

COMMITTEE ON STANDARDS, DESIGN AND METHODOLOGY AGENDA



March 6, 2025

11:20 am – 1:20pm EST

Space 1

11:20 am – 11:30 pm **Welcome and Updates**
Suzanne Lane, Chair

11:30 am – 12:30 pm **Progress Towards Enhanced Item Distribution
(CLOSED)** Attachment A
Enis Dogan, NCES

12:30 – 1:20 pm **Discussion of 2025 COSDAM Priorities –
Achievement Levels and Practical Significance** Attachment B
Suzanne Lane
Michelle Cantú-Wilson, Vice Chair
*Rebecca Dvorak, Assistant Director for
Psychometrics*

Progress Towards Enhanced Item Distribution

Closed Session, COSDAM Meeting

March 6, 2025

Goal

The purpose of this session is for staff from the National Center for Education Statistics (NCES) to provide the Committee on Standards, Design and Methodology (COSDAM) with updates on 2024 pilot studies examining efforts to increase the number of items at the low end of the performance scale.

Overview

Enis Dogan of NCES will provide a presentation highlighting the distributions of item difficulty for sets of items piloted for 2026 operational use for reading and mathematics at grades 4 and 8. COSDAM members are welcome to ask clarification questions, and to discuss implications towards shared goals to enhance information at the low end of the achievement scale.

Background

In recent years, COSDAM members have discussed issues related to students performing at the low end of the NAEP achievement scales – particularly those performing below the *NAEP Basic* achievement level. Discussions have focused on (a) the fact that a high percentage of students perform at below *NAEP Basic* on NAEP, (b) the difficulty in defining what students performing below *NAEP Basic* know and can do – rather, NAEP only provides information on what they do not know, and (c) concerns with the test-taking experience for low-performing students who may see few items they can correctly respond to.

Background materials prepared for the [March 2022](#) and [May 2022](#) COSDAM meetings describe past discussions about the lack of information at the low end of the NAEP achievement scale. At the conclusion of these discussions, the majority of COSDAM members expressed their continued support for the Governing Board's [achievement level policy](#) to maintain the three achievement levels, and that the priority should be to develop more items at the low end of the achievement scale. This priority has been shared by the Assessment Development Committee, and incorporated into their most recent NAEP Framework updates, and by NCES staff, who have worked with their item development contractors to increase the number of items at the low end of the scale in preparation for the 2026 operational assessments.

Discussion of 2025 COSDAM Priorities – Achievement Levels and Practical Significance

COSDAM Meeting

March 6, 2025

Goal

The purpose of this session is for COSDAM members to discuss committee priorities for 2025, related to the National Assessment Governing Board's legislative responsibilities regarding NAEP Achievement Levels and NAEP methodology, and to gather committee member input on the approach to achieve them.

Overview

Suzanne Lane, Chair of COSDAM, will provide a brief overview of recent activities and discussions related to the NAEP Achievement Levels and current methodology priorities for COSDAM. COSDAM members are asked to review the background information included in this document, and to consider the following questions:

1. Given our legislative responsibilities regarding achievement levels, and the tasks outlined in the Achievement Levels Work Plan, which of these efforts should be prioritized regarding the achievement levels? (e.g., internal validations, communications?)
 - a. Which of these priorities can be accomplished by staff without the need for contractor support?
2. If we were to create a one-pager with a research focus to help stakeholders understand statistical and practical significance at a general level, what would you recommend including?
 - a. What stakeholder types might be interested in this information, and what considerations need to be made to ensure it is understandable to them?
 - b. What research and tools might we link to (e.g., effect size calculators, [Matthew Kraft's research on effect size interpretations](#) for education)?

Background

NAEP Achievement Levels

The Governing Board is legislatively responsible for developing and maintaining the NAEP Achievement Levels. These duties are outlined on page 4 of the document prepared for the May 2024 quarterly Board meeting, [Legislative Roles and Responsibilities for the National Assessment of Educational Progress \(NAEP\)](#). The Governing Board defines three official achievement levels for NAEP: *NAEP Basic*, *NAEP Proficient*, and *NAEP Advanced*. These levels are defined for each content area

and grade assessed in the NAEP Assessment Frameworks by achievement level descriptions (ALDs). Information about the policies, procedures, and proper interpretations of the NAEP Achievement Levels can be found on the Governing Board's [website](#).

COSDAM is the committee within the Board with primary responsibility for the achievement levels. Achievement levels are set during a standard setting process in which groups of educators and others knowledgeable of the content come together and review the descriptions of each level and the assessment content and determine what minimum score is required to achieve each level. These minimum scores are called cut scores. Developing new cut scores generally occurs less frequently for NAEP compared to state assessments in part because of the Board's priority to maintain trend; maintaining the cut scores allows comparison of student performance across administration years. In addition to development, COSDAM is tasked with ensuring the achievement levels remain valid and meaningful over time.

