

# Benchmark Reports

## Minnesota Department of Education

### Science

Jim Wood

[Jim.wood@state.mn.us](mailto:Jim.wood@state.mn.us)

Dawn Cameron

[Dawn.cameron@state.mn.us](mailto:Dawn.cameron@state.mn.us)

### Math

Angela Hochstetter

[Angela.Hochstetter@state.mn.us](mailto:Angela.Hochstetter@state.mn.us)

Ann Page

[ann.page@state.mn.us](mailto:ann.page@state.mn.us)

### Reading

Angie Norburg

[Angela.m.norburg@state.mn.us](mailto:Angela.m.norburg@state.mn.us)

Tony Aarts

[Tony.aarts@state.mn.us](mailto:Tony.aarts@state.mn.us)

Holly Brunson 651-582-8563

[Holly.Brunson@state.mn.us](mailto:Holly.Brunson@state.mn.us)

# New MDE staff!

Holly Brunson

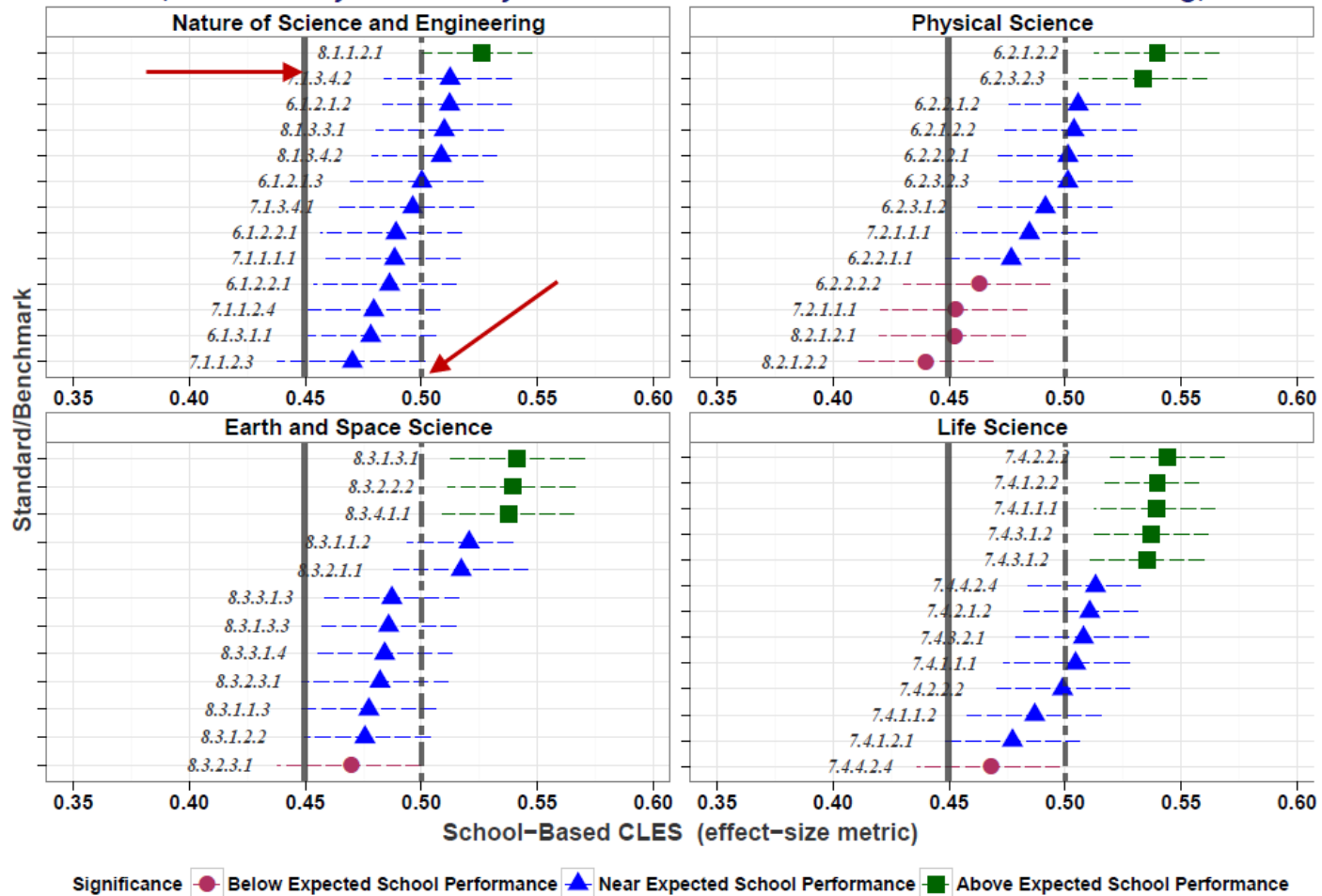
Outreach and Training Specialist

651-582-8563

[Holly.Brunson@state.mn.us](mailto:Holly.Brunson@state.mn.us)

