

## **Discussion of Proposed Revised Policy Statement on NAEP Achievement Level Setting**

For the past 1.5 years, COSDAM members have been discussing the need to update the 1995 Governing Board policy statement on [Developing Student Performance Levels for NAEP](#). The Board’s formal response to the November 2016 evaluation of the NAEP achievement levels (attached) noted that several of the report recommendations would be addressed through a revision of the Board policy statement. In particular, the Board’s response stated that the updated policy will specify a process and timeline for conducting regularly recurring reviews of the achievement level descriptions (ALDs) and will be explicit about the conditions that necessitate consideration of a new standard setting. In addition, one of the planned activities for the implementation of the Strategic Vision is to consider new approaches to creating and updating the achievement level descriptions in the revision of the Board policy on achievement levels.

Given that the policy is over 20 years old, there was also a need to revisit the operational guidance for setting achievement levels to ensure that it reflects currently accepted best practices.

Several activities have informed the draft revised policy statement, including ongoing COSDAM discussion, input from technical experts in standard setting, and several research efforts. The March and May 2018 COSDAM meetings were entirely devoted to this topic. Full Board discussions on this topic were held in August 2017 and August 2018, and additional discussion and action are planned for the upcoming November 2018 meeting.

The table below summarizes both prior and planned upcoming activities:

<b>Activity</b>	<b>Timeline</b>
Initial COSDAM discussions of planned activities and timeline	March – May 2017
Initial input from standard setting experts	March – April 2017
Initial full Board discussion about potential elements of policy revision	August 2017
Technical advisory panel to seek expert advice and debate on best practices in achievement level setting	January 2018
Literature review of best practices for creating and updating ALDs	February 2018
Technical memo on developing a validity argument for the NAEP achievement levels	February 2018
COSDAM discussion of using research to inform goals for policy revision	March 2018
COSDAM review and discussion of draft revised policy	May 2018
Joint COSDAM/R&D discussion on communicating achievement levels	May 2018
COSDAM call to finalize a consensus draft for full Board discussion	June 2018
Technical advisory panel to seek expert advice on policy and procedures for reviewing and revising achievement level descriptions	July 2018
Full Board review and discussion of draft revised policy	August 2018
Public comment on draft revised policy	August 30 - October 15, 2018
Calls to discuss additional proposed revisions to draft policy	October 2018
<b>Planned discussions and full Board action on revised policy</b>	<b>November 2018</b>

Compared to the current (1995) policy on [Developing Student Performance Levels for NAEP](#), the attached proposed revised policy statement reflects:

- Reorganization of principles, streamlining of language, minimization of redundancies
- Minor (non-substantive) edits to the NAEP policy definitions for clarity
- A change in terminology from *Proficient* to *NAEP Proficient* to better differentiate the NAEP achievement levels from other common uses of *Basic, Proficient, Advanced*
- A new principle on periodic review of achievement level descriptions and cut scores, prompted by the Board’s response to the evaluation of NAEP achievement levels
- A new principle to clarify participation of multiple stakeholders at various points throughout process
- A new principle to summarize the role of the Board
- Reference to an interpretative guide that would accompany the release of NAEP results and explain how the achievement levels should (and should not) be used
- Reference to multiple types of achievement level descriptions (ALDs), including reporting ALDs that would be created using empirical data and written in terms of what students *do know and can do* rather than what students *should know and be able to do*
- Clarification on the standard setting participants, in particular the non-educator group
- Additional details about the achievement level setting process, including some practices that have become institutionalized over time (e.g., the use of “impact data”)
- Removal of details on implementation directed to staff and contractors, which will instead be included in a “procedures manual”

The proposed policy revision is intended to be just one step towards implementing the Board’s formal response to the evaluation of NAEP achievement levels. Several other activities are underway and/or planned, and the overall status of all activities (both relevant to and distinct from the policy revision) will be discussed at the upcoming November COSDAM meeting.

### Public Comment

Following the full Board discussion in August 2018, the Governing Board sought public comment on the proposed revised policy via a dedicated page on its website from August 30 – October 15, 2018: <https://www.nagb.gov/news-and-events/calendar/public-comment-on-als-policy.html>. Notices were posted in the federal register<sup>1</sup> on September 10<sup>th</sup> and October 2<sup>nd</sup>.

Seventy-three comments were received and have been compiled (see attached). Some of the major themes noted in the comments are as follows:

---

<sup>1</sup> The initial public comment period was intended to take place from August 30 – September 30, 2018. However, because the federal register notice was not actually published until September 10, the deadline was extended until October 15<sup>th</sup>. The typical timeframe for public comments from federal register notices is 30-45 days.

- Discussion of the value or lack thereof of the NAEP achievement levels in general
- Discussion of the policy labels and/or definitions, including:
  - Concerns that *NAEP Basic*, *NAEP Proficient*, and *NAEP Advanced* will not result in achievement levels that are “reasonable, valid, and informative to the public”
  - Praise for the proposal to add NAEP in front of *Basic*, *Proficient*, and *Advanced* to better distinguish from other uses of these terms, and support for the reasonableness of the NAEP achievement levels in comparison with other rigorous assessments
  - Concern that the original labels of *Basic*, *Proficient*, and *Advanced* should be retained without adding NAEP in front
- Discussion of the rigor of the NAEP achievement levels, including both the importance of maintaining high standards and calls for lowering standards
- Praise by technical experts for reflecting many current best practices in standard setting
- Concerns and suggestions for better communicating the meaning of achievement levels and achievement level results to various stakeholders
- Suggestions to connect the NAEP achievement levels with external benchmarks of success, either for validation purposes or as an alternative method of setting standards
- Discussion of the Board’s commitment to periodically review the NAEP achievement level descriptions and cut scores, including endorsement of this idea and concerns that they should not be changed unless absolutely necessary
- Suggestions for additional procedural details to be included in the policy
- A suggestion to include information about additional work to examine alignment between NAEP frameworks and state content standards

COSDAM held two teleconferences in mid-October to discuss the comments received and potential additional revisions to the policy. The attached proposed revised policy document reflects the outcomes from those calls; both a clean copy and a “tracked changes” version (in comparison to the version of the policy discussed at the August 2018 Board meeting and submitted for public comment) are included.

# **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.



## **National Assessment Governing Board**

### **Developing Student Achievement Levels for the National Assessment of Educational Progress**

#### **Policy Statement**

It is the policy of the National Assessment Governing Board to conduct a comprehensive, inclusive, and deliberative process to develop and review student achievement levels for the National Assessment of Educational Progress (NAEP).<sup>1</sup> Achievement levels consist of general policy definitions for the *NAEP Basic*, *NAEP Proficient*, and *NAEP Advanced* levels, specific achievement level descriptions (ALDs) for each assessment, cut scores that demarcate adjacent levels, and exemplar items or tasks that illustrate performance at each level. This process shall be conducted according to widely accepted professional standards, to produce results that are reasonable, useful, and informative to the public.

The Governing Board, through its Committee on Standards, Design and Methodology (COSDAM), shall monitor the development and review of student achievement levels to ensure that the final Governing Board-adopted achievement level descriptions, cut scores, and exemplars comply with all principles of this policy.

The achievement level setting process shall be carried out by contractors selected through a competitive bidding process. The process shall be managed in a technically sound, efficient, cost-effective manner, and shall be completed in a timely fashion.

#### **Introduction**

Since its creation by Congress in 1988, the Governing Board has been responsible for developing student achievement levels for NAEP assessments. The Governing Board has carried out this important statutory responsibility by engaging with a broad spectrum of stakeholders to develop student achievement levels.

---

<sup>1</sup> According to current NAEP legislation, the Governing Board shall develop achievement levels for all NAEP assessments except for the Long-Term Trend assessment.

Under provisions of the National Assessment of Educational Progress Authorization Act of 2002 (P.L. 107-279), Congress authorized the Governing Board to, develop, “achievement levels that are consistent with relevant widely accepted professional assessment standards and based on the appropriate level of subject matter knowledge” (Section 303(e)(2)(A)(i)(II)).

Given this mandate, the Governing Board must ensure that all achievement level setting processes align with current best practices in standard setting, and that appropriate validity evidence is collected and documented to support the intended uses and interpretations of NAEP achievement levels.

The Governing Board has established the following **policy definitions** for the NAEP achievement levels, as expectations of what students should know and be able to do. They shall be consistent across all assessments in which achievement levels are set.

***NAEP Basic***

*This level denotes partial mastery of prerequisite knowledge and skills that are fundamental for performance at the NAEP Proficient level.*

***NAEP Proficient***

*This level represents solid academic performance for each NAEP assessment. Students reaching this level have demonstrated competency over challenging subject matter, including subject-matter knowledge, application of such knowledge to real world situations, and analytical skills appropriate to the subject matter.*

***NAEP Advanced***

*This level signifies superior performance beyond NAEP Proficient.*

The Governing Board engages multiple stakeholders throughout the achievement level setting process, including:

- |                             |                                   |
|-----------------------------|-----------------------------------|
| Teachers                    | Policymakers                      |
| Curriculum Experts          | Business Representatives          |
| Content Experts             | Parents                           |
| Assessment Specialists      | Users of Assessment Data          |
| State Administrators        | Researchers and Technical Experts |
| Local School Administrators | Members of the Public             |

This policy also complies with the documents listed below which express widely accepted technical and professional standards for achievement level setting. These standards reflect the agreement of recognized experts in the field, as well as the policy positions of major professional and technical associations concerned with educational testing.

In conjunction with this policy the Board shall maintain a procedures manual to establish and document additional details about how this policy is to be implemented. As professional standards evolve and new consensus documents are released, this policy and the procedures manual shall be updated to the extent that new professional standards require. Resources for this purpose shall include, but not be limited to the following:

*The Standards for Educational and Psychological Testing.* (2014). Washington, DC: American Educational Research Association, American Psychological Association, and National Council on Measurement in Education;

*Code of Fair Testing Practices in Education.* (2004). Washington, DC: Joint Committee on Testing Practices;

*Educational Measurement (4<sup>th</sup> ed.).* (2006). R.L. Brennan (Ed.). Westport, CT: Praeger; and

*National Center for Education Statistics (NCES) Statistical Standards.* (2012).

## **Principles for Setting Achievement Levels**

**Principle 1: Elements of Achievement Levels**

**Principle 2: Development of Achievement Level Recommendations**

**Principle 3: Validation and Reporting of Achievement Level Results**

**Principle 4: Periodic Review of Achievement Levels**

**Principle 5: Stakeholder Input**

**Principle 6: Role of the Governing Board**

## **Principle 1: Elements of Achievement Levels**

**The Governing Board is responsible for developing student achievement levels for each NAEP assessment. Achievement levels for each NAEP assessment consist of content achievement level descriptions (ALDs), cut scores that demarcate adjacent levels, and exemplar items or tasks that illustrate performance at each level.**

- a) Content achievement level descriptions (ALDs) translate the policy definitions into specific 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. Content ALDs reflect the range of performance that items and tasks should measure. During the achievement level setting process, the purpose of content ALDs is to 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.
- b) Cut scores mark the minimum threshold score, the lower bound, for each achievement level. Performance within a given achievement level begins at the cut score for that level and ends just below the cut score for the successive achievement level.
- c) Exemplar items or tasks, including student responses, illustrate student performance within each of the achievement levels. They provide specific examples to help the public better understand what students in each achievement level know and can do.

## **Principle 2: Development of Achievement Level Recommendations**

**The Governing Board shall develop student achievement levels for NAEP, consistent with relevant widely accepted professional assessment standards, based on the appropriate level of subject matter knowledge.**

- a) A Design Document shall be developed at the beginning of the achievement level setting process, to describe in detail the scope of the achievement level setting project being undertaken, including but not limited to all planned materials, procedures, and analyses needed for the project. The Design Document shall be posted for public review with sufficient time to allow for a response from those who wish to provide one.
- b) The development of content achievement level descriptions (ALDs) shall be completed initially through the process that develops the assessment frameworks. (See the Governing Board Policy on Framework Development for additional details). The Board may then review and revise content ALDs to advance the purposes they serve, whether that is guiding an achievement level setting or informing the public about the meaning of achievement levels. Whether revised or not, the ALDs that guide

achievement level setting shall be articulated in terms of what students *should know and be able to do*. There shall be no content ALDs developed for performance below the *NAEP Basic* level.

- c) An achievement-level setting panel of subject matter experts shall be convened to recommend achievement level cut scores and exemplars.
- i. Each panel shall reflect diversity in terms of gender, race/ethnicity, region of the country, urbanicity, and experience with students with disabilities and English language learners. To ensure that they are qualified to make the judgments required by the achievement level setting process, individual panel members shall have expertise and experience in the specific content area in which the levels are being developed, expertise and experience in the education of students at the grade under consideration, and a general knowledge of assessment, curriculum, and student performance.
  - ii. Each panel shall include teachers, non-teacher educators, and other interested members of the general public with relevant educational background and experience. Teachers shall comprise the majority of the panel, with non-teacher educators (e.g., curriculum directors, academic coaches, principals) accounting for no more than half the number of teachers. The remaining panelists shall be non-educators who represent the perspectives of additional stakeholders representing the general public, including parents, researchers, and employers.
  - iii. The size of the panels shall reflect best practice in standard setting and be operationally feasible while being large enough to allow for split panels. Most NAEP achievement level settings have historically included approximately 20-30 panelists per grade, divided into two comparable groups.
- d) Panelists shall receive training on all aspects of the achievement levels setting process to ensure that panelists are well-prepared to perform the achievement level setting tasks required of them. Panelists shall be instructed that their role is to make achievement level recommendations to the Governing Board. Training shall include but not be limited to: the purpose and significance of setting achievement levels for NAEP; the NAEP assessment framework for the given subject area; and administration of a sample assessment under NAEP-like conditions that students experience. It is important for panelists to arrive at a common conceptualization of *NAEP Basic*, *NAEP Proficient*, and *NAEP Advanced* based on the content ALDs. Panelists shall be trained on each element of the judgmental task they perform, including the selection of exemplar items. They should be led by capable *content facilitators* (who are content experts and have previous experience with achievement level setting) and *process facilitators* (who have background in standard setting and experience leading panelists through the achievement level setting process). Facilitators shall take a neutral stance and not attempt to influence panelist judgments.

- e) The achievement level setting method that generates cut score recommendations shall have a solid research base and be appropriate for the content area, item types, number of items, scoring rubrics, and mode of administration, as applicable.
- f) Evaluations shall be administered to panelists throughout the achievement level setting process, in accordance with current best practices. Evaluations shall be part of every major component of the process, and panelists shall be asked to confirm their readiness for performing their tasks. Evaluation data may be used for formative purposes (to improve training and procedures in future meetings); summative purposes (to evaluate how well the process was conducted and provide procedural validity evidence); and to inform the Governing Board of any relevant information that could be useful when considering cut score recommendations. The panelists shall have an opportunity to indicate to the Board whether they believe the recommended cut scores are reasonable.
- g) In accordance with current best practices, feedback shall be provided to panelists, including “impact data” (i.e., the implications of their selected cut scores on the reported percentages of students at or above each achievement level).
- h) The process shall consist of at least two achievement level setting meetings with distinct groups of panelists, a pilot study, and an operational meeting. The purpose of the pilot study is to conduct a full “dress rehearsal” of the operational meeting, including but not limited to: an opportunity to try out materials, training procedures, collection of panelist judgments, feedback given to panelists through the process, software used to conduct analyses, meeting logistics, and other essential elements of the process. The pilot study may result in minor changes to the procedures, as well as major changes that would need additional study before being implemented in an operational meeting. The pilot study provides an opportunity for procedural validity evidence and to improve the operational meeting. At the discretion of the Governing Board, other smaller-scale studies may be conducted prior to the pilot study or in response to issues raised by the pilot study. The criteria in Principle 2a apply to panelists of both meetings.
- i) The Governing Board shall ensure that a Technical Advisory Committee on Standard Setting (TACSS) is convened to provide technical advice on all achievement level setting activities. Technical advice provided by standard setting experts throughout the project is intended to 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. The Board or its contractor may also seek technical advice from other groups as appropriate, including NCES and the larger measurement community (e.g., the National Council on Measurement in Education).
- j) All aspects of the procedures shall have documentation as evidence of the appropriateness of the procedures and results. This evidence shall be made available to the Board by the time of deliberations about the achievement levels. A summary of the evidence shall be available to the public when the achievement level results are reported.



- k) Sample items and student responses known as exemplars shall be chosen from the pool of released items for the current NAEP assessment to reflect performance in the *NAEP Basic*, *NAEP Proficient*, and *NAEP Advanced* regions of the scale. The use of exemplars is intended to help the public better understand what performance in each achievement level represents for each subject and grade. When possible, exemplars may also be chosen that reflect performance at threshold scores. The collection of exemplars shall reflect the content found in the achievement level descriptions and the range of item formats on the assessment.
- l) The outcomes from the achievement level setting panel meetings (recommended cut scores, exemplars, and ALDs for use in reporting) shall be forwarded to the Board for their consideration.

### **Principle 3: Validation and Reporting of Achievement Level Results**

**The achievement level setting process shall produce results that have validity evidence for the intended uses and interpretations and are informative to policy makers, educators, and the public.**

- a) Professional testing standards require evidence to support the intended interpretations and uses of test scores. Among the sources of evidence supporting the validity of test scores is evidence bearing on the standard setting process and results. Standard setting is necessarily judgmental, and the Board shall examine and consider available evidence about the procedural integrity of the achievement level setting process, the reasonableness of results, and other evidence in order to support intended uses and interpretations.
- b) The Board shall examine and consider all evidence related to validity of the achievement level setting activities. These data shall include, but not be limited to: procedural evidence such as training, materials and panelist evaluation data; reliability evidence such as consistency across panelist type, subpanels, rounds, and meetings, if appropriate; and external comparisons to other similar assessments, if appropriate, with necessary caveats. The results from validation efforts shall be made available to the Board in a timely manner so that the Board has access to as much validation data as possible as it considers the recommendations regarding the final levels.
- c) NAEP achievement levels are intended to estimate the percentage of students (overall and for selected student groups) in each achievement level category, for the nation, and for states and trial urban districts (TUDAs) for some assessments. NAEP is prohibited by law from reporting any results for individual students or schools.
- d) In describing student performance using the achievement levels, terms such as “students performing at the *NAEP Basic* level” or “students performing at the *NAEP Proficient* level” are preferred over “*Basic* students” or “*Proficient* students”. The former implies that students have mastery of particular content represented by the achievement levels, while the latter implies an inherent

characteristic of individual students.

- e) In reporting the results of NAEP, the three achievement levels of *NAEP Basic*, *NAEP Proficient*, and *NAEP Advanced* refer to the three regions of the NAEP scale at and above each respective cut score. The remaining region that falls below the *NAEP Basic* cut score shall be identified as “below *NAEP Basic*” when a descriptor is necessary.
- f) In describing the *NAEP Proficient* level, reports shall emphasize that the policy definition is not intended to reflect “grade level” performance expectations, which are typically defined normatively and can vary widely by state and over time. *NAEP Proficient* may convey a different meaning from other uses of the term “proficient” in common terminology or in reference to other assessments.
- g) To facilitate valid uses of ALDs for the purpose of reporting, the Board shall ensure that the descriptions of performance for the achievement levels reflect what the empirical data reveal about the knowledge and skills demonstrated by students in that score range. To develop ALDs for reporting, following the achievement level setting the Board shall revisit and may revise content ALDs to ensure that they are consistent with empirical evidence of student performance. In particular, these “Reporting ALDs” chosen to illustrate the knowledge and skills demonstrated at different achievement levels shall be written to incorporate empirical data from student performance. Reporting ALDs shall describe what students at each level *do* know and *can* do rather than what they *should* know and *should* be able to do.
- h) An interpretative guide shall accompany NAEP reports, including specific examples of appropriate and inappropriate interpretations and uses of the results.

#### **Principle 4: Periodic Review of Achievement Levels**

**Periodic reviews of existing achievement levels shall determine whether new achievement level descriptions and/or cut scores are needed to continue valid and reliable measurement of current student performance and trends over time.**

- a) At least once every 10 years or 3 administrations of an assessment, whichever comes later, the Governing Board, through its Committee on Standards, Design and Methodology (COSDAM), shall review the alignment between the content ALDs and items, based on empirical data from recent administrations of NAEP assessments. In its review, COSDAM (in consultation with the Assessment Development Committee) shall solicit input from technical and subject matter experts to determine whether changes to the content ALDs are warranted or whether a new standard setting shall be conducted, making clear the potential risk of changing cut scores to trends and assessment of educational progress. Relevant factors may include but not be limited to: substantive changes in the item types or in the balance of item types; changes in the mode of administering assessments; advances in standard setting methodologies; and changes in the policy environment for using NAEP results.

- b) Within the period for a review of achievement level descriptions and cut scores, changes may occur to a NAEP framework. If a framework is replaced or revised for a major update, a new achievement level setting process may be implemented, except in circumstances where scale score trends are maintained. In this latter instance, COSDAM shall determine how to revise the ALDs and review the cut scores to ensure that they remain reasonable and meaningful.
- c) If there are major updates to a NAEP framework, the ALDs shall be updated by the Framework Visioning and Development Panel. (See the Governing Board Policy on Framework Development for additional details). Following an assessment administration under the revised framework, COSDAM shall use empirical data to revise content ALDs to align with the revised framework.
- d) As additional validation evidence becomes available, the Board shall review it and make a determination about whether the achievement levels should be reviewed and potentially revised.

