

# VIRGINIA'S PROPOSED STANDARD SETTING PLAN

Virginia Board of Education  
Work Session/Business Meeting  
February 26, 2025



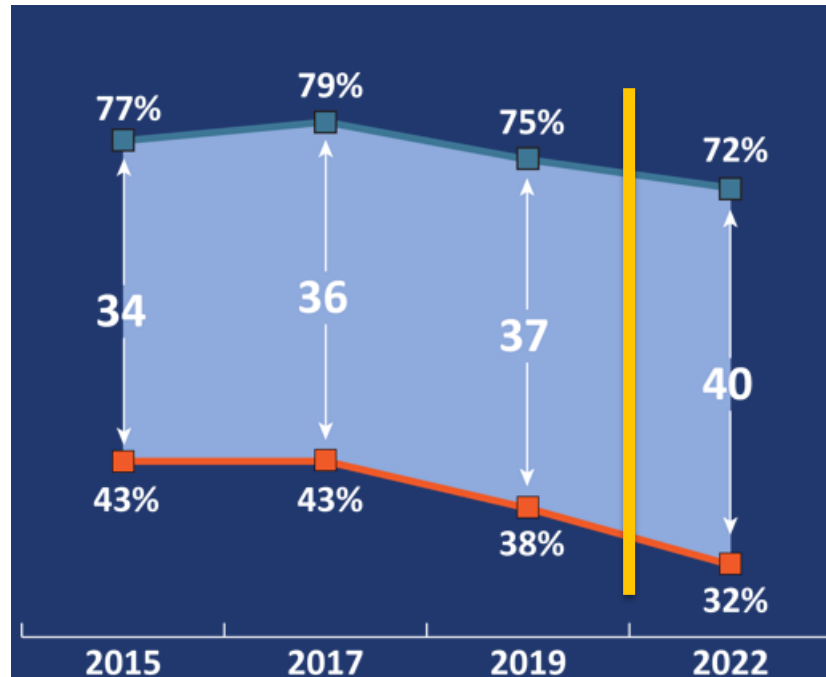
**VIRGINIA DEPARTMENT OF EDUCATION**

# GOVERNOR'S COMMITMENT

Governor Youngkin announced that Virginia's lowest-in-the-nation mastery standards would be raised "to align with the demands of the knowledge economy" as part of the [\*Our Commitment to Virginians\* event](#) in May 2022. Virginia Mastery will be benchmarked against employer/higher education knowledge and skill demands, NAEP proficiency, and the highest state definitions across the nation.

The Governor has committed to restoring excellence to education which includes more rigorous standards in the four core subjects, greater transparency and accountability, and a new assessment system. The evidence-based process requires standards to be raised first, assessment items to be aligned to content and rigor, and mastery setting based on the results of the new content and questions in the assessment.

# THE HONESTY GAP: VIRGINIA GRADE 4 NAEP PERFORMANCE 2017-2022



Grade-4 Reading: Percent Proficient and Above



Grade-4 Math: Percent Proficient and Above

■ Virginia SOL    
 ■ Virginia NAEP    
  Honesty Gap    
  Change in SOL Cut Scores

# NAGB PROCESS

The National Center for Education Statistics (NCES) has periodically published reports using results from the National Assessment of Educational Progress (NAEP) to compare the proficiency standards states and jurisdictions set for their students. The latest report highlights the results of mapping state proficiency standards onto the NAEP scales using state/jurisdiction assessment results from the 2021–22 school year and the 2022 NAEP assessments for public schools.

## DATA TOOL: Mapping State Proficiency Standards Onto NAEP Scales, 2007–22

[Back to Overview](#)

### Selected State Mapping Results: Virginia (VA)

193

Grade 4 Grade 8

Mathematics Reading

GRADE 4 | READING



For 2022 grade 4 reading, the NAEP scale equivalent score for Virginia was 193. Virginia had standards mapped at the *Below NAEP Basic level*.

### Selected State Mapping Results: Virginia (VA)

238

Grade 4 Grade 8

Mathematics Reading

GRADE 8 | READING



For 2022 grade 8 reading, the NAEP scale equivalent score for Virginia was 238. Virginia had standards mapped at the *Below NAEP Basic level*.

### Selected State Mapping Results: Virginia (VA)

224

Grade 4 Grade 8

Mathematics Reading

GRADE 4 | MATHEMATICS



For 2022 grade 4 mathematics, the NAEP scale equivalent score for Virginia was 224. Virginia had standards mapped at the *NAEP Basic level*.

