

Scoring Process for Texas English Language Proficiency Assessment System (TELPAS) Speaking Responses

TEXAS EDUCATION AGENCY
STUDENT ASSESSMENT DIVISION
March 2022



Overview

The Texas English Language Proficiency Assessment System (TELPAS) Listening and Speaking test is administered online every spring. The test has multiple speaking prompts (test questions) embedded throughout to measure a student’s English language speaking proficiency. The speaking portion of the test includes a variety of picture, passage, and text-based speaking prompts. The variety of question types gives emergent bilingual (EB) students, also referred to as English learners (ELs), the opportunity to show their comprehension and communication skills in various ways.

The purpose of this document is to explain how student speaking responses for TELPAS are scored by Pearson, one of the state’s contractors for TELPAS.

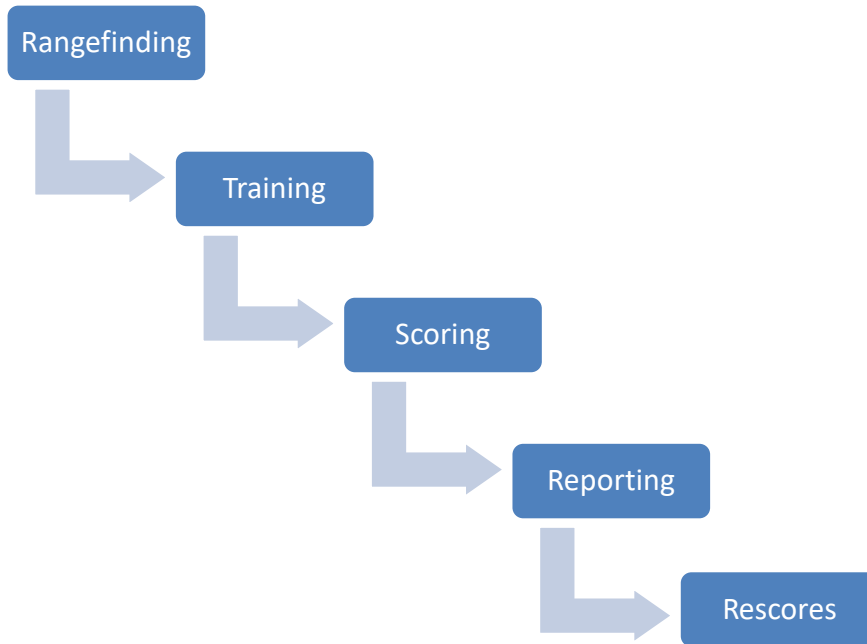


Scoring Responses

The scoring process is organized to ensure fairness and accuracy. Below is an overview of the five-step process. Each of the five steps will be further explained in this document.

1. **Rangefinding**—A selected group of student responses that were collected during field testing are scored using the rubrics. These responses are then used as model examples of different score points and in scorer training.
2. **Training**—The automated scoring engine and human scorers are trained with TEA- and Pearson-approved student responses.
3. **Scoring**— Scoring of student responses begins with the automated scoring engine. During the scoring window, a reliability and validity check is completed, and responses are also scored by human scorers. In addition, Pearson scoring directors rate responses previously scored by the automated scoring engine and human scorers.
4. **Reporting**—Student proficiency levels and scores are provided to districts through standard reports after all testing is completed.
5. **Rescores**— District testing coordinators may request a student’s speaking score be rescored for a fee. A Pearson scoring supervisor or scoring director will rescore the student’s speaking test. The results of the rescore will be shared with the requestor. If the score changes, the fee is waived.

Each step has an important role in the overall scoring process.



Pre-Scoring Activities

Before scoring begins, a few activities must occur to optimize how TEA assesses speaking.

- **Item Development**— These activities begin with content and assessment experts developing the TEA-approved prompts that will elicit the desired language from students. Committees of Texas educators (teachers, instructional specialists, district coordinators, principals, etc.) review potential speaking prompts and provide input and feedback. Once prompts are reviewed and edited accordingly, they are approved by TEA and are ready for field testing.
- **Field Testing**—TEA annually conducts a field test built-in to the base TELPAS test (the test for which students receive a score). During field testing, students taking the TELPAS Listening and Speaking test are given a pre-determined number of newly developed prompts and test questions to gather data. The answers and responses for the field-tested items are not included in the calculation of a student’s score.



Step 1: Rangefinding

Student responses from the field-tested prompts are taken to “rangefinding” meetings. The purpose of a rangefinding meeting is to apply the rubrics to the sample responses and set the scoring boundaries. During these meetings, selected student responses are scored by the group and given a consensus score (when all experts have come to an agreement on the score that the student response receives). A

selection of rangefinding responses are then used to train human scorers. The rangefinding responses are arranged into scorer training sets (anchor sets, practice sets, and qualifying sets) to ensure scorers are qualified to score the project.