# Example: Science Benchmark Reports

MDE Benchmark Report based on Effect-Size  
 MINNESOTA MIDDLE: G8 2015 MCA-III Science  
 (Created by MDE Psychometrics, Division of Statewide Testing)



# Sample Data Records for Item 1

## Responses at a School: Observed and Expected Counts of Correct Responses

Item ID	Student ID	Item Score	Theta Ability	Probability Correct
1	100007	1	0.50	0.72
1	100015	0	-0.25	0.48
1	123456	1	0.75	0.78
1	987654	1	0.10	0.60
1	654321	0	-1.35	0.10
1	456789	0	0.25	0.64
	<b>SUM:</b>	<b>3</b>		<b>3.32</b>



Observed  
Student  
Responses



Expected  
Student  
Responses

# Interpreting the Benchmark Reports

It is important to frame any interpretation in the context of the school's environment. Experience with the science curriculum, instruction, and data from other classroom assessments is critical to making meaningful inferences from this report.

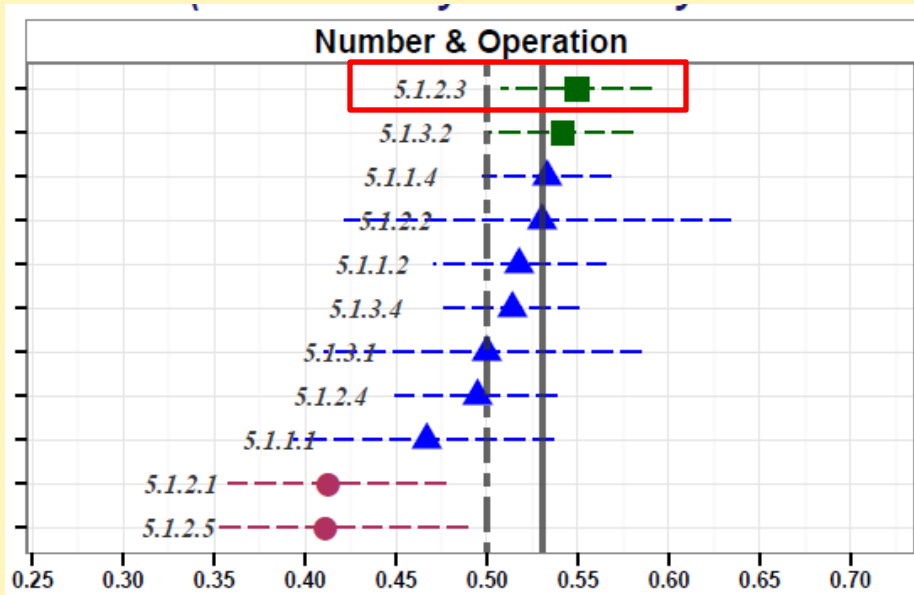
# CAUTION!!!

## What we can say and not say about the data

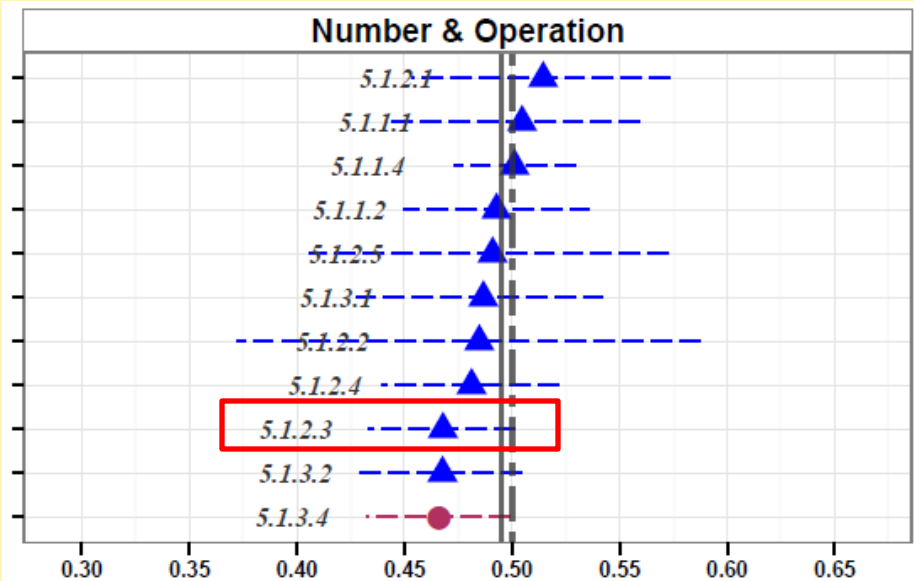
- **All data is relative to the expected school average**
- **Location of items on the scale do not provide information about the difficulty of the individual test questions**
- **Color codes and position of items in the graphs do not correspond to achievement levels**

## Things to watch out for

- **Keep an eye on the CLES metric when comparing schools. The horizontal axis is adjusted to fit each individual school's data.**
- **Format *does not* allow for comparison of schools, but deeper understanding within a school.**
- **Watch the dotted lines – different colors may overlap. In the cases where there is more than 1/2 overlap, the items may be considered statistically equivalent.**



Which school performed “better” on benchmark 5.1.2.3?

