

State of Texas Assessments of Academic Readiness (STAAR®) Standard Setting Questions and Answers

STAAR performance standards: What are performance standards and when will they go into effect?

1. What are academic performance standards?

Academic performance standards represent the degree to which students are learning the content and skills required to be taught, as demonstrated by performance on a test. On the STAAR assessments, there are three levels that describe student performance:

- Level III: Advanced Academic Performance
- Level II: Satisfactory Academic Performance
- Level I: Unsatisfactory Academic Performance

The academic performance standards are the cut scores on a test that divide students into these three levels. A student is considered to have passed a given STAAR assessment if he/she earned a score at least as high as the cut score for Level II: Satisfactory Academic Performance. The policy definitions for each performance level are described in more detail at http://tea.texas.gov/student.assessment/STAAR_Performance_Labels_and_Policy_Definitions.pdf.

2. When do the STAAR performance standards go into effect?

Performance standards for STAAR end-of-course (EOC) assessments took effect in spring 2012. Following the redesign of the English assessments, performance standards for STAAR English I and English II were set in January 2014 and took effect in spring 2014. The performance standards for STAAR 3–8 were first applied to the spring 2012 administrations.

3. Why are STAAR performance standards being phased in?

A phase-in period has been implemented for STAAR performance standards to provide school districts with time to adjust instruction, provide new professional development, and close knowledge gaps. A three-step phase-in for Level II is in place for all general STAAR assessments.¹

To be eligible to graduate from a Texas public high school, a student must achieve the Level II performance standard on the five STAAR EOC assessments. The STAAR EOC phase-in performance standard for Level II is based on the performance standard in place when the student takes his or her first EOC assessment. The same standard applies to all five EOC assessments. For example, for students who took STAAR Algebra I in spring 2013, the first phase-in standard for Level II: Satisfactory Performance would apply to STAAR Algebra I, English I, English II, biology, and U.S. history.

The STAAR 3–8 phase-in standards for Level II performance began with the 2012 test administration. Phase-in 1 performance standards for Level II are in effect for the 2011–2012, 2012–2013, 2013–2014, and 2014–2015 school years. Phase-in 2 standards will be in effect from 2015–2016 until 2017–2018. Phase-in 3 will last from 2018–2019 until 2020–2021, after which the final recommended Level II performance standards will be in effect.

¹ A two-step phase-in plan was initially set. However, after reviewing trends in STAAR performance data between 2012 and 2014, the commissioner of education chose to implement a three-step phase-in plan.

4. How were the phase-in cut scores determined?

Phase-in cut scores were determined empirically for each STAAR assessment based on the recommended Level II cut scores. For STAAR EOC, phase-in 1 cut scores for Level II were set at 1.0 standard deviation (SD) below the Level II recommended cut scores for the STAAR mathematics, science, and social studies assessments and at 0.5 SD below the Level II recommended cut scores for the STAAR English assessments. Phase-in 2 cut scores for Level II were set at 0.7 SD below the Level II recommended cut scores for the STAAR mathematics, science, and social studies assessments and at 0.35 SD below the Level II recommended cut scores for the STAAR English assessments. Phase-in 3 cut scores for Level II were set at 0.3 SD below the Level II recommended cut scores for the STAAR mathematics, science, and social studies assessments and at 0.15 SD below the Level II recommended cut scores for the STAAR English assessments.

For all STAAR 3–8 assessments, phase-in 1 cut scores for Level II were set at 1.0 SD below the Level II recommended cut scores, phase-in 2 cut scores were set at 0.7 SD below the Level II recommended cut scores, and phase-in 3 cut scores were set at 0.3 SD below the Level II recommended cut scores.

5. Why are there different phase-in standards for the STAAR English EOC assessments than for the mathematics, science, and social studies assessments?

Compared to the phase-in required for the mathematics, science, and social studies assessments, a smaller phase-in was appropriate for the STAAR English EOC assessments. This decision was made based on two sets of data:

- empirical study data that evaluated the relationship between performance on STAAR EOC English assessments and performance on other external assessments (e.g., ACT and SAT) and
- student performance data that, unlike the other EOC assessments, included student responses to multiple-choice questions, writing prompts, and short answer questions.

Both data sources indicated that the smaller phase-in for English would better align the performance standards across content areas.

STAAR performance standards: What performance standards were set and how were they set?

