

Assessment Standard Setting

Proficiency Levels and Cut Score Approvals

Introduction

The Tennessee Department of Education (Department) developed the TCAP assessments to measure the current mastery of students on the Tennessee state academic standards. To complete the reporting of achievement data, it is necessary to develop cut scores that will be used to assign students to performance levels on the assessments. These cut-scores are the basis of a criterion-referenced assessment, in which student performance is judged based on the expectations determined by content experts who are educators in Tennessee. During the week of July 13-17, 2026, panels of Tennessee educators convened in Chattanooga to review and recommend cut scores for the assessments.

After all meetings were concluded, the Department, in collaboration with psychometric experts at our assessment administration vendor (Pearson), reviewed all recommended cut scores and are presenting final recommendations for review and approval.

Today, the Department is presenting cut score recommendations for performance levels for the following assessments:

- Grades 3-8 TCAP Science
- TCAP Biology
- Grades 3-8 TCAP Alternate Science
- TCAP Alternate Biology

For each of the indicated assessments, student performance will be reported using performance categories, which require setting cut scores. The Department recommends the performance level cut scores shown in this report for adoption by the Tennessee State Board of Education. Of special note is that TCAP assessments classify students using four performance levels (e.g., exceeds expectations), and TCAP Alternate assessments classify students using three performance levels.

Standard Setting Process

Performance levels are used to classify and describe student performance on an assessment. In order to classify student performance into the different performance levels, the following components are generally required: 1) Policy Performance Level Descriptors, 2) Range Performance Level Descriptors, and 3) cut scores. Policy performance level descriptors provide general descriptions of what students at each performance level know and what they are able to do. The range performance level descriptors (PLDs) illustrate the performance levels in terms that are specific to a grade and subject. Cut scores represent the lowest boundary of each performance level on the scale used to score the assessments.

The process of recommending performance standards for the TCAP Science and TCAP Alternate Science assessments is based on industry best practice for standard settings. The standard setting methodology used is a modification of the Angoff method (Thorndike, 1971), a model commonly used for standard setting.

Policy Performance Level Definitions

Policy Performance Level Descriptors for the TCAP assessments are shown in Table 1. The titles and descriptions of the performance levels were defined to be part of a cohesive assessment system and provide general descriptions of student performance without regard to subject or grade-level. These policy-level descriptors are consistent across all grades and subject areas included in the full suite of TCAP assessments.

Table 1. Policy Performance Level Descriptors (general education)

Performance Level	Policy Performance Level Descriptors
Level 4: Exceeds Expectations	Performance at this level demonstrates that the student has an extensive understanding and has an expert ability to apply the grade/course-level knowledge and skills defined by the Tennessee Academic Standards.
Level 3: Meets Expectations	Performance at this level demonstrates that the student has a comprehensive understanding and has a thorough ability to apply the grade/course-level knowledge and skills defined by the Tennessee Academic Standards.
Level 2: Approaching	Performance at this level demonstrates that the student has an approaching understanding and has a partial ability to apply the grade/course-level knowledge and skills defined by the Tennessee Academic Standards.
Level 1: Below	Performance at this level demonstrates that the student has a minimal understanding and has a nominal ability to apply the grade/course-level knowledge and skills defined by the Tennessee Academic Standards.

Policy Performance Level Definitions

Policy Performance Level Descriptors for the TCAP Alternate assessments are shown in Table 2. The titles and descriptions of the performance levels were defined to be part of a cohesive assessment system and provide general descriptions of student performance without regard to subject or grade-level. These policy-level descriptors are consistent across all grades and subject areas included in the full suite of TCAP Alternate assessments.

