

Editor's note: Dave Clark, a Pearson administrator, told PolitiFact Texas on Sept. 30, 2014, that this Pearson document was last updated Feb. 19, 2013.

Pearson's Scoring Quality in Texas

This information relates to questions raised about Pearson's recruitment of test scorers via Craigslist and Pearson's delivery of fair, reliable, and transparent scoring of Texas student essays.

- Pearson conducts an exhaustive search for *the very best people to score student tests*. Pearson works with the same employment resources used by school systems across the state such as the Austin, Houston and San Antonio independent school districts to promote career opportunities.

- Pearson advertises broadly for qualified test scorers. The extensive search for test scorers in Texas includes advertisements with 21 different organizations and publications including the National Council of Teachers of English, the *Austin American-Statesman* newspaper and the Texas Workforce Commission. Here is a list of publications and organizations with which Pearson advertised for test scorers:
 - The Chronicle of Higher Education
 - The English Journal
 - The Council Chronicle
 - Education Week
 - American Association of Colleges for Teacher Education
 - National Council of Teachers of English
 - CareerBuilder.com
 - Austin American Statesman
 - Austin Job Finder
 - El Mundo
 - The University of Texas at Austin
 - St. Edwards University
 - Huston-Tillotson University
 - Texas State University
 - Texas Workforce Commission
 - TexasOnlineJobs.com
 - AllStarJobs.com
 - Facebook
 - Craigslist
 - ClassifiedAds.com
 - Austin Chronicle

- Texas school systems such as the Austin, Houston and San Antonio independent school districts, The University of Texas at Austin, Texas Tech University, and many other organizations also advertise career opportunities on Craigslist.

- All test scorers hired by Pearson must have at least a four-year degree and undergo very rigorous, state-approved training before they are allowed to begin work. As part of this rigorous training, applicants must complete and pass practice sets before being eligible to work. The rigorous training program for scorers was developed with the Texas Education Agency, and TEA approves all final training materials.
- Of the scorers in Texas, 2,911 (71%) have at least one year of teaching experience, 3,385 (83%) have at least a writing-related four-year college degree, and 1,114 (27%) have at least a four-year college degree in English.
- Scorers are closely monitored and supported by highly-qualified supervisors on a daily basis. The test scorers themselves must pass a daily test of their ability and consistency in evaluating prompts. Scorers are regularly given student responses which were previously scored for review to ensure accurate scoring throughout the day.
- Each student test is carefully reviewed by two scorers to ensure scores are accurate and reliable.
- Pearson categorically and strongly rejects assertions made in a book by Todd Farley, a former Pearson employee who has not worked at Pearson since before the turn of the century and has no experience with or internal knowledge of the Texas test scoring process.

STAAR Scorer Acceptance Rate

The Committee asked how many applicants for scorer positions end up scoring student essays from STAAR Writing Assessments.

Here are the numbers for 2012:

- From 37,468 applicants, 12,661 (34 percent) were offered a STAAR scorer position *pending qualification*.
- From 12,661 applicants offered a STAAR scorer position *pending qualification*, 4,169 (33 percent) scored student essays from STAAR Writing Assessments.
- From 37,468 applicants, **4,169 (11 percent) scored STAAR student essays.**

Applicants (2012)

Total applicants	37,468
Applicants offered to qualify to score STAAR	12,661
Scored live STAAR essays	4,169
Percent of applicants offered to qualify who eventually scored STAAR	33%
Percent of total applicants who scored STAAR	11%

STAAR Scorer Demographics

The Committee asked for demographic data on scorers for STAAR writing assessments.

A total of 4,169 full or part time scorers were assigned to the STAAR program in 2012. The overall demographic profile of this group is as follows:

Race/Ethnicity

African American	13%
Asian	3%
Caucasian	73 %
Hispanic	8%
Multiple	2%
Native American	1%
Pacific Islander	0%

Gender

Female	68%
Male	32%

Scoring Rate for STAAR Writing Assessments

The Committee asked for the rate at which a scorer scores student essays from STAAR EOC writing assessments.

In addition to multiple-choice questions, each STAAR EOC writing assessment requires a student to compose two essays.

Each essay is limited to the 26 lines provided on a single page in the answer document (for paper administrations) or 1750 characters (online administrations).

Scorers score an average of 19 assessments (or 38 essays) an hour.

All student essays are scored twice – by two separate scorers – with a third scorer brought in for all non-adjacent scores.

In Appendix A we have included two artifacts for your review:

1. The *STAAR English II Writing Released Test Questions* (available on TEA's website) shows both examples of the non-essay, multiple-choice items that contribute, along with essay scores, to a student's overall STAAR English II Writing assessment score and two example essay prompts.
2. The *Sample Essay Box* shows the 26-line page onto which students must compose their essays.

STAAR Item Bias Review

This information relates to concerns about potential bias in STAAR questions (i.e. items).

STAAR Item Review Committee Demographics

Item review committees composed of Texas educators review items to judge the appropriateness of item content and difficulty and to eliminate potential bias. Committee members discuss each test item and recommend whether the item should be field-tested as written, revised, recoded to a different eligible TEKS student expectation, or rejected.

All committee members conduct their reviews considering the effect on various student populations and work toward eliminating bias against any group. Specifically, educators are asked to answer three questions:

- Does the item or passage assume racial, class, or gender values or suggest such stereotypes?
- Might the item or passage offend any population?
- Are minority interests well represented in the subject matter and artwork?

A total of 2,426 Texas educators participated in STAAR Item Review Committees between January 1, 2011 and December 31, 2012.

Race/Ethnicity

African American	8%
Asian	1%
Caucasian	58%
Hispanic	33%
Multiple	2%
Native American	0%
Pacific Islander	0%

Gender

Female	79%
Male	21%

If item review committees found an item to be **inappropriate** after review and revision, it was removed from consideration for field-testing. If the committee found an item to be **appropriate** after review and revision, it was field-tested. Pearson conducts a set of statistical analyses to empirically evaluate the performance of different student groups on a test item once it has been field-tested.

NOTE: student performance on a field test item never counts toward a student assessment score. It is the item that is being tested, not the student.

After field-testing, TEA and Pearson curriculum and assessment specialists and psychometricians examine each test item with regard to objective/student expectation match, appropriateness, level of difficulty, and bias (economic, regional, cultural, gender, and ethnic). Differences in student performance across groups may indicate potential biases. It is only after this process that there is a recommendation to accept or reject each field-test item.

Items that pass all stages of development—item review, field testing, and data review—are placed in the item bank and become eligible for use on future test forms. Rejected items are identified and eliminated from use on any test.

APPENDIX A