MARYLAND STATE DEPARTMENT OF EDUCATION
EQUITY AND EXCELLENCE

WestEd
WestEd.org

# Using Data to Inform Standard-Setting Recommendations

## Maryland Integrated Science Assessment, Grades 5 & 8

Jennifer Judkins, MSDE | Matt Gaertner, WestEd

PART 1

# Background

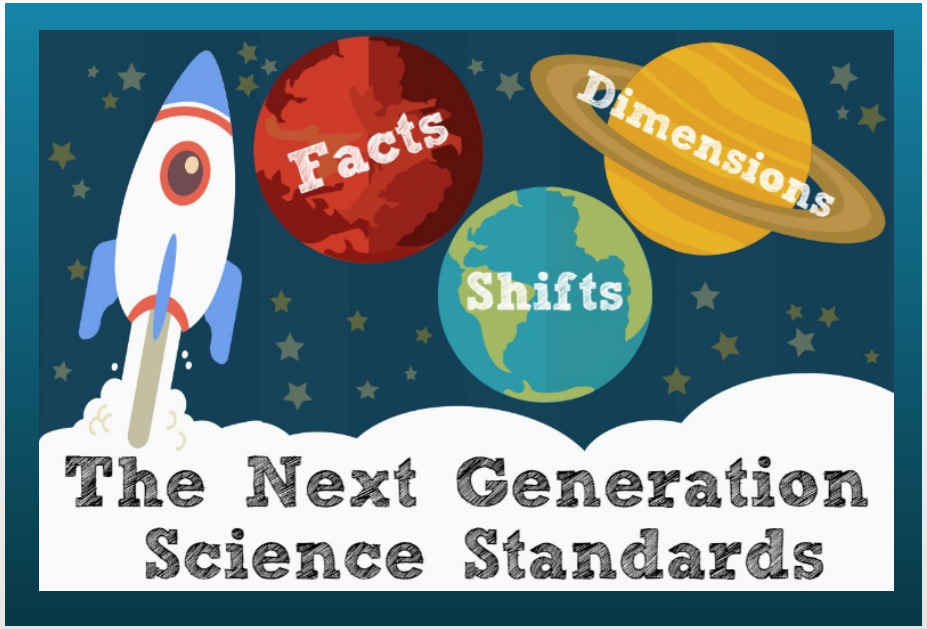
# Background

## 2013

The Maryland State Board of Education adopts the Next Generation Science Standards

## 2016-2017

The Maryland Integrated Science Assessment (MISA) replaces the Maryland School Assessment in science