6. What are the STAAR EOC performance standards?

Level II: Satisfactory Academic Performance

The phase-in 1 standard is

- 3500 for mathematics, science, and social studies EOC assessments
- 3750 for English EOC assessments

The phase-in 2 standard is

- 3650 for mathematics, science, and social studies EOC assessments
- 3825 for English EOC assessments

The phase-in 3 standard is

- 3850 for mathematics, science, and social studies EOC assessments
- 3925 for English EOC assessments

The recommended standard is

- 4000 for English, mathematics, science, and social studies EOC assessments

Level III: Advanced Academic Performance

The recommended standard is

- 4333 for Algebra I
- 4576 for biology
- 4440 for U.S. history
- 4691 for English I
- 4831 for English II

Prior to 2014, there were separate English I and II reading and writing assessments (i.e., English I reading, English I writing, English II reading, and English II writing) and unique performance standards for each of those assessments. However, to meet the requirements of House Bill 5, which was adopted by the 83rd Texas Legislature in June 2013, the separate reading and writing assessments were redesigned to combine reading and writing into a single measure with a single test score. To ensure continuity of the performance standards, the standard-setting committee used the separate reading and writing standards as points of reference during the January 2014 standard setting for the redesigned English I and English II assessments.

7. What are the STAAR 3–8 performance standards?

For STAAR 3–8 mathematics and reading, the phase-in 1, phase-in 2, phase-in 3, and recommended performance standards for Level II: Satisfactory Academic Performance will vary for each assessment because these assessments are reported on a vertical scale, as required by legislation. In addition, the recommended performance standards for Level III: Advanced Academic Performance will vary for all STAAR 3–8 assessments.

The Level II STAAR performance standards for grades 4 and 7 writing, grades 5 and 8 science, and grade 8 social studies are as follows:

- Phase-in 1: 3500
- Phase-in 2: 3650
- Phase-in 3: 3850
- Recommended: 4000

8. How were the performance standards for the STAAR EOC assessments determined?

Performance standards for STAAR EOC assessments are based on recommendations from standard-setting committees. These committees—convened in February and March 2012—were composed of both K–12 educators and higher education faculty, and each panelist was an expert in both the assessed content (e.g., biology) and the high school curriculum (i.e., the Texas Essential Knowledge and Skills [TEKS]). In addition, a policy committee was convened in early February 2012 to recommend reasonable score ranges within which performance standards should be set. This committee—composed of policy experts, legislative staff, business and workplace leaders, and secondary- and higher-education representatives—used the results of various studies to inform its recommendations. These studies established links between performance on STAAR and performance on other assessments and provided research-based anchors for setting meaningful and rigorous performance standards. A description of the standard-setting process used for STAAR EOC assessments can be found at <http://tea.texas.gov/student.assessment/staar/performance-standards/>.

Note that, in the time since performance standards were set in 2012, the STAAR English I and II assessments were redesigned. This change necessitated setting new performance standards for English in January 2014.

9. How were the performance standards for STAAR 3–8 assessments determined?

Performance standards for STAAR 3–8 assessments are based on recommendations from standard-setting committees. These committees—convened in October 2012—were composed of K–12 educators, and each panelist was an expert in both the assessed content (e.g., grade 5 mathematics) and the assessed curriculum (i.e., the Texas Essential Knowledge and Skills [TEKS]). Similar to the STAAR EOC standard-setting process, panelists were provided reasonable ranges within which performance standards should be set. The ranges were determined by considering the alignment of performance standards with EOC assessments and by using the results of various studies. The studies established links between performance on STAAR and performance on other assessments and provided research-based anchors for setting meaningful and rigorous performance standards. A brief description of the standard-setting process used for STAAR 3–8 assessments can be found at http://tea.texas.gov/student.assessment/STAAR_3-8_Standard_Setting_9_Steps.pdf.

10. Why were STAAR EOC performance standards set before high-stakes data were available?

A decision was made to set the performance standards prior to the first high-stakes administration (spring 2012) of the STAAR EOC assessments. The impact data used in the standard-setting process were based on student performance during the spring 2011 administration, before EOC assessments were part of graduation requirements for students. Several factors contributed to the decision to set standards without high-stakes, motivated data.