# HIGH LEVEL TIMELINE FOR BEST-IN-NATION EXPECTATIONS

<b>Date</b>	<b>Action</b>
October 2018	Board approved Science Standards of Learning (Test: 2023)
August 2023	Board approved Math Standards of Learning (Test: 2025)
March 2024	Board approved English Standards of Learning (Test: 2025)
Fall 2024	Decision to Ensure Evidence-Based Process
February 2025	Board Standard Setting Process Decision
Spring 2025	Assessment Reports Include Current Mastery Levels
Spring 2025	Standard Setting Occurs
June/July 2025	Board First and Final Review of New Mastery Levels

# BOARD PROPOSAL FOR MASTERY LEVEL STANDARD SETTING REVIEW COMMITTEE PHASES

Phase	Purpose	Timeline
Phase 1: Committee Selection	VDOE selects about 200 committee members of expert educators and community members to serve on about 20 committees.	March
Phase 2: Committee Training	All selected members will complete a series of trainings to ensure committee meetings produce valid and reliable recommendations.	April
Phase 3: Committee Convenings	Members review assessment items and data to make cut score recommendations.	May
Phase 4: Board and Stakeholder Feedback	Cut score recommendations receive feedback by the Board and public comment before final Board approval	June and July



# Setting Performance Expectations for the Virginia Standards of Learning (SOL) Assessments

**Chris Domaleski, Ph.D.**  
**Associate Director**  
**National Center for the Improvement of Educational Assessment**

# The Center for Assessment

---

- A NH-based non-profit technical and policy consulting organization established in 1998 with the mission of improving student learning through improved assessment and accountability practices
- Currently working with 35+ states, districts, and other entities to provide technical and design support for a range of assessment and accountability issues
- Non-partisan and independent of any governmental agency or testing company
- I currently serve as a member of Virginia's Technical Advisory Committee (TAC). The full TAC has reviewed this presentation.

# What is standard setting?

- A process used to determine the minimum performance required for each performance level. Minimum performance is operationalized as a “cut score.”
- Standard setting includes judgment from subject matter experts who know the test-taking population and assessed content.
- The credibility of cut scores is closely related to the process. It is important to design and implement an established process with a strong research base.

Virginia Performance Levels				
Grades 3-8 Mathematics and Reading SOL tests	Pass/Advanced	Pass/Proficient	Fail/Basic	Fail/Below Basic
End of Course, History & Social Science, and Science SOL tests	Pass/Advanced	Pass/Proficient	Fail/Does Not Meet	

# What influences the rigor of the performance standards?

---

## Content standards

- State content standards describe 1) what students should be able to do and 2) how well they should be able to do it.

## The assessment

- Assessment items are written to address the content standards. Item type and complexity matter (i.e., more innovative items designed to elicit higher-order thinking can support more rigorous performance standards).
- Form specifications and blueprints are created based on the measurable skills in the content standards. The assessment should reflect the breadth and depth of the standards.

## Policy priorities

- Education leaders may provide guidance to inform decisions about rigor.
- For example, Virginia has prioritized establishing state performance expectations that reflect the rigor associated with the National Assessment of Educational Progress (NAEP)

# Performance Level Descriptors

---

Performance Level Descriptors (PLDs) are an important part of standard setting. They describe knowledge skills and abilities associated with each performance level (i.e., Below Basic, Basic, Proficient, Advanced.)

## Different types of Performance Descriptors

- Policy Descriptors/ Policy Definitions PDs: High-level guidance, not directly linked to content, that describes the performance expected at each level.
- Range PLDs: Describes the spread of knowledge and skills within and across levels
- Threshold (or Target) PLDs: Describes minimum performance expectations (e.g., performance for students who just barely meet expectations). Threshold PLDs are important for standard setting.
- Reporting PLDs: Descriptions of performance often used on public reports for students, teachers, parents, etc.

# Common Standard Setting Methods

## *What is the central task for panelists?*

### Modified Angoff

- Panelists estimate the probability that a proficient student would answer items correctly OR determine if a proficient student is likely to respond correctly (yes/no).

### Bookmark

- Panelists identify the location (bookmark) in an ordered list of items that differentiate between those items a proficient student is or is not likely to answer correctly.

### ID Matching

- Panelists 'match' test items to specific performance levels based on the extent to which the knowledge and skills required to respond correctly are in the corresponding performance level.