## **Principle 5: Stakeholder Input**

**The process of developing student achievement levels is a widely inclusive activity. The Governing Board shall provide opportunities to engage multiple stakeholders throughout the achievement level setting process and shall strive to maximize transparency of the process.**

- a) The process of seeking nominations for the achievement level setting panels shall include outreach to relevant constituencies, such as: state and local educators; curriculum specialists; business representatives; and professional associations in a given content area.
- b) The Design Document (describing in detail all planned procedures for the project) shall be distributed for review by a broad constituency and shall be disseminated in sufficient time to allow for a thoughtful response from those who wish to provide one. All interested stakeholders shall have an opportunity to provide public comment.
- c) Achievement level setting panelists shall include teachers, non-teacher educators, and other interested members of the general public with relevant educational background and experience, including parents, researchers, and employers. Each panel shall reflect diversity in terms of gender, race/ethnicity, region of the country, urbanicity, and experience with students with disabilities and English language learners.
- d) All achievement level setting activities shall be informed by technical advice throughout the process. The Technical Advisory Committee on Standard Setting shall provide ongoing technical input from standard setting and assessment experts, and other groups with relevant technical expertise may be consulted periodically as needed.

- e) Ongoing input and coordination with staff and contractors from the National Center for Education Statistics (NCES) is necessary to ensure that all achievement level setting activities are carried out in a manner that is consistent with the design, analysis, and reporting of NAEP assessments.

## **Principle 6: Role of the Governing Board**

**The Governing Board, through its Committee on Standards, Design and Methodology (COSDAM), shall monitor the development and review of student achievement levels to ensure that the final achievement level descriptions, cut scores, and exemplars recommended to the Governing Board for adoption comply with this policy.**

- a) The Committee on Standards, Design and Methodology (COSDAM) shall be responsible for monitoring the development and review of achievement levels that result in recommendations to the Governing Board for any NAEP assessment under consideration. COSDAM shall provide direction to the achievement level setting contractor, via Governing Board staff. This guidance shall ensure compliance with the NAEP legislation, Governing Board policies, Department of Education and government-wide regulations, and requirements of the contract(s) used to implement the achievement level setting project.
- b) If there is a need to revise the initial achievement level descriptions (ALDs) created at the time of framework development for use in achievement level setting and/or reporting, the Governing Board shall take final action on revised ALDs based on recommendations from COSDAM.
- c) COSDAM shall receive regular reports on the progress of achievement level setting projects.
- d) COSDAM shall review and formally approve the Design Document that describes all planned procedures for an achievement level setting project.
- e) At the conclusion of the achievement level setting project, the Governing Board shall take final action on the recommended cut scores, exemplars, and ALDs for use in reporting. The Governing Board shall make the final determination on the NAEP achievement levels. In addition to the panel recommendations, the Board may consider other pertinent information to assess reasonableness of the results, such as comparisons to other relevant assessments.
- f) Following adoption by the Governing Board, the final ALDs, cut scores, and exemplars shall be provided to the National Center for Education Statistics (NCES) for reporting the results of the NAEP assessment(s) under consideration.
- g) Consistent with Principle 4 above, COSDAM shall periodically review existing achievement levels to determine whether it is necessary to revise achievement level descriptions or conduct a new standard setting.



**Adopted: TBD**

## **National Assessment Governing Board**

### **Developing Student Achievement Levels for the National Assessment of Educational Progress**

#### **Policy Statement**

It is the policy of the National Assessment Governing Board to conduct a comprehensive, inclusive, and deliberative process to develop and review student achievement levels for the National Assessment of Educational Progress (NAEP).<sup>1</sup> Achievement levels consist of general policy definitions for the *NAEP Basic*, *NAEP Proficient*, and *NAEP Advanced* levels, specific achievement level descriptions (ALDs) for each assessment, cut scores that demarcate adjacent levels, and exemplar items or tasks that illustrate performance at each level. This process shall be conducted according to widely accepted professional standards, to produce results that are reasonable, useful, and informative to the public.

The Governing Board, through its Committee on Standards, Design and Methodology (COSDAM), shall monitor the development and review of student achievement levels to ensure that the final Governing Board-adopted achievement level descriptions, cut scores, and exemplars comply with all principles of this policy.

The achievement level setting process shall be carried out by contractors selected through a competitive bidding process. The process shall be managed in a technically sound, efficient, cost-effective manner, and shall be completed in a timely fashion.

#### **Introduction**

Since its creation by Congress in 1988, the Governing Board has been responsible for developing student achievement levels for NAEP assessments. The Governing Board has carried out this important statutory responsibility by engaging with a broad spectrum of stakeholders to develop student achievement levels.

---

<sup>1</sup> According to current NAEP legislation, the Governing Board shall develop achievement levels for all NAEP assessments except for the Long-Term Trend assessment.

Under provisions of the National Assessment of Educational Progress Authorization Act of 2002 (P.L. 107-279), Congress authorized the Governing Board to, develop, “achievement levels that are consistent with relevant widely accepted professional assessment standards and based on the appropriate level of subject matter knowledge ~~for grade levels to be assessed~~” (Section 303(e)(2)(A)(i)(II)).

**Comment [RS1]:** Some comments state that the NAEP achievement levels should be abandoned.

The NAEP achievement levels are required by law and the Governing Board does not have the authority to eliminate them.

Given this mandate, the Governing Board must ensure that all achievement level setting processes align with current best practices in standard setting, and that appropriate validity evidence is collected and documented to support the intended uses and interpretations of NAEP achievement levels.

The Governing Board has established the following **policy definitions** for the NAEP achievement levels, as expectations of what students should know and be able to do. They shall be consistent across all assessments in which achievement levels are set.

**Comment [RS2]:** There were several comments related to the policy labels and definitions. COSDAM recommends the labels and policy definitions described in this policy revision and believes the Board should focus its efforts on better communicating these labels and implementing other activities related to the Board’s formal response to the Evaluation of NAEP Achievement Levels.

**NAEP Basic**

*This level denotes partial mastery of prerequisite knowledge and skills that are fundamental for performance at the NAEP Proficient level.*

**NAEP Proficient**

*This level represents solid academic performance for each NAEP assessment. Students reaching this level have demonstrated competency over challenging subject matter, including subject-matter knowledge, application of such knowledge to real world situations, and analytical skills appropriate to the subject matter.*

**NAEP Advanced**

*This level signifies superior performance beyond NAEP Proficient.*

The Governing Board engages multiple stakeholders throughout the achievement level setting process, including:

- |                             |                                   |
|-----------------------------|-----------------------------------|
| Teachers                    | Policymakers                      |
| Curriculum Experts          | Business Representatives          |
| Content Experts             | Parents                           |
| Assessment Specialists      | Users of Assessment Data          |
| State Administrators        | Researchers and Technical Experts |
| Local School Administrators | Members of the Public             |

This policy also complies with the documents listed below which express widely accepted technical and professional standards for achievement level setting. These standards reflect the agreement of recognized experts in the field, as well as the policy positions of major professional and technical associations concerned with educational testing.

In conjunction with this policy the Board shall maintain a procedures manual to establish and document additional details about how this policy is to be implemented. As professional standards evolve and new consensus documents are released, this policy and the procedures manual shall be updated to the extent that new professional standards require. Resources for this purpose shall include, but not be limited to the following:

*The Standards for Educational and Psychological Testing.* (2014). Washington, DC: American Educational Research Association, American Psychological Association, and National Council on Measurement in Education;

*Code of Fair Testing Practices in Education.* (2004). Washington, DC: Joint Committee on Testing Practices;

*Educational Measurement (4<sup>th</sup> ed.).* (2006). R.L. Brennan (Ed.). Westport, CT: Praeger; and

*National Center for Education Statistics (NCES) Statistical Standards.* (2012).

## **Principles for Setting Achievement Levels**

**Principle 1: Elements of Achievement Levels**

**Principle 2: Development of Achievement Level Recommendations**

**Principle 3: Validation and Reporting of Achievement Level Results**

**Principle 4: Periodic Review of Achievement Levels**

**Principle 5: Stakeholder Input**

**Principle 6: Role of the Governing Board**



## Principle 1: Elements of Achievement Levels

**The Governing Board is responsible for developing student achievement levels for each NAEP assessment. Achievement levels for each NAEP assessment consist of content achievement level descriptions (ALDs), cut scores that demarcate adjacent levels, and exemplar items or tasks that illustrate performance at each level.**

- a) Content achievement level descriptions (ALDs) translate the policy definitions into specific 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. Content ALDs reflect the range of performance that items and tasks should measure. During the achievement level setting process, the purpose of content ALDs is to 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.
- b) Cut scores mark the minimum threshold score, the lower bound, for each achievement level. Performance within a given achievement level begins at the cut score for that level and ends just below the cut score for the successive achievement level.
- c) Exemplar items or tasks, including student responses, illustrate student performance within each of the achievement levels. They provide specific examples to help the public better understand what students in each achievement level know and can do.

## Principle 2: Development of Achievement Level Recommendations

**The Governing Board shall develop student achievement levels for NAEP, consistent with relevant widely accepted professional assessment standards, based on the appropriate level of subject matter knowledge.**

- a) A Design Document shall be developed at the beginning of the achievement level setting process, to describe in detail the scope of the achievement level setting project being undertaken, including but not limited to all planned materials, procedures, and analyses needed for the project. The Design Document shall be posted for public review with sufficient time to allow for a response from those who wish to provide one.
- b) The development of content achievement level descriptions (ALDs) shall be completed initially through the process that develops the assessment frameworks. (See the Governing Board Policy on Framework Development for additional details). The Board may then review and revise content ALDs to advance the purposes they serve, whether that is guiding an achievement level setting or informing the public about the meaning of achievement levels. Whether revised or not, the ALDs that guide

**Comment [RS3]:** There are some comments suggesting that additional details be added about procedures, but COSDAM has previously decided that this level of detail is more appropriate for a procedures manual.

achievement level setting shall be articulated in terms of what students *should know and be able to do*. There shall be no content ALDs developed for performance below the *NAEP Basic* level.

- c) An achievement-level setting panel of subject matter experts shall be convened to recommend achievement level cut scores and exemplars.
- i. Each panel shall reflect diversity in terms of gender, race/ethnicity, region of the country, urbanicity, and experience with students with disabilities and English language learners. To ensure that they are qualified to make the judgments required by the achievement level setting process, individual panel members shall have expertise and experience in the specific content area in which the levels are being developed, expertise and experience in the education of students at the grade under consideration, and a general knowledge of assessment, curriculum, and student performance.
  - ii. ~~This Each~~ panel shall include teachers, non-teacher educators, and other interested members of the general public with relevant educational background and experience. Teachers shall comprise the majority of the panel, with non-teacher educators (e.g., curriculum directors, academic coaches, principals) accounting for no more than half the number of teachers. The remaining panelists shall be non-educators who represent the perspectives of additional stakeholders representing the general public, including parents, researchers, and employers.
  - iii. The size of the panels shall reflect best practice in standard setting and be operationally feasible while being large enough to allow for split panels. Most NAEP achievement level settings have historically included approximately 20-30 panelists per grade, divided into two comparable groups ~~with a subset of shared items~~.
- d) Panelists shall receive training on all aspects of the achievement levels setting process to ensure that panelists are well-prepared to perform the achievement level setting tasks required of them. Panelists shall be instructed that their role is to make achievement level recommendations to the Governing Board. Training shall include but not be limited to: the purpose and significance of setting achievement levels for NAEP; the NAEP assessment framework for the given subject area; and administration of a sample assessment under NAEP-like conditions that students experience. It is important for panelists to arrive at a common conceptualization of *NAEP Basic*, *NAEP Proficient*, and *NAEP Advanced* based on the content ALDs. Panelists shall be trained on each element of the judgmental task they perform, including the selection of exemplar items. They should be led by capable *content facilitators* (who are content experts and have previous experience with achievement level setting) and *process facilitators* (who have background in standard setting and experience leading panelists through the achievement level setting process). Facilitators shall take a neutral stance and not attempt to influence panelist judgments.

**Comment [RS4]:** There was a suggestion to delete this phrase or provide more information by saying that the item pool is also divided into two equivalent sets with a subset evaluated by both panel groups. The latter is information that seems more appropriate for a procedures manual.

- e) The achievement level setting method that generates cut score recommendations ~~may differ depending upon the specific assessment. Nevertheless, the method~~ shall have a solid research base and be appropriate for the content area, item types, number of items, scoring rubrics, and mode of administration, as applicable.
- f) Evaluations shall be administered to panelists throughout the achievement level setting process, in accordance with current best practices. Evaluations shall be part of every major component of the process, and panelists shall be asked to confirm their readiness for performing their tasks. Evaluation data may be used for formative purposes (to improve training and procedures in future meetings); summative purposes (to evaluate how well the process was conducted and provide procedural validity evidence); and to inform the Governing Board of any relevant information that could be useful when considering cut score recommendations. The panelists shall have an opportunity to indicate to the Board whether they believe the recommended cut scores are reasonable.
- g) In accordance with current best practices, feedback shall be provided to panelists, including “impact data” (i.e., the implications of their selected cut scores on the reported percentages of students at or above each achievement level).
- h) The process shall consist of at least two achievement level setting meetings with distinct groups of panelists, a pilot study, and an operational meeting. The purpose of the pilot study is to conduct a full “dress rehearsal” of the operational meeting, including but not limited to: an opportunity to ~~test-try~~ out materials, training procedures, collection of panelist judgments, feedback given to panelists through the process, software used to conduct analyses, meeting logistics, and other essential elements of the process. The pilot study may result in minor changes to the procedures, as well as major changes that would need additional study before being implemented in an operational meeting. The pilot study provides an opportunity for procedural validity evidence and to improve the operational meeting. At the discretion of the Governing Board, other smaller-scale studies may be conducted prior to the pilot study or in response to issues raised by the pilot study. The criteria in Principle 2a apply to panelists of both meetings.
- i) The Governing Board shall ensure that a Technical Advisory Committee on Standard Setting (TACSS) is convened to provide technical advice on all achievement level setting activities. Technical advice provided by standard setting experts throughout the project is intended to 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. The Board or its contractor may also seek technical advice from other groups as appropriate, including NCES and the larger measurement community (e.g., the National Council on Measurement in Education).
- j) All aspects of the procedures shall have documentation as evidence of the appropriateness of the procedures and results. This evidence shall be made available to the Board ~~by~~ at the time of deliberations about the achievement levels. A summary of the evidence shall be available to the public when the achievement level results are reported.

**Comment [RS5]:** The proposed edit is intended to address a concern that “differ” is vague and does not explain what the recommendations differ from.

- k) Sample items and student responses known as exemplars shall be chosen from the pool of released items for the current NAEP assessment to reflect performance in the *NAEP Basic*, *NAEP Proficient*, and *NAEP Advanced* regions of the scale. The use of exemplars is intended to help the public better understand ~~what students who are in performance in~~ each achievement level ~~actually know and are able to~~ ~~represent~~ for each subject and grade. When possible, exemplars may also be chosen that reflect performance at threshold scores. The collection of exemplars shall reflect the content found in the achievement level descriptions and the range of item formats on the assessment.
- l) The outcomes from the achievement level setting panel meetings (recommended cut scores, exemplars, and ALDs for use in reporting) shall be forwarded to the Board for their consideration.

**Comment [RS6]:** This proposed edit is intended to address a comment about a very strong notion that it is incorrect/inappropriate to refer to students in the achievement levels. Rather, one was to talk about the performance of students classified in the achievement levels and to stay that the performance represents the knowledge and skills required by the ALDs.

### Principle 3: Validation and Reporting of Achievement Level Results

**The achievement level setting process shall produce results that have validity evidence for the intended uses and interpretations and are informative to policy makers, educators, and the public.**

- a) Professional testing standards require evidence to support the intended interpretations and uses of test scores. Among the sources of evidence supporting the validity of test scores is evidence bearing on the standard setting process and results. ~~Although Standard setting is necessarily judgmental, and with no “true” or “correct” cut scores,~~ the Board shall examine and consider available evidence about the procedural integrity of the achievement level setting process, the reasonableness of results, and other evidence in order to support intended uses and interpretations.
- b) ~~The Board shall examine and consider all evidence related to validity of the achievement level setting activities. These data shall include, but not be limited to: procedural evidence such as training, materials and panelist evaluation data; reliability evidence such as consistency across panelist type, subpanels, rounds, and meetings, if appropriate; and external comparisons to other similar assessments, if appropriate, with necessary caveats. The results from validation efforts shall be made available to the Board in a timely manner so that the Board has access to as much validation data as possible as it considers the recommendations regarding the final levels.~~
- ~~b) NAEP achievement levels are intended to estimate the percentage of students (overall and for selected student groups) in each achievement level category, for the nation, and for states and trial urban districts (TUDAs) for some assessments. NAEP is prohibited by law from reporting any results for individual students or schools, so achievement levels do not apply to individual students or schools.~~
- c)
- ~~d) To facilitate valid uses of ALDs for reporting, the Board shall ensure that the~~

**Comment [RS7]:** This proposed edit is intended to address a comment that the Board should not apologize for the judgmental nature of standard setting – the reason there is a Governing Board is to make these important judgments.

**Comment [RS8]:** This paragraph has been re-ordered per COSDAM discussion

~~descriptions of performance for the achievement levels reflect what the empirical data reveal about the knowledge and skills of students in that score range. The Board shall revisit and may revise content ALDs following the achievement level setting to ensure that they are consistent with empirical evidence of student performance. In particular, when content ALDs are reported with results, they shall be written to incorporate empirical data from student performance. They shall describe what students at each level *do* know and *can* do rather than what they *should* know and *should* be able to do.~~

~~e) The Board shall examine and consider all evidence related to validity of the achievement level setting activities. These data shall include, but not be limited to: procedural evidence such as training, materials and panelist evaluation data; reliability evidence such as consistency across panelist type, subpanels, rounds, and meetings, if appropriate; and external comparisons to other similar assessments, if appropriate, with necessary caveats. The results from validation efforts shall be made available to the Board in a timely manner so that the Board has access to as much validation data as possible as it considers the recommendations regarding the final levels.~~

~~d) In describing student performance using the achievement levels, terms such as “students performing at the *NAEP Basic* level” or “students performing at the *NAEP Proficient* level” are preferred over “*Basic* students” or “*Proficient* students”. The former implies that students have mastery of particular content represented by the achievement levels, while the latter implies an inherent characteristic of individual students.~~

~~e) In reporting the results of NAEP, the three achievement levels of *NAEP Basic*, *NAEP Proficient*, and *NAEP Advanced* refer to the three regions of the NAEP scale at and above each respective cut score. The remaining region that falls below the *NAEP Basic* cut score shall be identified as “below *NAEP Basic*” when a descriptor is necessary.~~

~~f) In describing the *NAEP Proficient* level, reports shall emphasize that the policy definition is not intended to reflect “grade level” performance expectations, which are typically defined normatively and can vary widely by state and over time. *NAEP Proficient* may convey a different meaning from other uses of the term “proficient” in common terminology or in reference to other assessments.~~

g) To facilitate valid uses of ALDs for the purpose of reporting, the Board shall ensure that the descriptions of performance for the achievement levels reflect what the empirical data reveal about the knowledge and skills demonstrated by students in that score range. To develop ALDs for reporting, following the achievement level setting the Board shall revisit and may revise content ALDs to ensure that they are consistent with empirical evidence of student performance. In particular, these “Reporting ALDs” chosen to illustrate the knowledge and skills demonstrated at different achievement levels shall be written to incorporate empirical data from student performance. Reporting ALDs shall describe what students at each level *do* know and *can* do rather than what they *should* know and *should* be able to do.

**Comment [RS9]:** This paragraph has been re-ordered per COSDAM discussion

**Comment [RS10]:** The edits in this paragraph are intended to address a suggestion to refer to “Reporting ALDs” to be consistent with standard setting literature.

- h) An interpretative guide shall accompany NAEP reports, including specific examples of appropriate and inappropriate interpretations and uses of the results.

#### **Principle 4: Periodic Review of Achievement Levels**

**Periodic reviews of existing achievement levels shall determine whether new achievement level descriptions and/or cut scores are needed to continue valid and reliable measurement of current student performance and trends over time.**

- a) At least once every 10 years or 3 administrations of an assessment, whichever comes later, the Governing Board, through its Committee on Standards, Design and Methodology (COSDAM), shall review the alignment between the content ALDs and items, based on empirical data from recent administrations of NAEP assessments. In its review, COSDAM (in consultation with the Assessment Development Committee) shall solicit input from technical and subject matter experts to determine whether changes to the content ALDs are warranted or whether a new standard setting shall be conducted, making clear the potential risk of changing cut scores to trends and assessment of educational progress. Relevant factors may include but not be limited to: substantive changes in the item types or in the balance of item types; changes in the mode of administering assessments; advances in standard setting methodologies; and changes in the policy environment for using NAEP results.
- b) Within the period for a review of achievement level descriptions and cut scores, changes may occur to a NAEP framework. If a framework is replaced or revised for a major update, a new achievement level setting process may be implemented, except in circumstances where scale score trends are maintained. In this latter instance, COSDAM shall determine how to revise the ALDs and review the cut scores to ensure that they remain reasonable and meaningful.
- c) If there are major updates to a NAEP framework, the ALDs shall be updated by the Framework Visioning and Development Panel. (See the Governing Board Policy on Framework Development for additional details). Following an assessment administration under the revised framework, COSDAM shall use empirical data to revise content ALDs to align with the revised framework.
- d) As additional validation evidence becomes available, the Board shall review it and make a determination about whether the achievement levels should be reviewed and potentially ~~revised.~~

#### **Principle 5: Stakeholder Input**

**The process of developing student achievement levels is a widely inclusive activity. The Governing Board shall provide opportunities to engage multiple stakeholders throughout the achievement level setting process and shall strive to**

**maximize transparency of the process.**

- a) The process of seeking nominations for the achievement level setting panels shall include outreach to relevant constituencies, such as: state and local educators; curriculum specialists; business representatives; and professional associations in a given content area.
- b) The Design Document (describing in detail all planned procedures for the project) shall be distributed for review by a broad constituency and shall be disseminated in sufficient time to allow for a thoughtful response from those who wish to provide one. All interested stakeholders shall have an opportunity to provide public comment.
- c) Achievement level setting panelists shall include teachers, non-teacher educators, and other interested members of the general public with relevant educational background and experience, including parents, researchers, and employers. Each panel shall reflect diversity in terms of gender, race/ethnicity, region of the country, urbanicity, and experience with students with disabilities and English language learners.
- d) All achievement level setting activities shall be informed by technical advice throughout the process. The Technical Advisory Committee on Standard Setting shall provide ongoing technical input from standard setting and assessment experts, and other groups with relevant technical expertise may be consulted periodically as needed.
- e) Ongoing input and coordination with staff and contractors from the National Center for Education Statistics (NCES) is necessary to ensure that all achievement level setting activities are carried out in a manner that is consistent with the design, analysis, and reporting of NAEP assessments.

