



**Karen B. Salmon, Ph.D.**  
State Superintendent of Schools

---

**TO:** Members of the State Board of Education

**FROM:** Karen B. Salmon, Ph.D.

**DATE:** October 23, 2018

**SUBJECT:** Maryland Integrated Science Assessment (MISA) Standard Setting for Grades 5 & 8

---

**PURPOSE:**

The purpose of this agenda item is to provide information on the Maryland Integrated Science Assessment (MISA) Standard Setting for Grades 5 & 8 for which cut scores were established that place student test scores into performance levels of Exceeded Expectations, Met Expectations, Approached Expectations, and Partially Met Expectations.

**BACKGROUND:**

The Every Students Succeeds Act (ESSA) requires that states administer to all students annual assessments in science once in each grade span (3-5, 6-8 and HS). The MISA for Grades 5 & 8 was field tested in Maryland in 2016-2017 and administered operationally in 2017-2018. In August of 2018, the Maryland State Department of Education (MSDE) Assessment Department, along with technical support from WestEd, convened two panels of teachers to participate in standard setting for the two assessments. The panels were made up of Maryland teachers, instructional specialists/coaches, and administrators.

The state assessments in mathematics and English language arts (ELA) went through a similar standard setting process to establish the three (3) cut scores for the same four (4) performance levels. As with the mathematics and ELA assessments a modified Angoff method was used to train panelists to make judgments about the number of points that a student would need to earn to achieve a specific performance level.

Panelists began by experiencing the test items as a student would, then developing a collective definition of a “borderline” or minimally-competent student. Panelists went through three (3) rounds of judgments and were provided with additional information after each round to inform their decisions. After their first round of judgments, panelists were given item difficulty data (percentage of students who achieved each score point). After the second and third rounds of judgments, panelists were given impact data to show how their current judgments would reflect in the percentage of students performing at each level. The final step in the process was vertical articulation to ensure that cut scores made sense between grades and were not in opposition to the performance of students on the mathematics and ELA assessments.

The recommendations by the panel were reviewed by MSDE staff and a final adjustment and established cut scores were set for use of reporting on the MISA assessments.

**EXECUTIVE SUMMARY:**

Standard setting for the Maryland Integrated Science Program (MISA) for grades 5 & 8 took place in August of 2018 after the first operational administration of assessments. Maryland educators participated as panelists going through three (3) rounds of judgments using a modified Angoff method. After a final vertical articulation process took place to ensure the results were reasonable between grades, three cut-scores were recommended for each grade to place student scores into four (4) performance levels of Exceeded Expectations, Met Expectations, Approached Expectations, and Partially Met Expectations.

**ACTION:**

For information only. No action required.

**Attachments (2):**

Attachment I – Maryland Integrated Science Assessment (MISA) Standard Setting Overview  
Attachment II – PowerPoint Presentation

## Maryland Integrated Science Assessment (MISA) Standard Setting Overview

The purpose of standard setting is to establish cut scores that place students into performance levels for a given assessment. There are several methods for setting performance standards; these methods range from norm-referenced to criterion-referenced and from item-centered to person-centered. Most standard setting processes for state assessment programs use criterion-referenced methods, through which cut scores are determined based on the knowledge and skills required of students, according to state content standards and performance level descriptors (PLDs), which describe what students at each performance level know and are able to do.

The Maryland Integrated Science Assessment (MISA) has four performance levels: Exceeded Expectations, Met Expectations, Approached Expectations, and Partially Met Expectations. Therefore, the outcome of the MISA standard setting process for each grade (grades 5 and 8) was three recommended cut scores separating four performance levels. Each panelist set standards for one grade level, either grade 5 or grade 8. At the start of the standard setting process, panelists were guided through a detailed review of the assessment content for their grade level, the PLDs, and the method for setting performance standards. Prior to beginning their judgments, the grade-level panelists collectively defined the borderline, or minimally competent, student for each performance level. Definitions of borderline students serve as the foundation for the setting of performance standards, providing panelists with a common understanding of the minimum knowledge and skills required for a student to achieve each performance level on MISA.

As with standard setting for the PARCC assessments, a modified Angoff method was used for the MISA standard setting. Specifically, the Modified (Yes/No) Angoff method is a criterion-referenced and item-centered method in which trained panelists make judgments about the number of points that a student would need to earn on MISA to be placed into a specific performance level. Each panelist independently reviews each item on the test and decides the number of points that the borderline student for each performance level would earn on each item. For items worth one point, the judgment process involves answering the following yes-or-no question: Would the borderline student answer this item correctly? For items worth more than one point, the judgment process involves answering the following question: How many score points would the borderline student earn on this item? Each panelist's recommended cut score for each performance level is the sum of the individual item score judgments for the borderline student at each performance level. After all the grade-level panelists make their individual recommendations, the grade-level panelists' median cut scores serve as the recommended cut score for each performance level. For the MISA standard setting, this process was repeated for three rounds of judgments at each grade level.