Table 2. Policy Performance Level Descriptors (alternate education)

Performance Level	Policy Performance Level Descriptors
Level 3: Broad	<i>A student performing at this level demonstrates a broad understanding of the knowledge and skills defined by the Tennessee alternate assessment standards.</i>
Level 2: Developing	<i>A student performing at this level demonstrates a developing understanding of the knowledge and skills defined by the Tennessee alternate assessment standards.</i>
Level 1: Emerging	<i>A student performing at this level demonstrates an emerging understanding of the knowledge and skills defined by the Tennessee alternate assessment standards.</i>

Performance Level Descriptors (PLDs)

A multi-step iterative process was used in developing, reviewing, and approving the subject and grade-level Range PLDs. Prior to the standard setting committee, a draft set of PLDs representing a gradual increase in expectations across the performance levels was created by content staff from the department. The resulting draft PLDs were reviewed by Tennessee educators during a review committee. The committee reviewed for alignment to the Tennessee policy performance level descriptors, the Tennessee Academic Standards, and consistency of expectations across performance levels. Recommendations were made by the committee for edits and additions to the draft. Following applications of committee recommendations, a revised draft of the

PLDs was reviewed and finalized by content staff from the department. Panelists who participated in the PLD review committees were asked to participate in the eventual standard setting committee.

Standard Setting Meetings¹

From July 13 through July 17, 2026, following the spring 2026 operational administration, a standard setting committee meeting was conducted to provide cut score recommendations for the TCAP assessments for Science and Alternate Science. There were 14 committees, with each recommending cut scores for one assessment. Each committee was composed of between 8 and 22 individuals, including teachers and non-teacher Tennessee educators (e.g., administrators, curriculum specialists), for a total of 82 educators across all eight CORE regions of the state. The participants were selected for the standard setting committee to provide content and grade-level expertise during the committee meeting and be representative of the state teaching population, including geographic region, gender, ethnicity, educational experience, community size, and community socio-economic status. Many of the panelists who worked on the initial Tennessee Academic Standards for Science Standards Revision Committee were recruited to participate in both the PLD Review and Standard Setting meetings for those courses.

The Extended Modified (Yes/No) Angoff standard setting method was used for the standard setting meeting (Davis & Moyer, 2015; Plake, Ferdous, Impara, & Buckendahl, 2005). This is a content- and question-based method that leads participants through a standardized process through which they consider student expectations, as defined by PLDs, and the individual questions that could be administered to students to recommend cut scores for each performance level. The standardized process was used by the committees for each grade/subject.

The process started with participants reviewing the design of the TCAP assessments and assessment questions. Panelists then created borderline descriptions based on the performance level descriptions specific to the subject and grade level to create descriptors of the knowledge, skills, and abilities that students with performance at the borderline of the performance level would be expected to demonstrate. Panelists then began an iterative judgment process where they completed three rounds of judgments on each question included on the test. The cut score recommendation for each individual participant was the expected raw score a student with performance at the borderline of the respective performance level would likely earn, calculated as the sum of the individual question judgments. Each recommended cut score from the standard setting committee was the median of the recommendations from the individual participants in the committee.

Final committee recommended cut scores were then reviewed by the department, resulting in a small number of adjustments for some performance levels to ensure final performance level cuts reflected the expectation of the Tennessee educator panelists for similar distributions of performance levels across grade levels. This final, internal review of the cut score recommendations was conducted by TDOE stakeholders, including content experts, psychometricians and TDOE leadership, to ensure consistent expectations for performance levels across each grade-level assessment (i.e., vertical alignment). The adjustments to the recommendations made during the final review process were influenced by a desire to honor the content-based recommendations of the standard setting process, maintain high expectations for performance across the TCAP assessments grade-to-grade, and ensure the relationship among standards was coherent and defensible.

¹ Full standard setting report will be available upon completion.

Technical Advisory Committee Review

The plans for each standard setting committee meeting were reviewed, discussed, edited, and approved in advance of each meeting by the department's Technical Advisory Committee (TAC). The TAC includes nationally recognized assessment and psychometric experts. The TAC has approved the approach implemented in this standard setting process, which was of high quality, and the State Board of Education should feel confident in the panelists' resulting recommendations for cut scores.

Panelists

A total of 82 educators from the state of Tennessee participated as panelists in the standard setting committee meetings, with over 140 panel seats filled. The panelists were recruited for participation starting in the spring of 2026 and were selected using a comprehensive application process in which they demonstrated in-depth content knowledge of the Tennessee academic state standards. Panelists were selected, to the extent possible, so that they constituted a representative sample of the state teaching population in terms of region, gender, ethnicity, community size and socioeconomic status, and experience. The educators represented school districts across seven CORE regions of the state.

