

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 and their progress toward college and career readiness. In order 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 November 9-13, 2020, panels of Tennessee educators convened virtually to recommend cut scores for the assessments.

During the first two weeks of July 2021, committee members from the original content meetings were convened to complete standard setting with a final vertical articulation review and additional committee members were convened to conduct standard settings for alternate assessments. After all meetings have concluded, the department, in collaboration with psychometric experts at our assessment administration vendor (Pearson), reviewed all recommended cut scores and are presenting these recommendations for review and approval.

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

- Grades 3-8 Science
- Biology
- Grades 6-8 Social Studies
- U.S. History
- Grades 3-8 Alternate Science
- Alternate Biology

For each general education assessment, student performance will be reported using four performance categories, which require setting three cut scores. For each alternate assessment, student performance will be reported using three performance categories, which require setting two cut scores. The Department and Pearson recommend the performance level cut scores shown in this report for adoption by the Tennessee State Board of Education.

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) 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 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 social studies assessments is based on national best practice for standard settings. The standard setting methodology used is a modification of the well-known Angoff method (Thorndike, 1971).

Policy Performance Level Definitions

Policy Performance Level Descriptors for the TCAP assessments are shown in Table 1 (general education) and Table 2 (alternate assessment). 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: Mastered	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: On Track	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.

Table 2. Policy Performance Level Descriptors (alternate assessment)

Performance Level	Policy Performance Level Descriptors
Level 3: Broad	A student in Level 3 demonstrates a broad understanding of the grade-level knowledge and skills defined by the Tennessee Alternate Assessment Standards.
Level 2: Developing	A student in Level 2 demonstrates a developing understanding of the grade level knowledge and skills defined by the Tennessee Alternate Assessment Standards.
Level 1: Emerging	A student in Level 1 demonstrates an emerging understanding of the grade-level knowledge and skills defined by the Tennessee Alternate Assessment Standards.

Performance Level Descriptors (PLDs)

A multi-step iterative process was used in developing, reviewing, and approving the subject and grade-level specific 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 Tennessee Department of Education. The initial draft PLDs were reviewed and revised 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. The revised draft of the PLDs were reviewed and finalized by content staff from TDOE. Panelists who participated in the standard setting committees had the opportunity to provide suggestions and edits to the PLDs utilized during the standard setting meetings.

Standard Setting Meetings¹

From November 9 through 13, 2020, prior to the spring 2021 operational administration, a standard setting committee meeting was conducted to provide cut score recommendations for the TCAP general education assessments for science and social studies. There were 11 committees, with each recommending cut scores for one assessment. Each committee was composed of between 9 and 12 individuals, including teachers and non-teacher educators (e.g., administrators, curriculum specialists, postsecondary faculty), 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 socioeconomic status.

The Extended Modified (Yes/No) Angoff standard setting method was used for the general education 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.

A final review of the cut score recommendations was completed during the final vertical articulation meetings, held in July 2021, and involving a subset of educators of the original meetings (32 total). Question performance data from the Spring 2021 administration was used to consider if question judgements from the first round of meetings were in line with content expectations and if any adjustments to final recommended cut scores were warranted based on content. A final review was conducted to ensure consistent expectations for performance levels across each grade-level assessment (vertical alignment). The adjustments to the recommendations made during the articulation 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. Final committee recommended cut scores were then reviewed again by the Department, resulting

¹ Full standard setting report is available for your review upon request.

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.

For the alternate assessments, a standard setting committee meeting was conducted from July 12 through 15, 2021 to provide cut score recommendations for the TCAP Alternate assessments for science. There were seven sub-committees, with each recommending cut scores for one grade-level assessment. Each committee was composed of between 9 and 12 individuals from across the state of Tennessee, including classroom teachers that teach science and educators that specifically work with students in this population with the most severe cognitive disabilities. 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 socioeconomic status. In total, 41 panelists from seven of the state's eight CORE regions were involved, including some science educators who had previously served on the general education committees.

The Profile-Informed Extended Modified (Yes/No) Angoff standard setting method was used for the alternate assessment standard setting meeting. This is modification of the Extended Modified (Yes/No) Angoff standard setting method (Davis & Moyer, 2015; Plake, Ferdous, Impara, & Buckendahl, 2005) used for the general education assessments, with the inclusion of an additional step in which panelists reviewed student score profiles. Score profiles were recommended by the Technical Advisory Committee (TAC) to assist panelists in understanding how students in this special population engage with the assessment and arrived at potential score point totals.