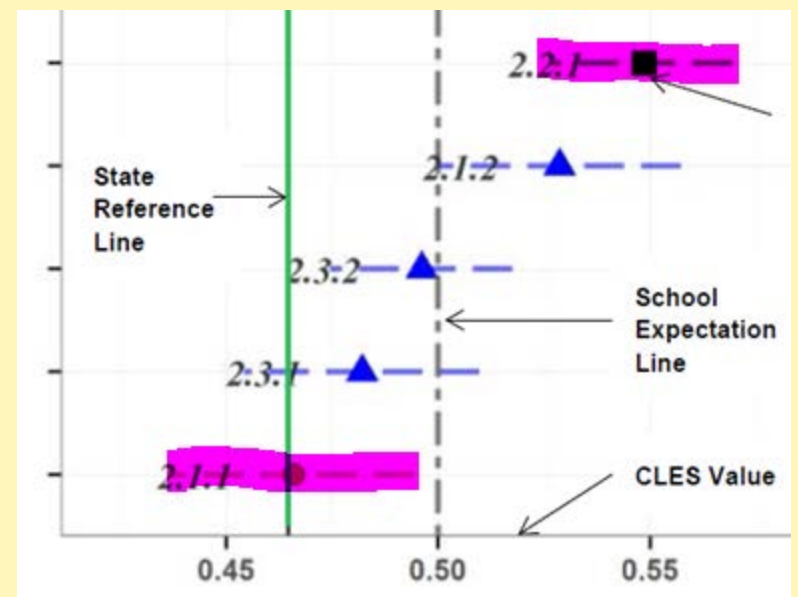
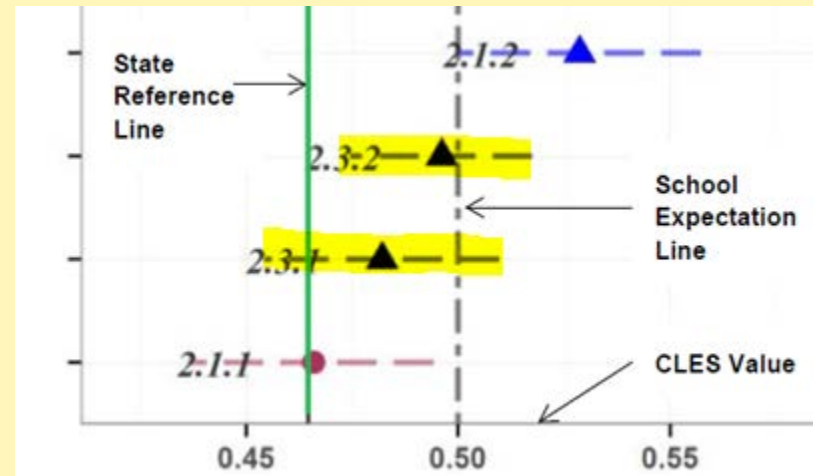


How do you know?

