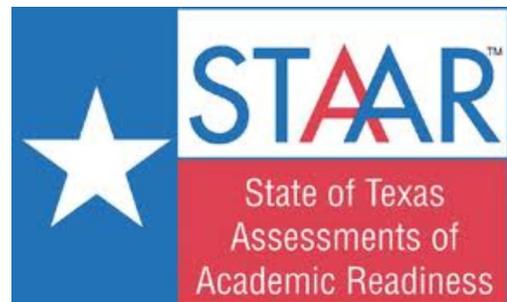


# Scoring the State of Texas Assessments of Academic Readiness (STAAR) Writing Essays

TEXAS EDUCATION AGENCY  
STUDENT ASSESSMENT DIVISION  
JANUARY 2019



The purpose of this document is to explain how student essays for the State of Texas Assessments of Academic Readiness (STAAR) are scored by the Texas Education Agency (TEA) and Educational Testing Service (ETS).

### Scoring Essays

The STAAR test is administered on paper or online. All student essays, whether submitted on paper or online, are scored by a group of trained individuals called raters. Raters score the essays in ETS's online scoring system. Online scoring allows for a large representative sample of raters from across the United States with varied professional backgrounds to accurately read and score student essays. All raters go through the same extensive training. Even though the process is as standardized as possible and all raters are trained using the same materials and sample essays, raters may differ in a score applied, especially when using a [holistic rubric](#). (A rubric is a guideline to provide assistance when assigning scores.)

The scoring process is organized to ensure fairness and accuracy. This process has 5 steps:

1. **Item and Form Development-** Content and assessment experts construct test items and writing prompts to be field tested.
2. **Field Testing and Rangefinding-** Writing prompts are field tested to determine if they are appropriate for use on an operational assessment. Once a prompt is selected, student responses for that prompt collected during the field test are then used to determine the scoring boundaries and definitions guided by the rubric.
3. **Rater Recruitment and Hiring-** ETS recruits and hires raters of various professional backgrounds that meet certain standards and hold certain credentials making them highly qualified to score assessment compositions.
4. **Rater Training, Scoring, and Monitoring-** ETS provides training that has been created with the assistance of TEA content and assessment experts. Raters that pass required calibration exercises score student compositions and must continue to pass subsequent calibrations as they score. If a rater does not maintain calibration standards, they are released from scoring and their scores are not used. Compositions are placed back into the scoring pool to receive a blind score (raters cannot see one another's scores) from a qualified rater.
5. **Rescores and Appeals-** District Testing Coordinators (DTC) may appeal an essay's score. Scoring Leaders (SLs) and/or Chief Scoring Leaders (CSLs) will blindly rescore the essay. Results are then shared with districts.

All 5 steps are interrelated, and each step has an important role in the overall scoring process.



### Step 1: Item and Form Development

The process begins with content and assessment experts developing potential multiple-choice questions and prompts. All items (questions or prompts) are reviewed and approved by TEA before being compiled onto a test form. Texas educators review multiple choice passages for reading, revising, editing, and writing prompts. As test forms are built, experts review the prompt with the other items, they work to ensure that there is cohesiveness between the items, and that the content does not overlap (i.e., questions are not repeated, questions do not provide clues to answers for other questions). New base forms are constructed to be similar to previously administered forms in content and difficulty and are aligned to the current Texas Essential Knowledge and Skills (TEKS). In addition, statistical procedures are used to ensure students' scale scores obtained from different test forms are comparable in meaning.

### Step 2: Field Testing and Rangefinding

TEA replenishes the writing prompts needed to construct the STAAR tests by periodically conducting a writing prompt field test. During field-testing, specific campuses representative of the state's demographics are selected to participate. Experienced raters are trained and score the field-test essays as they would for any other STAAR administration. TEA and ETS content experts analyze student performance and a variety of data from these items to determine if the items are eligible to be used on a future STAAR test or for various other assessment instruments.

