



National Assessment Governing Board Achievement Levels Work Plan

Overview

The National Assessment Governing Board has developed a comprehensive work plan (the Plan) to fully respond to the [National Academies of Sciences, Engineering, and Medicine \(NAS\) evaluation of NAEP achievement levels](#). The ultimate aim of the Plan is to develop a body of evidence that provides a sound basis for removal of the trial status of the NAEP achievement levels. Other related goals are to develop, for Governing Board members and other interested stakeholders, a summary of the validity evidence supporting the interpretation of NAEP achievement levels and to facilitate clear, accurate, and informative reporting of NAEP achievement level results to the public. The Plan described here includes a list of activities (and associated timelines) to be pursued in conjunction with the National Center for Education Statistics (NCES). COSDAM will provide oversight for the Plan's implementation, in conjunction with other committees and NCES, as appropriate.

Background

The Governing Board issued an initial response to the NAS evaluation in December 2016 (see Appendix A) and adopted a revised policy on [Developing Student Achievement Levels for NAEP](#) in November 2018. This Plan provides detail about how each of the seven recommendations from the evaluation will be addressed (using guidance from the revised policy statement, where appropriate), including roles and priorities for accomplishing the work. Supplementing this Plan is a statement of intended purpose and meaning of NAEP (see Appendix B).

As indicated above, a primary aim of the Plan is to develop a body of evidence that provides a sound basis for removal of the trial status of the NAEP achievement levels. According to the NAEP legislation (PL 107-279), "The achievement levels shall be used on a trial basis until the Commissioner for Education Statistics determines, as a result of an evaluation under subsection (f), that such levels are reasonable, valid, and informative to the public." The proposed Plan aligns to those priorities; the criteria "Reasonable," "Valid," and "Informative to the Public"

have been indicated in the proposed responses to the NAS Committee recommendations described below.

Input from NCES suggests that the criterion of “informative to the public” as particularly important, where “the public” is interpreted to be groups who are responsible for using NAEP results directly and/or communicating information about NAEP achievement levels to others, including, but not limited to, state and federal legislators, education administrators at all levels, researchers and policy makers who use NAEP data, and media who cover education).

The Standards for Educational and Psychological Testing (AERA, APA, & NCME, 2014; hereafter the *Standards*) comprise a collection of professional best practices for all aspects of assessment, including achievement level setting. The following Plan was informed by the guidance provided in the *Standards*.

Responding to Recommendations #1, 2, and 3 (Valid)

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.

Recommendation #2: Once satisfactory alignment among the frameworks, the item pools, the achievement-level descriptors, and the cut scores in NAEP mathematics and reading has been demonstrated, their designation as trial should be discontinued. This work should be completed and the results evaluated as stipulated by law: (20 U.S. Code 9622: National Assessment of Educational Progress: <https://www.law.cornell.edu/uscode/text/20/9622> [September 2016]).

Recommendation #3: To maintain the validity and usefulness of achievement levels, there should be regular recurring reviews of the achievement-level descriptors, with updates as needed, to ensure they reflect both the frameworks and the incorporation of those frameworks in NAEP assessments.

The first three recommendations of the evaluation are inter-related. Recommendation #1 is focused on reading and math and covers all of the ALDs throughout the process, whereas Recommendation #3 is more general and primarily focused on monitoring the reporting ALDs. To some extent, Recommendation #3 has already been substantially addressed by the recently updated and approved Governing Board policy on NAEP achievement level setting. One remaining element related to Recommendation #3 is the development of a timeline and process for reviewing ALDs, along with prioritization for content areas beyond reading and math—a task that the Governing Board is now pursuing. An Achievement Levels Procedures Manual to

address the implementation of the policy will include details about the process for conducting these studies.

The Governing Board does not have direct responsibility for Recommendation #2. The NCES Commissioner makes the decision about the trial status and is not required to adhere to this NAS recommendation.

