

Developing a Comprehensive Plan to Implement the Governing Board’s Response to the 2016 Evaluation of NAEP Achievement Levels

Public Law 107-279 states:

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.

Even after being in use for over 25 years and undergoing several evaluations, the NAEP achievement levels are still considered to be on a trial basis. The [2016 evaluation of NAEP achievement levels](#), conducted by the National Academies of Sciences, Engineering, and Medicine, focused on the NAEP mathematics and reading achievement levels for grades 4, 8, and 12. This evaluation report, presented to the Governing Board at its November 2016 meeting, stated, “During their 24 years [the achievement levels] have acquired meaning for NAEP’s various audiences and stakeholders; they serve as stable benchmarks for monitoring achievement trends, and they are widely used to inform public discourse and policy decisions. Users regard them as a regular, permanent feature of the NAEP reports” (page Sum-8). This evaluation included several recommendations, and the Board issued a formal response noting its planned actions in December 2016.

One important aspect of the Board’s response to the 2016 evaluation was a commitment to update the guidance provided in the Board policy statement on NAEP achievement levels. The Committee on Standards, Design and Methodology (COSDAM) began working to update this policy guidance in March 2017, and the revised policy on [Developing Student Achievement Levels for NAEP](#) was adopted by the Board in November 2018.

During the March 2019 Board meeting, Governing Board Chair Beverly Perdue established an Achievement Levels Working Group¹ to develop a comprehensive plan (including a list of activities for the Governing Board to pursue in conjunction with the National Center for Education Statistics) to fully respond to the evaluation. Over the past 8 months, the Working Group has convened by telephone and in person to produce the attached draft plan. Preliminary ideas were shared and discussed with NCES Commissioner Lynn Woodworth and Associate Commissioner Peggy Carr during the spring and summer of 2019, as well as with COSDAM and Reporting and Dissemination (R&D) Committee members during the August 2019 Board meeting.

During the upcoming November 2019 Board meeting, Achievement Levels Working Group Chair Gregory Cizek will present the proposed Achievement Levels Work Plan for input and discussion by the full Board. The goal is for the Board to take action on the plan during the March 2020 Board meeting.

¹ The Achievement Levels Working Group was comprised of the following members: Gregory Cizek (Chair), Father Joe O’Keefe, Fielding Rolston, Linda Rosen, and Joe Willhoft.



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). If adopted by the Board, COSDAM would 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 which has been under development by COSDAM (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, but the NAS evaluation did tie the trial status to addressing Recommendation #1. This implies that Recommendation #1 should be a priority, but we acknowledge that the NCES Commissioner makes the decision about the trial status and is not required to adhere to this 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	August 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	March 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 for this plan. The following table provides a draft of the activities, timeline and responsibilities for responding to Recommendation 4.

Proposed Activity	Responsibility	Timeline
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 (Center for Assessment)	Fall 2019
Governing Board deliberation on whether the Board should embark on a large research effort to support using NAEP assessments to make claims about being on track for a college-ready high school diploma for 8 th grade students and readiness for middle school for 4 th grade students	NAGB	Fall 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. COSDAM reviewed a draft statement at the May 2019 Board meeting and via email, and the full Board will discuss this document on Saturday morning at the upcoming November 2019 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) by the Working Group will be needed to address these recommendations, and the staff plans to convene an initial expert panel meeting and ultimately establish an ongoing advisory group on communication of NAEP achievement levels. Initial discussions in late 2019 or early 2020 could be used to refine and add details to this plan between November 2019 and March 2020. The following table provides a draft of the activities, timeline and responsibilities for responding to Recommendations #5 and #6.

Proposed Activity	Responsibility	Timeline
Convene expert panel on how to most effectively communicate information about NAEP achievement levels to generate additional ideas for this plan	NAGB	Late 2019 or early 2020
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.