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. TAC members not only reviewed and approved the standard setting methodology, processes, and materials in advance of the meetings, but members also sat in on committee meetings to ensure processes were implemented with fidelity by the department's assessment vendor and committee facilitators. The TAC has confirmed the process was high quality, was conducted as proposed, and the State Board of Education should feel confident in the panelists' recommendations for cut scores.

Panelists

A total of 109 educators from the state of Tennessee participated as panelists in the standard setting committee meetings across all grades and subject areas, including 84 participating in general education meetings and an additional 41 participating in alternate standard setting meetings (including 16 who served on both). The panelists were recruited for participation starting in the spring of 2020 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 47 school districts and 2 postsecondary institutions across all eight CORE regions of the state.

Recommended Cut Scores

The results from the standard setting meetings for the science and social studies committees, reviewed and finalized by the department and Pearson, are presented in the tables below.

Table 3: General Education Science Recommended Cut Scores

Grade	Performance Level			Max Points Available
	Level 2: Approaching	Level 3: On Track	Level 4: Mastered	
	Cut Score	Cut Score	Cut Score	
3	10	17	24	30
4	12	17	24	30
5	20	30	42	52
6	19	30	45	52
7	18	29	43	52
8	18	31	43	52
Biology	20	29	44	52

Table 4: General Education Social Studies Recommended Cut Scores

Grade	Performance Level			Max Points Available
	Level 2: Approaching	Level 3: On Track	Level 4: Mastered	
	Cut Score	Cut Score	Cut Score	
6	17	27	40	50
7	18	26	40	50
8	19	27	40	50
U.S. History	22	33	44	54

Table 5: Alternate Science Assessment Recommended Cut Scores

Grade	Performance Level		Max Points Available
	Level 2: Developing	Level 3: Broad	
	Cut Score	Cut Score	
3	14	25	32
4	14	24	32
5	15	24	32
6	15	24	32
7	13	21	32
8	15	24	32
Biology	13	22	31

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.
4. The cut scores on the reporting scale for the *Level 2* performance level should be the same across grades and tests.
5. The cut scores on the reporting scale for the *Level 3* performance level should be the same across grades and tests.
6. For general education assessments, the cut scores on the reporting scale for any of the performance levels should end in either a 0 or 5.

After the standards setting and vertical articulation processes 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:

General education scale

- The lowest obtainable scale score (LOSS) would be set at 200.
- The highest obtainable scale score (HOSS) would be set at 450.
- The cut score for the *Level 2 – Approaching* cut would be set at 300.
- The cut score for the *Level 3 – On Track* cut would be set at 330.

Alternate assessment scale

- The lowest obtainable scale score (LOSS) would be set at 100.
- The highest obtainable scale score (HOSS) would be set at 250.
- The cut score for the *Level 2 – Developing* cut would be set at 150.
- The cut score for the *Level 3 – Broad* cut would be set at 175.

The reporting scale was set using the two cut scores for *Level 2* and *Level 3*. For general education, the scale score for the *Level 4 – Mastered* cut is set empirically. While the cut scores were defined with the same scaled score cuts for the *Level 2* and *Level 3* levels and the same policy descriptions across the grades (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. The LOSS and HOSS for the alternate assessment are intentionally different than the general education assessment to prevent inaccurate comparisons across the two assessment types. The scaled scores and distributions of students resulting from the cuts were not designed for direct comparison.

Tables 6, 7, and 8 below present the results from the final scaling solutions for the science, social studies, and alternate science tests, respectively.

Table 6. Reporting Scale Cuts for TCAP Science

Grade	Performance Level		
	Level 2: Approaching	Level 3: On Track	Level 4: Mastered
3	300	330	365
4	300	330	370
5	300	330	370
6	300	330	375
7	300	330	370
8	300	330	365
Biology	300	330	375

Table 7. Reporting Scale Recommendations for TCAP Social Studies

Grade	Performance Level		
	Level 2: Approaching	Level 3: On Track	Level 4: Mastered
6	300	330	365
7	300	330	380
8	300	330	380
U.S. History	300	330	360

Table 8. Reporting Scale Cuts for TCAP-Alt Science

Grade	Performance Level	
	Level 2: Developing	Level 3: Broad
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.