Committee on Standards, Design and Methodology

November 16, 2021
2:00 – 4:00 pm ET (Virtual)
CLOSED

AGENDA

2:00 – 2:10 pm	**Welcome, Introductions, and Agenda Overview** *Suzanne Lane, Chair*	
2:10 – 2:50 pm	**Update: 2022 NAEP Administration (CLOSED)** *Enis Dogan, NCES*	Sent under separate cover
2:50 – 4:00 pm	**Briefing and Discussion: Improving Measurement and Reporting of Lower-Performing Students (CLOSED)** *Suzanne Lane* *Enis Dogan*	Attachment A
Information Item	**Review and Revision of NAEP Mathematics and Reading Achievement Level Descriptions**	Attachment B

Improving Measurement and Reporting of Lower-Performing Students

During the most recent Governing Board meeting, Ebony Walton of the National Center for Education Statistics (NCES) gave a presentation entitled, "A Decade of Monitoring Study Progress (or Lack Thereof) Through the Lens of NAEP." One of the main findings was that the scores of lower-performing students in math, reading, and science at grades 4, 8, and 12 declined nearly across the board from 2009 to 2019, while the scores of higher-performing students generally improved or held constant during the same time period. These patterns are sometimes referred to as divergent trend lines.

The August COSDAM discussion focused specifically on the growing percentage of students in the category below NAEP Basic and the need to better understand what lower-performing students know and can do. Some COSDAM members have expressed interest in exploring whether the Board should reconsider the current policy of defining only three achievement levels: *NAEP Basic*, *NAEP Proficient*, and *NAEP Advanced*. The policy specifies that, "The remaining region that falls below the *NAEP Basic* cut score shall be identified as 'below *NAEP Basic'* when a descriptor is necessary." The percentage of students in this lowest category is reported but there is no accompanying achievement level description (ALD).

One current challenge to reporting additional information for students performing below NAEP Basic is the relatively lower number of items at the lower end of the scale for some assessments. During the August COSDAM meeting, Enis Dogan, NCES, presented some item-person maps to demonstrate the alignment between student score distributions and the difficulty level of test items. Although there was variation by grade and subject, in general the item-person maps showed that there were more students than items at the lower end of the scale and more items than students at the upper end of the scale.

The NAEP frameworks generally contain a lot of rigor and cognitive complexity. NCES has several efforts underway to increase the number of items that lower-performing students can access. For example, in NAEP Mathematics a special effort was undertaken several years ago to provide additional clarifications and constraints for measuring basic skills in some of the framework objectives in the Mathematics Assessment and Item Specifications to produce easier items; the resulting items are known as "KaSA items," or Knowledge and Skills Appropriate items. Similar efforts could be applied to other subjects.

An alternative approach to understanding what lower-performing students can do was demonstrated by a special study recently released by NCES about the knowledge and skills possessed by students who performed below the *NAEP Basic* achievement level on the grade 4 reading assessment. This study focused on oral reading fluency, which is not part of the NAEP Reading Framework, but is a prerequisite skill to reading comprehension as defined by the framework.

Given the increasing divergence in performance, it may not be sufficient to develop more items at the lower end of the scale unless those items can also be administered to lower-performing students; this would require an adaptive testing model. The NAEP program has engaged in extensive research on adaptive testing, but it is not yet part of the operational assessment.

During the November COSDAM meeting, Enis Dogan will describe current NCES efforts to increase the number of items at the lower end of the scale and will briefly review previous and current efforts on adaptive testing for NAEP. COSDAM Chair Suzanne Lane will lead the Committee in a discussion of other efforts that the Board may want to consider to achieve the goal of improving measurement and reporting for lower-performing students.

This session is closed due to the inclusion of secure data on NAEP item pools.