~~f) The Governing Board may ask its standing groups representing various constituencies to provide input on the achievement level setting process.~~

**Comment [RS11]:** This proposed deletion is intended to address a comment that it is not necessary to state this.

**Principle 6: Role of the Governing Board**

**The Governing Board, through its Committee on Standards, Design and Methodology (COSDAM), shall monitor the development and review of student achievement levels to ensure that the final achievement level descriptions, cut scores, and exemplars recommended to the Governing Board ~~for-adopted achievement level descriptions, cut scores, and exemplars~~ comply with this policy.**

- a) The Committee on Standards, Design and Methodology (COSDAM) shall be responsible for monitoring the development and review of achievement levels that result in recommendations to the Governing Board for any NAEP assessment under consideration. COSDAM shall provide direction to the achievement level setting contractor, via Governing Board staff. This guidance shall ensure compliance with the NAEP legislation, Governing Board policies, Department of Education and government-wide regulations, and requirements of the contract(s)

used to implement the achievement level setting project.

- b) If there is a need to revise the initial achievement level descriptions (ALDs) created at the time of framework development for use in achievement level setting and/or reporting, the Governing Board shall take final action on revised ALDs based on recommendations from COSDAM.
- c) COSDAM shall receive regular reports on the progress of achievement level setting projects.
- d) COSDAM shall review and formally approve the Design Document that describes all planned procedures for an achievement level setting project.
- e) At the conclusion of the achievement level setting project, the Governing Board shall take final action on the recommended cut scores, exemplars, and ALDs for use in reporting. The Governing Board shall make the final determination on the NAEP achievement levels. In addition to the panel recommendations, the Board may consider other pertinent information to assess reasonableness of the results, such as comparisons to other similar-relevant assessments.
- f) Following adoption by the Governing Board, the final ALDs, cut scores, and exemplars shall be provided to the National Center for Education Statistics (NCES) for reporting the results of the NAEP assessment(s) under consideration.
- g) Consistent with Principle 4 above, COSDAM shall periodically review existing achievement levels to determine whether it is necessary to revise achievement level descriptions or conduct a new standard setting.





Council of the  
Great City Schools®

**MEMBER  
DISTRICTS**

- Albuquerque
- Anchorage
- Arlington, TX
- Atlanta
- Austin
- Baltimore
- Birmingham
- Boston
- Bridgeport
- Broward County
- Buffalo
- Charleston County
- Charlotte-Mecklenburg
- Chicago
- Cincinnati
- Clark County
- Cleveland
- Columbus
- Dallas
- Dayton
- Denver
- Des Moines
- Detroit
- Duval County
- El Paso
- Fort Worth
- Fresno
- Guilford County
- Hawaii
- Hillsborough County
- Houston
- Indianapolis
- Jackson
- Jefferson County, KY
- Kansas City
- Long Beach
- Los Angeles
- Miami-Dade County
- Milwaukee
- Minneapolis
- Nashville
- New Orleans
- New York City
- Newark
- Norfolk
- Oakland
- Oklahoma City
- Omaha
- Orange County, FL
- Palm Beach County
- Philadelphia
- Pinellas County
- Pittsburgh
- Portland
- Providence
- Richmond
- Rochester
- Sacramento
- San Antonio
- San Diego
- San Francisco
- Seattle
- Shelby County
- St. Louis
- St. Paul
- Toledo
- Tulsa
- Washington, D.C.
- Wichita

**Council of the Great City Schools®**

1331 Pennsylvania Avenue, NW, Suite 1100N, Washington, DC 20004

(202) 393-2427 (202) 393-2400 (fax) www.cgcs.org

September 29, 2018

National Assessment Governing Board  
U.S. Department of Education  
800 North Capitol Street NW – Suite 825  
Washington, DC 20002-4233

**Attention:** Sharyn Rosenberg, Assistant Director

**Comments on Revised Policy on Developing Student Achievement Levels for NAEP**  
**Document Citation: 83 FR 45618**

The Council of the Great City Schools (Council), the coalition of the nation’s largest central city school districts, submits the following comments on the revised policy on Developing Student Achievement Levels for the National Assessment of Educational Progress (NAEP) in response to the September 10, 2018 notice in the Federal Register. Over the years, the Council has worked closely with the National Assessment Governing Board (Governing Board) and the National Center for Education Statistics (NCES) on a variety of efforts to measure and improve student learning outcomes. Therefore, the Council supports the Governing Board’s efforts to periodically review and update NAEP policies and practices.

The Council is dedicated to the improvement of education for children in the nation’s inner cities. The Council and its member districts work to help our public-school children meet the highest standards and become successful and productive members of society. The organization and the 27 Trial Urban District Assessment (TUDA) participants regularly use results from NAEP to measure our progress and relative standing in achieving our goals. In fact, the Council of the Great City Schools initiated TUDA in 2000 as a way of holding ourselves and our students to the highest standards. As a result, we are heavily invested in any changes in policies and practices related to NAEP and other national measures of educational progress.

The Council’s comments in this letter are focused on retaining a rigorous assessment, maintaining current terminology in NAEP achievement levels, thoughtfully tagging NAEP performance levels to meaningful high-level expectations, and devoting additional effort to explaining to the public what NAEP is intended to measure and what it is not. Please let us know if clarification is needed on any of these comments.

Sincerely,

Michael Casserly  
Executive Director

## **COUNCIL OF THE GREAT CITY SCHOOLS COMMENTS ON THE REVISED POLICY ON DEVELOPING STUDENT ACHIEVEMENT LEVELS FOR THE NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS**

### **New Principle on Periodic Review of Achievement Levels and Cut Scores**

The Council of the Great City Schools agrees that the National Assessment Governing Board (NAGB) and the National Center for Educational Statistics (NCES) should periodically review the performance levels used in the National Assessment of Educational Progress (NAEP). These periodic reviews, however, should be tempered by the continuing need to have a national assessment that is consistently applied from state-to-state, a national assessment that allows comparisons across states and participating TUDA participants, and a national assessment that provides a clear trend line over time for the nation, states, and districts.

That said, the Council is also interested in seeing NAGB and NCES devote additional time and effort to benchmarking NAEP performance levels to such concrete high-level expectations as current college- and career-readiness standards, international measures of performance, average entrance requirements to competitive colleges, or other publicly understandable measures of excellence, rather than the judgment of expert panels, even if the result is somewhat aspirational. We believe the standard against which NAEP is pegged ought to be rigorous and not reflect the lowest common denominator as some have argued. Above all, NAEP is a measure of how the nation, its states, and many of its critical large city school systems are performing educationally; it is not an accountability tool or an assessment of individual attainment. To that end, NAEP should be pegged to the highest possible but specific standard of attainment.

We also encourage the Governing Board NOT to adjust cut scores or descriptions without extensive research or overwhelmingly compelling evidence to suggest misalignment with desired interpretations of NAEP results. The Council and our member districts that participate in the TUDA program find the current achievement level cut scores for proficiency to be a fair and accurate expectation of student outcomes. We believe our schools, teachers, and administrators should have the highest expectations for student achievement, and we believe the current level of proficiency reflects a realistic, although high, expectation for our students. Students are harmed when expectations are low, not high, a situation that urban schools are more aware of than many others.

While additional, more robust, research is needed on the current proficiency levels, a cursory review of post-secondary outcomes suggests current proficiency standards are basically sound. According to the 2017 Current Population Survey from the U.S. Census Bureau,<sup>1</sup> about 37 percent of adults 25 to 34 years old had a bachelor's degree or higher. On the 2007 eighth grade NAEP assessment, about the same time those 25-year olds would have been in eighth grade, approximately 32 percent the nation's students were proficient in reading and math.<sup>2</sup> By design or not, NAEP appears on its face to have some grounding in real-world attainment.

---

<sup>1</sup> Source: U.S. Census Bureau, Current Population Survey, 2017 Annual Social and Economic Supplement.

<sup>2</sup> Source: U. S. Department of Education. Institute of Education Sciences, National Center for Education Statistics.

Arguments in favor of shifting NAEP achievement level cut scores masks the real problem in educational attainment across the country. In 2017, over 57 percent of students not eligible for a free or reduced-price lunch were proficient or better on the fourth grade NAEP mathematics assessment compared to 24.7 percent of students eligible for a free or reduced-price lunch.<sup>3</sup> The disparities are even more discouraging when race and poverty are considered jointly. Advocates calling for the lowering of proficiency standards shift the nation’s attention away from the real issues about disparities in educational outcomes among the nation’s impoverished and traditionally under-represented student groups in favor of making more students looking artificially higher performing.

Finally, the Council does wonder whether there is a way to retain the current performance levels but expand use of NAEP’s 500-point scale. At present, scores routinely fall between points 200 and 300 on the scale, making it difficult to show movement—either upwards or downwards. A robust discussion about the wisdom of this seems prudent in the current review.

Recommendation: The Council recommends retaining NAEP’s high level of rigor in defining proficiency, but it also proposes that NAGB and NCES conduct additional, robust research to better tie those proficiency levels to college- and career-ready standards, some international benchmark, or to post-secondary success.

## Change in Terminology from Proficient to NAEP Proficient

The Council also does not think that the Governing Board should change the terminology used by NAEP from *Proficient* to *NAEP Proficient*, from *Basic* to *NAEP Basic*, or from *Advanced* to *NAEP Advanced*. We think the current labels should remain for several reasons. First, the summary of proposed revisions suggests that NAEP achievement levels should be “better differentiate[d]” from other common uses of the terms *Basic*, *Proficient*, and *Advanced*. For decades, however, NAEP has been, and should remain, the standard for these terms. Application of these terms from assessment-to-assessment have been made relative to NAEP definitions—even if they have not been faithfully applied. Changing the terminology suggests that NAEP should no longer be the standard upon which we understand student achievement.

Second, introducing NAEP-specific Basic, Proficient, and Advanced levels might introduce considerable confusion to the public’s understanding of student achievement across assessments. In a review of NAEP Achievement Levels conducted by the National Academies of Science, Engineering, and Medicine, they conclude 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; p. Sum-8)<sup>3</sup>. The public’s understanding of current terminology is well entrenched and already commonly understood. Parents, educators, and the public are better served when the educational community can consistently articulate student achievement outcomes. The Council encourages NAGB to retain current terminology without the modifying

---

<sup>3</sup> 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.

terminology, because NAEP should remain the standard against which these proficiency levels should be defined.

Recommendation: In 83 FR 45618 strike all references to “*NAEP Basic, NAEP Proficient, and NAEP Advanced*” in the document and restore the use of the terms *Basic, Proficient, and Advanced*.

## Communicating NAEP Terminology

While there is long-standing public and professional understanding of NAEP levels, there are also forces who would mislead the public about NAEP either in pursuit of their own agendas or because they were misinformed. There is little way for NAGB or NCES to prevent the deliberate misuse of NAEP results, but both organizations and their partners, including the Council, could do a better job of informing the public about what these performance levels mean and what they don’t mean. For instance, one routinely hears that NAEP levels are akin to grade-level scores. But, in the announcement of results every two years, there is little time devoted to revisiting or describing the definitions of the performance levels or how they were arrived at. Having the performance levels tagged to some external benchmark, as we suggested in the earlier recommendation, might help NAGB and NCES better describe what NAEP means. And spending some time during the release on what the performance levels mean—beyond examples of what students can do under each level—might help ward off some misunderstanding of the levels and protect against the deliberate misuse of terms.

Recommendation: Devote more time and attention during the release of NAEP results to what the proficiency levels mean and what they don’t mean beyond the examples that are often presented to illustrate performance.

October 15, 2018

National Assessment Governing Board  
U.S. Department of Education  
800 North Capitol Street NW – Suite 825  
Washington, DC 20002-4233  
**Attention:** Sharyn Rosenberg, Assistant Director (Psychometrics)

### **Comments on Revised Policy on Developing Student Achievement Levels for NAEP**

#### **Document Citation: 83 FR 45618**

The Council of Chief State School Officers (CCSSO) submits the following comments on the revised policy on Developing Student Achievement Levels for the National Assessment of Educational Progress (NAEP) in response to the September 10, 2018 notice in the Federal Register. CCSSO is a nonpartisan, nationwide, nonprofit organization of public officials who head departments of elementary and secondary education in the states, the District of Columbia, the Department of Defense Education Activity, the Bureau of Indian Education and the five U.S. extra-state jurisdictions. CCSSO has a positive and long-standing relationship with the National Assessment Governing Board (NAGB) and the National Center for Education Statistics (NCES).

State leaders recognize the importance of the National Assessment of Educational Progress (NAEP) as the longstanding common assessment metric used across the country. As a membership organization representing state leaders, CCSSO also greatly values this measure as it continues to serve as an important way for our organization to better understand and analyze student academic progress over time at the national level.

For these reasons, CCSSO is supportive of the revised principles as outlined in the notice published in the Federal Register. In particular, we believe the following proposed principles would be helpful to states, districts, and schools.

- Principle 1 (Elements of Achievement Levels) is clear and provides helpful descriptions of content level achievement level descriptions (ALDs), cut scores, and exemplar items and tasks. All of these, taken together, help to provide context and explain the meaning of the achievement levels used by NAEP.
- Principle 2 (Development of Achievement Level Recommendations) provides additional context about the NAEP achievement levels. We support the development of a design document to further explain who should be included on the achievement-level setting panels, particularly in regard to inclusion of diverse individuals in terms of gender, race/ethnicity, region of the country, urbanicity, and experience with students with disabilities and English language learners.
- Principle 4 (Periodic Review of Achievement Levels) commits NAGB to a review of existing achievement levels at least once every 10 years, or three test administrations, whichever comes later. We support this periodic review and believe that this type of examination is healthy for the assessment. We urge caution before proceeding with changes to the current “cut scores” or level of rigor attached to the current achievement levels. While there have been some difficulties in understanding the precise meaning of the current achievement levels, these levels provide an ongoing national trend and accurate state-by-state comparisons. CCSSO would like to ensure that any review maintain the appropriate balance and recognize the importance and value of NAEP as



a longstanding common assessment metric across the country while at the same time providing detailed information to states, districts and schools to use in decision-making.

- Principle 5 (Stakeholder Input) is an important component of NAEP, and we support inclusion of diverse stakeholders for the activities outlined in the principle, including nominations of achievement-level setting panelists, review of design documents, and review of achievement-level setting activities by the Technical Advisory Committee on Standard Setting. We have appreciated the inclusion of state chiefs through the NAGB and CCSSO State Policy Task Force and want to ensure this structure and similar activities continue for state chiefs and other stakeholders in the future.

In general, CCSSO supports the revised policy on Developing Student Achievement Levels for the NAEP as outlined in the notice published in the Federal Register. However, in addition to the principles already outlined, we believe the final policy also should address additional work to examine the alignment between the current NAEP frameworks and the content standards that are used in states. As an organization, CCSSO is dedicated to setting high expectations for all students through college- and career-ready standards. Since 2010, each state has improved their standards.

CCSSO recognizes and appreciates the robust and inclusive procedures by NAGB in support of a meaningful, valid, and reliable administration of the NAEP. At the same time, we support any additional work that can help to continue to explain the meaning and context surrounding NAEP's achievement levels. For this reason, we urge NAGB to continue to examine the alignment between the NAEP frameworks and the content standards states have adopted since 2010, as we know every state has updated its content standards in NAEP-tested subject areas significantly in recent years.

Mis-alignment presents two challenges to states and to NAEP. First, it is not reasonable for states – as well as districts, schools and teachers – to be held accountable if material that has not yet been taught at a particular grade level. Second, NAEP may not be providing accurate measurement if it fails to test material that has been taught in a particular grade level, but not yet covered by NAEP. We believe examining the alignment of the NAEP frameworks can be done in a way that maintains NAEP's long-term trend and continues its longstanding value as the 'Gold Standard' in assessments across the country, as evidenced in the *Evaluation of the Achievement Levels for Mathematics and Reading on the National Assessment of Educational Progress study* from 2016. CCSSO stands ready to support and assist throughout this process.

Thank you for your consideration of our comments on the revised policy on Developing Student Achievement Levels for the NAEP in response to the September 10, 2018 notice in the Federal Register.

Sincerely,



Carissa Moffat Miller  
Executive Director  
Council of Chief State School Officers





October 15, 2018

NAEP Achievement Level Setting Policy  
National Assessment Governing Board  
800 North Capitol Street, N.W., Suite 825  
Washington, DC 20002

RE: Revised Policy on Developing Student Achievement Levels for NAEP

Dear Dr. Rosenberg,

The National Assessment of Educational Progress (NAEP) is one of the most valued and recognized measures of educational quality and equity in the United States. Every two years, the assessment provides educators, policymakers, advocates, and the public with critical data on student achievement that are comparable over time and across states. As state standards and assessments continue to evolve, NAEP's consistency has made it an invaluable, trustworthy yardstick for gauging trends in educational outcomes for all students and, importantly, individual student groups.

**NAEP data are a critical tool for equity advocates, who use the results to draw attention to disparities in achievement both within their states and across the nation.** NAEP allows advocates and researchers to identify states that are making strong gains for historically underserved students and learn from their experiences. It allows advocates to maintain the push for high expectations for all students and to benchmark their state's results against top performers. NAEP achievement levels are crucial for understanding and communicating how well our education system is serving students, and their consistency over time is a major part of the assessment's utility.

**For the reasons above, we write in the interest of ensuring that the National Assessment Governing Board's (NAGB) final Achievement Level Policy maintains the rigor, comparability and transparency that have made NAEP such a valuable resource for education stakeholders.** We especially urge NAGB not to lower expectations for what it means to be "Proficient" or "Advanced" arbitrarily. If the Board deems it necessary to revise achievement level thresholds, we strongly recommend that revised definitions be based in the reality of what students need to know and be able to do at each tested grade level to be prepared for success in postsecondary education without the need for remediation by the time they leave high school.

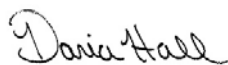
Our education system has a long history of telling students and their families — especially those from underserved communities — that their schools are preparing them for future success. Meanwhile, college remediation rates remain high and employers struggle to find candidates who have the necessary reading and math skills. One of the biggest "value adds" of NAEP has been its consistent high expectations. **If the revision of the Achievement Level Policy results in lower expectations for what it means to be "Proficient" or "Advanced" without solid justification for these changes, it could harm students across the country, with the highest risks for students who are already underserved in our**

**schools.** What’s more, in the absence of a clear rationale for the change, sudden increases or decreases in the percent of students who are “Proficient” may undermine faith in the measure.

Given the potential implications for the usefulness and transparency of NAEP data, **NAGB should only redefine NAEP achievement levels if there is strong evidence that the levels do not currently signify what their definitions entail.** Moreover, if NAGB deems it necessary to revise the levels or what they mean, it should **align the new achievement level thresholds and definitions to what students need to know and be able to do to maximize their educational and career options upon leaving high school (i.e., to be prepared for success in college or postsecondary training without the need for remediation).** In 2014, NAGB conducted [a series of studies](#) that showed that the 12th grade proficient cut score on NAEP was well aligned with college-readiness criteria. NAGB could use this research to identify appropriate cut scores for proficiency in eighth and fourth grade – levels that would indicate whether students are on track to meet the 12th grade proficiency (i.e., college-preparedness) expectations. This type of alignment would increase the accuracy of NAEP achievement levels, grounding them in the reality of what students need to know and be able to do, while maintaining transparency. If this alignment process results in different cut scores for NAEP achievement levels, the Nation’s Report Card should allow users to obtain percentages of students scoring at each performance level using previous and any new definitions both for the most recent and prior years’ results.

Any changes to NAEP achievement levels will pose a communications challenge, and changes that lack a clear rationale risk could detract from the utility of the results. We hope that if NAGB decides it is necessary to revise achievement levels or their definitions, it will do so in a way that yields more meaningful information about how well prepared students are for postsecondary success without the need for remediation.

Sincerely,



Daria Hall  
Vice President for Partnerships and Engagement  
The Education Trust



Ms. Peggy Carr  
NAEP Achievement-Level-Setting Program  
National Assessment Governing Board  
800 North Capitol Street, NW  
Suite 825  
Washington, D.C. 20002



October 15, 2018

Dear Ms. Carr,

On behalf of the undersigned organizations, representing public school superintendents, district administrators, and educational service agency administrators, we submit this letter in response to the proposed changes by the National Assessment Governing Board (NAGB) to the National Assessment of Educational Progress (NAEP) achievement benchmarks. We are opposed to the changes as proposed and are concerned the changes fail to meet the goal in the proposal's first paragraph: to produce achievement levels that are "reasonable, useful, and informative to the public."

We thank you for extending the comment period. While still short, the additional time reflects the reality of what it takes to review and adequately respond to the type of changes proposed in the rule and what it could mean for the nation's schools, the students they serve, and the communities to which they are accountable and in which they live.