- Without a defined course sequence, 9th grade students could have taken any of the EOC assessments in the 2011–2012 school year.
- Even after the first high-stakes administration, motivated data would have been available only for the EOC assessments typically taken by 9th grade students (English I, Algebra I, and biology).
- Performance standards needed to be established before the end of the 2011–2012 school year so that summer remediation could be scheduled for those students who were unsuccessful on the EOC assessments in the spring.
- Statute requires that EOC assessment performance standards be linked for the English and algebra assessments, so standards for these assessments had to be set at the same time.

Note that data from the spring 2013 high-stakes administrations of the separate English reading and writing EOC assessments were used to set performance standards for the redesigned STAAR English I and English II assessments in January 2014.

11. Why were the performance standards for STAAR EOC assessments set before the EOC administrations, but the STAAR 3–8 performance standards were set after the administrations?

Under an aligned set of standards, student performance at each level (i.e., Unsatisfactory, Satisfactory, or Advanced Academic Performance) within a content area should indicate whether each student is on track to be successful in the next grade or course.

In order to align the performance standards in this way, TEA started with STAAR EOC assessments at the high school level and worked backwards to grade 3. This means that performance standards for STAAR grades 3–8 could not be set until performance standards for STAAR EOC had been

determined. Based on the requirements in law that TEA determine STAAR EOC cut scores by looking at a variety of external data, the earliest the performance standards could be established was spring 2012. Given this fact, STAAR performance standards for grades 3–8 could not be set in time to report spring 2012 test scores in the regular time frame. STAAR 3–8 performance standards were first applied to spring 2012 test scores and reported in January 2013.

12. What does postsecondary readiness mean?

According to TEC §39.024(a), postsecondary readiness is the level of preparation a student must attain in English language arts and mathematics courses to enroll and succeed, without remediation, in an entry-level general education course for credit in that same content area for a baccalaureate degree or associate degree program or for certificates or credentials other than baccalaureate or advanced degrees. It should be noted, however, that the measurement of postsecondary readiness through the Algebra II and English III assessments will be only one piece of information that students, parents, and schools will have in making readiness determinations. Algebra II and English III are courses students typically take in grade 11; after students have taken these assessments and potentially met the Level II or Level III performance standards, they will need to continue to take higher-level courses in grade 12 to acquire content knowledge and fully prepare for postsecondary endeavors. Note that the STAAR Algebra II and English III assessments will not be offered during the 2013–2014 or 2014–2015 school years. As required by the Texas Education Code, those assessments will be offered again beginning in spring 2016 on an optional basis.

13. What research studies were used as part of the standard-setting process?

TEA conducted extensive research to support the standard-setting process. Studies focused on creating links between STAAR assessments and other measures of students' knowledge and skills. Some studies focused on comparisons between STAAR assessments and corresponding TAKS tests. Research was conducted to link STAAR grade 7 writing scores and grade 8 reading, mathematics, science, and social studies scores to first-year STAAR EOC assessment scores in the corresponding content areas. Additional studies linked STAAR assessments to established national and international assessments, such as SAT, ReadStep, ACT, EXPLORE, NAEP, and PISA. Finally, research was conducted to link STAAR EOC scores to corresponding grades in entry-level, credit-bearing college courses. To support reliable and meaningful score interpretations, links between two assessments were based on the same students taking STAAR and one of the assessments listed above, when data were available.

14. How were vertical scales for reading and mathematics used as part of the standard-setting process?

The vertical scales for reading and mathematics empirically link student performance on STAAR 3–8 assessments within the same subject area. Because a student's vertical scale score can be compared from grade to grade to gauge his or her academic progress in mathematics or reading across time, the vertical scale was used to inform the alignment of performance standards across assessments. The reasonable ranges for performance standards were informed using the alignment of the vertical scale across grades. In addition, the vertical scale allowed standard-setting panels to consider the progression of performance standards across grades for their specific grade in relation to previously recommended performance standards for higher grades. For example, the grade 5 reading committee considered the recommended performance standards for grades 6, 7, and 8 reading as one piece of information in recommending the grade 5 performance standards.

15. How were the performance standards for the STAAR Spanish 3–5 assessments determined?

Performance standards for the STAAR Spanish 3–5 assessments were set at the same time as the English assessments, following the same process. Because the Spanish assessments in mathematics and science contain items transadapted from the English assessments, only one set of performance standards was recommended for both English and Spanish in each grade and subject assessed. The Spanish reading and writing assessments have passages and questions developed uniquely in the Spanish language. Therefore, separate standard-setting committees convened to recommend performance standards specific to the Spanish reading and writing assessments.