Following the scoring and analysis of the field-test essays, ETS compiles summary performance of each item, focusing on factors such as:

- Variety of content seen in the student essays.
- Different approaches students took when writing their essay.
- Clarity of the wording of the prompt.
- Overall impression of the suitability of the prompt for use on a future STAAR test or other assessment instruments.

These summaries are presented to TEA to determine the most suitable prompts for an assessment. Then the chosen prompts and student responses are brought to “rangefinding” meetings. The purpose of a rangefinding meeting is to bring together a small group of experienced STAAR raters and content and assessment specialists from TEA and ETS and set the scoring boundaries and definitions used in the rubric. During these meetings, selected student essays from the field test are again independently and blindly scored by the group and given a consensus score (a consensus score is when all experts have come to an agreement on the score that the student essay receives). These scores typically reconfirm the scores that were given by two separate raters during field test scoring. Essays selected from the rangefinding meetings are included in the scoring guides and training materials for each test.

### Step 3: Rater Recruitment and Hiring

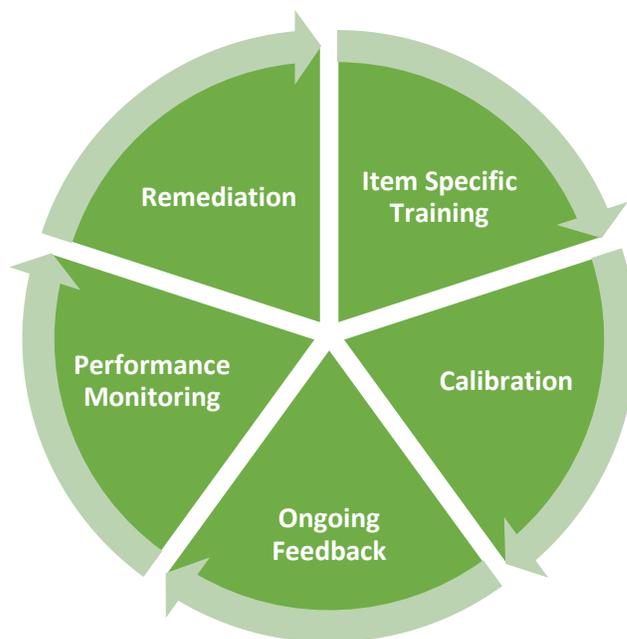
ETS leads the rater recruiting process for the scoring of STAAR essays. A rater must meet the following qualifications as required by TEA:

- Is eligible to work in the United States.
- Holds a bachelor's degree or higher.
- Has taught or is currently teaching writing (preferred).
- For Spanish scoring – raters must be proficient in reading, speaking, and writing Spanish.

ETS monitors the rater pool size throughout the year and maintains the number of raters required to successfully score each administration. Raters are required to take training on the online scoring system, essay training on both handwritten and typed sample student essays, and pass the STAAR certification test before scoring student essays.

### Step 4A: Rater Training and Monitoring

TEA and ETS work together to train and monitor raters throughout live scoring.



### **Item Specific Training**

After a rater has completed the STAAR certification test, he or she must complete 8 hours of self-paced interactive training related to the prompt that each rater will be scoring. Scoring Leaders complete 14 hours and Chief Scoring Leaders complete 16 hours of training. This training includes item specific hands-on practice sets of essays and interactive webinars plus live chat functionality with expert trainers on the scoring guide and benchmark papers that are representative of each rubric score point.

The training sets of essays have predetermined scores (as determined during rangefinding) and provide raters the opportunity for further explanation and rationale as to why and how the score was assigned. At the end of the training, raters are able to demonstrate a complete understanding of the rubric.

### **Calibration**

Upon completion of the training, every rater must pass a calibration exercise before he or she is able to score any student essays. The calibration set is comprised of 10 student essays 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 calibration set is a quality measure to ensure that raters are following the rubric and can accurately apply scores. Only raters who complete the training and successfully pass the calibration set can begin scoring STAAR student essays. Raters who cannot pass the calibration set are removed from the rater pool.