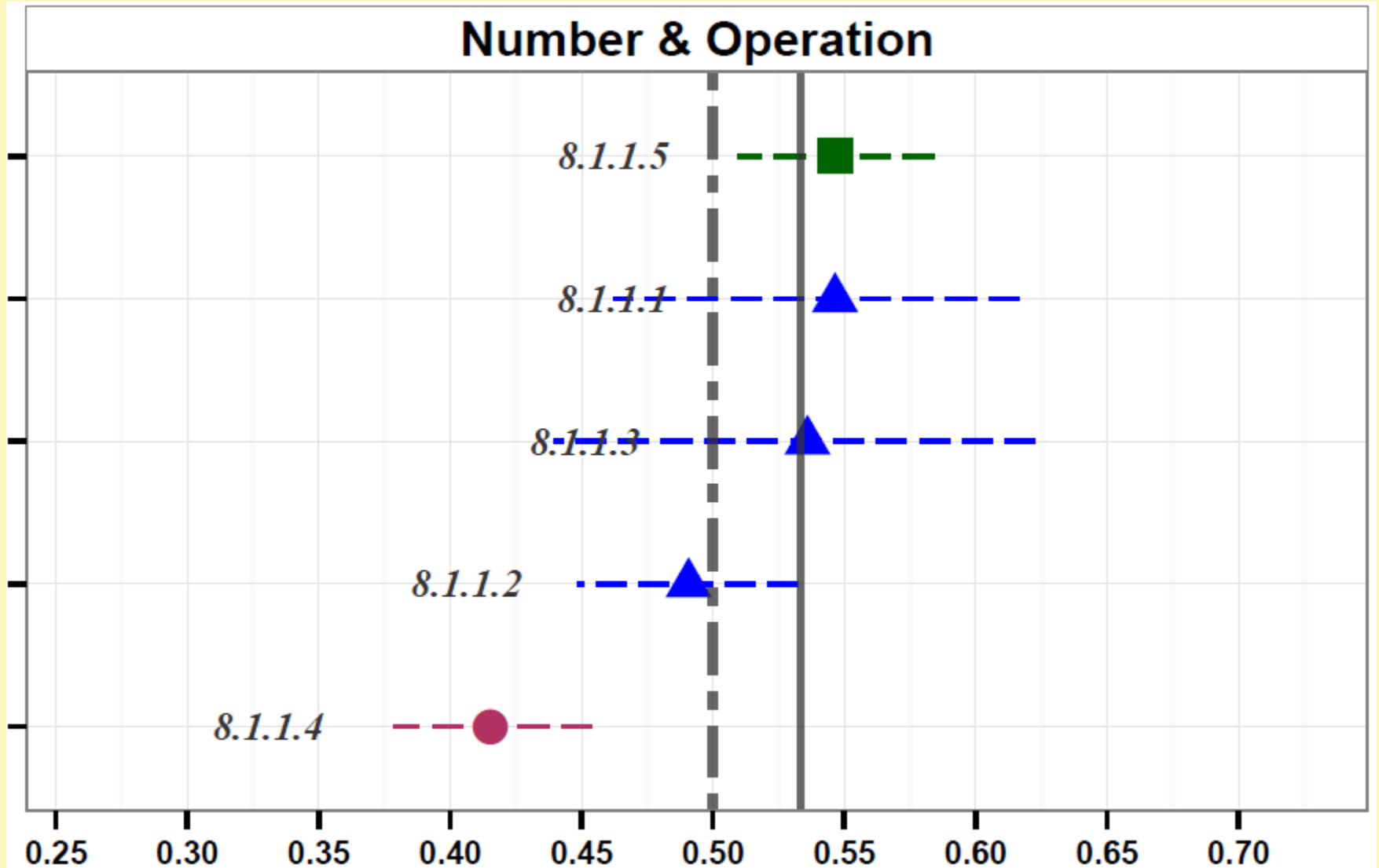


# Overlap of credible bands

- If the credible bands on two different benchmarks have **substantial overlap**, there is **little credible evidence** to suggest that actual performance was significantly different on the two benchmarks
- If the credible band across two benchmarks does **not overlap**, then there is **very clear evidence** of a significant difference in performance between the two benchmarks.



# Circle, Triangle, Square



# What can we conclude about 8.1.1.5?

- Since the observed performance indicator is **green** and the band does not overlap the dashed reference line, it shows performance above expectations
- Possible conclusions:
  - the benchmark may be included in the curriculum