Regarding Recommendation #1, there are general policy definitions that apply to all NAEP assessments. These policy ALDs are elaborated into several different types of content ALDs under the revised Board policy. Additional clarity on the labels and uses of different types of ALDs should be described in the Achievement Levels Procedures Manual, including:

- Content ALDs developed with an assessment framework (generally by content area) are used to inform item development.
- Content ALDs that apply to a framework overall (across content areas) are used to conduct standard setting. These ALDs may be created as part of the framework development process or by re-convening framework panels (or similar individuals) after the assessment has been administered, prior to standard setting.
- Reporting ALDs, as described in the Board's revised policy statement, will be created following administration of an assessment to communicate about what performance at each NAEP achievement level indicates about what students do know and can do.

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. A potential additional step is to examine and/or document the alignment between the item pools and the NAEP frameworks, including information about the extent to which each NAEP administration faithfully represents the NAEP frameworks. Finally, alignment of cut scores can be evaluated using item maps, as part of the work to review and revise the reporting ALDs. Frameworks should be taken as a given; validation of the frameworks is beyond the scope of this work and evidence for their validity results from the Board's framework development process.

The following table provides a draft of the activities, timeline and responsibilities for responding to Recommendations 1-3. Work will begin with reading and mathematics ALDs (based on 2019 data, to be used in reporting 2021 results). Reporting ALDs for other subjects will be reviewed and revised according to when they next appear on the NAEP Assessment Schedule. In accordance with Principle 4 of the Board policy on NAEP achievement level setting, reporting ALDs will be reviewed and revised on a regular basis (at least every 3 administrations or every 10 years, or when there is a major framework update). For example, the NAEP Mathematics and

Reading ALDs will need to be revisited following the 2025 administrations under the revised assessment frameworks.

Proposed Activity	Responsibility	Timeline
COSDAM approval of Achievement Levels Procedures Manual (described in policy statement)	COSDAM	May 2020
Conduct studies to examine and/or document alignment between NAEP Math and Reading Frameworks and item pools for grades 4, 8, 12	NCES	December 2020
Conduct studies to review and revise Math and Reading ALDs at grades 4, 8, and 12 Conduct studies to review and revise U.S. History, Civics, and Science ALDs at grade 8	NAGB	Contract awarded summer/fall 2020; complete by summer 2021 (reading/math); for other subjects the timeline will be determined by Assessment Schedule (ALDs updated in time for reporting of next administration after 2020)
Full Board action on revised Reporting ALDs	NAGB	Math/Reading at grades 4, 8, 12 (August 2021); for other subjects the timeline will be determined by Assessment Schedule (Board action will take place prior to release of results)
Conduct studies to examine and/or document alignment between NAEP U.S. History, Civics, Science, and TEL Frameworks and item pools for grade 8	NCES	December 2021

Responding to Recommendation #4 (Informative to the Public)

Recommendation #4: Research is needed on the relationships between the NAEP achievement levels and current or future performance on measures external to NAEP. Like the research that led to setting scale scores that represent academic preparedness for college, new research should focus on other measures of future performance, such as being on track for a college-

ready high school diploma for 8th-grade students and readiness for middle school for 4th-grade students.

Recommendation #4 is interpreted as articulating the need to provide context and relevance for NAEP results to show where NAEP fits in the constellation of other major assessments and external indicators of student achievement that are familiar to the general public, such as international assessments and indicators of postsecondary preparedness. Because NAEP and external indicators typically have different purposes, administration conditions, target populations, and other distinguishing characteristics, the purpose of this particular recommendation is not to make judgments about which results are “right” or “wrong” but to make the reporting of NAEP results more meaningful, useful, AND informative to the public.

This recommendation refers to both linking studies of NAEP and other measures of student achievement, as well as efforts to use NAEP to predict future performance. There are many different existing measures of student achievement, and we are aware of several efforts to link NAEP to various other measures, particularly in math, reading, and science. In order to consider what new studies might be pursued, it is important to better understand the resources that already exist, in addition to discussing how new efforts fit into the Board’s ongoing work and Strategic Vision.

The Governing Board’s work on reporting and dissemination includes the production of infographics and other descriptive reporting that describe student achievement in terms of several contextual variables. This work has typically been done using scale scores but could be expanded to include achievement level information, possibly including efforts to provide descriptive information about contextual factors associated with performance at the *NAEP Basic*, *NAEP Proficient*, and *NAEP Advanced* achievement levels.

