

Assessment Standard Setting

Proficiency Levels and Cut Score Approvals

Introduction: TCAP-Alternate Standard Setting

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 June 27-29, panels of Tennessee educators convened to recommend cut scores for the assessments. After these 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:

- Grade 6 Alternate Social Studies
- Grade 7 Alternate Social Studies
- Grade 8 Alternate Social Studies

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: TCAP-Alt

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 alternate 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 alternate assessments.

Table 1. Policy Performance Level Descriptors (alternate education)

Performance Level	Policy Performance Level Descriptors
Level 3: Broad	A student in the PLD 3 level demonstrates a broad understanding of the knowledge and skills defined by the TN alternate assessment standards.
Level 2: Developing	A student in the PLD 2 level demonstrates a developing understanding of the knowledge and skills defined by the TN alternate assessment standards.
Level 1: Emerging	A student in the PLD 1 level demonstrates an emerging understanding of the knowledge and skills defined by the TN 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 department. The initial draft PLDs were reviewed and revised by Tennessee educators during a review committee, which was convened May 2-4, 2022. 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 the department. 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 June 27 - 29, a standard setting committee meeting was conducted to provide cut score recommendations for the TCAP assessments for alternate social studies. There were 3 committees, with each recommending cut scores for one assessment. Each committee was composed of 6 individuals, including classroom teachers that teach social studies 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.

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. The Angoff method 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

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

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.

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 entered into 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. Finally, 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. Student performance data from the Spring 2022 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 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, with the intent to represent the most accurate representation of student performance against content expectations.

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 reviewed and approved the standard setting methodology, processes, and materials in advance of the meeting. 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 21 educators from the state of Tennessee participated as panelists in the PLD and standard setting committee meetings across all grades. The panelists were recruited for participation starting in the spring of 2022 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 15 different school districts across 7 of the 8 CORE regions of the state. Educators serving on these panels had significant experience in schools, all having served as an educator between 7 and 33 years, with the average years of experience being 18.

Recommended Cut Scores: TCAP-Alt

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

Table 2: Alternate Social Studies Recommended Cut Scores

Grade	Performance Level		
	Level 2: Developing	Level 3: Broad	Max Points Available
	Cut Score	Cut Score	
6	16	27	32
7	17	27	32
8	17	29	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.
4. The cut scores on the reporting scale for the *Level 2 - Developing* performance level should be the same across grades and tests.
5. The cut scores on the reporting scale for the *Level 3 - Broad* performance level should be the same across grades and tests.
6. The cut scores on the reporting scale for any of the performance level 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 Alternate reporting scale will have the following properties across all grades:

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*. While the cut scores were defined with the same scaled score cuts for these levels and the same policy descriptions across the grades, they are

not identical, and direct comparisons through averaging and aggregation across grades should not be made without study and/or statistical adjustments. The scaled scores and distributions of students resulting from the cuts were not designed for direct comparison.

Table 3 presents the results from the final scaling solutions for the alternate social studies tests.

Table 3. Reporting Scale Cuts for TCAP-Alt Social Studies

Grade	Performance Level	
	Level 2: Developing	Level 3: Broad
6	150	175
7	150	175
8	150	175

Introduction: MSAA Cut Score Recommendations, Overview of MSAA

The Multi-State Alternate Assessment (MSAA) assesses English Language Arts (ELA) and mathematics at grades 3–8 and 11 for students with the most significant cognitive disabilities and is aligned with Tennessee Academic standards and the MSAA Core Content Connectors (CCCs). The MSAA is a consortium that includes multiple states who partner together to expand the scale of assessments for this small student population. The MSAA is a computer-based, on-demand, two-stage adaptive assessment, consisting primarily of selected-response items, along with some constructed-response items and open-response writing prompts. These item types are written at distinct levels of complexity, representing different levels of skill and knowledge acquisition by students.