Between rounds, grade-level panelists were provided with data to help them analyze, discuss, and refine their judgments. After every round, the grade-level panelists were provided with the recommended cut scores for each panelist at their table, their table's median cut scores, and the entire grade-level group's median cut scores. After the first round, the grade-level panelists were also provided with item difficulty data (for each item worth one point, the percentage of students who answered the item correctly, and for each item worth more than one point, the percentages of students who achieved each score point). After the second and third rounds, the grade-level panelists were provided with impact data (the percentages of students in each performance level, based on the current cut-score recommendations). After each round, the panelists discussed their judgments with their fellow panelists (both at their table and within the entire grade-level group), and reflected on the reasonableness of their own recommendations, given their peers' judgments and the item difficulty and impact data.

Although impact data introduces a normative aspect into the process, its use is a common practice in standard setting for educational assessment programs. When developing statewide assessments, it is important that the full range of student abilities be assessed. Statewide assessments often include items that are very difficult, even for students who demonstrate a strong grasp of the skills and knowledge being assessed. Therefore, student performance on these assessments may differ from what is expected or observed on a classroom assessment.

After all judgments were completed for each grade level, all the panelists reconvened for “vertical articulation,” in which impact data for both grades were evaluated together and panelists discussed the appropriateness of the cut scores from the perspectives of students, parents, educators, and the public. Upon completion of vertical articulation, the recommended cut scores for each performance level and each grade were finalized.

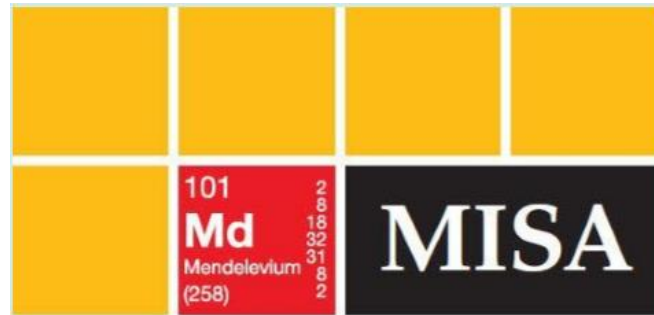
At the completion of the standard setting meeting, panelists were asked to provide background information about themselves. Of the 25 panelists who participated, 24 submitted responses.

<b>Question</b>	<b>Responses (<i>n</i> = 24)</b>
What is your current role?	Teacher (16) Instructional Specialist/Coach (5) Administrator (2) Other (1)
How many years have you been in your current role?	1–5 years (8) 6–10 years (9) 11–15 years (4) 15–20 years (1) 20+ years (2)
Select the student population(s) for which you have educational experience.	Students receiving special education services (24) Students who are English language learners (20) Students of low socioeconomic status (23)
What is your highest level of education?	Bachelor’s degree (5) Master’s degree (17) Doctoral degree (2)
Gender	Female (19) Male (5)
Race/Ethnicity	Black/African American (4) White (20)

The following references provide more information on methods for setting performance standards (Cizek) and the difference between the traditional and modified Angoff methods (Impara & Plake).

Cizek, G. J. (2012). *Setting performance standards: Foundations, methods, and innovations*. New York, NY: Routledge.

Impara, J. C., & Plake, B. S. (1997). Standard setting: An alternative approach. *Journal of Educational Measurement*, 34(4), 353–366.



# Maryland Integrated Science Assessment Standard Setting – State Board Presentation



October 23, 2018

# Purpose of Standard Setting Meeting

- Provide recommendations for cut scores for each of the performance levels.
- Cut score recommendations will be used to establish the final performance levels applied to student reports.



# Overview of Standard Setting Process

Experience the MISA Test

Understand the MISA PLDs and create  
borderline PLDs

Learn about the standard setting judgment  
process

Apply the standard setting judgment process

Review and discuss feedback data

# MISA Performance Levels Descriptors

## Exceeded Expectations

Exceeded expectations by demonstrating a full and complete understanding of grade appropriate Disciplinary Core Ideas (Life Science, Physical Science and Earth and Space Science), the Science and Engineering Practices, and the Cross Cutting Concepts.

## Met Expectations

Met expectations by demonstrating a general understanding of grade appropriate Disciplinary Core Ideas (Life Science, Physical Science and Earth and Space Science), the Science and Engineering Practices, and the Cross Cutting Concepts.

