beginning	1,046	1926.26
	Intermediate	5,363	2030.67
	Advanced	10,106	2126.20
	Advanced High	31,588	2294.60
4	Beginning	227	1942.81
	Intermediate	1,872	1951.38
	Advanced	12,082	2066.79
	Advanced High	17,771	2227.57
5	Beginning	126	1866.13
	Intermediate	842	1863.05
	Advanced	7,747	1979.32
	Advanced High	20,359	2144.04
6	Beginning	101	1878.06
	Intermediate	601	1902.32
	Advanced	9,411	2071.07
	Advanced High	11,925	2251.25
7	Beginning	105	1826.76
	Intermediate	458	1892.50
	Advanced	6,217	1984.06
	Advanced High	8,363	2120.08
8	Beginning	85	1867.80
	Intermediate	397	1877.38
	Advanced	6,392	1987.76
	Advanced High	10,702	2138.38
9	Beginning	860	1833.50
	Intermediate	2,042	1886.57
	Advanced	9,599	2012.16
	Advanced High	6,245	2139.76
10	Beginning	363	1929.51
	Intermediate	986	1952.64
	Advanced	5,371	2035.87
	Advanced High	4,650	2133.16
11	Beginning	245	1943.15
	Intermediate	737	1951.93
	Advanced	4,022	2024.26
	Advanced High	3,896	2129.65

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

Second Edition of RPTE

The second edition of RPTE, to be referred to as TELPAS reading tests, will be implemented in the 2007–2008 school year. The new tests are being developed to add a grade 2 assessment and to ensure that the tests measure the academic English language proficiency that enables ELLs to understand and process the English encountered in mathematics and science written materials as well as materials encountered in language arts instruction. In addition, the grade clustering of TELPAS reading tests will be different. RPTE provided grade cluster tests for grades 3, 4–5, 6–8, and 9–12. The grade clusters for the new tests are grades 2, 3, 4–5, 6–7, 8–9, and 10–12.

In the 2004–2005 school year, educator committees convened to review proposed assessment objectives, TEKS student expectations, and prototype items for the redesigned assessment. The prototype test items were piloted in spring 2005. In summer 2005 Texas educators convened to review the prototype items in conjunction with the student performance data gathered during the pilot. Using information from the pilot, the test developer generated large numbers of items to be field-tested.

In spring 2006 and spring 2007, large-scale field testing was conducted. Stand-alone field testing was conducted for grade 2. Field-test items for grades 3–12 were embedded in the operational RPTE test booklets. Because certain grade clusters for the new assessment are different, the test booklets were reconfigured to allow both the RPTE operational tests and embedded field-test items to be administered to the appropriate students. Field-test data review committees composed of Texas educators convened to review these items and make determinations about which items to place in TELPAS reading item banks.

A fall study was conducted in 2006 for TELPAS reading. The goals of the study were to gather validity evidence for the assessment and to evaluate whether it is appropriate to administer the test in grade clusters rather than by individual grades. These goals were achieved by comparing the performance of LEP and non-LEP students on TELPAS reading items and regular TAKS reading items. As such, the target population for this study was LEP and non-LEP students in Texas within the various grades or grade clusters. The sampling procedure was similar across the various grades. Schools across Texas were randomly selected to participate in the study based on their estimated LEP student enrollment. The type and number of students asked to participate in the study differed by grades. More detailed information about the study can be found in the “RPTE II Fall Study” report in the 2007 Texas Education Agency Technical Report Series, which can be found at <http://www.tea.state.tx.us/student.assessment/resources/techdig07/index.html>.