The original achievement levels were developed in a rushed process, and resulted in levels that continue to confuse educators, citizens, and policymakers. The levels have been described as 'wishful thinking' more than 'reasonable' or 'common sense', and the latest research linking NAEP's benchmarks to international assessments reveals that the majority of students in most nations cannot clear NAEP's proficiency bar.

Our organizations support high standards, and this response is not a request to lower the standards of NAEP. Rather, we are requesting that should NAGB move forward with a proposal to revise the definitions of NAEP terms, that the final product be an improvement over the current definitions and ensure that parents and the community can understand what the NAEP standard terms mean and how those levels do or do not relate to other academic scores, including assessments and grade level. We align with the recommended changes submitted by the National Superintendents Roundtable<sup>1</sup>, as they appear in the section starting at the bottom of page six of their letter. Their recommendations revise the proposed policy definitions and represent changes that improve efforts to better communicate information that is reasonable and accessible to educators and the general public.

We thank you for reviewing our submitted comment today and urge you to revise the policy definitions for the standards. We reiterate that the proposed policy definitions for NAEP achievement levels miss the mark and are unresponsive to criticism of the levels. These standards and benchmarks must be revised so as to be expressed in terms that parents and the public can understand. We welcome the opportunity to work with you as this process moves forward. Please direct any questions to Noelle Ellerson Ng ([nellerson@aasa.org](mailto:nellerson@aasa.org)).

Sincerely,

Daniel A. Domenech  
AASA, The School Superintendents Association

Joan Wade  
Association of Educational Service Agencies

---

<sup>1</sup> Letter from National Superintendents Roundtable (September 18, 2018)  
<http://files.constantcontact.com/d6ed868c001/b63f4624-49ad-40a1-afd8-660913afe0cd.pdf>

**Bob Wise**  
**Governor of West Virginia, 2001–2005**  
**Member of U.S. House of Representatives, 1983–2001**



October 15, 2018

National Assessment Governing Board  
800 North Capitol Street, N.W., Suite 825  
Washington, DC 20002

Dear Governing Board Members,

The Alliance for Excellent Education (All4Ed) appreciates the opportunity to provide feedback regarding the National Assessment Governing Board's (NAGB) proposed revision to its policy on *Developing Student Achievement Levels for the National Assessment of Educational Progress (NAEP)*. NAGB's policy on setting achievement levels for NAEP was last updated in 1995. As such, it is pertinent to review and revise the policy as necessary. All4Ed is pleased that the current proposal does so in such a way that maintains NAEP's critical role as an honest gauge of student performance across the nation.

It is critical that NAEP's achievement levels continue to be rigorous. While the nation has yet to provide the majority of students with sufficient support to reach the levels of NAEP Proficient or NAEP Advanced, this does not mean that these levels should be lowered. The 21<sup>st</sup> century economy demands that more students acquire greater levels of education than ever before – content mastery, critical thinking, the ability to apply knowledge to solve problems, and other “deeper learning” competencies. These competencies, demonstrated at the levels of NAEP Proficient and NAEP Advanced, should be the goal to which the education system aspires.

Moreover, NAEP's results are consistent with the results of other nationally administered assessments. This demonstrates that NAEP's achievement levels do not hold students to an unreasonable standard. When comparing the percentage of students performing at the level of NAEP Proficient or higher with the percentage of students receiving a four or higher on the Partnership for Assessment of Readiness for College and Careers (PARCC) assessment, the results are within a modest margin (roughly 34 to 40 percent). In addition, the percentage of students who are prepared for college according to NAEP, the SAT, and the ACT are also within a modest margin (37 to 46 percent). Thus, NAEP is setting a similar bar to assessments that are widely acknowledged to measure college readiness.

For decades, NAEP has been the standard-bearer for academic performance. The growing diversity of statewide assessments, indicators of college and career readiness, and accountability systems increases the importance of NAEP as a consistent state and national measure to which other assessments can be compared. NAGB's proposed policy allows NAEP to continue this critical role. Lowering expectations would be a disservice to students and the nation at large.

Sincerely,

A handwritten signature in black ink that reads 'Bob Wise'.

Bob Wise  
President

"Go West...." Horace Greeley reportedly ( arguably) told Josiah Grinnell to "go west" in an instance of some of the most important directional advice ever given. The advice was not "go west 182 degrees for 500 miles and then 187 degrees west" with 95 percent confidence in that direction. And yes , Josiah Grinnell said " It is a wide country , but I do not know just where to go." However , there was an important, compelling rationale for "going west" and the precise direction for going west in this "wide country" was less important and not entirely mapped out. But Greeley's advice was given in the mid-nineteenth century , not in the 1990s when NAEP achievement levels were being developed and when latitude and longitude measures would have been expected.

The comparison to "go west" and to chart your educational direction with a NAEP compass where "Basic" and "Proficient" are the most relevant markers is not a perfect comparison , and yet few comparisons are .

As a participant in the original National Assessment Governing Board process to establish achievement levels, I can say that I was certain beyond a reasonable doubt that creating achievement levels was the right direction for NAEP and the nation. I was less certain that the efforts to describe quite precise "latitude/longitude" achievement level measures could be superimposed on a judgment decision ... a statutorily described judgment decision.

It is important , I would assert vital, to remember where our nation was in educational measurement when NAEP achievement levels were produced. In 1984 I reported to the Southern Regional Education Board and its chair , Governor Lamar Alexander, that virtually all of the southern states had student achievement scores "at or above the national average" on the most widely recognized and used "national" education tests. Yes , southern states were among the nation's poorest by most economic measures and were usually cited among the lowest in various educational measures , and almost all of these states were reporting to their residents ( including parents of course) that the state's student test results were "at or above the national average" in some subjects and at some grades.

I never claimed that this was a conspiracy by state and district education leaders. A West Virginia physician did claim that it was an unethical conspiracy a year or so later with a report asserting that every state was reporting that its students' achievement was "at or above the national average". So in the 1980s we had state reports showing student achievement in virtually every state "at or above the national average" ( not to mention that there were "national norms" and "urban national norms") and the U.S. Department of Education's widely publicized "Wall Chart" with a very different and equally misleading picture of American education. Then add to this mix the numerous state "minimum competency tests" that had virtually by definition low standards and high percentages of students meeting these "competency" standards.

There was certainly plenty of legitimate criticism of the "Wall Chart" and the misleading state SAT scores it featured that created bizarre state rankings. There were also legitimate criticism of the minimum competency tests, but often the criticism was more about the high rate of

failure for economically disadvantaged students. This was not only a concern about testing , but I would argue a major , fundamental issue for our nation and our educational system. What "passing" state minimum competency tests meant , and didn't mean, for students and the nation was also a significant problem.

Shortly after the first NAEP reports based on achievement levels were available I published an SREB report "Setting Education Standards High Enough" (1996). At that time there were eleven states that I could identify with their own eighth grade mathematics standards that were NOT described as "minimum" expectation standards . The percentages of students meeting their own state's standard ranged from 13 percent to 84 percent. Therefore , in one state residents ( including parents) were being told that 87 percent of their students DID NOT MEET the standard and in another state residents ( including parents) were being told that 84 percent of their students DID MEET their state standard.

If this is confusing , add that when the first eighth grade NAEP achievement level results were available in the mid-1990s it was possible to look at the results for the state where 13 percent of students met its standard ( Delaware) and a state where 83 percent met its standard ( Georgia) . And what did one find ? The eighth graders in Delaware had higher NAEP mathematics scores than did Georgia eighth graders.

This was the situation -- I would argue very confusing situation for the American public ( and parents)-- in which the NAEP achievement levels were created.

If a judgment decision ( achievement levels) about a measurement device ( NAEP) is to be seen as strictly or overwhelmingly a psychometric matter then the achievement levels may never be deemed "correct". My view , in the 1990s when helping establish the achievement levels ... and now... is that this process is more about direction than precision and the results should be used in that way. I never argued that the achievement levels were the defining measures for states. I did , and do, believe that state education and government leaders should continually seek and analyze state-NAEP results to determine if their "official" student assessment results and their state-NAEP results are pointing in the same direction. If so , what does this mean? If not , what does this mean? If there is a gap in these results , what does this mean?

Yes , state -NAEP results can be misused and state leaders and the press should strongly oppose this. This possible misuse is no greater than the obvious misuse of non-NAEP results before the 1990s , and no student has ever failed ( or passed) a NAEP test and had to deal with the consequences. I am not denying the possibility of misuse , intentional or unintentional but NAEP achievement levels did not create the misuse of test results.

My greatest concern about the possible misuse of NAEP achievement level results is the one about which I have the most serious disagreement with some achievement level critics. Some critics claimed , and from the most recent report apparently still do, that the achievement levels were created to discredit public education by producing very low measures of student achievement. When National Association Governing Board members totally committed to public education said "No" to this charge , critics claimed that the Board was duped by one or

two persons. None of these critics were at all of the Board's discussions of achievement levels. The Governing Board had some "super smart" folks and even if there were some "super dupers" they could not have , and did not , hoodwink the Governing Board . Some persons might have sought to misuse the results but it was not the Governing Board's intent .

For NAEP and its achievement levels, "precision" measures are a legitimate touchstone , but it is a judgment process with direction , stability ... and transparency... being vitally important factors that are not given to precise measures. "Transparency" means showing the American public ( and of course parents) NAEP tests and many NAEP test questions/answers so that over time there are ways they can participate in the "judgment decision" about NAEP achievement levels. This NAEP transparency and sharing , not equating studies with other nations results , is the most important missing input about NAEP achievement levels.

Mark Musick  
former member and chair  
National Assessment Governing Board

October 11, 2018

Sharyn Rosenberg  
NAEP Achievement Level Setting Policy  
National Assessment Governing Board  
800 North Capitol Street NW, Suite 825  
Washington, DC 20002

Document Number: 2018-21451

Dear Ms. Rosenberg:

Thank you for the opportunity to comment on the revised policy on Developing Student Achievement Levels for the National Assessment of Educational Progress (NAEP). Wisconsin appreciates the effort to more clearly define NAEP-Proficient and we welcome more transparency in standard setting procedures. The comments that follow below are specific to draft Principle 3.

Item g, Principle 3

In describing the NAEP-Proficient level, reports shall emphasize that the policy definition is not intended to reflect “grade level” performance expectations, which are typically defined normatively and can vary widely by state and over time. NAEP-Proficient may convey a different meaning from other uses of the term “proficient” in common terminology or in reference to other assessments.

*Comment:* Wisconsin appreciates the effort to more clearly define for users the meaning of NAEP-Proficient and to differentiate it from other common uses of the term proficient. However, we find that the contrasts with grade level expectations and norms are potentially confusing.

Although the policy achievement level descriptors (ALD) don’t explicitly mention a grade level, the NAEP achievement levels by grade do and are likely to be referenced by users. In addition, in the current assessment and accountability climate, state definitions of proficient are also standards-based and not norm-referenced.

It would be clearer to describe what NAEP-Proficient does mean, rather than, or in addition to, what it does not mean. For instance, it would be helpful to draw the connection between the ALDs, the grade-level NAEP frameworks, and item exemplars.

Item h, Principle 3

An interpretive guide shall accompany NAEP reports, including specific examples of appropriate and inappropriate interpretations and uses of the results.

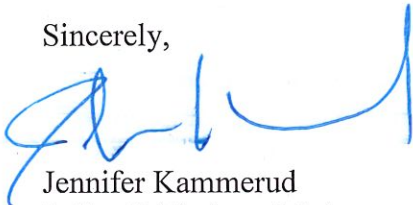
Sharyn Rosenberg  
October 11, 2018  
Page 2

*Comment:* We appreciate how comprehensive NAEP reporting is, and that there is a great deal of interpretive information available. However, we agree with the evaluators that the interpretive guidance can sometimes be difficult to find and that some reports may lead users toward inappropriate interpretations of the data.

We recommend that this item be expanded to reflect its importance and to more clearly describe the National Assessment Governing Board's policy goals related to the means of providing guidance and validating appropriate interpretations.

If you have any questions about our comments, please contact Visalakshi Somasundaram at [Visalakshi.Somasundaram@dpi.wi.gov](mailto:Visalakshi.Somasundaram@dpi.wi.gov).

Sincerely,



Jennifer Kammerud  
Policy Initiatives Advisor

JK:po

I reviewed the ALD policy documents and note the following comments:

- Within the proposed ALDs, the challenge states have with interpretation is in regards to specific content requirements that are assessed in each grade or content area assessment. This challenge is important for state insight as policymakers often wish to leverage NAEP benchmarks in relationship to states' standard setting process.
- With the pilot exercise noted before the formal standard setting process is undertaken, will any data or impact information be discussed by the pilot panel? It may be helpful to consider potential insight into challenges in interpreting the item map or ordered item book where panelists may struggle with the outcome.
- It may also be valuable for NAEP to consider four proficiency levels and not three. This has been a recent change in many state systems, where the level of not proficient are segmented further to provide insight into weaknesses. Again, if states are often urged to use NAEP reporting levels, this better guides these conversations regarding benchmarks.

**Dr. Charity Flores**

*Director, Assessment*

**Indiana Department of Education**



Dear Members of the NAGB:

As a former member of the National Assessment Governing Board (NAGB), I want to express my support of the proposed changes to the policy regarding "Developing Student Achievement Levels for the National Assessment of Educational Progress." In particular, I appreciate the insertion of the acronym "NAEP" prior to the labels of Basic, Proficient and Advanced. I also fully support the fact that Principle 3, g. describes how "reports shall emphasize that the policy is not intended to reflect "grade level" performance expectations..." While this will not satisfy all critics of the high aspirational nature of the NAEP Proficient level, it will continue to remind the public that it is not equivalent to grade level, especially in light of the fact that "grade level" performance is not defined the same way in all states in this country.

Sincerely,

*Louis M. Fabrizio*

-----  
Louis M. (Lou) Fabrizio, Ph.D.  
Chief Executive Officer  
Fabrizio Education Consulting, L.L.C.

## Alignment of NAEP Achievement Levels

I hope that NAGB and NCES might discontinue reporting student achievement using achievement level percentages.

A recommendation in the mandated 2017 evaluation study suggested that when satisfactory alignment among the frameworks, item pools, achievement-level descriptors, and cut scores in NAEP mathematics and reading has been demonstrated, their designation as trial should be discontinued. However, I do not believe an evaluation of the newly aligned achievement levels will demonstrate them to be reasonable, valid, and informative to the public. They will just be an updated, and fatally flawed example of the percent-above-cut-score metric minus the “trial” designation.

The 2017 evaluation study (page 208) noted that the “percent above cut score metric” is not really useful for following student achievement over time (one of two NAEP stated purposes):

\*\* One of the most common and unwarranted inferences [using achievement level percentages] involves assessing the amount of progress students have made over time, particularly by population groups. For instance, news reports often focus not only on how students are doing at a particular time, but the extent to which the percentage of students scoring Proficient or above has (or has not) improved over successive NAEP assessments. When these comparisons are based on the scale scores, they provide useful information. When they are based on the “percentage Proficient or above” metric and used to compare progress across groups, they can be misleading.

\*\* Holland (2002) focused on misinterpretations associated with using the “percent above a cut score” metric. Although this metric is widely used (for NAEP and many other achievement tests), there are serious limitations to the inferences that [the percent above cut score metric] can support, particularly when evaluating trends over time, gaps among groups, or trends in gaps....

Intended or not, the primary use of NAEP’s achievement level percentages since 2002 has been to accuse state public education officials of racing to the bottom, or dumbing down state tests, or having an “honesty gap” in order to hide the extent that state public schools are failing. I was NAEP State Coordinator for Idaho from 2002 to retirement in 2012. Idaho was always listed as one of the “greatest offenders.” All of this because there were two federal definitions of “proficient,” one by NAEP, and one by NCLB. Moreover, just looking at NAEP: the Proficient Achievement Level was not the same thing as NAEP’s “proficiency in a subject.”

In case, I failed to make my two points. First, achievement level percentages are a poor choice to examine student achievement over time.

Moreover, an alignment among the frameworks, item pools, achievement-level descriptors, and cut scores will not change anything.

Second, I would like to see NAEP eliminate the main weapon that today’s “school reformers” have used to undermine public confidence in the nation’s public schools.

Bert Stoneberg

this is nonsense spending. nobody in america needs this finroatmion. there is no reason to fund this federal dept for this kind of survey. the taxpayers of this country pay for education at these levels:

local  
county  
state  
federal

and we pay for other countties education costs too and military education costs.its time to stop all this spending at the federal level. nobody needs this survey. its just wasteful spending that tehe fat cat bureaucrts want to make sure they still have a job that never ends. we need to downsize and this can be done away with. it i9s in no way necessary for elducation. its just fat cat bureaucracy. this coment is for the pubilc record. please receipt. jb ker [bk1492@aol.com](mailto:bk1492@aol.com)



**CALIFORNIA DEPARTMENT  
OF EDUCATION**

**TOM TORLAKSON**  
STATE SUPERINTENDENT OF  
PUBLIC INSTRUCTION

1430 N STREET, SACRAMENTO, CA 95814-5901 • 916-319-0800 • WWW.CDE.CA.GOV

September 28, 2018

NAEP Achievement Level Setting Policy  
National Assessment Governing Board  
800 North Capitol Street, N.W., Suite 825  
Washington, DC 20002  
[NAEPALSpolicy@ed.gov](mailto:NAEPALSpolicy@ed.gov)

To Whom It May Concern:

The California Department of Education (CDE) wishes to express support for the revised policy in *Developing Student Achievement Levels for the National Assessment of Education Progress* (NAEP). The CDE recognizes many hours of expert, dedicated work in these updated guidelines, and believes the proposed amendments improve and advance the way student achievement on the NAEP is measured and described.

Specifically, the CDE supports:

- The change in terminology from *proficient* to *NAEP proficient* – this will help minimize confusion between various assessment standards and clarify the definition of proficiency on NAEP.
- The suggestion that students be referred to as “achieving at a NAEP Basic level,” for example, rather than “NAEP Basic.” This terminology focuses more on the assessment and the shared responsibility of the state, district, school, teacher, student, parents, and community in contributing to the achievement of individual students rather than ascribing qualities to students.
- The Governing Board’s commitment to periodically review the NAEP achievement level descriptions and cut scores. As education standards and curricula change, student achievement will also develop and change. At all levels, educators must continually evaluate what constitutes proficiency, fluency, and mastery of subject matter.

The CDE endorses these changes and appreciates the opportunity to comment. Please contact Julie Williams at 916-319-0408 or by email at [julwilli@cde.ca.gov](mailto:julwilli@cde.ca.gov) if you wish to discuss these comments further.

Sincerely,

/s/

Michelle Center, Director  
Assessment Development and Administration Division

MC:jw

Comments on draft NAGB revised policy on Developing Student Achievement Levels for the National Assessment of Educational Progress (NAEP).

Submitted by Chester E. Finn, Jr. (short bio available at <https://edexcellence.net/about-us/fordham-staff/chester-e-finn-jr>)

September 22, 2018

I appreciate the opportunity to comment on the draft policy. It was on my watch as (first) NAGB chairman that achievement levels were first developed, and we now have more than a quarter century of experience with them.

Though they've been endlessly criticized (mostly for being too high, or not being scientific enough, or not having proven validity), they've also become indispensable to American K-12 education, widely used, cited, calibrated with, and depended upon in so many ways.

NAEP's achievement levels are, in effect, the only rigorous and revealing national standards or benchmarks that we have for K-12 education, at least the only ones that are universal (unlike SAT/ACT), that yield state-level (and in the case of TUDA municipal-level) data (unlike PISA, for example), and that apply to the elementary and middle grades as well as high school (unlike AP, for example).

Vitality, they yield comparative data—from state to state, state to nation, sometimes city to city and city to state and nation—and they also yield trend data so that changes over time in student achievement can be tracked. Sure, there are other ways of doing that (e.g. ups and downs in scale scores for the top and bottom deciles) but they're basically beyond the easy understanding of parents, most policymakers and many educators. Achievement levels are crystal clear, readily intelligible.

The ability to compare NAEP results across jurisdictions with the help of achievement levels is more important than ever now that the PARCC and Smarter Balanced assessments are more or less imploding, with almost every state—even those with common standards—devising and deploying its own “branded” assessment. Had those coalitions stuck together, we would have had other sources of inter-state comparison but, sadly, we don't.

Are achievement levels set too high, as some allege (e.g. the “How High the Bar” attack on both Common Core standards and NAEP achievement levels)? What so many critics refuse to acknowledge is that the achievement levels—at least the proficient and advanced levels—are aspirational, and meant to be, much like scoring 4 or 5 on PARCC, or 4 or 5 on an AP exam. (PARCC is on my mind because we're still using it in Maryland where I'm on the state board of education.) Everybody knows that most kids aren't there yet, but there is where they *ought* to be, and where the education system should

be doing its utmost to get them. (It's also where many, many more kids are in other countries, including many kids who are as "diverse" as young Americans. We know this in part from successful efforts to equate NAEP and PISA.)

The "too high a bar" criticism was pretty much knocked out of the water by NAGB's own careful research demonstrating (to oversimplify a bit) that 12<sup>th</sup> grade NAEP proficient is approximately equal to true college readiness. That's about as solid a validity check as we're ever going to see, at least for the "proficient" level! (I strongly suspect that NAEP "advanced" could be linked to PISA 5/6—research worth undertaking.)

The other issue that keeps arising—and it's more visible than I think desirable in the new draft policy statement—is the ceaseless desire of technical folks to turn achievement level-setting into a technical or scientific process. At its heart, it isn't and cannot be, for at its heart it's a *judgmental statement of aspiration*. Aspirations are almost never "scientific" and while judgments can and should be informed by technical analysis, in the end they're judgments, not "findings."