Studies to Review and Revise NAEP Achievement Level Descriptions (ALDs) for Mathematics, Reading, and Other Subjects

Background

On September 24, 2020, the National Assessment Governing Board (Governing Board) awarded contract# 91995920C0004 to Pearson (as a result of a competitive bidding process) for conducting studies to review and revise NAEP achievement level descriptions (ALDs) in mathematics and reading using the 2019 NAEP assessments at grades 4, 8, and 12[1]. This work is intended to address the first recommendation of the evaluation of NAEP achievement levels that was conducted by the National Academies of Sciences, Engineering, and Medicine:

Recommendation #1: Alignment among the frameworks, the item pools, the achievement-level descriptors, and the cut scores is fundamental to the validity of inferences about student achievement. In 2009, alignment was evaluated for all grades in reading and for grade 12 in mathematics, and changes were made to the achievement-level descriptors, as needed. Similar research is needed to evaluate alignment for the grade 4 and grade 8 mathematics assessments and to revise them as needed to ensure that they represent the knowledge and skills of students at each achievement level. Moreover, additional work to verify alignment for grade 4 reading and grade 12 mathematics is needed.

The Board committed to conducting studies to review and revise the NAEP ALDs in its initial response to the evaluation that was formally adopted and sent to the Secretary of Education and Congress in December 2016. The Board's Achievement Levels Work Plan, adopted in March 2020, further describes the intention for this work: "Addressing Recommendation #1 should focus on the current reporting ALDs for mathematics and reading at grades 4, 8, and 12. The methodology will be similar to what was done to evaluate the alignment and revise the 2009 NAEP Reading ALDs for grades 4, 8, and 12 (Donohue, Pitoniak, & Beaulieu, 2010) and the 2009 NAEP Mathematics ALDs for grade 12 (Pitoniak, Dion, & Garber, 2010). This process will generate new reporting ALDs that comply with the revised Board policy statement" (p. 3).

According to Principle 1a of the Board policy on Developing Student Achievement Levels for NAEP, "Content achievement level descriptions translate the policy definitions into specific

[1] The base period of this contract includes the review and revision of ALDs in mathematics and reading at grades 4, 8, and 12; in addition, an option may be exercised for a second phase of the contract focusing on review and revision of ALDs in U.S. history, civics, science, and technology and engineering literacy (TEL) at grade 8 based on data from the most recent administrations of those assessments in 2018 and 2019.

expectations about student knowledge and skills in a particular content area, at each achievement level, for each subject and grade. Content ALDs provide descriptions of specific expected knowledge, skills, or abilities of students performing at each achievement level. They reflect the range of performance that items and tasks should measure. When setting achievement levels, the content ALDs provide consistency and specificity for panelist interpretations of policy definitions for a given assessment. During reporting, content ALDs communicate the specific knowledge and skills represented by *NAEP Basic*, *NAEP Proficient*, and *NAEP Advanced* for a given assessment" (p. 5).

Principles 3g and 4a of the Board policy apply specifically to this project of reviewing and revising the current ALDs and creating reporting ALDs (based on empirical data) that indicate what students at each achievement level *do* know and *can* do rather than what they *should* know and *should* be able to do[2]. Additional details for carrying out the work described by principles 3g and 4a are included in the Achievement Levels Procedures Manual.

The basis for the evaluation of NAEP achievement levels (and subsequently for this project) is the *existing* NAEP frameworks and item pools, not the new NAEP Mathematics Framework or NAEP Reading Framework currently scheduled for implementation in 2026. In accordance with principle 4b of the Board policy, the achievement levels and/or ALDs will need to be reviewed again once the new frameworks are implemented. Such work is beyond the scope of this project.

Project Overview

Dr. Eric Moyer is the project director at Pearson and Dr. Jennifer Galindo is the assistant project director at Pearson. Pearson will conduct a pilot study and an operational meeting using scale anchoring studies where panels of content experts judge the alignment of the current mathematics and reading ALDs and produce a set of recommended reporting ALDs for the Governing Board to consider in reporting the results from the next regular administration of the NAEP reading and mathematics assessments at grades 4, 8, and 12. The Governing Board is expected to take action on the reporting ALDs for mathematics and reading at grades 4, 8, and 12 in advance of the next release of these results.