  - students may have a better understanding of a green benchmark than a red benchmark
  - students may have a lot of exposure to the benchmark
  - the benchmark may be taught right before testing

# What can we conclude about 8.1.1.4?

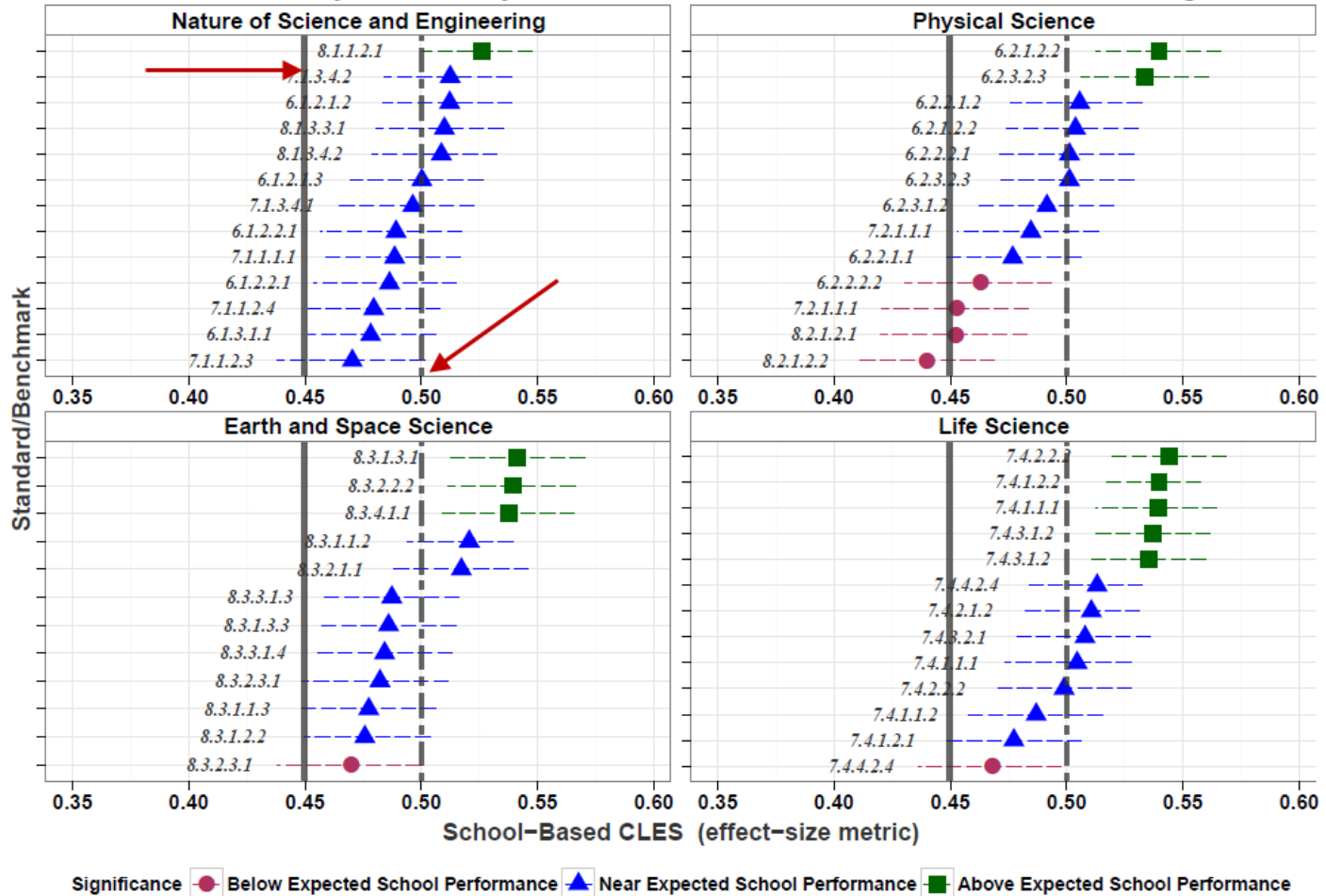
- **Since the observed performance indicator is red and the band does not overlap the dashed reference line, it shows performance below expectations**
- **Possible conclusions:**
  - the benchmark might not be included in the curriculum
  - students do not have as strong an understanding of this benchmark as a green benchmark
  - students might have less exposure to the benchmark
  - the benchmark might be taught after testing
  - the benchmark may be a good starting point for strengthening instruction and curriculum the next year