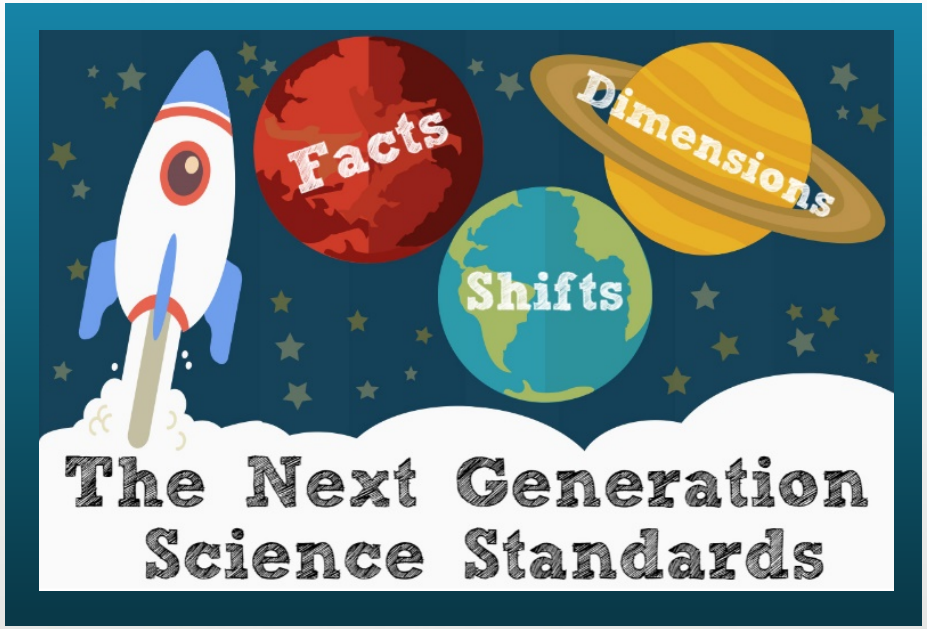
# Background

## Spring 2018

The MISA is administered for the first time operationally

## Summer 2018

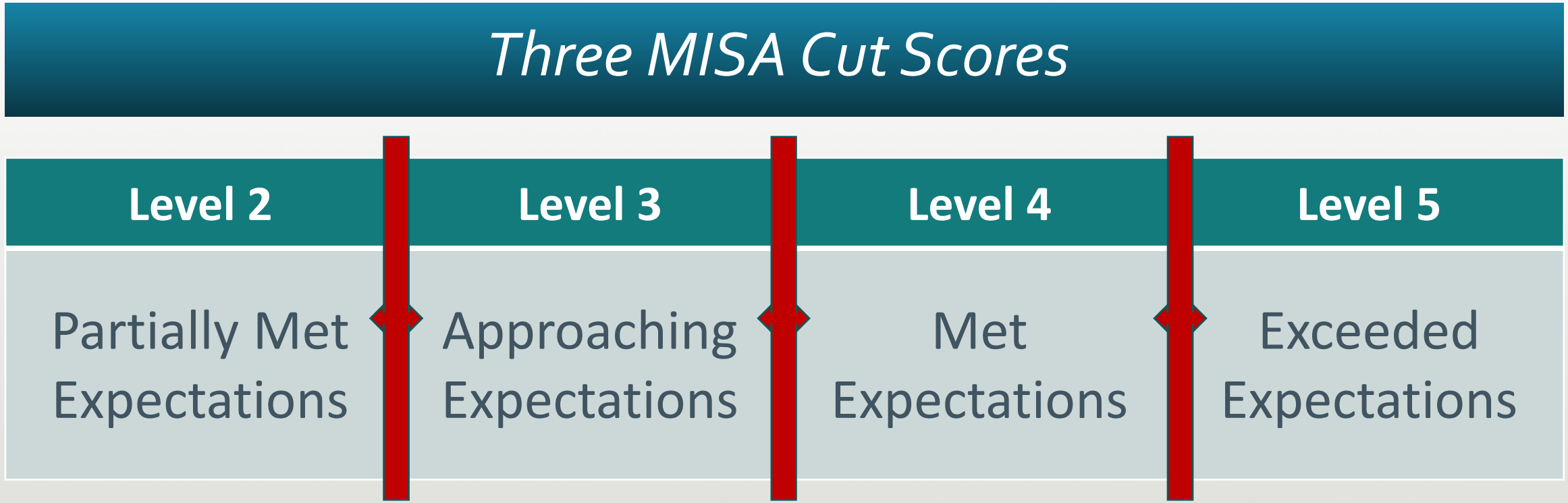
WestEd and MSDE conduct standard-setting meetings for grade 5 and grade 8 assessments



# Background

| *Four MISA Performance Levels* | | | |
|---|---|---|---|
| **Level 2** | **Level 3** | **Level 4** | **Level 5** |
| Partially Met Expectations | Approaching Expectations | Met Expectations | Exceeded Expectations |

# Background


Three MISA Cut Scores
Level 2
Level 3
Level 4
Level 5
Partially Met Expectations
Approaching Expectations
Met Expectations
Exceeded Expectations

PART 2

# Meeting Preparation

# Meeting Preparation

## Key Considerations for Using Data During Standard Setting

- **What** data should be shared with panelists?
- **When** should those data be shared?
- **How** should those data be shared?