My only real criticism of the new draft is that it's not nearly as clear as the old one on this matter of judgment. It alludes to it, but in a semi-apologetic way, instead of bluntly and candidly declaring that it's NAGB's job to exercise judgment and nowhere is that more important than in determining the "cut points" for the achievement levels. The reason there's a Governing Board, comprised as it is, is to make such judgments! You can conjure up all the complex procedures, analyses and technical advisors you like, but in the end they cannot transform an act of judgment into science, and you shouldn't even hint that they might!

Having said that, technical expertise is valuable in a number of ways such as, for example, keeping achievement levels calibrated from assessment to assessment, else the trend information would be useless. You don't want your yardstick bending, stretching or shrinking! In the end, however, just as a judge and jury hear from all sorts of expert witnesses, it's judgment that they must exercise. It's judgment that NAGB must exercise, and I wish the revised statement of principles were clearer on that point and less self-conscious about all the trappings of process, technique, and expertise.

One last point. In my experience (both on NAGB, as assistant secretary of education and in sundry other roles), the non-educator members of panels such as you convene for standard-setting are arguably more valuable than the educators, for they have a better sense of the level of performance that the "real world" wants to see from kids emerging from the K-12 system. Employers, newspaper editors, instructors of first-year college students, authors of articles and blogs, military recruiters—I could go on. Please don't stint on incorporating a good cross-section of them on the level-setting panels!

Thanks again for this opportunity to weigh in.

Dear Ms. Rosenberg and members of NAGB:

Thank you for very much for the opportunity you are providing to the draft policy regarding achievement levels.

As a former board member for NAGB representing the role of a public high school principal from Colorado (at that time), I feel compelled to respond to the ongoing debate about whether achievement levels are benchmarks that are too high or not useful or valid.

I was fortunate to serve on the Achievement Level Committee and remember well the rigorous discussions and debates with math experts, researchers and a wide range of faculty from across the country. The need for well-prepared students of all backgrounds and income levels was well documented then, as it is now. All students in public, private or faith-based schools all across the country deserve to be challenged and supported to reach their highest potential. Having worked in public education during my k-12 and now higher education career, I know how much schools care and challenge students.

Achievement Levels are and have been about the educational aspirations we have for all students, about our competitiveness, about our talent development to build the workforce and economic and cultural development worthy of the human spirit.

I urge NAGB to stay the course, to continue to challenge our students, our schools and indeed our country to the highest levels of achievement. And I urge NAGB to continue to refine and improve assessments as new technologies and protocols are developed—however to NOT give up on the need of high and rigorous standards and achievement levels.

I can tell you that in the last 15 years in the community college sector, I see on daily basis the need for our students to increase their mathematical skill—and their persistence is paying off as they prepare for careers and professions in STEM and the challenges of our technological society.

Please feel free to call on me—I would be happy to say more.

Sincerely,

**Christine Johnson, PhD**  
**Chancellor**

Community Colleges of Spokane  
501 N. Riverpoint Blvd. Suite 110  
PO Box 6000, MS 1001  
Spokane, WA 99217

Concerning the Achievement Level Setting Program,

I am writing today to express my urgent plea that the current NAEP Achievement Level system be restructured and renamed effective immediately. For years, the Levels have confused, confounded, and misled the American Public, the media, politicians, and policy makers to the detriment of our nations schools. Teachers have been fired and schools have closed because of the misuse of NAEP data and its leveling system. America's children have been made out to be 'failing' when they score below Proficient. when in reality the passing mark is out of reach and always will be. The National Center for Education Statistics (NCES) has clearly stated that 'proficient' is not synonymous with grade level performance. But when a metric is so clearly misused, misunderstood, and abused it is clearly time for an immediate restructuring. That time is now.

Emily Maurek  
Teacher





September 29, 2018

National Assessment Governing Board  
800 North Capitol Street, N.W., Suite 825  
Washington, DC 20002

Dear Governing Board Members,

Thank you for the opportunity to provide feedback regarding your draft policy, “Developing Student Achievement Levels for the National Assessment of Educational Progress.”

On behalf of the trustees of the Thomas B. Fordham Institute, I would like to commend the draft for its clarity and rigor. NAEP’s achievement levels have served a vital purpose for almost thirty years, clearly communicating to the nation the proportion of students who achieve at a basic, proficient, or advanced level, and setting a high, aspirational standard for the nation’s schools and educators.

The revisions to the previous policy that are under consideration—such as clarifications regarding the meaning of the various performance levels and guidelines for how they are communicated to the public and the press—are measured, and appropriately so.

It would be an enormous mistake to follow the advice of some advocates who want NAGB to describe the proficient level as “extremely demanding” and to equate the basic level with “roughly analogous to performance at grade level.” This would represent a lowering of the bar, plain and simple.

To be sure, the proficient standard is a high one—indicating “solid” performance and competency in “challenging” material, in the words of your draft policy. It would indeed be foolish to expect one hundred percent of students—in the United States or anywhere—to be able to meet such a high bar. And it’s unfair when some critics of the nation’s schools imply that anything less than universal proficiency is tantamount to failure.

Yet, as NAGB’s own studies have demonstrated, the proficient standard is not much, if any, higher than the aspirations of America’s moms and dads. Research from Learning Heroes and others indicates that almost all parents expect their children to go to college, and most children share those aspirations as well. Setting aside whether we actually need everyone to go to college (or even attend “postsecondary education”), it’s undeniable that higher education is now part of the American Dream. And NAGB has determined that in reading at least, twelfth graders who are proficient are also well-prepared for college. (For math, the preparedness level is between basic and proficient.)

It makes sense, then, for proficient to be the goal, and for Americans to continue to receive reports regarding how many students are attaining it.

Happily, most state policymakers appear to agree. Analyses by *Education Next* and other indicate that in recent years states have dramatically raised the rigor of their own annual assessments, so that

“proficient” on state tests is now much closer to “proficient” on NAEP. In some cases it’s even higher. This is evidence, from America’s laboratories of democracy, that “proficient” is a reasonable if challenging goal.

And again, it would be inappropriate for states to consider schools that do not get 100 percent of students to the proficient level to be failures. But there is comforting news on that front as well; under the Every Student Succeeds Act, the majority of states have moved aggressively to weight student progress over time much more heavily than the proportion of students attaining proficiency. That too indicates that proficiency is seen as an aspirational goal—one that may never be attained by everyone, but one that is a marker of solid performance.

Your draft policy reflects a good-faith effort to be responsive to the valid concerns made about NAEP’s standard-setting processes, without lowering the bar. Please don’t make significant changes that would diminish this reasonable compromise.

Cordially,

A handwritten signature in black ink, appearing to read "Michael Petrilli". The signature is fluid and cursive, with a large initial "M" and "P".

Michael J. Petrilli  
President  
Thomas B. Fordham Institute

CC: David Driscoll, chair, Thomas B. Fordham Institute Board of Trustees  
Chester E. Finn, Jr., president emeritus, Thomas B. Fordham Institute

I have just one point - but a crucial one.

Performance standards that are describe linguistically are inevitably highly ambiguous.

Rich exemplification of:

- \* the types of task students will be asked to perform

- \* how performances will be scored

- \* the presumed degree of unfamiliarity of the task (not the genre, of course) makes the standards clear.

I and my Shell Centre colleagues would be happy to exemplify such exemplification!

It also has the effect of providing learning goals for the classroom.

Hugh Burkhardt

Mathematics Assessment Project

map.mathshell.org

see also

<https://www.mathunion.org/icmi/awards/past-reipients/2016-icmi-award-winners>

The National Assessment Governing Board should change the achievement benchmarks for the National Assessment of Educational Progress. These standards and benchmarks must be expressed in terms that parents and the public can understand.

The current achievement levels do not do that. Those levels, instead, have led to confusion in the news media and among teachers, parents, and the general public. It appears that very high aspirations ruled their development, instead of realistic conclusions based on sound data.

The consequences of the current approach are shown in recent research linking NAEP's benchmarks to international assessments: the majority of students in most nations cannot clear NAEP's proficiency bar. A disregard over many years of such studies done by respectable organizations showing the deficiencies in the processes used to develop these achievement benchmarks is the reason for these unrealistic results.

The National Assessment Governing Board has been a good steward of NAEP, with this major exception. I urge NAGB to establish a sound process to rewrite these achievement benchmarks. In addition, the public should not only be given more time to comment on these measures, but the advice given should also be considered with an open mind. I would expect nothing less from the Board.

I realize that this message is being sent after the [September 30] deadline due to my being out of the country for a while. I feel so strongly about this issue that I am sending this anyway. Thank you.

**Jack Jennings**

this project should not be funded. we need a much smaller federal dept of education. this dept is too large, it is a spendthrift with no validity for the excessive trillions of dollars it spends. it is letting students take out huge loans for nothing at all. they are never productive. they never intend to pay back. they never will pay back. the entire system is out of control. it sucks. this agency deserves an f minus for its work. it needs to be shut down. we can do without 4 levels of education depts.

we pay for local, county, state and federal levels for education. no other country in the world requires 4 different gov't levels for education. they do a better job with 1 or 2 levels. f minus for this dept and its socialism, its corruption, its first class bureaucracy. nothing good comes from the pedantry bureaucracy of this dept. this comment is for the public record. please receipt. jean public1 [jean public1@yahoo.com](mailto:public1@yahoo.com)

I strongly encourage you to set "Proficient" for grade 12 at the level required to earn a "B" in entry-level credit-bearing courses at moderately competitive American colleges and universities. This standard could be set by examining student work, assigned texts, and assessments at a sampling of these universities/courses. This would make standard setting more objective and less political and would provide clarity so that the average American high school student and parent could understand the results.

*Kimberly Fleming, PhD*

[Subscribe to Core Education's Newsletter](#)

[www.CoreEducationLLC.com](http://www.CoreEducationLLC.com)

Blog: [www.CoreEducationLLC.com/blog2](http://www.CoreEducationLLC.com/blog2)

Twitter: @CoreEducation1

*The following pieces constitute Marc Tucker, President and CEO of the National Center on Education and the Economy's comments for the National Assessment Governing Board's policy statement on Student Achievement Levels for the National Assessment of Educational Progress (NAEP). They originally appeared in his Education Week blog, Top Performers.*

Blog #365

## NAEP Must Get Its Standards Right This Time

Forgive me if I am a little cynical about the eternal dance between measurement and accountability when it comes to reporting on the progress and achievement of American school children. From the beginning, the leaders of our state education systems have invited testing experts to help them set the cut points for passing or not passing the state tests. They listen gravely to the advice of the experts, then ask them how many students will fail at the recommended cut point and set a new one at a point that is politically tolerable.

The heads of municipal school systems for a long time picked the test they would use to report student performance from vendors who offered to compare the performance of their students to that of any of many different student bodies elsewhere. The superintendent would pick the comparison that would make their district look the best. All the insiders knew that was how it worked. Only the public was fooled

When George W. Bush became President, he wanted to hold every state, district and school to a common national standard. He couldn't get what he wanted, but he did the next best thing. He required all the states to participate in the National Assessment of Educational Progress, "The Nation's Report Card." For the first time, the performance of all the states could be compared on a common metric.

Well, that was interesting. President Bush's signature education program, No Child Left Behind, required each state to set its own standards for student performance and then commit to reaching those standards by 2014. These standards, of course, could be different. It turned out, rather famously, that the states claiming to make the most progress toward reaching their standards were those that performed the worst on NAEP. The governors of the states that ended up with egg on their face were among the strongest supporters of the development of the Common Core State Standards. They did not want to be embarrassed that way again.

But, as we all know now, not all state versions of the Common Core are the same and there are a number of states that have not embraced the Common Core in any form. And the state consortia formed to create common assessments of the Common Core have withered on the vine, so there is now no prospect that the Common Core and its associated tests will enable all schools, districts and states to compare their performance to all the others on a common, honest metric

That leaves NAEP. The way NAEP is done permits the observer to compare scores among states for the grade levels and subjects it assesses. But what do those scores mean? In an attempt to answer that question, the NAEP Governing Board, starting a quarter century ago, settled on three distinct levels of performance: NAEP Basic, NAEP Proficient and NAEP Advanced. The Board's [policy statements define](#) those terms and describe the process by which the Board will decide on the cut scores that demarcate the boundaries between performance levels.

What makes this moment special is that NAEP is now, today, engaged in the first major revision of the procedures by which those standards are set. You have a chance to weigh in. [Look here to see how you can make your views known](#) to the Board.

This policy is really important, because the views that Americans have about the performance of their schools are significantly affected by the generous press attention that the NAEP reports routinely get. But what those reports mean has a lot to do with how NAEP defines performance. That is what this blog is about.

Let's take the meaning of the word "proficient." The new draft standards say proficient means "...solid academic performance for each NAEP assessment. Students reaching this level have demonstrated competency over challenging subject matter, including subject-matter knowledge, application of such knowledge to real world situations, and analytical skills appropriate to the subject matter."

Hmmm... What does "solid" mean? Who defines what it means to be "competent"? What is terribly "challenging" to one child might be super easy for another. This definition is quite obviously a matter of judgment. And that is the way the issue is treated by the draft. It says a panel of "subject matter experts will be convened to recommend achievement level cut scores...." What really counts here is their opinion.

And then, of course, it says that these subject matter experts do not decide on the cut scores, but instead make recommendations to the full NAEP Board. It explicitly directs that the Board have information on the effects of setting the cut scores at different levels—that is, how many students are likely to be found proficient. That strongly suggests that political judgment will play a decisive role in cut score setting, just as it has always done at the state level.

But then the document says that these judgments need to be "valid." You would think that would mean coming up with empirical data showing that students said to be proficient actually are, in some commonsense meaning of the word, proficient. But it does not mean that. It means that there is empirical evidence that students who are said to be proficient do in fact have the capacities specified in the definition. But that is circular. What do you mean by challenging?" Answer: "Whatever I have measured."

What would make it uncircular? Answer: Knowing whether a student is proficient or not would have some meaning for me if I knew whether that student could do something in particular that is important to me or the student. For example, whether that student is ready for college or



ready for a career. Now, you say, fine, I can go with that, but can you?

Ready for which college? The fly-by-night “institution” down the street that offers no instruction but is ready to take your college loan money this afternoon or Michigan State? Ready for which career? A career as a cashier in a fast food chain or a career in finance on Wall Street? The Department of Education’s draft provides no guidance on any of these points.

Nor does the history here give this reader confidence that the process described will get this right unless such guidance is provided. The NAEP Governing Board has done a good job of sponsoring research that correlates scores on NAEP with certain college outcomes and workers’ incomes. But that does nothing to tell students, parents or teachers or even state policy makers what students need to study or how well they have to do to be college and career ready. If the members of the Board think that one can only be proficient in mathematics at the 12<sup>th</sup>-grade level if one has demonstrated a thorough command of the topics typically included in an Algebra II course, then the people who construct the test will include a lot of Algebra II questions on the test and policy makers will tell the schools everyone has to take Algebra II. If the Governing Board says that, in their judgement, that is what proficient means, who is to say they are wrong?

Actually, me. One does not have to have [mastered the content of the typical Algebra II course](#) to succeed in college or career. This is how I know that:

First, a very large fraction of high school students going to college in the United State either do all their college work in a community college or take their first two years of a four-year college program in a community college.

Second, the nation’s primary provider of career training, meaning vocational education and training, is the nation’s community colleges.

Third, successfully completing the first year of a typical community college program is a good predictor of the likelihood that the individual will successfully complete a two-year degree program or acquire an occupational certificate of value to an employer.

Fourth, it follows from “1,” “2” and “3” above that, if one cannot succeed in the first year of a typical community college program, one’s chances of succeeding in further college or career are slim, and the converse is also true. So, one could reasonably say that whether or not the high school student is ready for success in the first year of a typical community college program is a very good measure of the degree to which the student is “ready for college and career.” It does not mean the students has a high probability of success in the first-year program at Stanford or in the first year of a program designed to train medical technicians to administer and interpret sonograms, but it does specify a standard of proficiency that is specific and broadly applicable, a standard that would have intuitive appeal to millions of American students, parents and college admissions officers. In most advanced industrial countries, there are one or more high school leaving credentials that are matched to the requirements for going

on to university or to advanced occupational training. Would it not make sense for “proficient” to mean just that, using the demands of credit-bearing courses in the first year of our community colleges as the benchmark standard? To do this right, the National Assessment Governing Board would have to know not just what cut score to use on a general test of mathematics, but what topics in mathematics would have to be mastered to what level to enable the student to succeed in College Mathematics or College Algebra.

Our organization [has done the research](#) needed to establish the benchmarks for such a standard, at least with respect to reading, writing and numeracy. We did not do it the way NAEP has done it, by asking “experts” what they think the standard ought to be. Industrial psychologists found out years ago that that approach almost never actually ends up with descriptions of what is required to do a particular job or the level and kind of education and training needed to do that job. Instead they study the job itself, the way real people do it and then use that information to figure out what sort of education and training they need.

That’s what we did. We gathered the most widely used texts used in the most commonly taken initial credit-bearing courses in a randomly selected set of community colleges and asked leading reading experts to determine their reading level. We asked for graded writing samples from typical assignments given to students along with scored exams and had those reviewed by leading writing experts. And we reviewed the texts for the courses called College Mathematics and College Algebra and had them reviewed by the nation’s leading mathematics experts.

It turns out that College Mathematics and College Algebra are mostly topics covered in Algebra I, and a little geometry, statistics and probability. Students leaving high school do not need to be proficient in Algebra II in order to study Algebra I. The first-year texts are mostly written at the 12<sup>th</sup>-grade level, but a large fraction of our high school graduates cannot comprehend what is written in them. Many of the community college instructors told us that they do not assign writing to their students because the students cannot write and the instructors do not think they were hired to teach basic writing. Yes, some of the NAEP benchmark standards for mathematics are well above any reasonable definition of “proficient.” But a standard of proficiency that was based on what it would actually take to succeed in the first year of community college would be way below the global standard for college level work in the advanced industrialized nations.

If I were on the NAEP Board, I would press for setting proficiency standards based on what empirical data—not anyone’s “expert opinion”—tell us about the content and performance requirements for success in the first year of the typical community college. I would urge my fellow board members to adopt a policy for reporting to the American people on how many high school students reach that standard at the end of high school and how many students are on a trajectory to reach that standard in elementary and middle school. I would push NAEP to tell the American people that this benchmark *should* be used by the states to set a target for what their students should achieve by the end of *tenth* grade, because that would represent a level of achievement for students of that age comparable to the level achieved by most

students in the top-performing countries by that time and there is no reason why we should expect less.

And lastly, and most importantly, I would tell them that NAEP is the last redoubt, the last remaining hope that the United States will have an instrument that we can use to get an honest measure of how our students are doing. If we lose that check, if anyone can say whatever they like about how our students are doing, ignorance will not be bliss.

Blog #366

## Setting NAEP Performance Standards: How We Got Here

Last week, [I wrote a blog](#) commenting on the draft policy on performance standards recently issued by the National Assessment Governing Board. In it, I called for performance standards based not on people's opinions about what constitutes basic, proficient and advanced performance at three grade levels but on what the evidence shows it takes to succeed in the first-year program of a typical community college, on the grounds that, for at least half of our high school graduates, our community colleges are the gateway to both careers requiring occupational certificates and to two-year and four-year college degrees.

Among the responses I got was one from Jim Pellegrino, who sent me two papers he thought might interest me. One was the chapter on setting performance standards in a book that he, Lee Jones and Karen Mitchell had done for [the National Academies in 2009](#) titled *Grading the Nation's Report Card: Evaluating NAEP and Transforming the Assessment of Educational Progress*. The [other was a paper](#) by Albert Beaton, Robert Linn and George Bohrnstedt written for the National Center for Educational Statistics in 2012 titled *Alternative Approaches to Setting Performance Standards for the National Assessment of Educational Progress (NAEP)*. They make for very interesting reading for those of us who care about the usefulness of The Nation's Report Card.

Reports from the National Academies are usually written in the most measured scholarly language imaginable. Not this one. The smoke rises from its pages. NAEP's standard-setting process, it says, is "fundamentally flawed."

Reviewing critiques of the NAEP standard-setting process offered through the 1990s, the chapter from the National Academies book quotes analysts who described the process of making judgements a "nearly impossible task" for the raters, pointed out that the process produced different cut scores for different kinds of test items (e.g., open ended vs. multiple choice) and said the cut scores had been set at levels that were simply not credible when compared to evidence from other well-regarded assessments.

All these issues came to a head with the [1996 science assessment](#). The results showed that, at all three grade levels, very low percentages of students scored proficient, and, at the high

school level, hardly any students had made it into the advanced level. The obvious conclusion was that students who had earned good grades on the Advanced Placement tests in science were not considered by NAEP to have achieved at advanced levels. One would also have to conclude that students who had done very well on the Trends in International Mathematics and Science Study (TIMSS) were not doing advanced work either.

These results were not all that unusual. In other cases, too, the NAEP findings on the performance levels of American students seemed to be way off in both directions. The standards were just not credible.

So the Board adjusted the cut scores to make them more credible. But then the critics noticed that the new cut scores did not match up with the performance descriptions for the standards. What was defined as proficient work in the descriptions was not what was tested by the items used in the proficient range. So the Board took out the definitions! Well that was one way to deal with the discrepancy. Then one could say a student was proficient if he or she scored in a certain score range, but one could no longer say what that meant in terms of what the student could do.

The Board empaneled another group to write new descriptions of what the performance levels meant. This time they wrote descriptions not based on what a student should know to be proficient, but rather on descriptions of what they currently know and can do. "...[I]nstead of reporting achievement results relative to an established standard of performance... the science report presented results that were based on NAGB's a priori judgment as to what constituted reasonable percentages of students at the three achievement levels."