# Differences between Benchmark Reports by Content Area

	Content		
	Science	Math	Reading
Test type	Fixed form	Adaptive form	Adaptive form
Tested grades	Grades 5, 8 and high school	Grades 3-8 and grade 11	Grades 3-8 and grade 10
What the data describes	Each item on the test	Each benchmark given to a school or district	Each substrand

# Science Benchmark Reports

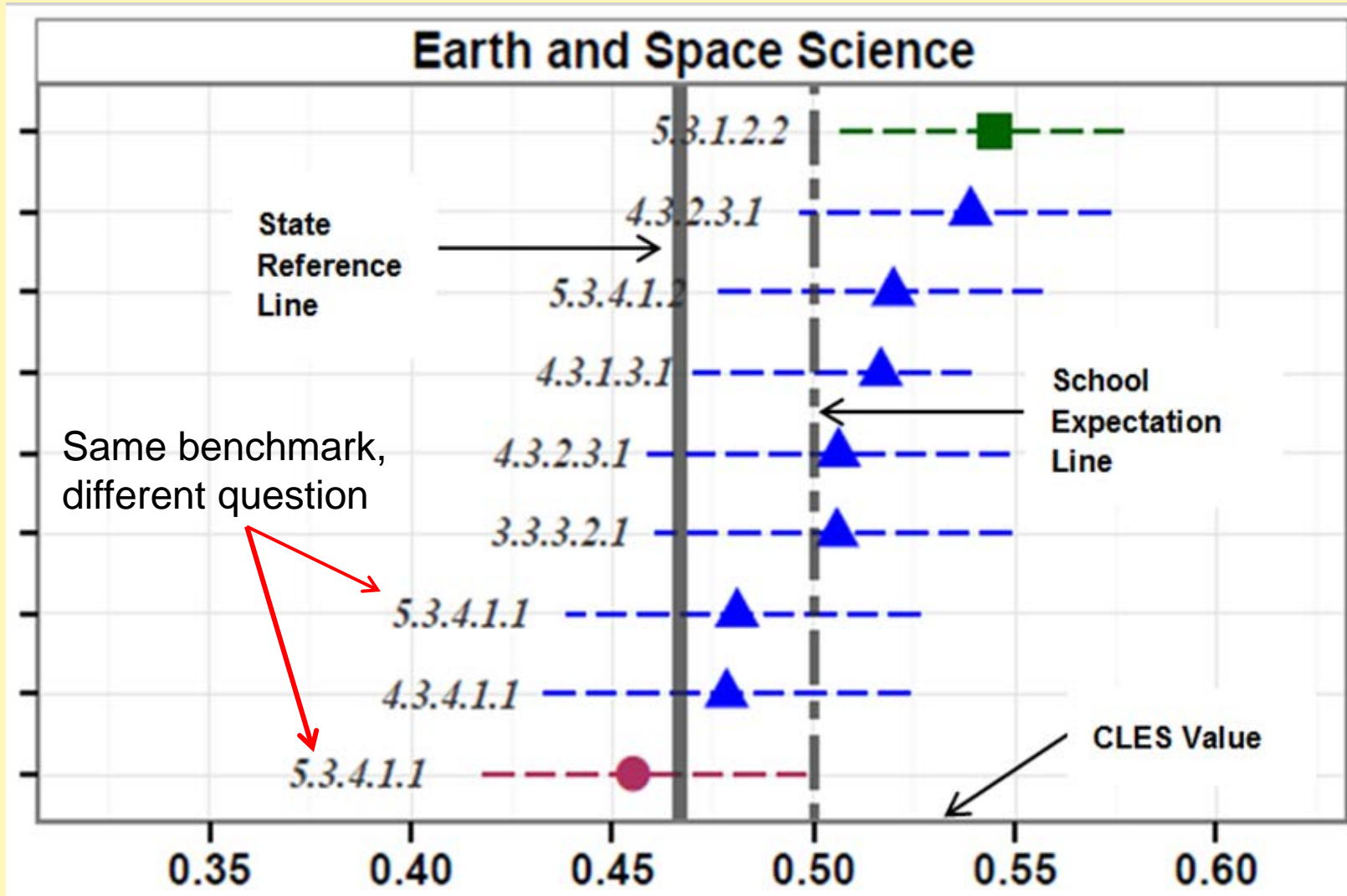
MDE Benchmark Report based on Effect-Size  
 MINNESOTA MIDDLE: G8 2015 MCA-III Science  
 (Created by MDE Psychometrics, Division of Statewide Testing)