## Approached Expectations

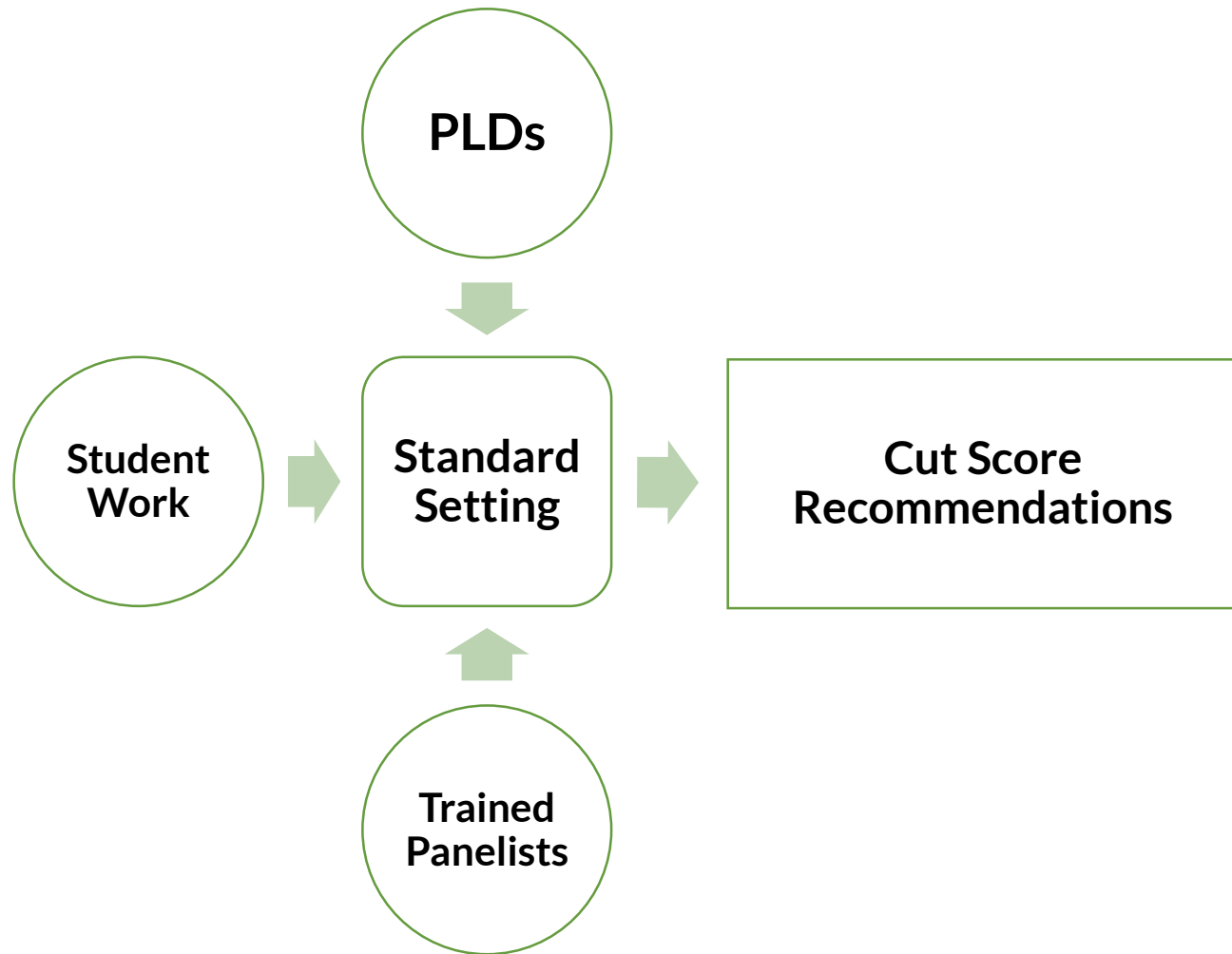
Approached expectations by demonstrating a basic understanding of grade appropriate Disciplinary Core Ideas (Life Science, Physical Science and Earth and Space Science), the Science and Engineering Practices, and the Cross Cutting Concepts.