In 2017, the National Academies of Sciences, Engineering, and Medicine (NASEM) conducted an independent evaluation of the NAEP Achievement Levels, and offered recommendations to strengthen support of the levels through collection of new validity evidence, to synthesize available evidence, and to communicate the meaning of achievement levels to a wide audience. In response, the Governing Board adopted the [Achievement Levels Work Plan](#) in 2020, identifying tasks to address the recommendations. Some of the key activities conducted to date include:

- Enhancing internal validity evidence by conducting achievement level description (ALD) review studies to examine the alignment between the content included on the NAEP assessments for NAEP [Reading and Math](#) all grades, and for [U.S. History, Civics, and Science](#) at grade 8. The study findings indicated high alignment between what the NAEP Frameworks claim students know and can do, and how they performed on the assessments. These studies also resulted in detailed descriptions of what students performing at each achievement level likely know and can do, referred to as Reporting ALDs.
- A Linking Studies Working Group, comprised of COSDAM and Research and Dissemination (R&D) committee members, was convened for just over a year in 2022 and 2023 to address linking NAEP to external measures. The working group resulted in the Board's [Resolution to Encourage Prioritization of NAEP Linking Studies](#). This resolution acknowledged linking studies offers valuable insights into understanding NAEP and its achievement levels, and sought collaboration with NCES to prioritize future linking studies, and to disseminate information and datasets from existing and future studies, to enhance understanding of NAEP data as it relates to external sources.
- The Governing Board has taken efforts to enhance the communications regarding proper interpretations of the NAEP Achievement Levels. For example, the Board published documents intended to facilitate interpretability of NAEP Achievement Levels for the [2022 NAEP Reading and Math](#) and the [2022 NAEP U.S. History and Civics](#) releases. The Governing Board also adopted [The](#)

[Intended Meaning of NAEP Results](#) in 2020 to help communicate interpretations of NAEP, including the achievement levels.

- Another major milestone activity was the completion of the NAEP [Achievement Levels Validity Argument](#). This report synthesizes available achievement level validity evidence – procedural, internal, and external – and provides information on appropriate and inappropriate interpretations of the levels based on the available evidence.

Developing and maintaining valid achievement levels is an ongoing effort; the following activities may need to be completed in the coming years to ensure achievement levels remain valid. COSDAM should consider how best to accomplish specific activities, including which activities can be accomplished fully at the staff level with committee guidance and which would require contractor assistance in the future. The following activities have been identified as priorities related to achievement levels:

- Board policy calls for updated Reporting ALDs every three administrations or following use of a new framework; math and reading are to be assessed from a new framework in 2026.
- A new NAEP Science Framework was recently adopted with significant changes compared to the prior framework. The Governing Board will need to work with NCES to identify if trends can be maintained, or if new cut scores will need to be developed for the 2028 administration.
- COSDAM, in collaboration with the R&D committee, may want to consider how to continue to improve communications around the achievement levels so they are useful and accessible to a wide audience.
- COSDAM may want to consider new opportunities to learn from linking to external data to understand how performance at each NAEP Achievement Level corresponds to outcomes on external academic and life outcome measures.

Practical Significance

COSDAM members have noted limitations in describing NAEP score differences using statistical significance only. Statistical significance is highly impacted by sample size. National sample sizes incorporating all student groups tend to be very large, and so even a one-point difference will likely be identified as significant, whereas for comparisons using smaller samples, a difference of two- or three-points, or even more, may be required before it is deemed statistically significant. This is because larger samples create greater confidence that an observed difference represents a true difference in the population of all students for which that sample represents. Statistical significance does not indicate whether a detected difference should be interpreted as small, moderate, or large.

COSDAM members discussed the potential of including effect sizes to help aid in the interpretations of NAEP score differences in August 2023. Members looking to understand what effect sizes are should review the relevant [August 2023 materials](#). At a high level, computing effect sizes for NAEP could provide a way to measure the size of

a difference between two scores. While COSDAM members have expressed interest in the past in having effect sizes included in NAEP reporting, the R&D committee and NCES staff have cautioned that they are difficult for non-researchers to understand. COSDAM members have acknowledged that if this data were to be included, we would need to provide some guidance regarding how to interpret them in the context of NAEP.

In May of 2024, COSDAM and the R&D committee held a joint meeting to discuss developing an interpretive guide to help aid in the interpretation of NAEP scores. The group identified the following topics for inclusion: (1) magnitude - bringing meaning to the size of score differences; (2) issues of statistical significance - helping distinguish between a difference which does not achieve statistical significance and a finding of no difference; (3) achievement levels - linking levels to skills and knowledge; (4) highlighting purposes of NAEP relative to state assessments and where the two intersect; and (5) perhaps providing sample sizes and standard deviations so specific audiences can calculate effect sizes.

Since this meeting, R&D committee members and staff have worked with communications contractors to best relay this information; it was determined that digging into statistical significance and practical significance as described at the joint meeting was more technical than average stakeholders would likely use, and so was not included as a priority for inclusion with the 2024 release. In addition, the achievement levels information is already current using the communications documents developed for the 2022 releases. However, COSDAM may want to lead efforts to develop a tool for researchers or other stakeholders who may be interested in understanding statistical significance and effect sizes.