Online Field Testing

It is anticipated that the grades 2–12 TELPAS reading tests will be administered exclusively online by spring 2009. The spring 2006 grade 2 field test included an online pilot component designed to examine the feasibility of online administrations for LEP second graders. On a small number of volunteer campuses, grade 2 students were randomly assigned to either an online or paper version of the TELPAS reading field test. As part of the study, test

administrators, test coordinators, and technology specialists completed a voluntary survey about online test administration procedures and their perceptions about how well grade 2 students interacted with the online test mode. In addition, student performance data on the online and paper versions of the items were compared. Item-level analyses indicated that item statistics were generally similar across online and paper testing. This pilot, though small in scale and voluntary, generated findings that were positive enough for the state to move forward with plans for larger scale online field testing. For more information about this pilot, see Appendix 9 of the *2005–2006 Technical Digest*, which may be accessed at <http://www.tea.state.tx.us/student.assessment/resources/techdig06/index.html>.

In spring 2007, large-scale online field-testing of TELPAS reading items for grades 2–12 took place. Participating districts and campuses were randomly selected so that data were collected from a representative sample of students. The stratified sampling procedure that was used took into account factors such as students’ proficiency levels in the prior year, campus size, and geographical location. In an effort to evenly spread out field-test participation across campuses with LEP students, TEA selected campuses that did not participate in the 2006 Reading Proficiency Fall Study. The online field test was administered during the same test window as the live paper tests, March 5–April 13, 2007. The spring 2007 online field-test items were selected from items field-tested on paper in spring 2006. Table 3 below gives information about the scope of the online field-test administration.

Table 3. 2007 Online TELPAS Reading Field-Test Administration

Grade Level	Number of Forms	Districts	Campuses	Students
2	10	437	1,223	29,659
3	6	390	1,015	22,146
4–5	6	440	1,233	26,981
6–7	6	452	848	20,053
8–9	6	478	876	19,629
10–12	6	430	549	12,167

This online field test was administered to a total of 130,635 students. As part of the field test, a voluntary survey was made available to test coordinators, test administrators, and technology specialists so that they could provide feedback on test administration procedures, the feasibility of large-scale online testing with this assessment, and the ability of the students to be assessed meaningfully and appropriately via computer. A technical report summarizing the results of the spring 2007 survey and analyses of the relationship of item performance in the paper and online modes will be prepared and posted at <http://www.tea.state.tx.us/student.assessment/resources/techdig07/index.html>.

Ongoing Item Development

Each year, new TELPAS reading test items will be written, reviewed by Texas educators, and field-tested in the spring as part of the ongoing test development needed to support annual test construction and the release of these tests to the public.

TELPAS Holistically Rated Assessments

In spring 2004 TEA implemented holistically rated TELPAS components to assess reading in K–2 and listening, speaking, and writing in K–12. These TELPAS components were referred to as the Texas Observation Protocols (TOP) through the 2005–2006 school year.

These assessments enable teachers 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.

TOP was developed by TEA in collaboration with test development contractors, bilingual/ESL consultants, and members of an English language learner focus group composed of teachers, bilingual/ESL directors, assessment directors, campus administrators, and university professors. Like TELPAS reading, these assessments are aligned with the Texas Essential Knowledge and Skills (TEKS) for English language proficiency and are designed to assess the English communication skills that ELLs need to engage meaningfully and successfully in the learning of the academic content required by the TEKS. The holistically rated assessments draw upon second language acquisition research, research-based standards, the experience of Texas practitioners, and observational assessment processes used in other states. As of the 2007–2008 school year, the grade 2 reading component of TELPAS will be assessed with the new multiple-choice reading test rather than through the holistic rating process.

TEA chose to implement an observation protocol assessment process for the following reasons:

- This assessment method, which is integrally linked with sound classroom instructional practices, has the potential for making a significant, positive impact on student learning. The training teachers receive in the holistic rating and classroom observation process helps them better understand and meet the needs of English language learners.
- TOP does not require annual field testing.
- The administration of TOP does not require students to lose instructional time.
- TOP does not have the logistical drawbacks associated with tape-recorded or individually administered speaking and listening assessments required by other types of English language proficiency tests.