## Partially Met Expectations

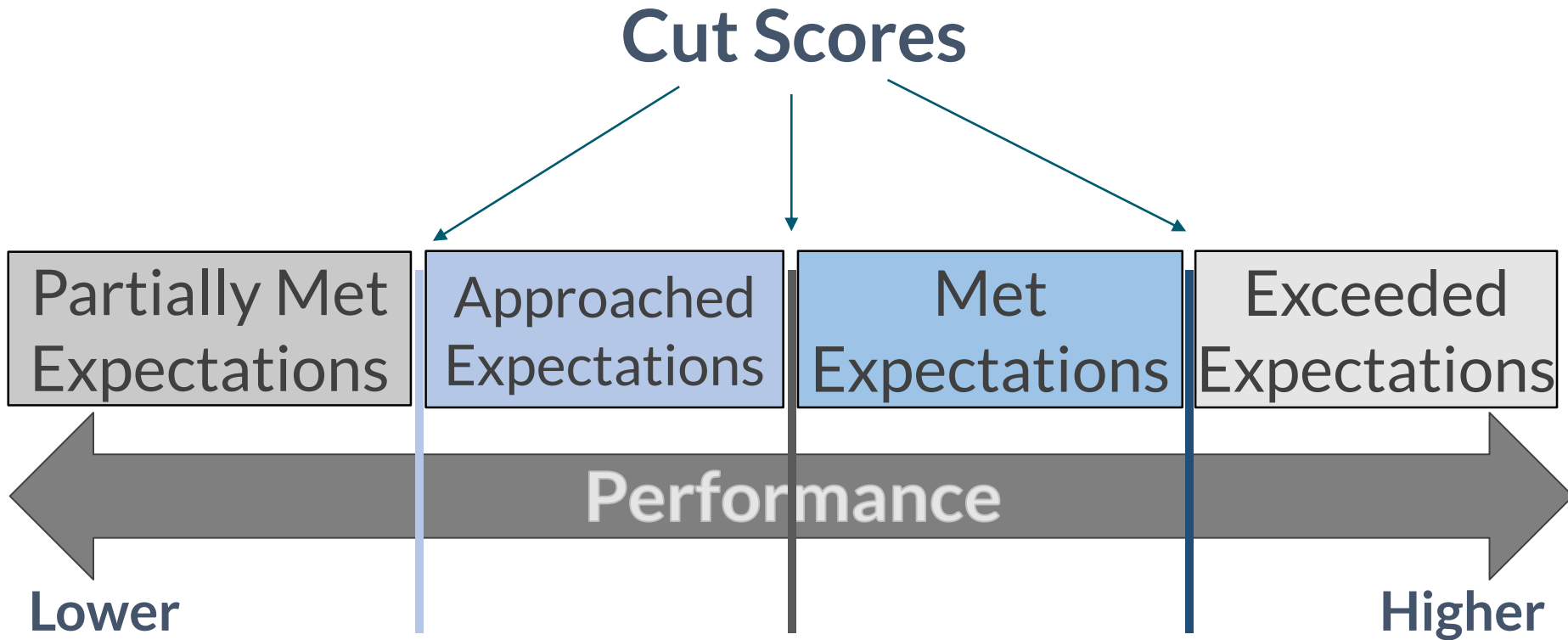
Partially met expectations by demonstrating a minimal understanding of grade appropriate Disciplinary Core Ideas (Life Science, Physical Science and Earth and Space Science), the Science and Engineering Practices, and the Cross Cutting Concepts.