Recommended Cut Scores

The results from the standard setting meetings for the TCAP Science and TCAP Alternate Science committees, reviewed and finalized by the department, are presented in the tables below.

Table 3: TCAP Science Recommended Cut Scores

Grade	Performance Level			Max Points Available
	Level 2: Approaching	Level 3: Meets	Level 4: Exceeds	
	Cut Score	Cut Score	Cut Score	
3	10	16	25	30
4	11	18	26	30
	15	27	39	45
6	14	26	39	45
7	15	26	40	45
8	15	29	40	45
Biology	14	24	38	45

Table 4: TCAP Alternate Science Recommended Cut Scores

Grde	Performance Level		Max Points Available
	Level 2: Developing	Level 3: Broad	
	Cut Score	Cut Score	
3	14	23	32
4	15	23	32
5	15	23	32
6	15	24	32
7	15	23	32
8	15	25	32
Biology	16	25	32

Reporting Scale

The process of determining the transformation rules from the Item Response Theory (IRT) scale to the final reporting scale will be guided by several principles in order to ensure consistent understanding and ease of interpretation of the scale scores by students, parents, educators, and leaders across the state.

1. The final cut scores determined while selecting the final scaling solution should respect the cut score recommendations from the standard setting committee as closely as possible.
2. The scaling solution should involve a single linear transformation, from the IRT scale to the reporting scale.
3. The reporting scaled score range should be the same across grades and tests.

After the standards setting and finalization of recommendations were complete, the procedures for transforming student raw scores from the administration to the reporting scale were found. To ensure consistent interpretation and use of scale scores, it has been determined that the TCAP reporting scale will have the following properties across all grades:

TCAP Science Reporting scale

- The lowest obtainable scale score (LOSS) would be set at 200.
- The highest obtainable scale score (HOSS) would be set at 450.

The reporting scale was set using the two fixed scores for the Lowest (LOSS) and Highest (HOSS) obtainable scale scores at 200 and 450, respectively. The scale score for all cut scores is set empirically. While the cut scores were defined with the same policy descriptions across the courses (for each type of assessment), they are not identical, and direct comparisons through averaging and aggregation across grades should not be made without study and/or statistical adjustments.

Table 5 below presents the results from the final scaling solutions for the TCAP Science tests.

Table 5: Reporting Scale Cuts for TCAP Science

Grade	Performance Level		
	Level 2: Approaching	Level 3: On Track	Level 4: Mastered
	Cut Score	Cut Score	Cut Score
3	300	330	374
	300	330	376
5	300	330	361
6	300	330	362
7	300	330	367
8	300	330	358
Biology	300	330	367

TCAP Alternate Science Reporting scale

- The lowest obtainable scale score (LOSS) would be set at 100.
- The highest obtainable scale score (HOSS) would be set at 250.

The reporting scale was set using the two fixed scores for the Lowest (LOSS) and Highest (HOSS) obtainable scale scores at 100 and 250, respectively. The scale score for all cut scores is set empirically. While the cut scores were defined with the same policy descriptions across the courses (for each type of assessment), they are not identical, and direct comparisons through averaging and aggregation across grades should not be made without study and/or statistical adjustments.

Table 6 below presents the results from the final scaling solutions for the TCAP Alternate Science tests.

Table 6: Reporting Scale Recommendations for TCAP Alternate Science

Grade	Performance Level	
	Level 2: Approaching	Level 3: On Track
	Cut Score	Cut Score
3	150	175
4	150	175
5	150	175
6	150	175
7	150	175
8	150	175
Biology	150	175

Bibliography

Davis, L. L. & Moyer, E. L. (2015). PARCC performance level setting technical report. Available from Partnership for Assessment of Readiness for College and Careers (PARCC), Washington, D.C.

Plake, B. S., Ferdous, A. A., Impara, J. C., & Buckendahl, C. W. (2005). Setting Multiple Performance Standards Using the Yes/No Method: An Alternative Item Mapping Method. Meeting of the National Council on Measurement in Education. Montreal, Canada.

Thorndike, R. L. (Ed.). (1971). Educational measurement (2nd ed.). Washington, D.C.: American Council on Education.