# Common Standard Setting Methods

*What are some advantages and disadvantages of each method?*

## Modified Angoff

- **Pros:** Widely used and well-researched
- **Cons:** It can be difficult for panelists to make judgments about item difficulty, individual items can have an outsized influence, some concerns about CAT

## Bookmark

- **Pros:** Widely used and researched, well suited for multiple cuts, considers the collection of items
- **Cons:** It can be difficult to make judgments based on probabilities, requires trustworthy item statistics to create the ordered item book

## ID Matching

- **Pros:** Widely used and researched, panelist 'matching' task is very clear and connected to content
- **Cons:** Requires trustworthy item statistics to create the ordered item book

# Characteristics of Defensible Methods

---

There is not a consensus “best” method for standard setting as long as it is research-based and upholds established professional practices such as:

- The panelists are well-qualified, broadly representative, and follow the procedures with fidelity
- The panelists receive high-quality training
- The process centers judgments on content-based expectations described in the Performance Level Descriptors
- The procedures include efforts to promote coherence
- The methods and rationale are well documented and reviewed by appropriate technical experts and policy leaders

# Transition from Angoff to ID Matching

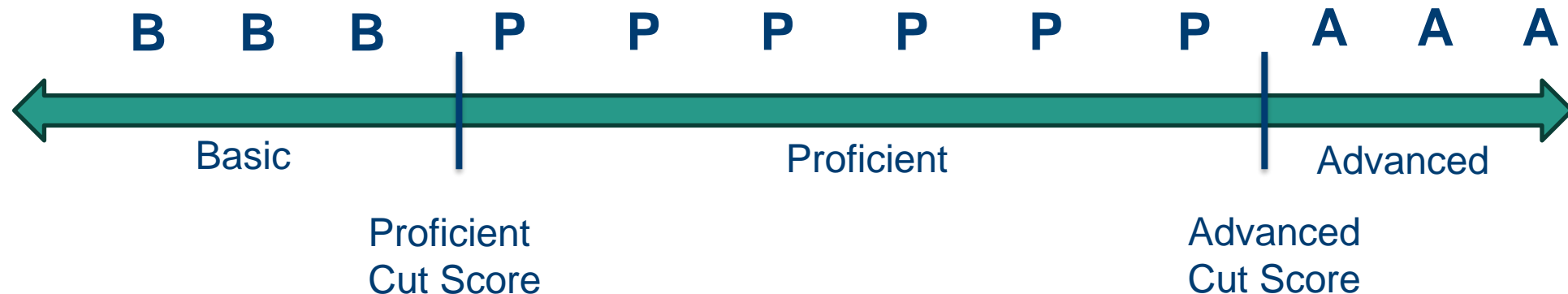
---

- In the past, Virginia has implemented a modified Angoff procedure.
- However, we will transition to ID Matching for the upcoming standard setting.
- This transition is fully supported by the TAC. The rationale includes:
  - It meets all the quality characteristics previously reviewed.
  - This method more prominently centers the content represented on the assessment and Virginia's performance expectations.
  - The panelists' task is very clear and straightforward to implement.
  - ID Matching is growing rapidly in use among state assessment programs, as is the supporting research base.