The problem with that, of course, is that the public might reasonably think that a student who was rated proficient on a subject at a certain grade level by NAEP was able to do what a student needed to do to be successful according to expert judgment, but it did not mean any such thing.

What had really happened was that a complex technical process that was supposed to produce findings about student achievement against common sense standards had failed badly. The process had produced findings showing that, in some cases, the standards were much too high by any reasonable measure and in other cases much too low. In the end, the NAEP Board did what such bodies had always done before in such cases. It adjusted the results to produce a politically palatable result without a solid rationale for its decision. It had been and was still the case that it was very unclear what the performance standards were or what they ought to be.

The problem, as I pointed out in my last blog, is that the nation was much more focused on finding accurate, unfudgeable measures of student performance with which to measure the performance of state education systems, districts and schools than ever before. The issue of what the performance standards meant and how they should be developed would not go away.

You might reasonably assume that a slashing attack like this from such a distinguished group of critics would have led the NAEP Governing Board to respond to the critique with a fix for its standard-setting problems. But that did not happen. Years later, the process for setting the standards had not changed substantially, so another, no-less-distinguished band of scholars took up where Pellegrino and his colleagues had left off. Beaton, Linn and Bohrnstedt were members of the NAEP Validity Studies Panel, now chaired by George Bohrnstedt himself. Their paper was written as part of that program of studies.

In it the authors explore three possible alternatives to the process for setting NAEP achievement levels I described above.

The first alternative would be to make the achievement cut points predictive. The cut point for the end of elementary school assessment would predict the likelihood of success in middle school, the cut point for the end of middle school assessment would predict the likelihood of student success in high school, and the purpose of the end of high school assessment would be to predict the likelihood of success in college and career.

The second alternative would be to “benchmark the achievement levels against international standards.”

Their third alternative was to use percentile rankings to set base-year norms against which progress could be measured in succeeding years. The authors acknowledge that hybrids of these approaches could be developed, too.

I was astonished when I read this paper. Taken together, its proposals mirror the plan I described, in more detail, in last week’s blog. But those ideas are not new. I first proposed them years ago when my organization created a program called *Excellence for All* to put them to the test in the field in several states including Kentucky, Arizona and Mississippi. The assessments we used were not the NAEP assessments—they were not designed to be used as census assessments of all the students in a school—but the International General Certificate of Secondary Education exams offered by Cambridge Assessment International Education of Cambridge, England.

The idea was to set a high standard for all high school students based on what it would take to succeed in the first year of a typical community college program and design the program of the high school so that most students would reach that standard by the end of grade 10 and almost all would get there by the time they graduated high school. Students who reached the standard by the end of grade 10 would be able to enroll in a demanding upper division high school program like International Baccalaureate, a whole program of AP courses or the Cambridge Diploma program, all of which are designed to qualify students who get good grades on those programs into the world’s most selective colleges and universities. Or they could take a full program of community college courses and wind up either with a strong vocational credential or two years of college credit, ready to transfer at the end of high school straight into the junior year of a state college or university.

We needed a team of top education researchers to help us set the right pass points on the Cambridge exams to make this program work. We asked Jim Pellegrino and Howard Everson to chair the Technical Advisory Committee. The other members were Catherine Snow, Phil Daro, Bob Linn, Richard Duran, Ed Haertel, Dylan Wiliam, Joan Herman and Lloyd Bond.

Beaton, Linn and Bohrnstedt recommend that NAEP do an empirical study to benchmark college and career readiness. We did that, with a research plan approved by the members of this Technical Advisory Committee (TAC) and drawing on the services of several of its members. Beaton, Linn and Bohrnstedt recommended that the NAEP high school performance standards be set to predict the likelihood of student success against the empirically determined college and career benchmark. We did that, using the Cambridge exams, rather than the NAEP assessments, also under the supervision of the members of our TAC and using their services. Beaton, Linn and Bohrnstedt recommended that NAEP benchmark international student performance standards and design NAEP performance standards to predict student success on those benchmarks, too. We did that, too, using instead of the NAEP assessments the Cambridge assessments.

The only difference between what we did and what was recommended in the paper was that we substituted Cambridge examinations for NAEP assessments. There is no reason why NAEP could not replicate what we did substituting NAEP assessments for Cambridge assessments.

Not only that, but, because we have already done the work, we know what an empirical study of college and career readiness, defined as readiness for success in the first year of a typical community college program, requires. We also know that this is pretty much the same performance standard that is met by the typical student who is entering gymnasium in Europe or beginning to take “A”-level exams in England.

The point is that a very good model exists and has been tested for producing performance standards for NAEP that would address and resolve the problems in setting performance levels that have dogged NAEP for years.

This is no esoteric matter. For all the reasons I advanced in last week’s blog, NAEP is this country’s last redoubt for honest comparisons of the performance of a state’s education system to that of other states. It is certainly true that performance could just be reported as a number on a scale. But what does that number mean? What does it say about what the students know and can do? About whether they are ready for the next stage of their education or to begin a rewarding career? How their performance compares to the performance of students of the same age in other countries? If you care about the answers to these questions, you should care about the way NAEP sets its performance standards.

The changes now proposed by NAEP to the standard setting process do not address these issues. If you think the issues I have raised are important, [write to the NAEP Board](#) with your comments.

## Comments on NAEP Standard Setting

The draft policy statement on NAEP standard setting is clear and generally reflective of current best practices. I offer the following comments and suggestions for the Board's consideration.

**p. 6: There shall be no content ALDs developed for performance below the *NAEP Basic* level.**

If the only purpose of the ALDs is to guide standard setting panelists in setting cut scores, it would be reasonable not to have ALDs below the Basic level. However, since ALDs may also be used to guide item development, via the creation of range ALDs, there should be at least some indication of what Below Basic looks like so that items assessing the knowledge and skills of students performing at that level could be created and matched to an appropriate range ALD.

**p. 7: h) The process shall consist of at least two achievement level setting meetings with distinct groups of panelists, a pilot study, and an operational meeting.**

This is a great idea. Experience has shown that in high-stakes situations such as NAEP, a pilot study eliminates all manner of bugs that could appear in the operational standard setting. The Board has the resources to conduct split-group meetings and should do so. Individual standard setting plans should indicate how cut scores will be compared/resolved/aggregated across groups.

**p. 8: c)...The Board shall revisit and may revise content ALDs following the achievement level setting to ensure that they are consistent with empirical evidence of student performance.**

Modifying ALDs after standard setting is a time-honored practice but a dangerous one. If the content ALDs are sufficiently vetted, and if the range ALDs have been properly developed, vetted, and applied in item development, and if threshold ALDs faithfully reflect the intent of the range ALDs, then any adjustment at the end of standard setting should be to align cut scores with fixed ALDs, not the other way around. By the time standard setting is conducted, ALDs will be highly visible and will have taken on a life of their own. Modifying them will be much more visible than modifying cut scores. If necessary, reporting ALDs may be fine tuned to reflect specific outcomes of standard setting, so long as they can be legitimately aligned to the policy and threshold ALDs. Please consider revising this portion of the plan.

**p. 8: d) The Board shall examine and consider all evidence related to validity of the achievement level setting activities.**

In many situations, cut score validation now includes review of ALDs and cut scores by content matter experts from higher grades (e.g., middle school educators for grade 4 standards, postsecondary educators for grade 12 standards) as well as predictive validity with respect to performance in subsequent grades. The Board may wish to consider extending validation of NAEP cut scores by including on the front end a review of all ALDs, test items, test forms, and

cut scores by educators beyond the grades for which the NAEP forms are designed and on the back end predictive validity studies using school performance in subsequent grades (e.g., relevant course grades for fifth graders who took the fourth grade exam the previous year). [See also Principle 4, section d.]

### **General**

Principle 4 sets forth excellent guidance for ongoing validation and updating of cut scores and integration of standard setting with all other aspects of test development, administration, and reporting.

Thank you for the opportunity to participate in this review.

Michael B. Bunch, Ph.D.  
Senior Vice President  
Measurement Incorporated



I helped write Oregon's academic standards and standardized tests in the 90's, based on national and international expectations.

In addition, I was a guest teacher in Europe where I learned about their national standards, the PISA and exams for college prep seniors.

I can tell you right now that the current US standards based on the NAEP are unwieldy, inconsistent and completely unrealistic for students who aren't college bound.

No other nation requires all students to complete Algebra II. That alone is a major reason for America's high dropout rate. I know teachers forced to teach special education students to mindlessly copy quadratic equations just to prove the entire school is working toward the Common Core and NAEP standards.

Finally, setting the standards at NAEP's mastery level instead of at grade level creates a generation of kids that feel like failures. Regrettably, I was part of the creation of the rubric scale of 1-4 designed to match the GPA system, whereby 3's or B's equate with on-track for college. But internationally the mastery level is only expected of the top third of students, not everyone.

Let's acknowledge that the NAEP holds up an ideal unattainable by the average student under the best of circumstances. Let's return US standards to grade level expectations, whereby average students will be rewarded for passing the tests based on those standards and the stars will always shine with A's and B's.

Rachel Rich

Retired High School Teacher  
Past President of OATG  
Past Board Member of PNCFL, COFLT, AATG  
Goethe Institute Teacher Trainer  
Workshop Presenter

**Richard G. Innes**

2836 Deerfield Drive

Villa Hills, KY 41017

October 2, 2018

NAEP Achievement Level Setting Policy  
National Assessment Governing Board  
800 North Capitol Street NW, Suite 825  
Washington, DC 20002  
(Via e-mail: [NAEPALSpolicy@ed.gov](mailto:NAEPALSpolicy@ed.gov))

Dear Sirs:

I offer the following information as you consider changes to the NAEP Achievement Level scoring system.

Based on my comparison of NAEP data to data from ACT Grade 8 EXPLORE testing conducted in Kentucky from 2007 to 2015 (See Attachment 1), it appears the current NAEP "Proficient" achievement level is already well aligned to identify students on track as of the eighth grade to be ready for college upon high school graduation. Therefore, I encourage NAGB to make no changes to the scoring for Proficient on NAEP.

Sincerely yours,

(Signed)

Richard G. Innes

1 Atch

Comparison of Kentucky's Benchmark Score Results from ACT's EXPLORE with NAEP Grade 8 Proficiency Rates

**Attachment 1**  
**Comparison of Kentucky's Benchmark Score Results from ACT's EXPLORE with NAEP Grade 8 Proficiency Rates**

This research takes advantage of same cohort testing in Kentucky with the NAEP and the ACT Incorporated's EXPLORE assessments to explore the relationship between the percentage of students scoring at or above the NAEP "Proficient" level and the percentage of students scoring at or above the readiness "Benchmark" scores on the EXPLORE assessments.

The Kentucky wide data is available for years from 2007 to 2015.

Because the Jefferson County Public School District in Kentucky also participated in the NAEP Trial Urban District Assessments from 2009 onward, an additional set of NAEP to EXPLORE comparison data is available.

For those who are not familiar with the EXPLORE, the ACT, Incorporated expended considerable effort to insure the EXPLORE test Benchmark Scores are closely linked to the College Readiness Benchmark Scores for the ACT college entrance test. Those ACT test Benchmark Scores are the result of empirical studies that relate those scores to the real performance of college freshmen in related college courses. Thus, the EXPLORE has a strong relationship to what students need for college and careers as of the eighth grade.

**Data Sources:**

The Kentucky and Jefferson County Public Schools EXPLORE Benchmark results are available in an Excel Spreadsheet available here.

<http://openhouse.education.ky.gov/Data/Download?file=EXPLORE%20Benchmarks%20Listing%202014-2015.xls&path=Assessment>

NAEP percentages of students scoring at or above "Proficient" in math and reading for years prior to 2015 were obtained from the NAEP Data Explorer on June 5, 2015. The NAEP 2015 data were obtained from the NAEP Report Cards supporting Excel spreadsheets on October 28, 2015.

**Comparisons of NAEP and EXPLORE Results**

The four figures below compare the percentage of students scoring at or above NAEP Proficient to the percentage of the same cohort of students scoring at or above the EXPLORE Readiness Benchmark Score. Note the close agreement throughout.

Figure 1 – Kentucky Reading

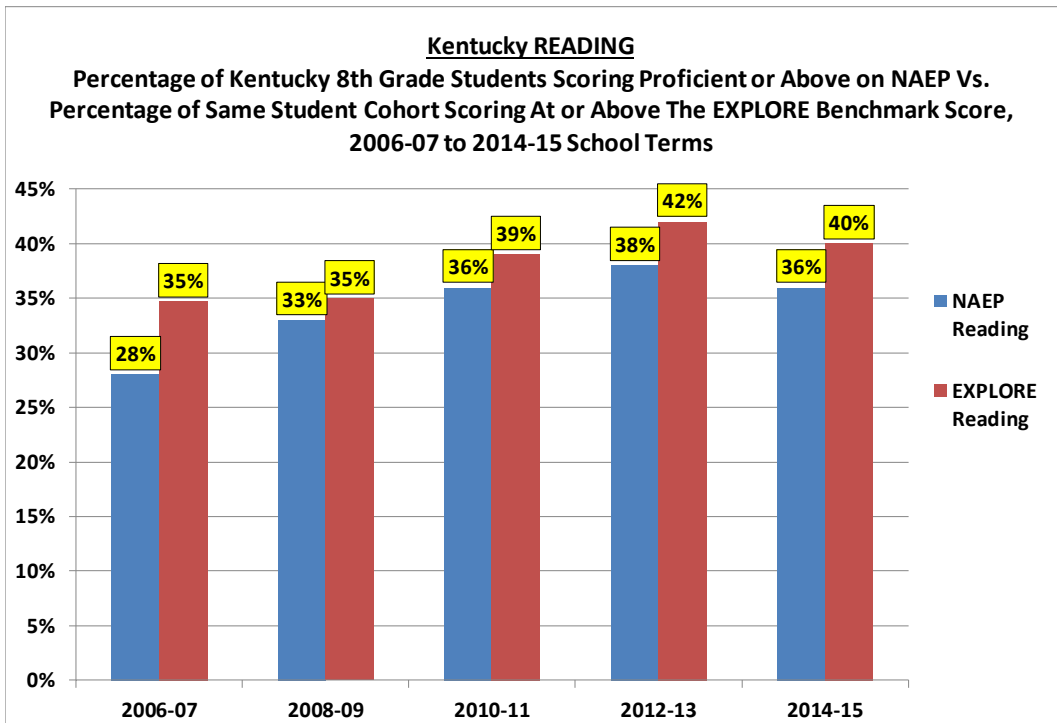
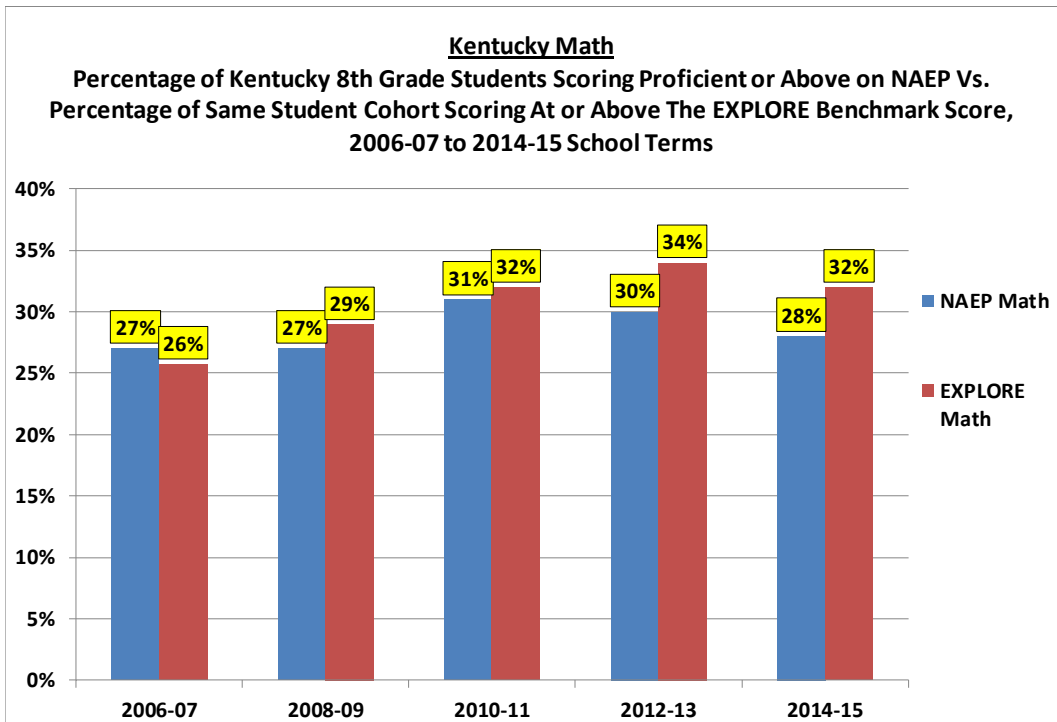
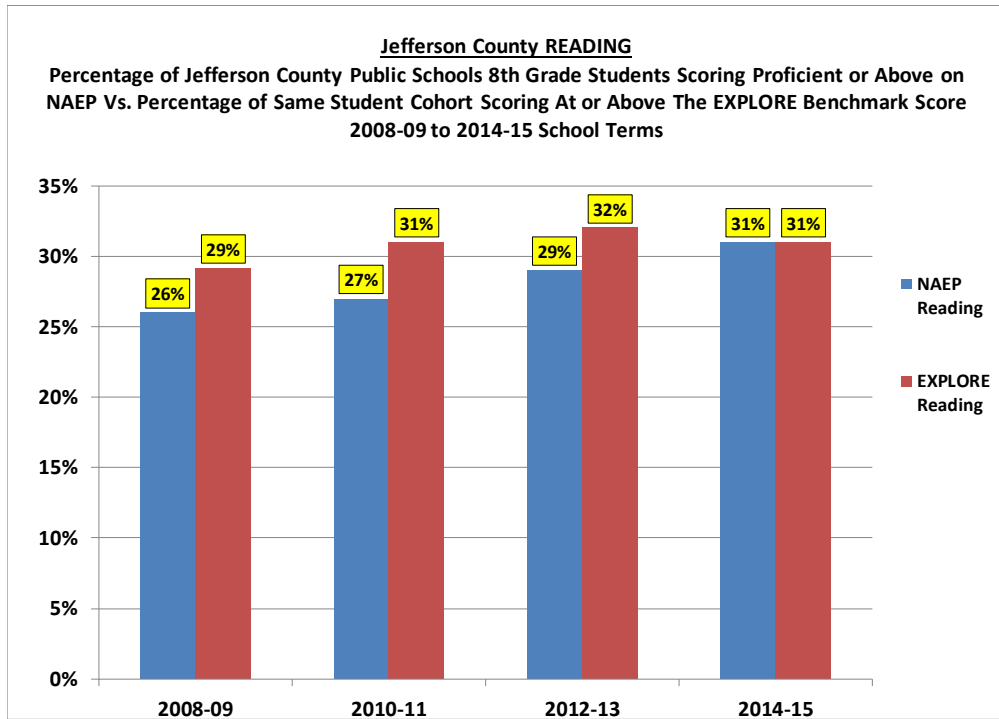


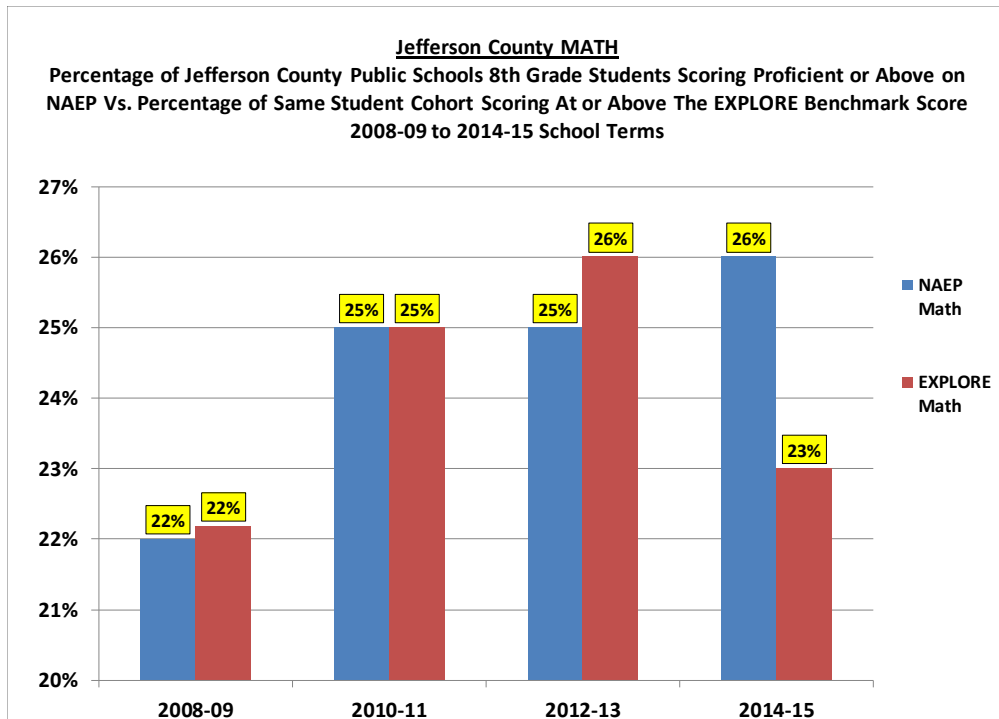
Figure 2 – Kentucky Math



**Figure 3 – Jefferson County Reading**



**Figure 4 – Jefferson County Math**



### **Observations on NAEP to EXPLORE - Math**

The Kentucky NAEP to EXPLORE math agreement shows only a 1-point differential in 2007 and 2011, a 2-point spread in 2009 and a 4-point spread in 2013 and 2015.