# What is Standard Setting?

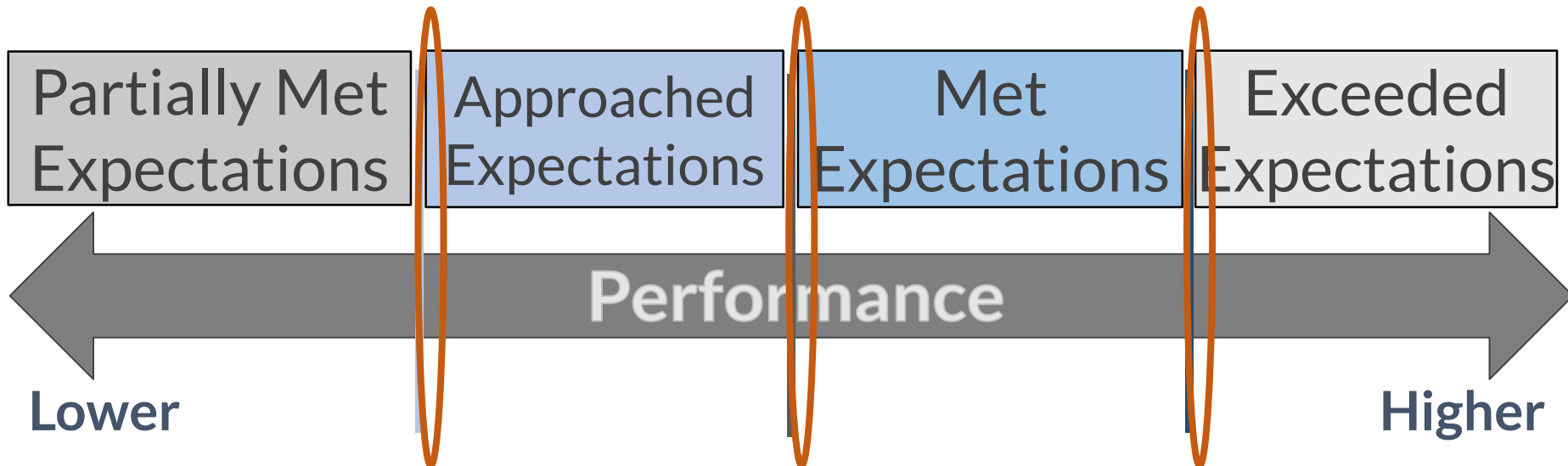


# What is Standard Setting?



# Overview of the Modified Angoff Standard-Setting Method

1. Panelists defined the borderline student for each performance level



# Overview of the Modified Angoff Standard-Setting Method

1. Panelists defined the borderline student for each performance level.
2. Each panelist individually reviewed each item and decided (for each borderline student):
  - Would this student answer this item correctly (Yes / No)?  
*(For correct/incorrect items)*
  - How many score points would this student earn on this item?  
*(For multi-part items or items with a scoring rubric)*
3. For each of the three borderline students, all the “Yes” judgments and score points were summed, producing three cut scores to separate four performance levels.
4. The grade-level panelists’ median cut score was used to determine the recommended cut score.

# Overview of the Modified Angoff Standard-Setting Method

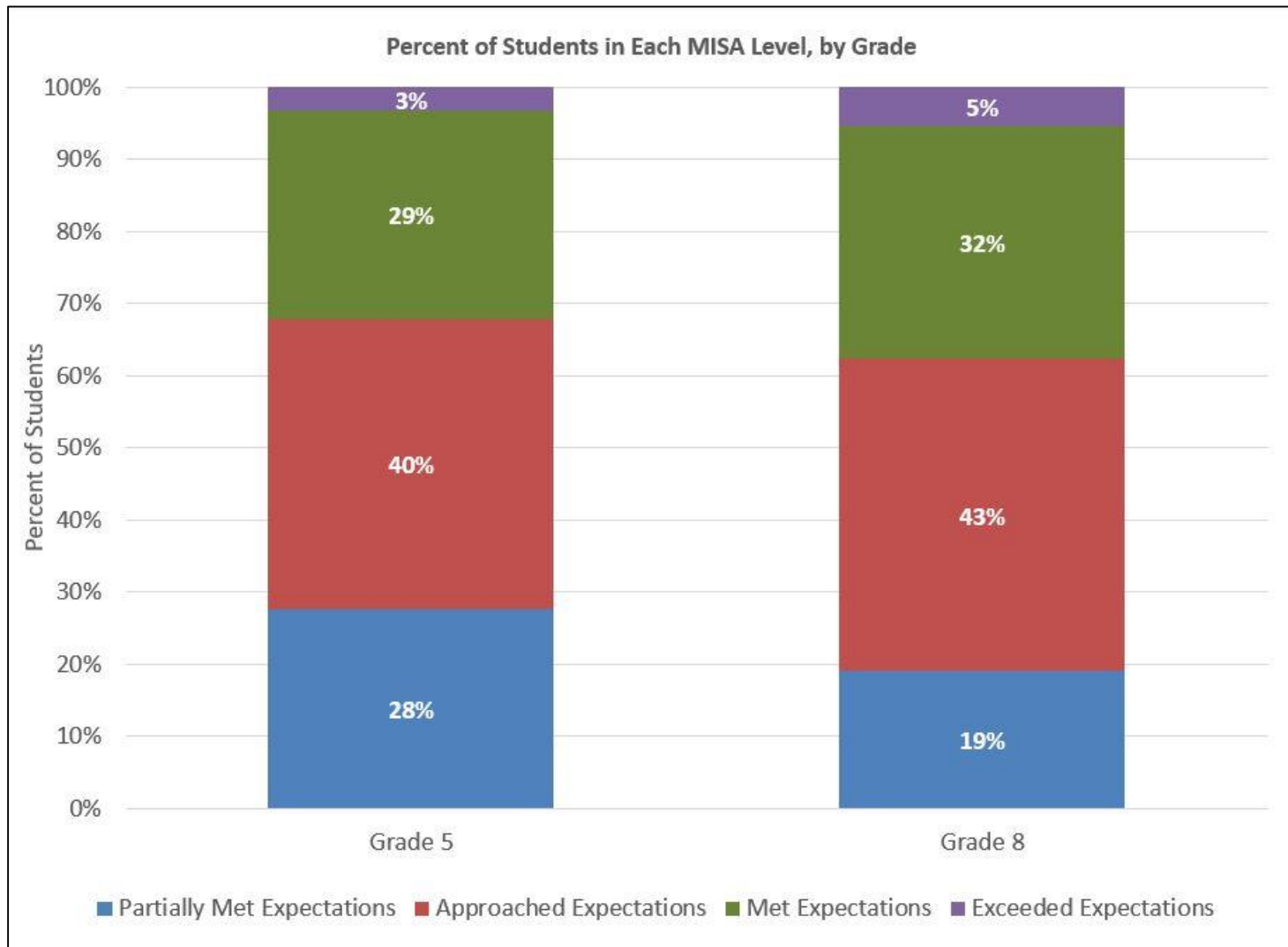
- There were three rounds of judgments.
- After each round, the panelists were provided data:
  - Rounds 1, 2, and 3 – Each panelist’s recommended cut scores (at the table), the table’s median cut scores, and the room’s median cut scores.
  - Round 1 – Item statistics (i.e., item difficulty and distribution of students at each score point)
  - Rounds 2 and 3 – Impact data (i.e., the percentage of students who would be in each performance level)
- Panelists discussed the data and their judgments after each round, before beginning the next round.
- Round 3 concluded the grade-level judgments.