# Science Benchmark Reports

- **Science MCA-III is a fixed form test, so every data point represents an item.**
- **There may be more than one item assessing a particular benchmark**
- **The number of items on each report corresponds to the number of items on the assessment for each grade**

# Science Benchmark Reports

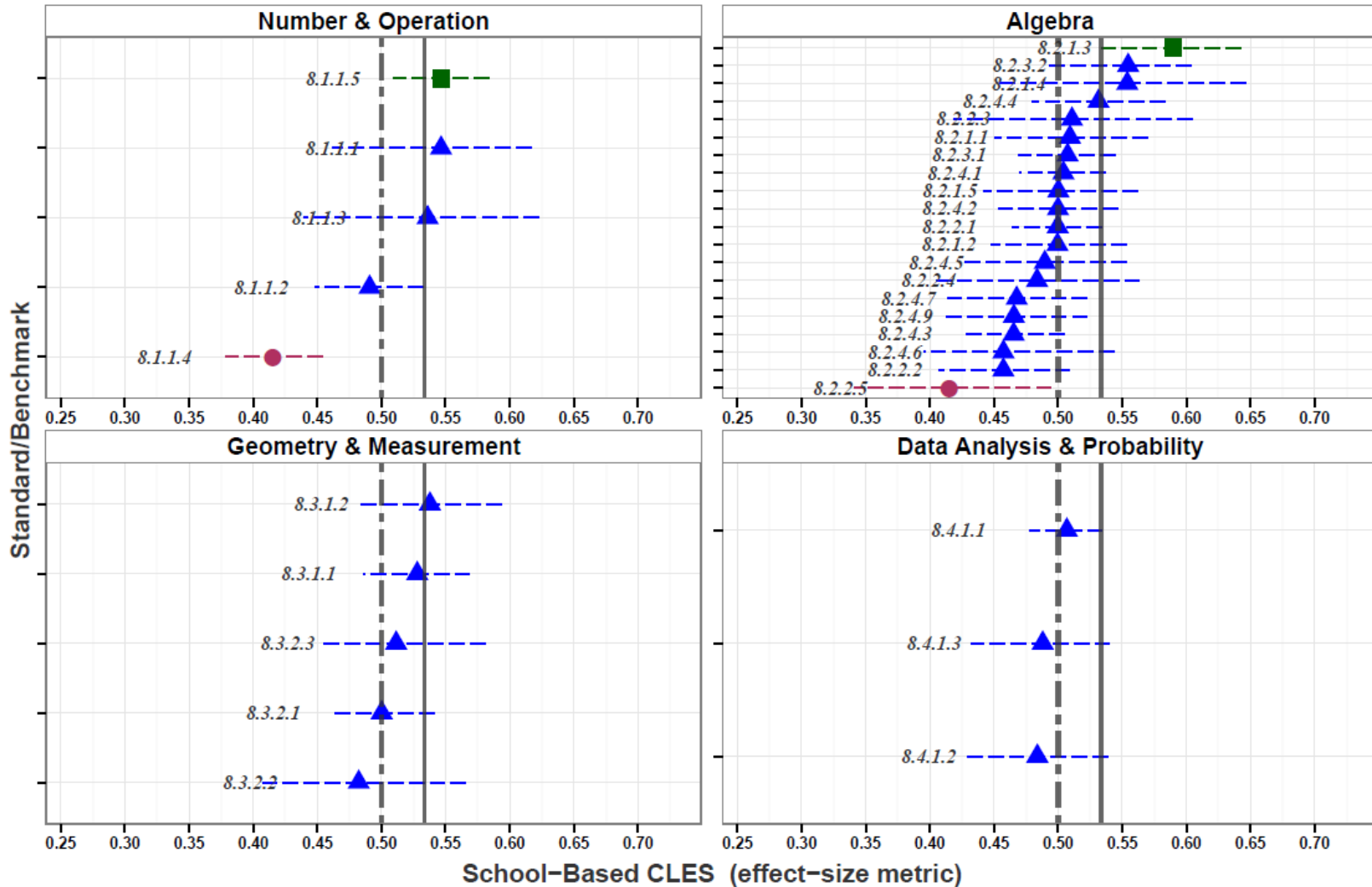




# Math Benchmark Reports

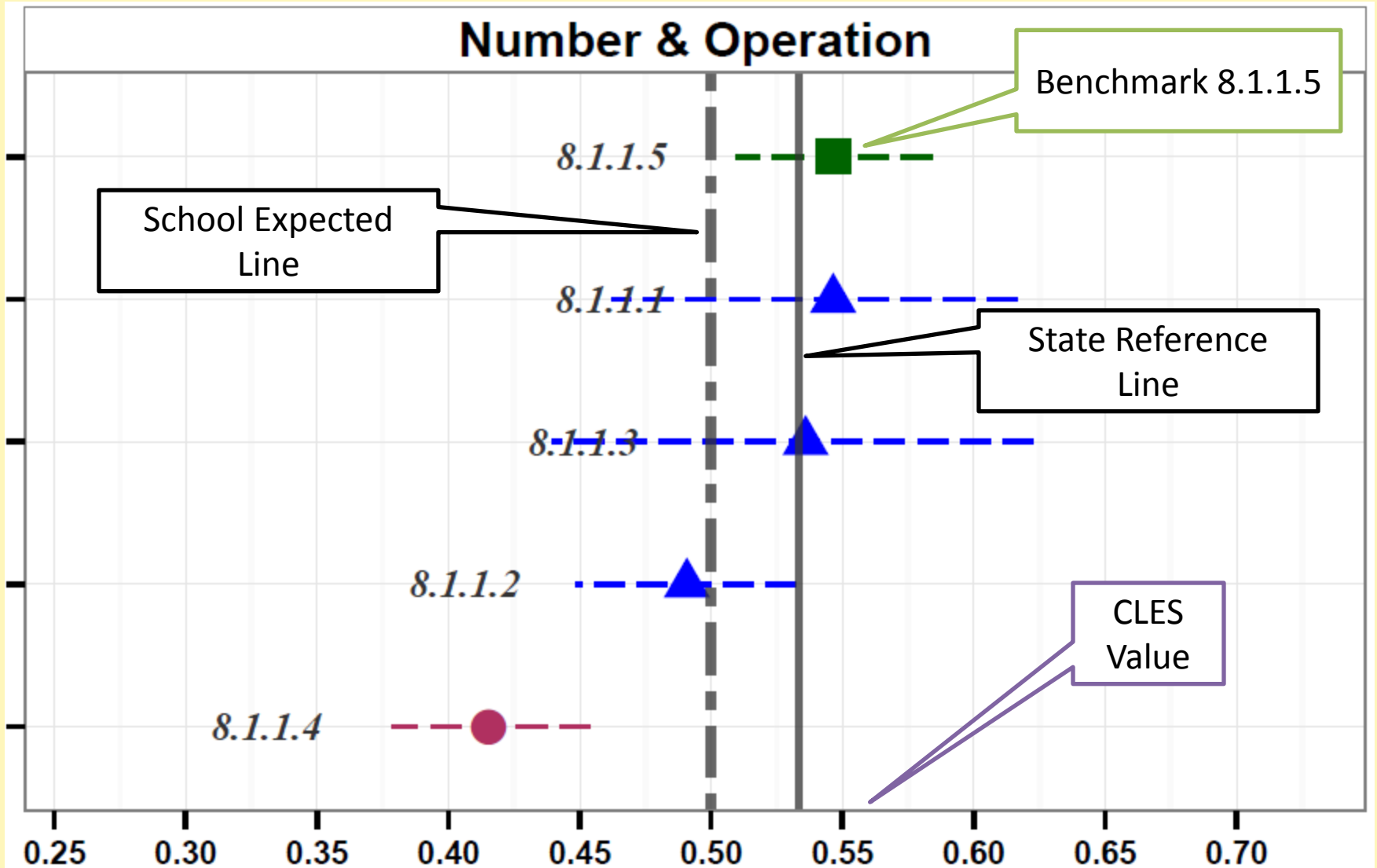
MDE Benchmark Report based on Effect-Size  
 [Redacted] G8 2015 MCA-III Math

(Created by MDE Psychometrics, Division of Statewide Testing)



Significance ● Below Expected School Performance ▲ Near Expected School Performance ■ Above Expected School Performance