Students with the most significant cognitive disabilities often need materials and instructional strategies that are substantially adapted and scaffolded, providing built-in supports to meet their individual needs. When students begin to learn a new skill or acquire new knowledge, they need more support and scaffolding. As students learn and develop mastery of that skill or knowledge, they need less support. The MSAA levels of complexity are designed to follow instructional practices. The test items are developed with many scaffolds and supports embedded within the items. Students are provided additional support based on their individual requirements, including other allowable ways for Test Administrators (TAs) to present each item.

Standard Setting Process: MSAA

Cut scores for MSAA in ELA and mathematics were originally set in a standard setting process that took place in August 2015. Details of the standard setting procedures can be found in the standard setting report (Measured Progress, 2015). In July 2018, MSAA's test development and administration vendor (Cognia) and the MSAA Psychometric Subcommittee conducted a standards validation. Standards validation does not change the scale; its purpose is only to determine whether adjustments to the cut scores are needed. The standards validation process for the 2018 MSAA was necessary to ensure that cut scores, set in 2015 for the assessments, continue to provide valid interpretation of ELA and mathematics performance using the Performance Level Descriptors (PLDs). The validation process for the ELA performance standards was necessitated by the addition of the open-response writing prompt scores to the existing ELA score scale in 2018. After review of historical documentation this year, the department realized that these updated

standards and cut scores were not previously brought to the Tennessee State Board of Education for official review and approval.

A complete description of the standards articulation and validation processes appears in the 2018 MSAA Standards Validation Report (see Appendix M of the [2018 MSAA Technical Report](#). Final cut scores, after mathematics and ELA vertical articulation and ELA standards validation for the 2017-18 MSAA, appear below.

Recommended Cut Scores: MSAA

The results from the standards validation process, reviewed and finalized by MSAA partner states, are presented in the tables below.

MSAA Assessment Scale

- The lowest obtainable scale score (LOSS) would be set at 1200.
- The highest obtainable scale score (HOSS) would be set at 1290.

The ELA standards validation process involved five steps: (a) vertically articulating the performance level 3 cut scores for both ELA and mathematics, to update the performance standards and provide a coherent basis for interpreting 2018 scores and performance and validating the ELA cut scores; (b) expanding the ELA PLDs to include references to direct open-response writing prompt knowledge and skills; (c) expanding the existing ELA score scale by linking the open-response writing prompt scores to the scale; (d) reviewing and possibly adjusting the existing cut scores to align the response demands of all items, including the open response writing prompt scores, and knowledge and skill requirements in the expanded PLDs; and (e) reviewing the expanded ELA PLDs with stakeholders from various schools and districts to confirm the writing knowledge and skills were clear and included language around the open-response writing prompt expectations in each of the performance levels. Members of the MSAA Psychometric Subcommittee, including MSAA ELA Content Specialists, comprised the ELA standards validation panel. In addition, two members of the TAC attended to monitor the process and provide advice and support as needed. These cut scores for mathematics and ELA were brought to each member state for thorough review and approval prior to finalizing the cut scores.

Table 4 and 5 reporting scale score cuts for MSAA ELA and Math.

Table 4. Reporting Scale Cuts for MSAE ELA

Grade	Performance Level		
	Cut 1	Cut 2	Cut 3
3	1234	1240	1254
4	1234	1240	1259
5	1232	1240	1256
6	1231	1237	1251
7	1236	1240	1255
8	1230	1238	1250
11	1236	1240	1255

Table 5. Reporting Scale Cuts for MSAE Math

Grade	Performance Level		
	Cut 1	Cut 2	Cut 3
3	1235	1242	1254
4	1232	1239	1251
5	1232	1240	1253
6	1233	1239	1251
7	1234	1240	1254
8	1234	1240	1251
11	1235	1240	1250

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.

Measured Progress Department of Psychometrics and Research (2015). National Center and State Collaborative (NCSC) standard setting report. Chapter 7:

http://ncscpartners.org/Media/Default/PDFs/Resources/NCSC15_NCSC_TechnicalManualNarrative.pdf

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.