National Assessment Governing Board’s Response to the National Academies of Sciences, Engineering, and Medicine 2016 Evaluation of NAEP Achievement Levels

Legislative Authority

Pursuant to the National Assessment of Educational Progress (NAEP) legislation (Public Law 107-279), the National Assessment Governing Board (hereafter the Governing Board) is pleased to have this opportunity to apprise the Secretary of Education and the Congress of the Governing Board response to the recommendations of the National Academies of Sciences, Engineering, and Medicine evaluation of the NAEP achievement levels for mathematics and reading (Edley & Koenig, 2016).

The cited legislation charges the Governing Board with the authority and responsibility to “develop appropriate student achievement levels for each grade or age in each subject area to be tested.” The legislation also states that “such levels shall be determined by... a national consensus approach; 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; ... [and] shall be updated as appropriate by the National Assessment Governing Board in consultation with the Commissioner for Education Statistics” (Public Law 107-279).

Background

NAEP is the largest nationally representative and continuing assessment of what our nation’s elementary and secondary students know and can do. Since 1969, NAEP has been the country’s foremost resource for measuring student progress and identifying differences in student achievement across student subgroups. In a time of changing state standards and assessments, NAEP serves as a trusted resource for parents, teachers, principals, policymakers, and researchers to compare student achievement across states and select large urban districts. NAEP results allow the nation to understand where more work must be done to improve learning among all students.

For 25 years, the NAEP achievement levels (*Basic*, *Proficient*, and *Advanced*) have been a signature feature of NAEP results. While scale scores provide information about student achievement over time and across student groups, achievement levels reflect the extent to which student performance is “good enough,” in each subject and grade, relative to aspirational goals.

Since the Governing Board began setting standards in the early 1990s, achievement levels have become a standard part of score reporting for many other assessment programs in the US and abroad.

Governing Board Response

Overview

The Governing Board appreciates the thorough, deliberative process undertaken over the past two years by the National Academies of Science, Engineering, and Medicine and the expert members of the Committee on the Evaluation of NAEP Achievement Levels for Mathematics and Reading. The Governing Board is pleased that the report concludes that the achievement levels are a meaningful and important part of NAEP reporting. The report states that, “during their 24 years [the achievement levels] have acquired meaning for NAEP’s various audiences and stakeholders; they serve as stable benchmarks for monitoring achievement trends, and they are widely used to inform public discourse and policy decisions. Users regard them as a regular, permanent feature of the NAEP reports” (Edley & Koenig, 2016; page Sum-8). The Governing Board has reviewed the seven recommendations presented in the report and finds them reasonable and thoughtful. The report will inform the Board’s future efforts to set achievement levels and communicate the meaning of NAEP *Basic*, *Proficient*, and *Advanced*. The recommendations intersect with two Governing Board documents, the Strategic Vision and the achievement levels policy, described here.

On November 18, 2016, the Governing Board adopted a Strategic Vision (<https://www.nagb.org/content/nagb/assets/documents/newsroom/press-releases/2016/nagb-strategic-vision.pdf>) to guide the work of the Board through 2020, with an emphasis on innovating to enhance NAEP’s form and content and expanding NAEP’s dissemination and use. The Strategic Vision answers the question, “How can NAEP provide information about how our students are doing in the most innovative, informative, and impactful ways?” The Governing Board is pleased that several of the report recommendations are consistent with the Board’s own vision. The Governing Board is committed to measuring the progress of our nation’s students toward their acquisition of academic knowledge, skills, and abilities relevant to this contemporary era.

The Governing Board’s approach to setting achievement levels is articulated in a policy statement, “Developing Student Performance Levels for the National Assessment of Educational Progress” (<https://www.nagb.org/content/nagb/assets/documents/policies/developing-student-performance.pdf>). The policy was first adopted in 1990 and was subsequently revised in 1995,

with minor wording changes made in 2007. The report motivates the revision of this policy, to add clarity and intentionality to the setting and communication of NAEP achievement levels.

The seven recommendations and the Governing Board response comprise a significant research and outreach trajectory that the Governing Board can pursue over several years in conjunction with key partners. The Governing Board will implement these responses within resource constraints and in conjunction with the priorities of the Strategic Vision.