# Math Benchmark Reports

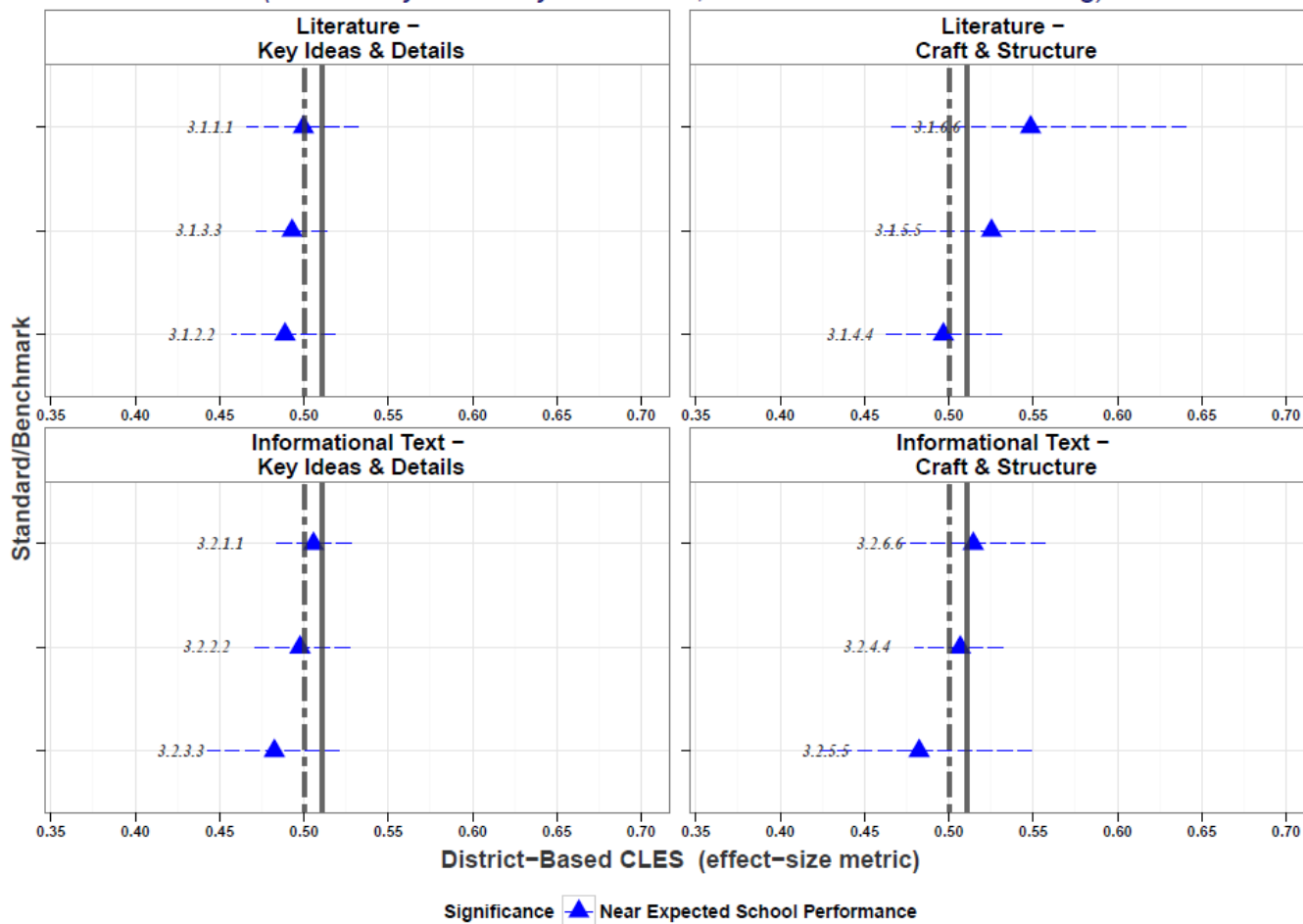


# Math Benchmark Reports

- **Math MCA-III is an adaptive test, so every data point represents a benchmark.**
- **Not every student sees every benchmark.**
- **There may be more than one item assessing a particular benchmark for a student.**
- **Credible bandwidths vary more than science because of these factors.**

# Reading Benchmark Reports

MDE Benchmark Report based on Effect-Size  
 [Redacted] G3 2015 MCA-III Reading  
 (Created by MDE Psychometrics, Division of Statewide Testing)



# Reading Benchmark Reports

What can we conclude about the Reading benchmark reports?

- Reading MCA is now adaptive. This means each data point for a benchmark represents a group of items associated to that benchmark. Therefore, a comparison of 2015 performance to 2016 performance is not feasible.
- Make a comparison of standards within a skill domain from one substrand to the other (Literature and Informational Text).
- If observed performance indicators are coded **blue**, the school is performing as expected. To investigate further, it would be helpful to study other common assessments at the school.

# Digging into the Reports

## Use the Test Specifications

- Account for Test Design
- Know the Content Specifications

## Look for patterns in data

- Do certain standards show a pattern within the year and over time?

## Start asking questions

- Classroom and District Level Assessments
  - Does the data match what is happening in the classroom?

# How do I get my school's benchmark reports?

- **Available on PearsonAccessNext**
- **Contact your District Assessment Coordinator to download the reports**
- **User guides for each assessment are also available**

# Get involved!

**Assessment Advisory Panels—ongoing**

**Register to be invited to panels at**

**<http://education.state.mn.us>**

**→ Educator Excellence**

**→ Testing Resources**

**→ Register for Advisory panels**