Step 2: Training

Automated Scoring Engine

Advantages to automated scoring include the ability to account for various accents, standardization of scoring procedures, and more objective, efficient scoring. To begin the process of training the automated scoring engine, a selection of student responses from the field-tested prompts are transcribed and scored by Pearson. Transcriptions are used to train the scoring engine to accurately recognize students' speech and to learn how to score responses appropriately. Human ratings are used to train the scoring engine to accurately predict human scores based on the content and linguistic features of the responses. In addition, an ample amount of spoken responses produced by tested EB students/ELs with a variety of accents, who are representative of the demographics of TELPAS test takers, are also used to train the scoring engine. When the scoring engine has been given sufficient data at the outset, it is able to interpret a great variety of accents in English and score responses appropriately.

Human Scorers

Pearson and TEA work together to develop training materials and train scorers throughout live scoring. All scorers go through the same extensive training. This process is as standardized as possible, and all scorers are trained using the same materials and rubrics.



Training

Two different rubrics, a 2-point rubric and a 4-point rubric, are used to score different types of speaking items. Both rubrics are derived from the TELPAS proficiency level descriptors (PLDs) which are part of the state's English language proficiency standards (ELPS). Scorers are trained on how to use and apply the 2-point and 4-point rubrics in one to two grade bands. Bands include grades 2-3, 4-5, 6-8, and 9-12. Scorers are trained using anchor sets which contain student responses of each score point. Scorers then complete practice sets where they are asked to score student responses for a designated grade band(s). Both the anchor and practice sets include rationales as to why and how the score was assigned. Scoring directors and scoring supervisors are available for clarification and guidance during the process and support the scorers in their training as needed. After practice sets are complete, scorers must then complete

qualification sets. At the end of the training, scorers are able to demonstrate a complete understanding of the rubrics.

Qualification

Upon completion of the training, every scorer must pass qualification exercises before being able to score any student responses. There are two qualification sets. Scorers must achieve 80 percent perfect agreement (which means that they must score 80 percent of responses exactly the same as the score determined in rangefinding) on one of the two sets in order to qualify to enter live scoring. Each qualification set is comprised of 10 student responses that were selected during the rangefinding meeting. The set contains samples that represent all score points and are arranged randomly to mimic the scoring environment. The qualification set is a quality measure to ensure that scorers are following the rubrics and can accurately apply scores. Only scorers who complete the training and successfully pass the qualification set can begin scoring TELPAS student responses. Scorers who cannot pass the qualification exercise are removed from the scorer pool.



Step 3: Scoring

Pearson leads the scoring and communicates with the scoring services team throughout the assessment process.

All responses are initially scored by the automated scoring engine. Responses that are “not scorable” are identified and assigned to scorers. These are responses the system does not have confidence in assigning a score. The “not scorable” responses may have loud background noise (e.g., school bell rings, static sound in recording), mumbled or faint speech, the presence of a language other than English, the volume of the recorded response is too low and difficult to score or are responses that did not have similar enough representation in the materials used to train the scoring engine.

During the scoring window, Pearson conducts checks to ensure that the automated scoring engine is scoring reliably and begins the reliability and validity check of 10 percent of the responses. Responses are scored on a scale of 1-2 or 1-4 depending on the score point value of the prompt. Scorers have the ability to defer a response to a scoring supervisor for assistance if needed. Additionally, directors score a selection of responses scored by the engine and human scorers for reliability and validity. Data comparing human-to-human agreement and human-to-automated scoring engine agreement ensures the scoring process is reliable.

Ongoing Feedback During Scoring

Once a scorer has completed training, passed qualification, and begun scoring, he or she will continue to receive feedback and monitoring from a scoring director or supervisor each day. Scorers who pass the initial qualification set exercise are required to maintain consistent

scoring quality throughout the project. Blind validity student responses (student responses previously scored by Pearson) are delivered to scorers during the scoring window. Scorers must maintain a 70 percent average agreement for 4-point items and a 90 percent average agreement for 2-point items on validity responses. If validity falls below standards during the project, a targeted calibration set will be administered. If a scorer cannot achieve the minimum standard (70 percent or 90 percent, respectively), the scorer will be released from the project. Responses scored by the removed scorer are put back into the scoring pool to be scored by a scorer maintaining the required performance standard. Scoring directors and supervisors will also backread (listen behind and score) as another quality measure. Backreading occurs continually throughout the TELPAS scoring window. These measures, as well as others, help to ensure consistent scoring and high scorer agreement (agreement between the scoring engine and scorers).



Step 4: Reporting

Results from the TELPAS assessment are sent to districts after testing. Results for a student include a proficiency level rating, the total number of points earned for all spoken responses (raw score), and a scale score.

Standard reports are provided automatically to districts. Information contained in standard reports satisfies mandatory reporting requirements. For more information about reporting of the TELPAS results, refer to the TEA's [Interpreting Assessment Reports](#) webpage.

Step 5: Rescores

Scores are provided to the district based on the dates noted in the Calendar of Events. If the district has questions about a student's speaking score, the district can request that the student's responses be rescored. The rescore request must be submitted in the Test Information Distribution Engine (TIDE) by a district coordinator within a specific window for score code changes, rescores, and language changes. Specific window dates can be found on the applicable school year *Calendar of Events* in the [District and Campus Coordinator Resources](#).

For assistance in submitting score code changes, rescores, and language changes in TIDE, contact the Texas Testing Support, 1-833-601-8821 or TexasTestingSupport@cambiumassessment.com.