*Sharing data with panelists: What, when, and how?*

# Meeting Preparation

## What data?

- When convergence is the goal, panelists need to understand how their judgments compare to their colleagues' judgments
- Panelists have a right to know the impact of their recommendations
- Panelists are science educators, but NGSS assessments are new, so student performance is difficult to forecast

*Sharing data with panelists: What, when, and how?*

# Meeting Preparation

## When?

- Standard setting for criterion-referenced tests should begin with content – students' knowledge and skills relative to the standards
- Panelists judgments converge incrementally through deliberation – informed but not dictated by data
- Impact data (e.g., projected proficiency rates) should be a reality check, not a starting point

*Sharing data with panelists: What, when, and how?*

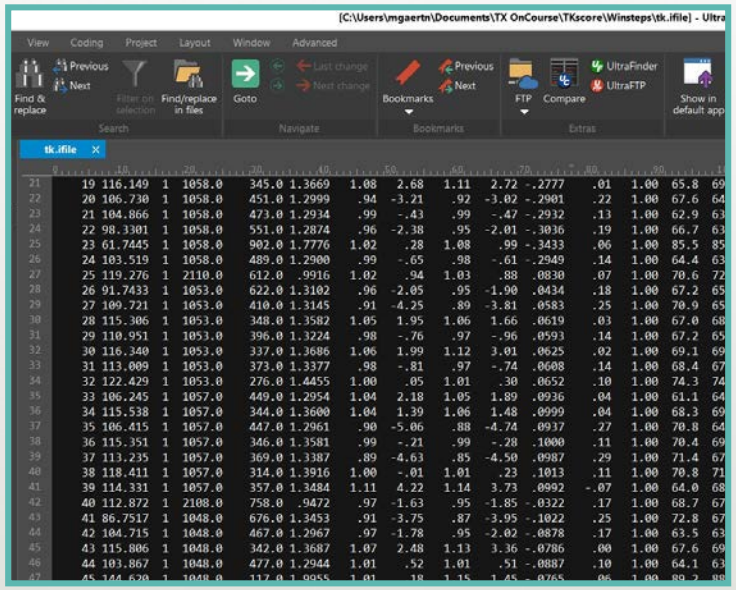
# Meeting Preparation

## How? (… and how much?)

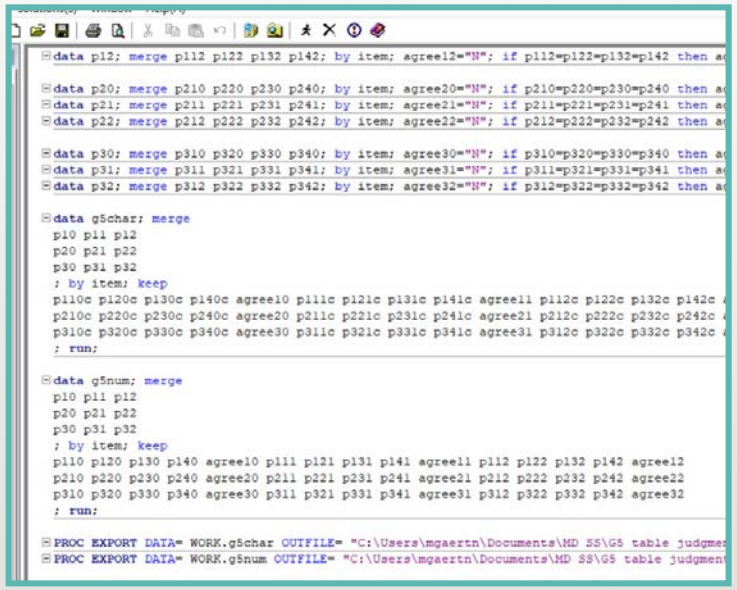
- Without explanation and context, summary statistics (e.g., p-values) are ambiguous at best
- Standard-setting data (just like items on a test) should produce the intended interpretations
- With new standards, new assessments, and three cut scores, panelists have a lot to consider. Data overload is a real concern.