# Vertical Articulation

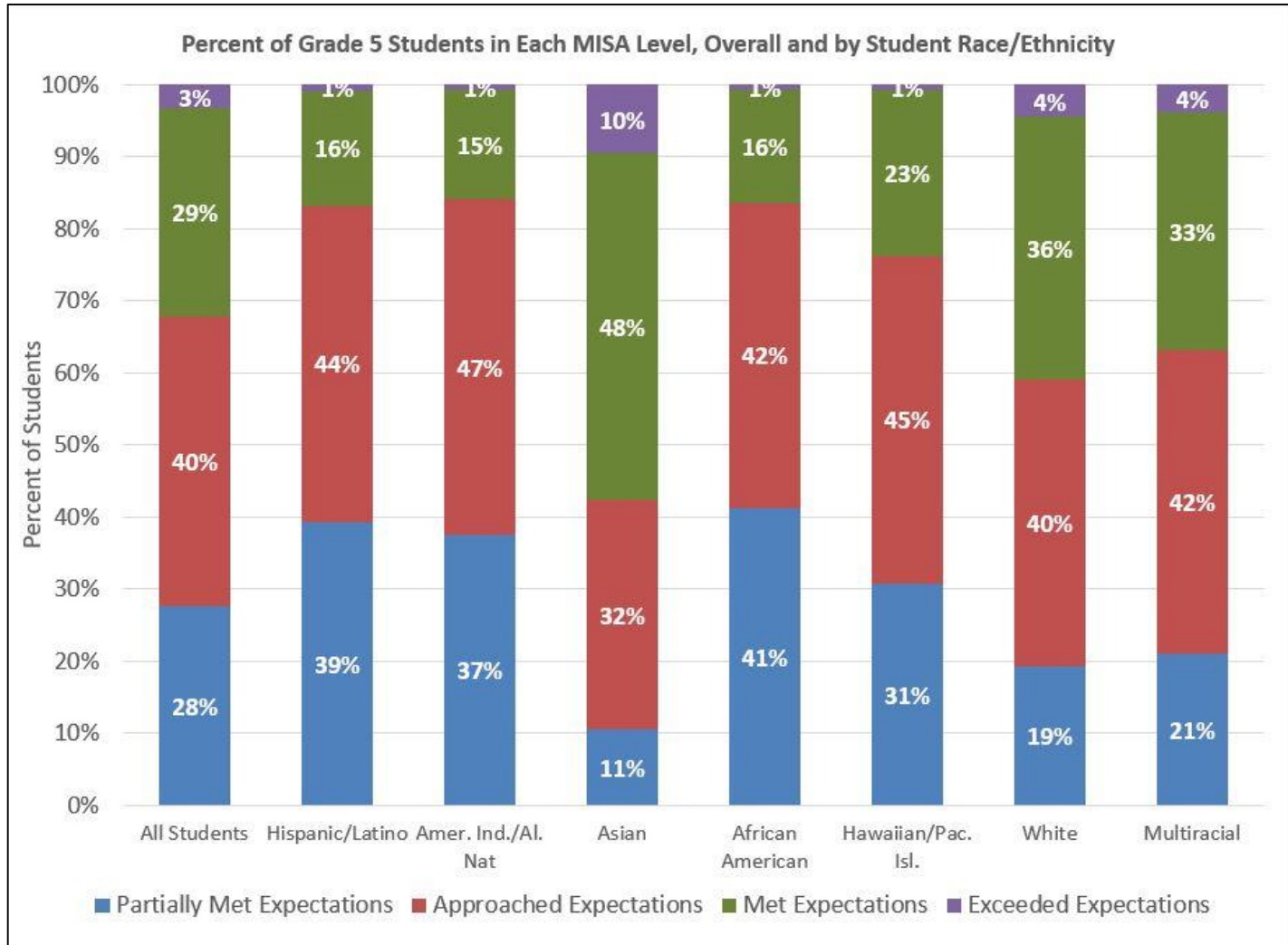
After all the judgments were completed for grade 5 and grade 8, all panelists reconvened and looked at the impact data for both grades together. Panelists considered the appropriateness of the cut scores, using the questions below to prompt the discussion.