Based on careful review of the history of ALD development, review, and revisions for NAEP mathematics and reading, a model-based anchored approach for reviewing the alignment of the ALDs for NAEP mathematics and reading will be used. The methodology for this alignment review study is based on that of previous studies, including the ALD development and review meeting held in 2009. The methodology was specified by the Board's Achievement Levels Work Plan and was selected to reduce the potential for possible inconsistencies from the use of different methods. The process of the model-based anchored approach will result in organizing

[2] According to the Board policy, ALDs will continue to describe what students *should* know and *should* be able to do for the purposes of item development and standard setting; only the reporting ALDs will be written in terms of what students *do* know and *can* do.

specific NAEP items by achievement level, which will serve as a key referent for panelists in reviewing and revising the current ALDs.

The model-based anchored approach includes three stages. The first stage will involve conducting statistical analyses to determine the items from the subject and grade that are anchored to a level corresponding to the score range within cut scores set to represent the achievement level descriptors (ALDs). The second stage relies on panels of content experts for each individual assessment. The panelists individually review the items that are anchored to each performance level and create summary descriptions of what students in each level are expected to know and be able to demonstrate based on the knowledge and skills measured by the items. In the final stage, the panelists compare the current ALDs for the respective assessment with their summary descriptions. The panelists note the similarities and differences, making a recommendation regarding whether the current ALDs accurately describe what students in each level are expected to know and be able to demonstrate or if revisions to the current ALDs are needed to improve alignment. The final alignment judgment will be used to report whether the panels determined that there exists alignment between the current ALDs and student expectations. The final panel summary descriptions will be used to revise the current ALDs to create reporting ALDs that indicate what students at each achievement level do know and can do.

During the May 2021 COSDAM meeting, the final Design Document for the NAEP ALD Review study was discussed by the Committee members and there were no recommended changes.

There is a technical advisory committee (TAC) consisting of the following experts in ALDs:

Dr. Karla Egan (Principal, EdMetric)

Dr. Ellen Forte (CEO and Chief Scientist, edCount)

Dr. Susan Loomis (Independent Consultant)

Dr. Marianne Perie (President, Measurement in Practice)

Dr. Mark Reckase (University Distinguished Professor Emeritus, Michigan State University)

Dr. Lauress Wise (Principal Scientist, Human Resources Research Organization)

The TAC is scheduled to meet for more than 100 hours (approximately 4 hours per month, with additional meeting time following the pilot and operational meetings) to provide technical advice on all aspects of the project to review and revise the mathematics and reading ALDs; this is intended to help ensure that all procedures, materials, and reports are carried out in accordance with current best practices, providing additional validity evidence for the process and results. In addition to frequent meetings and reviews of materials, two TAC members will attend the pilot and operational meetings to observe and provide feedback on the process.

In response to previous COSDAM discussions, the project schedule was modified to account for attempting to conduct the panel meetings in person (if public health conditions allow) in late 2021 and early 2022. The pilot meeting took place virtually during the week of October 25,

2021, and the operational meeting will take place during the week of February 21, 2022. The resulting ALDs will be presented for Board discussion at the May 2022 Board meeting and Board action at the August 2022 Board meeting. The intention is for the ALDs from this project to be used in the reporting of NAEP results in fall 2022.

Project Update (October 2021)

Following the August 2021 COSDAM meeting and additional discussion with COSDAM leadership about the risks of conducting in-person panel meetings in the wake of the Delta variant of COVID-19, a contract modification was issued to revert the panel meetings back to a virtual format and extend each panel meeting from 4 days to 5 days to account for the change in format.

Project work since the last COSDAM update in early August has focused on adapting to the virtual format and preparing for the pilot study, including: recruiting panelists, preparing materials, training facilitators, organizing meeting logistics, and conducting quality control checks. The TAC met on August 23 and September 28-29 to review and discuss materials and other preparations for the pilot study.

The pilot study took place from October 25-29; following review by the TAC in November and early December, lessons learned and any proposed modifications for the operational meeting will be discussed with COSDAM via a focused webinar in mid-December or early January.