Evaluating the Alignment of NAEP Achievement Level Descriptors

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 report's primary recommendation is to evaluate the alignment, and revise if needed, the achievement level descriptors for NAEP mathematics and reading assessments in grades 4, 8, and 12. The Governing Board intends to issue a procurement for conducting studies to achieve this goal. The Governing Board has periodically conducted studies to evaluate whether the achievement level descriptors in a given subject should be revised, based on their alignment with the NAEP framework, item pool, and cut scores. The Governing Board agrees that this is a good time to ensure that current NAEP mathematics and reading achievement level descriptors align with the knowledge and skills of students in each achievement level category. In conjunction with the response to Recommendation #3, the updated Board policy on NAEP achievement levels will address the larger issue of specifying a process and timeline for conducting regular recurring reviews of the achievement level descriptions in all subjects and grades.

The Governing Board agrees strongly with the recommendation that, while evaluating alignment of achievement level descriptors is timely, it is not necessary to consider changing the cut scores or beginning a new trend line at this time. The NAEP assessments are transitioning from paper-based to digital assessments in 2017, and current efforts are focused on ensuring comparability between 2015 and 2017 scores. The Governing Board articulated this in the 2015 Resolution on Maintaining NAEP Trends with the Transition to Digital-Based Assessments (<https://www.nagb.org/content/nagb/assets/documents/policies/resolution-on-trend-and-dba.pdf>).

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]).

Ultimately, the Commissioner of Education Statistics is responsible for determining whether the “trial” designation is removed. The Governing Board is committed to providing the Commissioner with the information needed to make this determination in an expedient manner.

Regular Recurring Reviews of the Achievement Level Descriptors

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 Board’s current policy on NAEP achievement levels contains several principles and guidelines for *setting* achievement levels but does not address issues related to the continued use or reporting of achievement levels many years after they were established. The revised policy will seek to address this gap by including a statement of periodicity for conducting regular recurring reviews of the achievement level descriptors, with updates as needed, as called for in this recommendation. The Governing Board agrees that it is important to articulate a process and timeline for conducting regular reviews of the achievement level descriptors rather than performing such reviews on an ad hoc basis.

Relationships Between NAEP Achievement Levels and External Measures

Recommendation #4: Research is needed on the relationships between the NAEP achievement levels and concurrent 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.

In addition to the extensive work that the Governing Board has conducted at grade 12 to relate NAEP mathematics and reading results to academic preparedness for college, the Governing Board has begun research at grade 8 with statistical linking studies of NAEP mathematics and reading and the ACT Explore assessments in those subjects. This work was published while the evaluation was in process and was not included in the Committee’s deliberations. Additional studies in NAEP mathematics and reading at grades 4 and 8 are beginning under contract to the National Center for Education Statistics (NCES). The Governing Board’s Strategic Vision includes an explicit goal to increase opportunities for connecting NAEP to other national and

international assessments and data. Just as the Board's previous research related grade 12 NAEP results in mathematics and reading to students' academic preparedness for college, the Governing Board anticipates that additional linkages with external measures will help connect the NAEP achievement levels and scale scores to other meaningful real-world indicators of current and future performance.

Interpretations and Uses of NAEP Achievement Levels

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.

The Governing Board's Strategic Vision emphasizes improving the use and dissemination of NAEP results, and the Board's work in this area will include achievement levels. The Governing Board recognizes that clarity and meaning of NAEP achievement levels (and scale scores) are of utmost importance. The Governing Board will issue a procurement to conduct research to better understand how various audiences have used and interpreted NAEP results (including achievement levels). The Governing Board will work collaboratively with NCES to provide further guidance and outreach about appropriate and inappropriate uses of NAEP achievement levels.

Guidance for Inferences Made with Achievement Levels versus Scale Scores

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 Governing Board understands that improper uses of achievement level statistics are widespread in the public domain and extend far beyond the use of NAEP data. Reports by the Governing Board and NCES have modeled appropriate use of NAEP data and will continue to do so. This recommendation is also consistent with the goal of the Strategic Vision to improve the dissemination and use of NAEP results. The Governing Board will continue to work with NCES and follow current research to provide guidance about inferences that are best made with achievement levels and those best made with scale score statistics.