In Jefferson County, the agreement was exact in both 2009 and 2011 and differed by only 1 point in 2013. The Jefferson County difference grew to three points in math in 2015.

### **Observations on NAEP to EXPLORE - Reading**

The Kentucky NAEP to EXPLORE reading difference is 7 points in 2007, but this was reduced to only a 2-point spread in the 2009 and increased slightly to a 3-point difference in 2011 and a 4-point difference in 2013 and 2015.

In Jefferson County, the agreement in 2009 was also quite close with just a 3-point differential, which increased to a 4-point difference in 2011 and then reduced to a 3-point spread in 2013, finally becoming a perfect tie in 2015.

### **Implications**

The close agreement for NAEP and EXPLORE for both math and reading is remarkable given the fact that the development process for these two assessments and their scoring schemes are quite different.

Also, there is sampling error in the NAEP. After that sampling error is considered, most scores reported here from the EXPLORE are essentially equivalent to the NAEP.

The findings here help further inform long-standing concerns about the accuracy and meaning of the NAEP Achievement Level Scores, at least for the meaning of NAEP "Proficient" determinations at the eighth grade level in math and reading.

If a good psychometric case can be made that the NAEP 4th grade achievement level scores are well-linked to the eighth grade math and reading scores, then we would also have the ability to reach down into elementary grades to gain insight into how many students in each state are on track at that early level to be college and career ready in those subjects. That is obviously a valuable piece of information.

It would be very worthwhile to tie the NAEP achievement level scores to something with great meaning to the public and the education community. The findings here offer hope that this may be possible, and that the NAEP is already providing important information that is currently unrecognized.

Richard G. Innes  
October 2, 2018  
70224.434@compuserve.com

Good afternoon-

I am writing as a private professional in educational measurement. My comments should be interpreted only as my own and should not be construed as representing my current employer. Disclosure: I have served in several capacities on NAEP projects over that last 25 years and was a member of the advisory panel convened by the Human Resources Research Organization (HumRRO) in July this year. I do not speak for other members of that panel. My comments follow.

First and foremost, I endorse the recommendations of the advisory panel and the revised ALs policy principles.

The seemingly small addition of the work "Content" to "Achievement Level Descriptors" provides important emphasis that these ALDs explicate the content area knowledge and skills targeted in NAEP assessments and required of examinees.

Principle 2.b emphasizes a modern concept in the validation of intended score interpretations and uses, that ALDs should be developed at the same time as assessment frameworks so that the assessment framework and all other design and development decisions and outcomes explicitly support the intended score interpretations and uses defined by the ALDs. This is the core of principled approaches to assessment design, development, and implementation and is fundamental to making rigorous validity arguments for the interpretation and use of NAEP assessment results.

Principle 2.c.i-iii and 2.d clarify nicely the idea that, while general public representation and input on NAEP ALs is crucial and ultimately democratic, representatives of the general public must be selected and trained so that they are well informed about the content area and students at the grade level under consideration in order to provide useful, well informed, and supported recommendations on AL cut scores.

Principle 3 in general, regarding validity evidence to support use of NAEP ALDs and data to interpret performance on NAEP assessments, is crucial to supporting such interpretations and uses.

However, principle 3.c may cause some confusion by stating that ALDs should reflect what students *do know and can do*, versus principle 2.b, which states that Content ALDs that guide AL setting should articulate what students *should know and be able to do*. Principle 3.c seems to be a reference to what are called Reporting ALDs in the standard setting literature. If so, to avoid confusion, why not use the term "Reporting ALDs" to distinguish them and their intended interpretations from Content ALDs.

Please let me know if you have questions about these comments.

Thank you,  
Steve Ferrara, PhD

This review of NAEP achievement levels policies and practices is long overdue. These changes provide a more comprehensive policy while simplifying and clarifying the policy. A tremendous thanks to all who have contributed to the document.

Here are my suggestions. I hope they will be of help in finalizing this document.

Page 5, Principle 1, part a-- I recommend adding this statement: "The content ALDs are the statement of the standard for each subject, grade, and level."

Page 5, Principle 1, part c—insert "know" so that the last sentence in that part reads: "what students in each achievement level know and can do."

Page 5, Principle 2, part a—There should be a provision for coordination between the framework development group and the Assistant Director for Psychometrics who is responsible for the ALS process for developing the ALDs.

Page 6, Principle 2, part c. ii—Need to specify that each panel must meet these requirements." The current statement is "This panel . . ." Also, I think you need a minimum number of general public panelists specified. I think it is a bad idea to have fewer than 1 GP per table group, for example. As currently stated, the policy would allow as few as one or two GP panelists per grade panel.

Page 6, Principle 2, part c. iii—I would either omit the reference to shared (common) items or provide more information by saying that the item pool is also divided into two equivalent sets with a subset that is evaluated by both panel groups.

Page 6, Principle 2, part d—I feel pretty strongly that content facilitators should have experience with NAEP, specifically the NAEP framework for the subject area. That expertise is the way to assure that the content facilitators have a "claim" to authority over the panelists who may also be content experts. In addition, I feel strongly that the process facilitators must have some background and training in quantitative analysis. I have experienced the negative effects of having a process facilitator lacking that background and not being a successful process facilitator, despite being an excellent facilitator of groups in general.

Page 7, Principle 2, part e—The first sentence needs clarification—differ from what?

Page 7, Principle 2, part h—I think "testing out" is a little misleading. I think it should be made clear that any new features of the process should be tested prior to the pilot study, and any research questions must be evaluated prior to the pilot study.

Page 7, Principle 2, part i--TACSS has always been appointed by the contractor, with Board approval; and TACSS "reports" to the contractor—not the Board. This is important, and the current wording of the policy may be thought to suggest that the TACSS is appointed by the Board.



Page 7, Principle 2, part j, line 3—I suggest changing “at the time” to “by the time”

Page 8, Principle 2, part k, first sentence—I don't know if the Board still upholds this principle, but there was once a very strong notion that it was incorrect/inappropriate to refer to students in the achievement levels. Rather, one was to talk about the performance of students classified in the achievement levels and to say that the performance represents the knowledge and skills required by the ALDs.

Page 8, Principle 3, part b—The first sentence needs to be re-written to clarify that national results are available for all assessments and state and TUDA results for only some assessments. Also, I suggest omitting the last part of the second sentence because it does not really follow logically from the first part. I don't think that the prohibition on reporting necessarily means that the achievement levels do not apply to individual students or schools.

Page 8, Principle 3, part c, sentence 3—I think there will always be ALDs reported with results, so I suggest the sentence be edited to say: “In particular, the content ALDs for reporting results shall be written to incorporate. . .”

Should the policy indicate who will write them and anything about the procedures required for that process? How to vet them, etc.

Page 9, Principle 3, part f—At and above each cut score would result in overlapping regions! I think the levels need to be described as performance within the score ranges demarcated by each of the three cut scores with no upper limit to performance at the Advanced level.

Page 9, Principle 3, part g—I think this would be clearer and perhaps more accurate to say that NAEP Proficient may describe performance that is different from that in other assessments and that is not necessarily suggested by the common usage of “proficiency.” (I think it is “proficiency” that gets things off-track!)

Page 9, Principle 4, part a—I note that this statement indicates that more than one NAEP assessment is to be involved in the review. I am pleased to see that!

Page 9, Principle 4, part b—Technically, this is fine, as written, but I feel some concern that this might imply that COSDAM, per se, will be expected to revise the ALDs. I think this statement should be clarified.

Page 9, Principle 4, part c—Again, I think it should be clear that this is to be coordinated with (actually, under the supervision of) the Assistant Director for Psychometrics because the ALDs must be developed with an understanding of how they are to be used in standard setting.

Page 9, Principle 4, part d—I think “revised” or “reset” sounds better than “redone.”

Page 10, Principle 5, part a—Suggest change to add some qualification about business representatives, such as: “business representatives in fields related to the content area.”

Page 10, Principle 5, part d—There is no indication of the qualifications required for TACSS members (for example, experience with NAEP, psychometric expertise, whether content knowledge is a consideration, representation of state assessment staff, representative of DAR contractor, etc.) nor the number of members for TACSS. There was also a requirement that at least one TACSS member have served on TACSS previously.

Page 11, Principle 5, part f—I am not sure this is needed since the Governing Board can ask anyone for input on the ALS process.

Page 11, Principle 6 –general statement—Suggest modification to state: “. . . ensure that the final achievement level descriptions, cut scores, and exemplars recommended to the Governing Board for adoption comply with this policy.”

Page 11, Principle 6, part b—This does not specify that COSDAM will and take action and make recommendations to the Board.

Page 11, Principle 6, part e, last line—Perhaps change “similar” to “relevant”.

Thank you for the opportunity to provide public comment on the review and possible revision of the NAGB policy on achievement levels.

As Commissioner of Education in Kentucky, I served on NAGB from 2011-2015 representing state superintendents.

I would encourage NAGB to maintain current cut scores and achievement levels to maintain the trend lines that have been established over the many years that NAEP has been administered. Given the many changes in education at the state and national levels over the past decade, it is critical that the "gold standard" of NAEP be maintained.

NAGB could certainly improve the communication about the achievement levels and provide clarity that proficient levels are set at the level of college and career ready expectations. Given the changing economy in our nation, it is critical that more high school graduates reach college and career readiness levels. NAEP is the "gold standard" to inform the public about our nation's performance and hold state assessments to the expectation that state achievement levels are set at rigorous levels that report on the preparation of students for college and career readiness.

Best wishes in your deliberation.

Terry Holliday

September 18, 2018

NAEP Achievement-Level-Setting Program  
National Assessment Governing Board  
800 North Capitol Street, NW  
Suite 825  
Washington, D.C. 20002

*Response to request for comments on "Draft Policy Statement on Developing Student Achievement Levels for the National Assessment of Educational Progress," posted Monday, September 10, 2018 by the National Assessment Governing Board, U.S. Department of Education in Notices of the Federal Register, Vol. 83, No. 175*

These comments are authorized by the Steering Committee of the National Superintendents Roundtable, representing 75 school superintendents responsible for the education of some five million K-12 students, and 25 former superintendents.

### **Summary**

The Roundtable appreciates the opportunity to comment on the draft policy statement. Although initially inclined to commend the National Assessment Governing Board (NAGB) for this effort to respond to the 2016 report of the National Academy of Sciences,<sup>1</sup> the Roundtable concludes that the proposed statement is unlikely to attain the goal defined in its first paragraph, namely to produce achievement levels for the National Assessment of Educational Progress (NAEP) that are "reasonable, useful, and informative to the public."

The Roundtable supports high standards. While it believes NAEP misuses the term "Proficient," it does not ask for a lowering of standards but for a clearer definition of what they mean in terms parents and the public can understand. In pursuit of that objective, we make the following observations: The original achievement levels were developed in a rushed process. Those levels have produced results have confused educators, citizens, and policymakers. Instead of being reasonable, the benchmarks represent "wishful thinking" and defy "reason" and "common sense," according to knowledgeable experts. The latest research linking NAEP's benchmarks to international assessments reveals that the majority of students in most nations cannot clear NAEP's proficiency bar. Finally, the proposed modifications to the policy definitions for NAEP achievement levels are unresponsive to the history of criticisms of the achievement levels and to the latest research.

We conclude with suggestions for rewriting the policy definitions and an appeal to extend the comment period from the currently contemplated 20 days to six months.

---

<sup>1</sup> Christopher Edley, Jr. & Judith A. Koenig (eds.). *Evaluation of the Achievement Levels for Mathematics and Reading on the National Assessment of Educational Progress*. Washington: National Academies Press, 2017.

### **Rushed Initial process**

Public Law 107-279, the *Augustus F. Hawkins-Robert T. Stafford Elementary and Secondary School Improvement Amendments of 1988*, created the National Assessment Governing Board and charged it with identifying appropriate achievement goals, which were established as “Basic,” “Proficient,” and “Advanced.”<sup>1</sup> An advisory panel appointed by NAGB in June 1990 to agree on the levels, and the proportion of students at each level who could be expected to answer questions correctly, reached agreement on these issues in November,<sup>2</sup> a remarkably brief period, given the importance and complexity of the issues involved. NAGB rejected the advice of experts to slow down the benchmarking process.<sup>3</sup> NAGB board member Chester E. Finn, Jr. later dismissed the value of technical expertise: “I get fed up with technical experts [who] . . . take an adversarial stance toward some of the things that are most important the views of those operating NAEP, such as setting standards.”<sup>4</sup>

### **The results have confused educators, the public, and policymakers**

Understandably the public, many commentators, and policymakers are confused by these benchmarks (achievement goals).

It is clear, as NCES states emphatically that “Proficient is not synonymous with grade-level performance.” According to Dr. Peggy Carr of NCES if the public wants to understand how many US students are performing on grade level it is more accurate to look at Basic scores, than Proficient results.<sup>5</sup>

Yet, on 2016, former television commentator Campbell Brown, newly leading an education advocacy organization called *The 74* released a video on *Slate* arguing that “Two out of three eighth graders in this country cannot read or do math at grade level.” Asked for evidence of this claim she cited NAEP, apparently confusing the term Proficient with being at grade level.<sup>6</sup>

Major organizations such as Achieve Inc. and Students First misuse the term consistently in their publications, while Achieve and NCES issue analyses lamenting the fact that state definitions of proficiency fall far short of the NAEP Proficient benchmark.<sup>7</sup>

---

<sup>1</sup> Jack Jennings, *A Paper on the NAEP Long-Term Trend Assessments Prepared for the National Assessment Governing Board*, February 13, 2017.

<sup>2</sup> M.A. Vinovskis, *Overseeing the Nation’s Report Card: The Creation and Evolution of the National Assessment Governing Board*. Washington: NAGB, 1998.

<sup>3</sup> Vinovskis, M.A., *op. cit.*

<sup>4</sup> Chester E. Finn, Jr., “An Interview with Chester E. Finn, Jr.,” in Lyle V. Jones and Ingram Olkins (eds.), *The Nation’s Report Card: Evolution and Perspectives*. Bloomington: Phi Delta Kappa Educational Foundation.

<sup>5</sup> Story Hinckley, “National Testing: What Does it Mean for a Student to be ‘Proficient’?”, *Christian Science Monitor*, April 12, 2018.

<sup>6</sup> Tom Loveless, “The NAEP Proficiency Myth”. *The Brown Center Chalkboard*, the Brookings Institution, June 13, 2016.

<sup>7</sup> See for example: Achieve, Inc: *Proficient vs. Prepared 2018: Disparities between State Tests and the 2017 National Assessment of Educational Progress*. Washington: Author, May 2018; also Bendeira de Mello, Rahman, T., and Park, B.J., *Mapping State Proficiency Standards onto NAEP Scales: Results from the 2015 NAEP Reading and Mathematics Assessments*. Washington, National Center for Education Statistics, 2018.

Earlier this month, *Education Week* released its latest “Quality Counts” report.<sup>8</sup> In this effort to “grade the states” on school quality, the report included state-by-state data on poverty gaps, high school graduation rates, and Advanced Placement experience. But the foundation of its achievement index is the proportion of students meeting the NAEP proficiency definitions in fourth- and eighth-grade reading and mathematics.

Oddly, the term Proficient as used by NAGB does not even mean proficient. As former NAGB officials wrote in 2001:<sup>9</sup>

Nor is performance at the Proficient level synonymous with ‘proficiency’ in the subject. That is, students who may be considered proficient in the subject, given the common usage of the term, might not satisfy the requirements for performance at the NAEP achievement level.

In short, although most citizens are understandably interested in performance at grade level, many confuse performance at grade level with NAEP’s proficiency benchmark. Understandably, journalists and many members of the public ignore the benchmark of Basic, interpreting it as an indication that large numbers of students are barely scraping by. This interpretation does great damage to public perceptions of public schools and perceptions of the ability and skills of the next generation of young Americans. To add to the confusion, officials associated with NAGB have openly acknowledged that NAEP’s definition of proficiency is unrelated to common usage of the term.

### **Benchmarks represent “wishful thinking”**

The rushed process produced a scientific debate that lasts to this day.

NAGB hired a team of evaluators in 1990 to study the process involved in developing the three levels. A year later the evaluators were fired after their draft report concluded that the process “must be viewed as insufficiently tested and validated, politically dominated, and of questionable credibility.”<sup>10</sup>

In 1993, the U.S. General Accounting Office labeled the standard-setting process as “procedurally flawed” producing results of “doubtful accuracy.”<sup>11</sup>

In 1999, the National Academy of Sciences reported the achievement-level setting procedures were flawed: “difficult and confusing . . . internally inconsistent . . . validity evidence for the cut scores is lacking . . . and the process has produced unreasonable results.”<sup>12</sup>

Shortly after *No Child Left Behind* was signed into law in 2001, Robert Linn, past president of the American Educational Association and of the National Council on Measurement in

---

<sup>8</sup> “Quality Counts 2018: K-12 Assessment & Chance for Success. Grading the States.”, *Education Week*, September 5, 2018.

<sup>9</sup> Loomis, S.C. and Bourque, M.L. (eds.). *National Assessment of Educational Progress Achievement Levels, 1992-1998 for Reading*. Washington: National Assessment Governing Board.

<sup>10</sup> Jack Jennings, *op. cit.*

<sup>11</sup> U.S. General Accounting Office. *Educational Achievement Standards: NAGB’s Approach Yields Misleading Interpretations*. GAO/PEMD-993-12. Washington, DC: Author. (ERIC database: ED359268)

<sup>12</sup> Pellegrino, J.W., Joners, L.R., and Mitchell, K.J. *Grading the Nation’s Report Card*. Washington, DC: National Academy Press.

Education, and former editor of the *Journal of Educational Measurement* called the “target of 100% proficient or above according to the NAEP standards is more like wishful thinking than a realistic possibility.”<sup>13</sup>

In 2007, researchers concluded that fully a third of high school seniors who completed calculus, the best students with the best teachers in the country, could not clear the proficiency bar. Moreover, they added, fully 50 percent of those who scored “basic” in twelfth grade math had achieved a bachelor’s degree<sup>14</sup> (a proportion comparing favorably with four-year degree rates at public universities).

The Buros Institute, named after the father of *Mental Measurements Yearbook*, criticized the lack of a validity framework for NAEP assessment scores in 2009 and recommending continuing “to explore achievement level methodologies.”<sup>15</sup>

Fully 30 percent of 12<sup>th</sup>-graders who completed calculus were deemed to be less than proficient, said a Brookings Institution scholar in 2016, a figure that jumped to 69 percent for pre-calculus students and 92 percent for students who completed trigonometry and Algebra I. These data “defy reason” and “refute common sense,” he concluded.<sup>16</sup>

Finally, the NAS study to which the proposed rule responds took note in 2016 of the “controversy and disagreement around the achievement levels, noting that Congress has insisted since 1994 that the achievement levels are to be used on a trial basis until on objective evaluation determined them to be “reasonable, reliable, valid, and informative to the public.”<sup>17</sup>

In the Roundtable’s judgment, such an objective evaluation has yet to be completed and a determination that the achievement levels are “reasonable, reliable, valid, and informative to the public” has yet to be seen.

### **Linking studies conclude most students in most nations cannot clear “proficiency” bar**

The Roundtable points also to research studies dating from 2007 to 2018 indicating NAEP’s proficiency bar is beyond the reach of most students in most nations. When Gary Phillips of the American Institutes of Research (and former Acting Commissioner of NCES) asked how students in other nations would perform if their international assessment results were expressed in terms of NAEP achievement levels, his results were sobering.<sup>18</sup> The results demonstrated that just three nations (Singapore, the Republic of Korea, and Japan) would have a majority of their

---

<sup>13</sup> Cited in Loveless, 2016.

<sup>14</sup> Scott, L. A., Ingels, S. J., and Owings, J. A. (2007). *Interpreting 12th-graders' NAEP-scaled mathematics performance and postsecondary outcomes from the National Education Longitudinal Study of 1988 (NELS:88)*. U.S. Department of Education, National Center for Education Statistics, (NCES 2007-328).

<sup>15</sup> Buckendahl, C.W. et al, (2009). *Evaluation of the National Assessment of Educational Progress*. Lincoln, NE: Buros Center for Testing, University of Nebraska.

<sup>16</sup> Loveless, T., “The NAEP Proficiency Myth,” (2016). Brookings Institution: Brown Center Chalkboard, June 13, 2016.

<sup>17</sup> Edley & Koenig (eds). Op. cit.

<sup>18</sup> Phillips, G. W. (2007). *Expressing international education achievement in terms of U.S. performance standards: Linking NAEP achievement levels to TIMSS*. Washington, DC: American Institutes for Research.

students clear the NAEP bar in 8<sup>th</sup>-grade mathematics, while Singapore alone could meet that standard (more than 50% of students clearing the bar) in science.<sup>19</sup>

Subsequently Hambleton, Sireci, and Smith (2007) and also Lim and Sireci (2017) reached conclusions similar to those of Phillips.

An analysis conducted by Gonulates and Harvey in 2017<sup>20</sup> and reported by the Horace Mann League and the National Superintendents Roundtable in 2018<sup>21</sup> applied Phillips' methodology to international fourth-grade reading results as defined by the Progress in International Reading and Literacy Survey (PIRLS). PIRLS is conducted by the International Association for the Evaluation of Educational Achievement (IEA), the organization that administers the TIMSS mathematics and science assessments. Not a single nation participating in PIRLS can demonstrate that a majority of its students clear the NAEP reading proficiency bar in Grade 4. Singapore leads the way, with 39 percent of its students clearing the bar, followed by the Russian Federation (37 percent), Finland (36 percent), England (32 percent), and the United States (31 percent).

In light of these findings, it defies logic and common sense, as Dr. Loveless noted, to act on the belief that NAEP's definition of proficiency is a reasonable and reliable guide to policy action.