To address the issue of how best to synthesize and report information about how NAEP relates to other assessments and indicators, the Governing Board has commissioned a technical memo on recommendations for synthesizing relevant findings from multiple studies in ways that are informative to a general audience. The purpose of this effort is to explore how to place NAEP in a meaningful context of other familiar assessments and indicators, and to generate additional ideas. The following table provides a draft of the activities, timeline and responsibilities for responding to Recommendation 4.

Proposed Activity	Responsibility	Timeline
Review of technical memo on various ideas (including pros/cons) for synthesizing and representing findings about how other assessments and external indicators of student performance relate to NAEP (including a summary of existing linking studies) and what the findings mean for NAEP.	NAGB	Spring 2020
As the Governing Board works to develop its next Strategic Vision, deliberations will take place as part of that effort to determine how to approach the goal of making NAEP more relevant by connecting NAEP results to important real world indicators of student achievement.	NAGB	August 2020
Decision on additional studies that should be pursued to connect NAEP to other assessments and external indicators of student performance	NAGB/NCES	November 2020

Responding to Recommendations #5 & #6 (Reasonable, Valid, Informative)

Recommendation #5: Research is needed to articulate the intended interpretations and uses of the achievement levels and collect validity evidence to support these interpretations and uses. In addition, research to identify the actual interpretations and uses commonly made by NAEP's various audiences and evaluate the validity of each of them. This information should be communicated to users with clear guidance on substantiated and unsubstantiated interpretations.

Recommendation #6: Guidance is needed to help users determine inferences that are best made with achievement levels and those best made with scale score statistics. Such guidance should be incorporated in every report that includes achievement levels.

The *Standards* clearly indicate that any validation plan should begin with specifying the intended interpretations and uses of test scores. It is important to recognize that NAEP ALDs do not make claims about the achievement levels predicting performance on other current or future criteria (e.g., college readiness); however, strong claims *are* asserted about mastery of the content covered by relevant NAEP frameworks. Therefore, Recommendations #5 and #6 are related and should be considered together. The Governing Board is currently working on developing a statement of intended purpose and meaning for NAEP, which includes intended interpretations and uses for scale scores and achievement levels at a general level. The full Board discussed this document at the November 2019 Board meeting and is expected to take action during the upcoming March 2020 Board meeting (Appendix B). The Reporting and Dissemination Committee and COSDAM have provided initial guidance on an interpretative guide for the NAEP achievement levels.

After the Board reaches consensus about the intended interpretations and uses, the next step is to gather and document the evidence that exists related to those interpretations and to identify areas where additional evidence may be needed. This would take the form of building validity arguments to document the evidence that exists to support intended interpretations and uses; separate activities would be appropriate for supporting NAEP *scale scores* and NAEP *achievement levels*.

Gathering and summarizing validity evidence regarding interpretations of NAEP scale scores should primarily be a responsibility of NCES. This may be a matter of gathering and synthesizing documentation of existing NCES procedures that provided validity evidence for NAEP interpretations (e.g., qualifications of item writers, procedures for reviewing items, pilot testing, cognitive labs, etc.). This activity would also help to uncover areas where more research and evidence is needed.

Gathering and summarizing validity evidence regarding interpretations of NAEP achievement levels is a responsibility of the Governing Board (via COSDAM). Research undertaken to address Recommendation #1 should also provide evidence to address part of Recommendation #5, because the ALDs represent the intended meaning of NAEP achievement level categories.

In contrast to the established traditions for validating score meaning (e.g., the *Standards*), broadly endorsed procedures or criteria for gathering and evaluating evidence regarding score (or achievement category) use do not yet exist. Nonetheless, the interpretative guide contemplated by COSDAM and R&D would be one source of evidence to address Recommendations #5 and #6.