STAAR performance standards: What do they mean?

16. Why are the STAAR performance standards presented as scale scores rather than raw scores?

As with many standardized assessments, the STAAR program uses scale scores to communicate information about performance levels. A scale score is a more exact way to determine subject mastery than a raw score because a scale score considers the difficulty level of each individual test question in addition to whether or not a student answers the question correctly.

The basic score on any test is the raw score, which is the number of questions answered correctly regardless of difficulty level. A scale score is a conversion of the raw score onto a scale that takes into account the difficulty level of the specific set of questions used on a test in any given year. A scale-score system allows every test to have exactly the same passing standard, or level of performance required, even though the raw score needed to pass the test may vary slightly from year to year.

When building new tests each year, it is not always possible to select questions that have exactly the same difficulty as questions on previous versions of the test. Maintaining the passing standard (but not necessarily the raw score needed to pass) from year to year is important to ensure that students passing in one year will have exactly the same rigorous testing requirements as students passing in a subsequent year, even though the test questions differ from one year to the next.

It is not appropriate to compare raw scores or percent of questions answered correctly across test administrations, school years, or tests within the same content area (e.g., English I and English II). When looking at passing standards, one cannot make the assumption that a lower raw score (or percent of questions correct) passing standard on one test necessarily means that the test is easier than another test with a higher raw score passing standard. For example, on one administration of STAAR English II, the passing standard might be at 60% of the total possible points, while for that same administration the passing standard on STAAR English I might be at 63% of the total possible points. This difference in percent of questions answered correctly does not signify that it is easier to pass STAAR English II than it is to pass STAAR English I.

17. Why isn't the passing standard for each STAAR assessment set at 70% of the questions correct?

While many people believe a raw score that is equal to 70% of the questions correct should qualify as passing, a score that is simply the percentage of questions correct does not take into account the difficulty of the questions on a test. A student that gets 50% of the questions correct on a very difficult test will likely demonstrate a higher mastery of subject matter or course content than a student that gets 90% of the questions correct on a very easy test.

Consider the following scenario as another way to think about this. You are given a ten-question test on calculus, and you answer 7 out of 10 questions correctly, which equals 70%. Another person is given a ten-question test on multiplication and answers 7 out of 10 questions correctly, which equals 70%. Although you both answer 70% of the questions correctly, it would not be accurate to say that both of you demonstrate the same level of mathematics proficiency. Your test covered more difficult content—calculus as compared to multiplication.

Scale scores are a better indicator of a student's mastery of test content. While raw scores on STAAR will be available to students, parents, and teachers, it is important to understand that answering fewer than 70% of the questions correctly on a test does not necessarily indicate poor performance either in terms of scale scores or mastery of the assessed content.

18. Which STAAR assessments are reported on a vertical scale?

Under TEC §39.036, TEA is required to develop a vertical scale for assessing student performance in grades 3–8 for reading and mathematics. A vertical scale allows for a student's scale scores to be compared across different grades for the same subject area. The changes in the student's vertical scale scores indicate the academic progress the student has made over time. The assessments for which vertical scales were developed are STAAR grades 3–8 mathematics and reading in English and STAAR Spanish grades 3–5 mathematics and reading.

19. If a student attains a vertical scale score in the current grade that is higher than the passing score at a future grade, does this mean the student has met the Level II standard at the future grade?

No. While it is appropriate to compare vertical scale scores for the same student across grades to evaluate how much progress that student has made, it is not appropriate to compare a vertical scale score for a student in one grade to the passing standard in a grade in which that student has not yet received instruction. The passing scores on the vertical scale are based on the assumption that a student will have received instruction in the grade-specific curriculum in that subject area.

20. Which STAAR assessments are reported on a horizontal scale?

For any STAAR assessment that is not reported on a vertical scale, test results are reported on a horizontal scale. Horizontal scale scores were developed for STAAR grades 4 and 7 writing, grades 5 and 8 science, grade 8 social studies, and EOC assessments. A horizontal scale converts a raw score onto a scale that allows for comparisons across test forms from year to year for a specific assessment. Similar to vertical scales, horizontal scales maintain the passing standard that students are required to meet in order to reach the Level II or Level III performance categories. However, unlike vertical scales, horizontal scale scores cannot be compared to scale scores for other grades in the same subject area.

Additional information about the STAAR program can be found at <http://tea.texas.gov/student.assessment/staar/>.