Raters who pass the initial calibration set exercise are required to recalibrate (pass a calibration set again) every third day that he or she is scheduled to score. Raters who do not pass subsequent calibration sets have their scoring shift canceled for that day. Raters who do not pass for a second time are removed from the scoring pool and their scores are not used, but instead diverted back into the rater pool to be blindly scored by raters who maintain calibration standards.

### **Ongoing Feedback**

Once a rater has completed training, passed calibration, and begun scoring, he or she will continue to receive feedback and monitoring from a Scoring Leader each day he or she is scheduled to score. SLs will backread (read behind) raters as another quality measure. Backreading occurs continually throughout STAAR scoring. These measures, as well as other reporting functions in the STAAR scoring system, help to ensure consistent scoring and high rater agreement (agreement between separate raters on a score for an essay).

### **Performance Monitoring and Remediation**

Rater performance is monitored daily. If necessary, some raters will be required to review the item specific training sets and pass a remediation calibration set. If the rater is unable to pass the remediation calibration set, his or her essay scores are invalidated and he or she is removed from the rating pool. Essays scored by the removed rater are put back into the scoring pool to be scored by a rater maintaining the required rater performance.

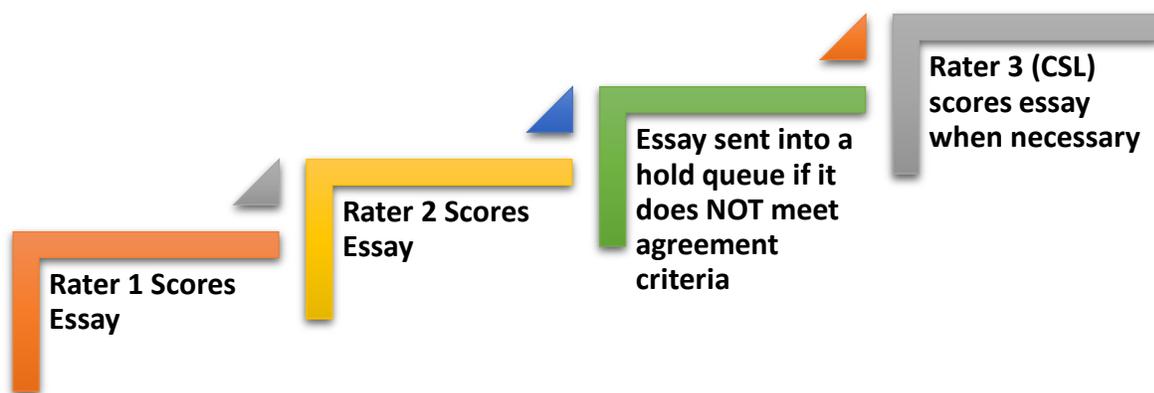
#### Step 4B: Scoring

After the student essays are loaded into the STAAR scoring system, raters log in to the system and score the essays on a 1–4 scale or identify the response as not scorable. When a student’s essay is determined to be nonscorable (see possible reasons for a nonscorable rating at the end of this document), the essay is assigned a rating of zero. Raters also have the ability to place a student essay on a temporary hold to await further instruction from an SL or defer an essay to the SL for assistance.

All student essays are scored using an “adjacent agreement scoring model.” This means that two different raters assign a score to a student essay. The raters are unable to see each other’s scores. One of the following three scoring scenarios can occur:

1. Exact agreement – this occurs when rater 1 and rater 2 have assigned the same scores.
2. Adjacent agreement – this occurs when the scores that rater 1 and rater 2 assigned differ by *no more than 1* score point.
3. Discrepant agreement – this occurs when the scores that rater 1 and rater 2 assigned differ by *more than 1* score point. When scores are discrepant, the student essay is automatically moved into a hold queue where it will be reviewed by a Chief Scoring Leader. The CSL will review the student essay and provide the final score.