# Meeting Preparation

## 1. Student Data




## 2. A Bit of Code



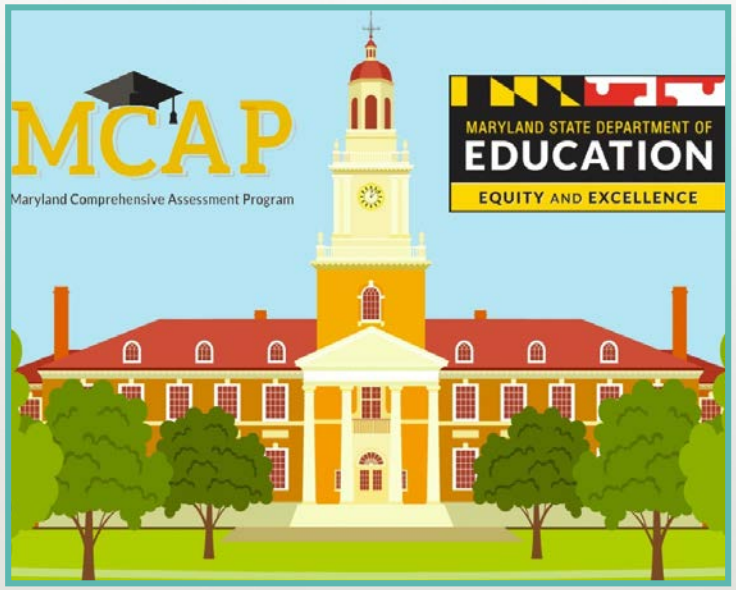
```
data p12; merge p112 p122 p132 p142; by item; agree12="N"; if p112=p122=p132=p142 then a
data p20; merge p210 p220 p230 p240; by item; agree20="N"; if p210=p220=p230=p240 then a
data p21; merge p211 p221 p231 p241; by item; agree21="N"; if p211=p221=p231=p241 then a
data p22; merge p212 p222 p232 p242; by item; agree22="N"; if p212=p222=p232=p242 then a
data p30; merge p310 p320 p330 p340; by item; agree30="N"; if p310=p320=p330=p340 then a
data p31; merge p311 p321 p331 p341; by item; agree31="N"; if p311=p321=p331=p341 then a
data p32; merge p312 p322 p332 p342; by item; agree32="N"; if p312=p322=p332=p342 then a

data g5char; merge
 p10 p11 p12
 p20 p21 p22
 p30 p31 p32
 ; by item; keep
 p110c p120c p130c p140c agree10 p111c p121c p131c p141c agree11 p112c p122c p132c p142c
 p210c p220c p230c p240c agree20 p211c p221c p231c p241c agree21 p212c p222c p232c p242c
 p310c p320c p330c p340c agree30 p311c p321c p331c p341c agree31 p312c p322c p332c p342c
 ; run;

data g5num; merge
 p10 p11 p12
 p20 p21 p22
 p30 p31 p32
 ; by item; keep
 p110 p120 p130 p140 agree10 p111 p121 p131 p141 agree11 p112 p122 p132 p142 agree12
 p210 p220 p230 p240 agree20 p211 p221 p231 p241 agree21 p212 p222 p232 p242 agree22
 p310 p320 p330 p340 agree30 p311 p321 p331 p341 agree31 p312 p322 p332 p342 agree32
 ; run;

PROC EXPORT DATA= WORK.g5char OUTFILE= "C:\Users\mgaertn\Documents\MD SS\G5 table judgmen
PROC EXPORT DATA= WORK.g5num OUTFILE= "C:\Users\mgaertn\Documents\MD SS\G5 table judgmen
```

## 3. Item/Pra Software



PART 3

# Standard Setting Meeting

# Standard Setting Meeting

## *Three Rounds of Judgment*

### Round 1 Data

Judgments from your table and your grade

### Round 2 Data

Judgments + item p-values

### Round 3 Data

Judgments + p-values + impact data (overall and by demographic groups)

# Standard Setting Meeting

## *Vertical Articulation*

After all the judgments were complete, grade 5 and grade 8 panelists reconvened and reviewed the cut scores together, with two fundamental questions in mind…

# Standard Setting Meeting

## *Vertical Articulation*

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

# Standard Setting Meeting

## *Vertical Articulation*

The full group considered three relevant benchmarks in addition to MISA impact data:

- **Impact data from similar states with similar NGSS-aligned assessments**
- **NAEP Science impact data**
- **PARCC Mathematics and English / Language Arts impact data**

# Thank you!


## Jennifer Judkins

*Assistant State Superintendent*

*Division of Assessment, Accountability, and Information Technology*

*Maryland State Department of Education*

## Matthew Gaertner

*Director of Research*

*WestEd*