The Board recognizes that some stakeholders may hold misconceptions of the achievement levels. For example, legislators or education writers have sometimes confused performance at the *NAEP Proficient* level with grade-level performance. To respond to these misconceptions, we propose to work to create and provide materials and to conduct new outreach activities. The first step to addressing the misconceptions is to better understand how various stakeholder groups are interpreting and using the NAEP achievement levels. The NAS evaluation included reviews of existing materials and conversations with multiple audiences to begin to understand and articulate how various stakeholder groups are interpreting and using the NAEP achievement levels. The NAS evaluation did uncover several existing misconceptions about the NAEP achievement levels, and the Board will need to conduct additional work to more fully understand actual interpretations and uses of the NAEP achievement levels. We will need to develop and refine additional materials in formats most relevant to targeted audiences, (e.g., print, video, workshops) to address existing misconceptions and promote appropriate use. It would also seem desirable to engage in a companion evaluation effort to assess the effectiveness of these new materials and outreach activities.

Recommendation #6 (need for explicit guidance about when to use scale scores versus achievement levels) appears to be very narrow in scope, referring specifically to the

inappropriateness of using the percentage above a cut score to describe changes over time and across groups. To best address Recommendation #6, the interpretative guide should explicitly include information about which inferences are best made with scale scores versus achievement levels.

Effective communication of the NAEP achievement levels is an important aspect of Recommendations #5 and #6. There is a need to better understand how users interpret the policy definitions and ALDs for *NAEP Basic*, *NAEP Proficient*, and *NAEP Advanced*. For example, what does “solid academic performance” mean, and is it possible to describe this educational goal more effectively?

Further development of these ideas (and others) will be needed to address these recommendations, and the staff plans to convene an ongoing advisory group on communication of NAEP achievement levels. The following table provides a draft of the activities, timeline and responsibilities for responding to Recommendations #5 and #6.

Proposed Activity	Responsibility	Timeline
Convene ongoing advisory group to discuss and provide feedback on the development of materials for communicating NAEP achievement levels	NAGB/NCES	Spring 2020 – Spring 2021
Collect information about current uses of NAEP achievement levels via focus groups and evaluate appropriateness of interpretations and uses that are not directly intended	NAGB	Spring/summer 2020
Adopt statement of intended purpose and meaning of NAEP (Appendix B)	NAGB	March 2020
Improve communications of what NAEP frameworks and achievement levels represent	NAGB/NCES	Ongoing
Develop and finalize interpretative guide for NAEP achievement levels; iterative drafts will be discussed by COSDAM and R&D	NAGB	Spring 2020-Spring 2021
Collect and document validity evidence to support intended interpretations and uses of NAEP achievement levels <ul style="list-style-type: none"> • Collect and summarize validity evidence to support intended interpretations and uses of NAEP scale scores 	NAGB NCES	Spring 2021 Spring 2021

Responding to Recommendation #7 (Valid)

Recommendation #7: NAEP should implement a regular cycle for considering the desirability of conducting a new standard setting. Factors to consider include, but are not limited to: substantive changes in the constructs, item types, or frameworks; innovations in the modality for administering assessments; advances in standard setting methodologies; and changes in the

policy environment for using NAEP results. These factors should be weighed against the downsides of interrupting the trend data and information.

Recommendation #7 has been addressed by inclusion in the revised policy statement (Principle 4). It will be necessary to develop a process for carrying out a cut score review, but this should occur under COSDAM's purview as part of the development of the Achievement Levels Procedures Manual.

References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education [AERA, APA, & NCME]. (2014). *Standards for educational and psychological testing*. Washington, DC: AERA.
- Donahue, P., Pitoniak, M., & Beaulieu, N. (2010). *Final report on the study to draft achievement-level descriptions for reporting results of the 2009 National Assessment of Educational Progress in reading for grades 4, 8, and 12*. Princeton, NJ: Educational Testing Service.
- National Academies of Science, Engineering, and Medicine. (2017). *Evaluation of the Achievement Levels for Mathematics and Reading on the National Assessment of Educational Progress*. Washington, DC: The National Academies Press.
- Pitoniak, M., Dion, G., & Garber, D. (2010). *Final report on the study to draft achievement level descriptions for reporting results of the 2009 National Assessment of Educational Progress in mathematics for grade 12*. Princeton, NJ: Educational Testing Service.