TELPAS Holistic Rating and Training Process

For the administration of the holistically rated components of TELPAS, districts identify their K–12 LEP students and designate teachers to observe and rate the students' English language proficiency. This rating process is designed to capture an overall level of English language acquisition and is a holistically scored assessment rather than an assessment of isolated skills. Teachers are trained to use holistic rating rubrics called proficiency level descriptors (PLDs) to assign one of four proficiency ratings—beginning, intermediate, advanced, or advanced high—in each domain assessed. The ratings are submitted to the state's testing contractor, and performance reports are generated and sent back to the districts. The proficiency level descriptors can be found online at <http://www.tea.state.tx.us/student.assessment/resources/perlevel/index.html>.

Since spring 2004, TELPAS raters have received annual holistic rating training through a training-of-trainers model. In response to field input, the TELPAS training model was changed during 2005–2006 to move the training of new raters to the fall of the year. The goals of fall training are to

- train teachers to use the TELPAS classroom observational processes throughout the school year to maximize student learning,
- help ensure that teachers receive a thorough holistic rating training, and
- reduce the amount of training time required in the spring.

Training materials include annotated examples of student performance and authentic student samples for use in training raters to apply the holistic rating rubrics.

Since 2005–2006, the TELPAS training model has included a qualifying component to ensure that teachers are adequately prepared to rate the writing collections of students in grades 2–12. All raters, except those who rated K–1 students only, are assigned particular grade clusters in which to qualify. Individuals successful on the qualifying activities, which are administered online, receive a training certificate qualifying them to rate student writing collections independently. Also since 2005–2006, campuses have been required to implement and document state-approved procedures they use during the TELPAS administration to ensure the validity and reliability of the rating process.

Beginning in the spring of 2007, individuals previously trained in the TELPAS holistic rating process have received spring training via online refresher courses in which they practice and calibrate their rating skills. An online training course has also been developed to help districts train raters in procedures for assembling the grades 2–12 writing collections and verifying that the collections contain the appropriate types and number of writing samples.

Spring 2007 TELPAS Audit

Since the 2005–2006 administration, annual state audits of selected student writing collections and other TELPAS assessment processes have been conducted to monitor the

effectiveness of the training and application of the holistic rating rubrics. As part of the spring 2007 audit, 551 total writing collections originally rated by district raters were re-rated by state-trained scorers. District ratings of the audit writing collections were in perfect agreement with the state ratings 76% of the time. This overall perfect agreement rate of 76% was found to be acceptable by the state. According to the International Standards Organization (ISO), a perfect agreement rate of 70% is considered acceptable for a holistically scored writing assessment that uses a 4-point rating scale. Furthermore, the audit revealed an adjacent agreement rate of 98%, which means that 98% of the time the district and state audit ratings were either the same or one proficiency level apart.

Teachers who rated the audited writing collections completed survey questions about the TOP rating process for all language domains. Results of the survey indicated that raters reported that they had enough information about student’s abilities to make proficiency level judgments. In particular, the table below shows the breakout by language domain of the percentage of responses indicating that raters felt they had sufficient information to judge the English language proficiency levels of their students.

Table 4. Percentage of Raters Reporting That They Had Sufficient Information to Rate Students

Language Domain	Percentage
Listening, grades 2-12	95
Speaking, grades 2-12	96
Writing, grades 2-12	98
Reading (grade 2 only*)	95

*TELPAS Reading assesses the reading

The rater survey also asked about adequacy of training activities. Results indicated that 95% of raters who attended a face-to-face training and 95% of raters who took an online refresher course indicated that the training was sufficient to prepare them for rating students. TEA plans to conduct TELPAS audits annually as part of an ongoing process to monitor the validity and reliability of the TELPAS holistic rating process. See the “TELPAS Audit Analysis” in the 2007 Texas Education Agency Technical Report Series, which can be found at <http://www.tea.state.tx.us/student.assessment/resources/techdig07/index.html>.

Additional Sources of Information

More information about the TELPAS assessments can be found in the following publications:

- *TELPAS Reading Preliminary Information Booklet*
- *LPAC Decision-Making Process for the Texas Assessment Program.*

Both resources are available online at <http://www.tea.state.tx.us/student.assessment/resources/techdig07/index.html>.