Regular Cycle for Considering Desirability of Conducting a New Standard Setting

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.

When the Board’s achievement levels policy was first created and revised in the 1990s, the Board was setting standards in each subject and grade for the first time and had not yet considered the need or timeline for re-setting standards. To address this recommendation, the Governing Board will update the policy to be more explicit about conditions that require a new standard setting.

Board’s Commitment

The Governing Board remains committed to its congressional mandate to set “appropriate student achievement levels” for the National Assessment of Educational Progress. The Board appreciates the report’s affirmation that NAEP achievement levels have been set thoughtfully and carefully, consistent with professional guidelines for standard setting, and based on extensive technical advice from respected psychometricians and measurement specialists. The Board also takes seriously the charge to develop the current achievement levels through a national consensus approach, involving large numbers of knowledgeable teachers, curriculum specialists, business leaders, and members of the general public throughout the process. This is only fitting given the Governing Board’s own congressionally mandated membership that explicitly includes representatives from these stakeholder groups.

The Governing Board remains committed to improving the process of setting and communicating achievement levels. The Governing Board is grateful for the report recommendations that will advance these aims.

Reference

Edley, C. & Koenig, J. A. (Ed.). (2016). *Evaluation of the Achievement Levels for Mathematics and Reading on the National Assessment of Educational Progress*. Washington, DC: The National Academies Press.

The Intended Meaning of NAEP Results

The primary purpose of the National Assessment of Educational Progress (NAEP), also known as the Nation’s Report Card, is to measure the educational achievement and progress of the nation’s students in established grades, ages, and subject areas. This primary purpose also facilitates the use of NAEP results for comparisons of what representative students know and can do in relation to the content measured by the NAEP frameworks, among states and jurisdictions, among various demographic groups, and over time.

The authorizing legislation for NAEP and the National Assessment Governing Board states that the purpose of the NAEP program is broadly to, “conduct a national assessment and collect and report assessment data, including achievement data trends, in a valid and reliable manner on student academic achievement in public and private elementary schools and secondary schools...” (Public Law 107-279, Section 303(b)(2)(B)). That legislation also prohibits NAEP from maintaining any system of personally identifiable information. Thus, NAEP assesses the educational progress of groups of representative students, not individuals.

The National Center for Education Statistics (NCES) develops numerical score scales for each NAEP subject. NAEP scale scores convey the degree to which students have mastered the content described in the NAEP assessment frameworks, with higher scores indicating greater levels of mastery.

The Governing Board is charged with setting NAEP achievement levels and has established general policy definitions for *NAEP Basic*, *NAEP Proficient*, and *NAEP Advanced*, [here](#). Percentages at or above achievement level cut scores indicate the percentage of students in a group who *meet or exceed* the knowledge and skills represented by specific content achievement level descriptions. These specific descriptions are found in the NAEP [assessment frameworks](#) and [reports](#). Additional information about the NAEP achievement level descriptions can be found in the Governing Board policy statement on [Developing Student Achievement Levels for NAEP](#).

NAEP results describe educational achievement for groups of students at a single point in time, progress in educational achievement for groups of students over time, and differential educational achievement and progress among jurisdictions and subpopulations.

There are several features of NAEP that distinguish it from many other assessment programs, such as state-level or international assessments. For example:

- 1) NAEP produces results for the nation and participating states and districts. NAEP does not produce results for individual students or schools.
- 2) NAEP measures progress based on successive cohorts of students. NAEP does not produce results about the growth of individual students or groups of students over time.
- 3) NAEP results measure achievement and progress; however, NAEP results alone cannot indicate either why or how progress has occurred. Educational policies and practices that concur with NAEP progress may have caused this progress or been coincidental.
- 4) NAEP assessments are based on independent assessment frameworks developed through a national consensus approach described [here](#). NAEP frameworks do not represent specific state or local curricula. NAEP results cannot—and are not intended to—directly indicate how well students are performing in relation to state and local curricula nor state achievement levels.