All essays receive a blind score and follow a quality review process to ensure accuracy of scoring.



To see how the composition scores apply to the overall raw score for the test, please review the [STAAR Writing Weighting Charts](#).

After rater 1 has scored a student essay, the essay is routed into rater 2’s queue for a second score. Once an essay has received two ratings and the ratings meet exact or adjacent agreement, the ratings are added together, and the student receives a total essay score of 2–8. If the ratings do not meet score agreement, the essay is routed into a hold queue. Only CSLs score essays in the hold queue. Occasionally, there is a fourth reading of a student essay if the original two raters and the third (CSL) rater all differ in their scores by more than the agreement criteria allow. For example, if an essay is given a score of 1 by rater 1, a score of 3 by rater 2, and a score of 4 by the CSL, a fourth rating is required. When this occurs, the essay is placed in a separate queue and scored only by ETS content experts.

## **Step 5: Rescores and Appeals**

Scores are provided to the district in the STAAR Report Card. If the district has questions about the score assigned to an essay, the district can request that the essay be rescored. The request for a rescore must be submitted through the STAAR Assessment Management System. Scoring Leaders and/or Chief Scoring Leaders (high qualified raters) will conduct a blind review of all essays submitted for rescore, meaning the new rater cannot see the initial score provided. If the leaders' new score agrees with the initial score, then the initial score is verified and there is no score change. If the leaders' score does not agree with the initial score, then the essay is rescored by two additional raters. After rescoring, the new set of ratings are compared to the old set of ratings and the higher of the two sets of ratings will become the student's score. If there is a score change, ETS will post an updated STAAR Report Card in the STAAR Assessment Management System and updated scores in the Student Portal. If the score does not change, no update will be made to the student record. The fee to rescore a STAAR essay is \$25 per essay. The fee to rescore writing multiple choice items is \$15. Districts may choose to rescore the essay, multiple choice, or both (possible fee of \$40). The district will be charged the fee if there is no change to the student score. If the student's score changes, the fee is waived.

**STAAR COMPOSITIONS**  
**NONSCORABLE RESPONSES—GRADES 4 AND 7 AND ENGLISH I, II, AND III**

The rubrics used to evaluate STAAR compositions are based on four score points (1–4), with 1 being the lowest score and 4 being the highest. Every attempt is made to provide a score to a student essay; however, if a composition cannot be scored, it is assigned a score of 0 (nonscorable). Responses are considered nonscorable if they fall into one of the following categories.

**BLANK**

The student writes nothing on the answer document in response to the prompt.

**OFF TOPIC**

The student writes on a topic other than that prescribed by the prompt. In some of these responses, the writer pays brief attention to the specified topic before shifting to another topic. In other responses the writer writes on an entirely different topic, making no mention of the specified topic.

**INDECIPHERABLE**

The student’s response is indecipherable because the student strings letters, words, and/or phrases together in a meaningless fashion.

**ILLEGIBLE**

The student attempts to respond but has handwriting and/or spelling issues that are so severe that it is impossible to interpret the writing.

**WRITTEN IN A LANGUAGE OTHER THAN TESTED LANGUAGE**

The student writes primarily in a language other than English when taking the test in English, or the student writes primarily in a language other than Spanish when taking the test in Spanish.

**INSUFFICIENT RESPONSE**

The student’s response is so brief that the reader cannot make an accurate judgment about the quality of the writing.

**LACKS ANY ORIGINAL WRITING**

The student merely copies or paraphrases the prompt, content from another portion of the test, or content from a previously published work, such as the lyrics to a song.

**DOES NOT WRITE IN PROSE**

The student’s response to the prompt is in the form of a poem, play, or song instead of in standard written prose.

**REFUSES TO WRITE**

The only writing on the answer document is the student’s explicit refusal to respond to the prompt.