From the perspectives of students, parents, educators, and the general public, do the results make sense?  
If not, how should the cut scores be adjusted?

# Final Impact Data

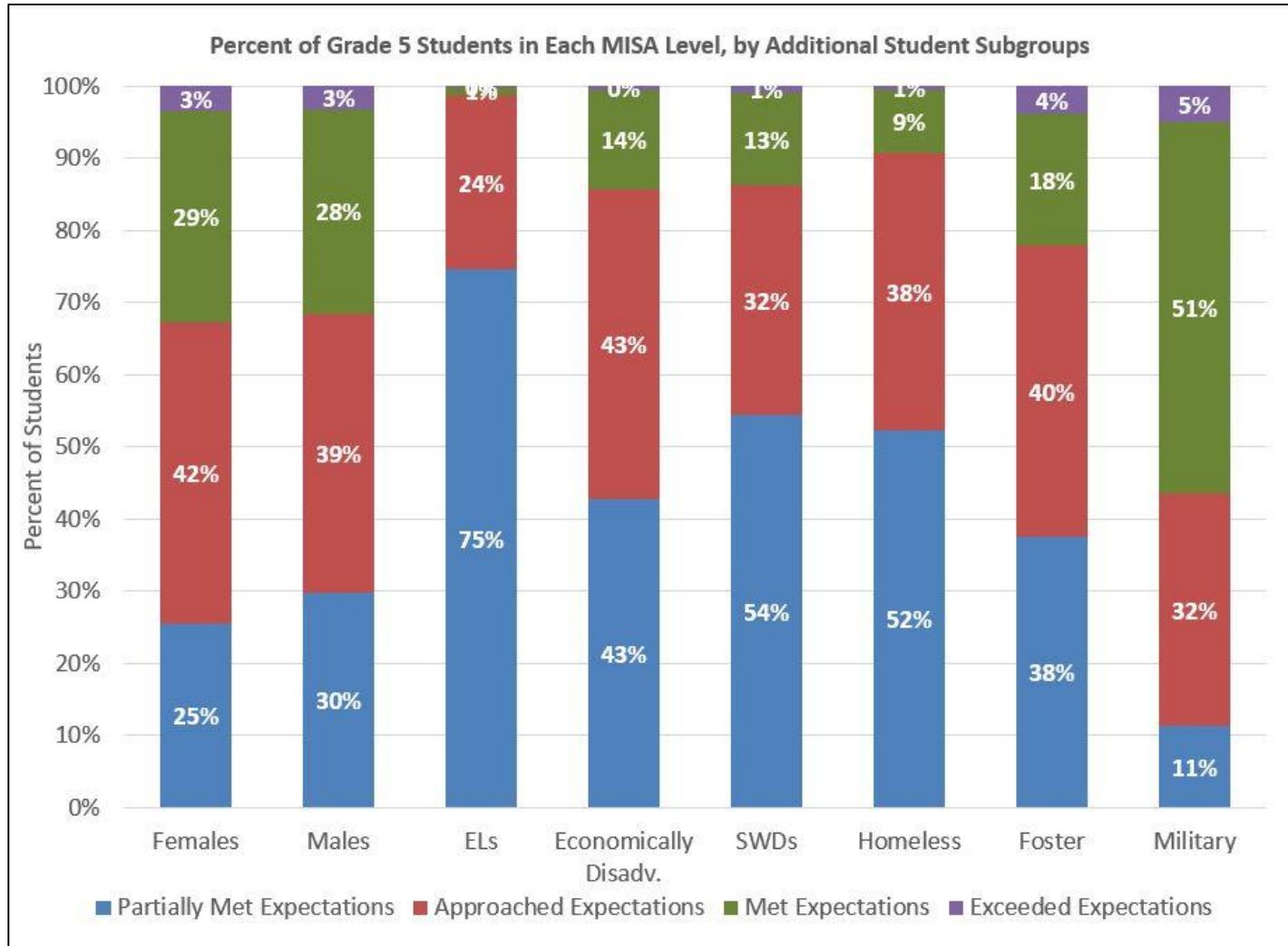