# Virginia's Proposed Standard Setting Method: ID Matching Procedure

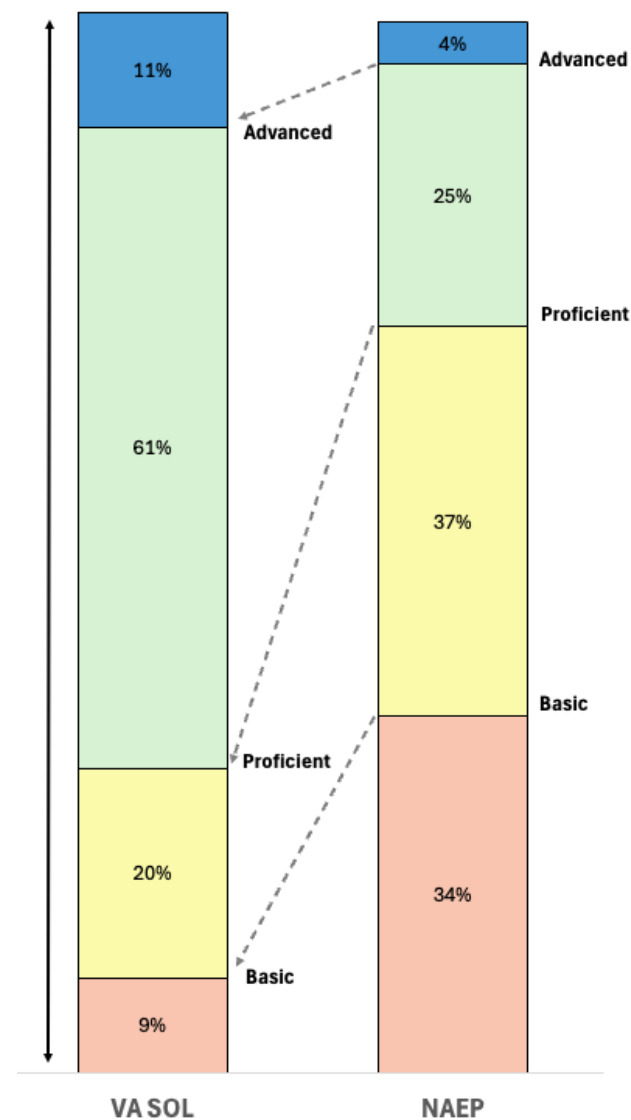
## General procedures

- Review test and PLDs; receive training
- In multiple rounds, participants review an Ordered Item Book (OIB) and assign items to a performance level based on the extent to which the item matches the expectations in the PLDs
- Ratings are refined as participants evaluate and discuss the results and consider impact data
- Cut scores are determined by identifying the threshold region
- Examine results across grades and adjust as needed
- Document process and results, submit for review



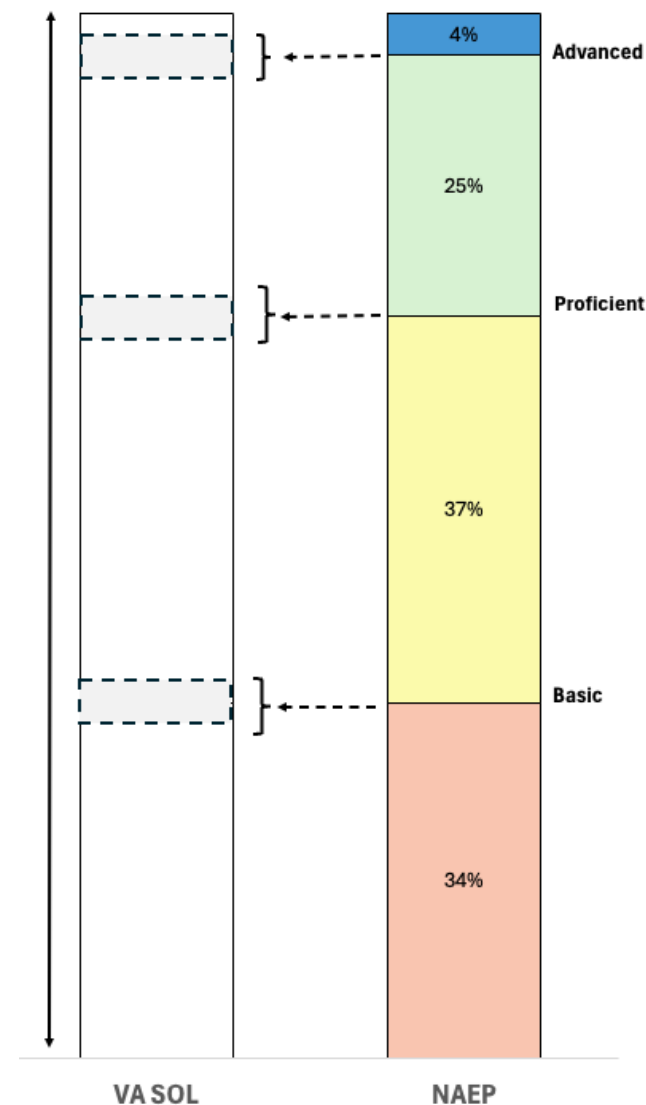
# Supporting Policy Priorities with NAEP Data

- The graph at right compares the percent of students scoring in each performance level for the 2024 Virginia SOL and 2024 NAEP in 8<sup>th</sup> grade reading.
- Clearly, the performance level cut scores are more rigorous on NAEP compared to the current Virginia SOL cut scores.
- How can they be brought into greater alignment?



# Supporting Policy Priorities with NAEP Data

- To strengthen the relationship between the rigor of performance standards, NAEP results can be used in standard setting to establish ‘reasonable ranges.’
- This involves mapping NAEP to Virginia SOL scores that produce similar impact.
- This information can be provided to panelists so they will be able to see what cut scores are likely to produce results similar to NAEP. This information is combined with the PLDs to inform their decision making.



# Timeline Considerations

---

- Allows multiple reviews of the process and results by the Board and the Technical Advisory Committee
- Uses assessment data collected under operational conditions
- Communicates expectations in advance of the academic year
- Positions the department to meet required professional standards including federal peer review

# Discussion

---

- Does the Board have any questions or feedback about the proposed process for setting standards?



[www.nciea.org](http://www.nciea.org)

Chris Domaleski  
[cdomaleski@nciea.org](mailto:cdomaleski@nciea.org)



This work is licensed under a  
Creative Commons Attribution 4.0  
International License. 21