# Final Impact Data – Grade 5 Subgroups (1)

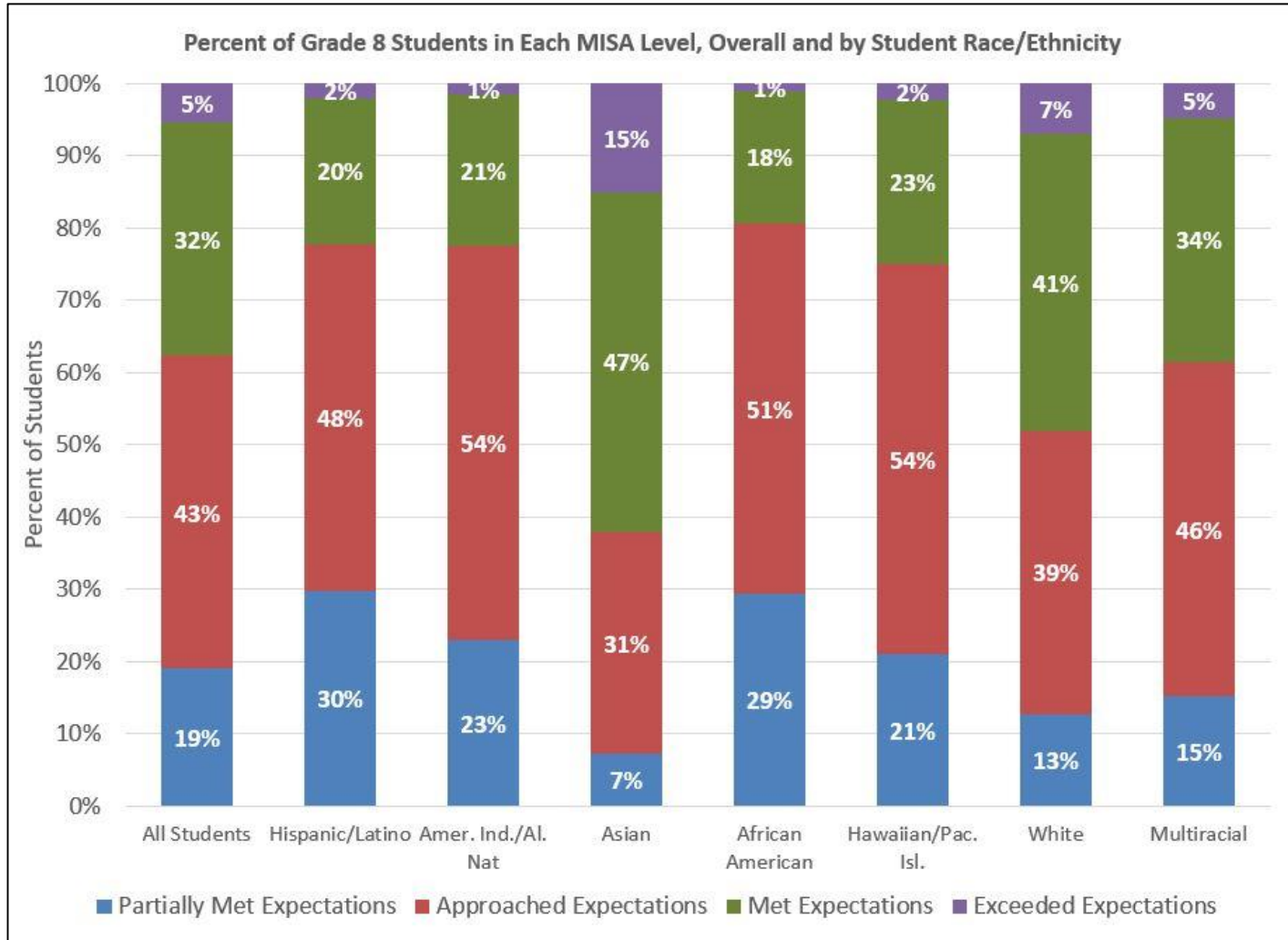




## Final Impact Data – Grade 5 Subgroups (2)



# Final Impact Data – Grade 8 Subgroups (1)



## Final Impact Data – Grade 8 Subgroups (2)