**Proposed new policy definitions define anything less than "NAEP Proficient" as deficient**

Against that backdrop, it is surprising to see that NAGB proposes not to modify its policy definitions to clarify the term "Proficient" but to tinker at the margins with the definitions and add the word "NAEP" in front of each of the benchmarks. The justification offered in the Federal Register is to "better differentiate the NAEP achievement levels from other common uses of Basic, Proficient, and Advanced." But the effect is to define anything less than NAEP Proficient as deficient.

The table on the following compares the existing policy definitions of the terms Basic, Proficient, and Advanced, with the policy definitions attached to the new terms "NAEP Basic," "NAEP Proficient," and "NAEP Advanced." Like the existing policy definitions, the proposed definitions point inexorably to Proficient as the ideal to which states, the nation, and major demographic groups should, on average, aspire.

In the Roundtable's judgment these minor modifications are in no way responsive to the major criticisms that have been leveled at the NAEP benchmarks over the years. To retreat behind the claim that the Proficient benchmark is an aspirational standard is deceptive and evasive. In light of the results of the international linking studies completed between 2007 and 2018, far from being aspirational, the Proficient benchmark is delusional. In Dr. Linn's phrase, the proficiency benchmark represents wishful thinking.

---

<sup>19</sup> It should be noted that no nation (or city or province) could demonstrate that its students met the NAEP definition of Advanced. (Cities and provinces, such as Chinese Taipei, Hong Kong are generally ignored for purposes of the analysis in this section.)

<sup>20</sup> See Appendix C of *How High the Bar?: "Applying NAEP Benchmarks to PIRLS results."*

<sup>21</sup> The National Superintendents Roundtable and the Horace Mann League: *How High the Bar? How Would Other Nations Perform if Their Students Were Judged by Common Core or NAEP Benchmarks?* Seattle: Authors, 2018.



<b>Benchmarks</b>	<b>Current Policy Definitions</b>	<b>Proposed Policy Definitions</b>
<b>Basic</b>	This level denotes partial mastery of prerequisite knowledge and skills that are fundamental for proficient work at each grade.	This level denotes partial mastery of prerequisite knowledge and skills that are fundamental for performance at the NAEP Proficient level.
<b>Proficient</b>	This level represents solid academic performance for each grade assessed. Students reaching this level have demonstrated competency over challenging subject matter, including subject-matter knowledge, application of such knowledge to real-world situations, and analytical skills appropriate to the subject matter.	This level represents solid academic performance for each NAEP assessment. Students reaching this level have demonstrated competency over challenging subject matter, including subject matter knowledge, application of such knowledge to real world situations, and analytical skills appropriate to the subject matter.
<b>Advanced</b>	This higher level signifies superior performance.	This level signifies superior performance beyond NAEP Proficient.

Repairing the damage inflicted on public schools and their students by NAEP's misleading terminology is all the more critical today as groups such as Achieve, Inc. openly advocate for NAEP's definition of Proficient as the benchmark to be employed in state assessments under the *Every Student Succeeds Act*.

### Recommendations

Against the backdrop of the discussion above, the Roundtable makes two recommendations:

1. Revise the proposed policy definitions so that they are more informative to the general public (one of the principal goals of the NAS analysis) We propose the following definitions:

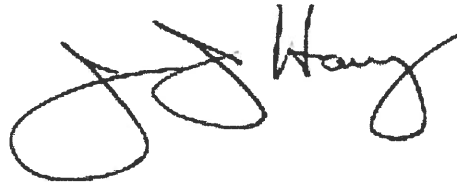
NAEP Basic	This level is roughly analogous to performance at grade level in each NAEP assessment. It denotes partial mastery of prerequisite knowledge and skills that are fundamental for performance at the NAEP Proficient level.
NAEP Proficient	This level represents extremely demanding academic performance, a level beyond the abilities of most students in most nations. Students reaching this level have demonstrated competency over challenging subject matter, including subject matter knowledge, application of such knowledge to real world situations, and analytical skills appropriate to the subject matter.
NAEP Advanced	This level signifies superior performance beyond NAEP Proficient.

2. Extend the comment period on the draft policy statement for six months, during which time NAGB should hold between four and six hearings around the nation to solicit the advice of educators, parents, and other stakeholders. George Santayana's oft-quoted aphorism is usefully

born in mind here: Those who cannot remember the past are condemned to repeat it. The rushed process that produced the NAEP benchmarks in 1990 provides a cautionary tale for us today. NAEP has existed for nearly half a century. The benchmarks have been in use for over 25 years. Taking a few months to get it right on this occasion will be time well spent.

We commend NAGB for inviting comments and appreciate the opportunity to offer our insights. We stand ready to answer questions or help in any way with this process as it moves forward.

Sincerely,

A handwritten signature in black ink, appearing to read "James J. Harvey". The signature is fluid and cursive, with the first name "James" being more prominent and larger than the last name "Harvey".

James J. Harvey, Executive Director  
(On behalf of Roundtable Steering Committee)

c.c. Peggy Carr, NCES

by email to [NAEPALSpolicy@ed.gov](mailto:NAEPALSpolicy@ed.gov)

&

[Peggy.Carr@ed.gov](mailto:Peggy.Carr@ed.gov)

And by overnight mail to:  
NAEP Achievement-Level-Setting Program  
National Assessment Governing Board  
800 North Capitol Street, NW  
Suite 825  
Washington, D.C. 20002

I have been following the NAEP data for 30 years as a teacher, principal and superintendent. The data is always misrepresented in the media and to the public. I am very concerned that the term "proficiency" will again be changed, and misrepresented again!

I do not believe the changes proposed by the National Assessment Governing Board to the achievement benchmarks for the National Assessment of Educational Progress will produce achievement levels that are "reasonable, useful, and informative to the public," the goal defined in the first paragraph of the proposed policy statement.

These standards and benchmarks must be expressed in terms that parents and the public can understand. To that end, I point out the following: The original achievement levels were developed in a rushed process. Those levels have produced results that have confused educators, citizens, and policymakers. Instead of being reasonable, the benchmarks represent "wishful thinking" and defy "reason" and "common sense," according to knowledgeable experts. The latest research linking NAEP's benchmarks to international assessments reveals that the majority of students in most nations cannot clear NAEP's proficiency bar. Finally, the proposed modifications to the policy definitions for NAEP achievement levels are unresponsive to the history of criticisms of the achievement levels and to the latest research.

NAGB should rewrite the policy definitions for its standards and extend the comment period from the completely inadequate 20 days now contemplated to three months or more to give educators time to comment.

Frank Ohnesorgen  
Superintendent  
Pond USD  
29585 Pond Road  
Wasco, CA 93280

To whom it may concern,

I do not believe the changes proposed by the National Assessment Governing Board to the achievement benchmarks for the National Assessment of Educational Progress will produce achievement levels that are "reasonable, useful, and informative to the public," the goal defined in the first paragraph of the proposed policy statement.

These standards and benchmarks must be expressed in terms that parents and the public can understand. To that end, I point out the following: The original achievement levels were developed in a rushed process. Those levels have produced results that have confused educators, citizens, and policymakers. Instead of being reasonable, the benchmarks represent "wishful thinking" and defy "reason" and "common sense," according to knowledgeable experts. The latest research linking NAEP's benchmarks to international assessments reveals that the majority of students in most nations cannot clear NAEP's proficiency bar. Finally, the proposed modifications to the policy definitions for NAEP achievement levels are unresponsive to the history of criticisms of the achievement levels and to the latest research.

NAGB should rewrite the policy definitions for its standards and extend the comment period from the completely inadequate 20 days now contemplated to three months or more to give educators time to comment.

Sincerely,

Dr. Geoff Thomas  
Superintendent  
Madison 321

To Whom It May Concern,

I do not believe the changes proposed by the National Assessment Governing Board to the achievement benchmarks for the National Assessment of Educational Progress will produce achievement levels that are "reasonable, useful, and informative to the public," the goal defined in the first paragraph of the proposed policy statement.

These standards and benchmarks must be expressed in terms that parents and the public can understand. To that end, I point out the following: The original achievement levels were developed in a rushed process. Those levels have produced results that have confused educators, citizens, and policymakers. Instead of being reasonable, the benchmarks represent "wishful thinking" and defy "reason" and "common sense," according to knowledgeable experts. The latest research linking NAEP's benchmarks to international assessments reveals that the majority of students in most nations cannot clear NAEP's proficiency bar. Finally, the proposed modifications to the policy definitions for NAEP achievement levels are unresponsive to the history of criticisms of the achievement levels and to the latest research.

NAGB should rewrite the policy definitions for its standards and extend the comment period from the completely inadequate 20 days now contemplated to three months or more to give educators time to comment.

Respectfully Submitted,

James H. Erinakes, II  
Superintendent of Schools  
Exeter-West Greenwich Regional School District  
Rhode Island

Dear Ms. Carr,

I do not believe the changes proposed by the National Assessment Governing Board to the achievement benchmarks for the National Assessment of Educational Progress will produce achievement levels that are "reasonable, useful, and informative to the public," the goal defined in the first paragraph of the proposed policy statement.

These standards and benchmarks must be expressed in terms that parents and the public can understand. To that end, I point out the following: The original achievement levels were developed in a rushed process. Those levels have produced results that have confused educators, citizens, and policymakers. Instead of being reasonable, the benchmarks represent "wishful thinking" and defy "reason" and "common sense," according to knowledgeable experts. The latest research linking NAEP's benchmarks to international assessments reveals that the majority of students in most nations cannot clear NAEP's proficiency bar. Finally, the proposed modifications to the policy definitions for NAEP achievement levels are unresponsive to the history of criticisms of the achievement levels and to the latest research.

NAGB should rewrite the policy definitions for its standards and extend the comment period from the completely inadequate 20 days now contemplated to three months or more to give educators time to comment

Thanks for considering.

Jamey Harvey

CEO

Agilian.com

To Whom It May Concern,

I do not believe the changes proposed by the National Assessment Governing Board to the achievement benchmarks for the National Assessment of Educational Progress will produce achievement levels that are "reasonable, useful, and informative to the public," the goal defined in the first paragraph of the proposed policy statement.

These standards and benchmarks must be expressed in terms that parents and the public can understand. To that end, I point out the following: The original achievement levels were developed in a rushed process. Those levels have produced results that have confused educators, citizens, and policymakers. Instead of being reasonable, the benchmarks represent "wishful thinking" and defy "reason" and "common sense," according to knowledgeable experts. The latest research linking NAEP's benchmarks to international assessments reveals that the majority of students in most nations cannot clear NAEP's proficiency bar. Finally, the proposed modifications to the policy definitions for NAEP achievement levels are unresponsive to the history of criticisms of the achievement levels and to the latest research.

NAGB should rewrite the policy definitions for its standards and extend the comment period from the completely inadequate 20 days now contemplated to three months or more to give educators time to comment.

Sincerely,  
Jessica Essenter  
Third Grade Teacher in Connecticut

Dear Department of Education:

I do not believe the changes proposed by the National Assessment Governing Board to the achievement benchmarks for the National Assessment of Educational Progress will produce achievement levels that are "reasonable, useful, and informative to the public," the goal defined in the first paragraph of the proposed policy statement.

These standards and benchmarks must be expressed in terms that parents and the public can understand. To that end, I point out the following: The original achievement levels were developed in a rushed process. Those levels have produced results that have confused educators, citizens, and policymakers. Instead of being reasonable, the benchmarks represent "wishful thinking" and defy "reason" and "common sense," according to knowledgeable experts. The latest research linking NAEP's benchmarks to international assessments reveals that the majority of students in most nations cannot clear NAEP's proficiency bar. Finally, the proposed modifications to the policy definitions for NAEP achievement levels are unresponsive to the history of criticisms of the achievement levels and to the latest research.

NAGB should rewrite the policy definitions for its standards and extend the comment period from the completely inadequate 20 days now contemplated to three months or more to give educators time to comment.

**Kim Fry**  
Superintendent  
Rochester School District



NAEP Achievement-Level-Setting Program  
National Assessment Governing Board  
800 North Capitol Street, NW  
Suite 825  
Washington, D.C. 20002

RE: Response to draft policy statement on developing student achievement levels for the National Assessment of Educational Progress

To Whom It May Concern:

I do not believe the changes proposed by the National Assessment Governing Board to the achievement benchmarks for the National Assessment of Educational Progress will produce achievement levels that are "reasonable, useful, and informative to the public," the goal defined in the first paragraph of the proposed policy statement.

These standards and benchmarks must be expressed in terms that parents and the public can understand. To that end, I point out the following: The original achievement levels were developed in a rushed process. Those levels have produced results that have confused educators, citizens, and policymakers. Instead of being reasonable, the benchmarks represent "wishful thinking" and defy "reason" and "common sense," according to knowledgeable experts. The latest research linking NAEP's benchmarks to international assessments reveals that the majority of students in most nations cannot clear NAEP's proficiency bar. Finally, the proposed modifications to the policy definitions for NAEP achievement levels are unresponsive to the history of criticisms of the achievement levels and to the latest research.

NAGB should rewrite the policy definitions for its standards and extend the comment period from the completely inadequate 20 days now contemplated to three months or more to give educators time to comment.

Thank you for considering my comments.

Kurt H. Hilyard  
District Superintendent  
Ocosta School District #172  
2580 South Montesano Street  
Westport, WA 98595

There is a lot of research out there that shows the current NAEP benchmarks are way above what are normal expectations. The research proves over and over again that no country can get most of their students to achieve at those benchmarks. The current benchmarks reflect more of a gifted standard which is only 10-15% of the population, not an expectation for the general populace. Also, the NAEP assessments are only given to a random sampling of students. They should never be used to rate or compare state or local education systems.

I do not believe the changes proposed by the National Assessment Governing Board to the achievement benchmarks for the National Assessment of Educational Progress will produce achievement levels that are "reasonable, useful, and informative to the public," the goal defined in the first paragraph of the proposed policy statement.

These standards and benchmarks must be expressed in terms that parents and the public can understand. To that end, I point out the following: The original achievement levels were developed in a rushed process. Those levels have produced results that have confused educators, citizens, and policymakers. Instead of being reasonable, the benchmarks represent "wishful thinking" and defy "reason" and "common sense," according to knowledgeable experts. The latest research linking NAEP's benchmarks to international assessments reveals that the majority of students in most nations cannot clear NAEP's proficiency bar. Finally, the proposed modifications to the policy definitions for NAEP achievement levels are unresponsive to the history of criticisms of the achievement levels and to the latest research.

NAGB should rewrite the policy definitions for its standards and extend the comment period from the completely inadequate 20 days now contemplated to three months or more to give educators time to comment.



**Laurie Vent**  
**Superintendent** Upper Sandusky E. V. Schools

p: 419-294-2306 p/f: 419-294-6891 a: 800 N. Sandusky Ave., Suite A Upper Sandusky, OH 43351

w: [www.usevs.org](http://www.usevs.org) e: [laurie\\_v@usevs.org](mailto:laurie_v@usevs.org)

Good afternoon:

My name is Matthew Montgomery and I am the superintendent of Revere Local Schools in Bath, Ohio.

I do not believe the changes proposed by the National Assessment Governing Board to the achievement benchmarks for the National Assessment of Educational Progress will produce achievement levels that are "reasonable, useful, and informative to the public," the goal defined in the first paragraph of the proposed policy statement.

These standards and benchmarks must be expressed in terms that parents and the public can understand. To that end, I point out the following: The original achievement levels were developed in a rushed process. Those levels have produced results that have confused educators, citizens, and policymakers. Instead of being reasonable, the benchmarks represent "wishful thinking" and defy "reason" and "common sense," according to knowledgeable experts. The latest research linking NAEP's benchmarks to international assessments reveals that the majority of students in most nations cannot clear NAEP's proficiency bar. Finally, the proposed modifications to the policy definitions for NAEP achievement levels are unresponsive to the history of criticisms of the achievement levels and to the latest research.

NAGB should rewrite the policy definitions for its standards and extend the comment period from the completely inadequate 20 days now contemplated to three months or more to give educators time to comment.


Sincerely,

Matthew Montgomery

---

Matthew L. Montgomery  
Superintendent  
Revere Local Schools

Twitter: @SuptMontgomery

 Revere Local Schools

I do not believe the changes proposed by the National Assessment Governing Board to the achievement benchmarks for the National Assessment of Educational Progress will produce achievement levels that are "reasonable, useful, and informative to the public," the goal defined in the first paragraph of the proposed policy statement.

These standards and benchmarks must be expressed in terms that parents and the public can understand. To that end, I point out the following: The original achievement levels were developed in a rushed process. Those levels have produced results that have confused educators, citizens, and policymakers. Instead of being reasonable, the benchmarks represent "wishful thinking" and defy "reason" and "common sense," according to knowledgeable experts. The latest research linking NAEP's benchmarks to international assessments reveals that the majority of students in most nations cannot clear NAEP's proficiency bar. Finally, the proposed modifications to the policy definitions for NAEP achievement levels are unresponsive to the history of criticisms of the achievement levels and to the latest research.

NAGB should rewrite the policy definitions for its standards and extend the comment period from the completely inadequate 20 days now contemplated to three months or more to give educators time to comment.

While I have cut and pasted the suggested comments, The current framework is not working and because all the scores are so low across the board, the reporting lacks value to students, parents, and staff.

Best regards,

*Matthew G Scoggins*  
*Rangely School District Re-4*

I do not believe the changes proposed by the National Assessment Governing Board to the achievement benchmarks for the National Assessment of Educational Progress will produce achievement levels that are "reasonable, useful, and informative to the public," the goal defined in the first paragraph of the proposed policy statement.

These standards and benchmarks must be expressed in terms that parents and the public can understand. To that end, I point out the following: The original achievement levels were developed in a rushed process. Those levels have produced results that have confused educators, citizens, and policymakers. Instead of being reasonable, the benchmarks represent "wishful thinking" and defy "reason" and "common sense," according to knowledgeable experts. The latest research linking NAEP's benchmarks to international assessments reveals that the majority of students in most nations cannot clear NAEP's proficiency bar. Finally, the proposed modifications to the policy definitions for NAEP achievement levels are unresponsive to the history of criticisms of the achievement levels and to the latest research.

NAGB should rewrite the policy definitions for its standards and extend the comment period from the completely inadequate 20 days now contemplated to three months or more to give educators time to comment.

--

Michael J. Hoose, Superintendent

Dear Dr. Carr,

I do not believe the changes proposed by the National Assessment Governing Board to the achievement benchmarks for the National Assessment of Educational Progress will produce achievement levels that are "reasonable, useful, and informative to the public," the goal defined in the first paragraph of the proposed policy statement.

These standards and benchmarks must be expressed in terms that parents and the public can understand. To that end, I point out the following: The original achievement levels were developed in a rushed process. Those levels have produced results that have confused educators, citizens, and policymakers. Instead of being reasonable, the benchmarks represent "wishful thinking" and defy "reason" and "common sense," according to knowledgeable experts. The latest research linking NAEP's benchmarks to international assessments reveals that the majority of students in most nations cannot clear NAEP's proficiency bar. Finally, the proposed modifications to the policy definitions for NAEP achievement levels are unresponsive to the history of criticisms of the achievement levels and to the latest research.

NAGB should rewrite the policy definitions for its standards and extend the comment period from the completely inadequate 20 days now contemplated to three months or more to give educators time to comment.

Sincerely,  
Oren Pizmony-Levy

----

Oren Pizmony-Levy, PhD  
Assistant Professor of International and Comparative Education  
Department of International and Transcultural Studies  
Teachers College, Columbia University  
525 West 120th Street  
370 Grace Dodge Hall  
Box 55  
New York, NY 10027

I do not believe the changes proposed by the National Assessment Governing Board to the achievement benchmarks for the National Assessment of Educational Progress will produce achievement levels that are "reasonable, useful, and informative to the public," the goal defined in the first paragraph of the proposed policy statement.

These standards and benchmarks must be expressed in terms that parents and the public can understand. To that end, I point out the following: The original achievement levels were developed in a rushed process. Those levels have produced results that have confused educators, citizens, and policymakers. Instead of being reasonable, the benchmarks represent "wishful thinking" and defy "reason" and "common sense," according to knowledgeable experts. The latest research linking NAEP's benchmarks to international assessments reveals that the majority of students in most nations cannot clear NAEP's proficiency bar. Finally, the proposed modifications to the policy definitions for NAEP achievement levels are unresponsive to the history of criticisms of the achievement levels and to the latest research.

NAGB should rewrite the policy definitions for its standards and extend the comment period from the completely inadequate 20 days now contemplated to three months or more to give educators time to comment.

Richard Myles  
Superintrndent

Scottsbluff (NE) Public Schools  
1722 First Avenue  
Scottsbluff, NE 69361

Every coach sets the goal of being undefeated on the end of the season, but it just isn't possible for almost every team. I do not believe the changes proposed by the National Assessment Governing Board to the achievement benchmarks for the National Assessment of Educational Progress will produce achievement levels that are "reasonable, useful, and informative to the public," the goal defined in the first paragraph of the proposed policy statement.

These standards and benchmarks must be expressed in terms that parents and the public can understand. To that end, I point out the following: The original achievement levels were developed in a rushed process. Those levels have produced results that have confused educators, citizens, and policymakers. Instead of being reasonable, the benchmarks represent "wishful thinking" and defy "reason" and "common sense," according to knowledgeable experts. The latest research linking NAEP's benchmarks to international assessments reveals that the majority of students in most nations cannot clear NAEP's proficiency bar. Finally, the proposed modifications to the policy definitions for NAEP achievement levels are unresponsive to the history of criticisms of the achievement levels and to the latest research.

NAGB should rewrite the policy definitions for its standards and extend the comment period from the completely inadequate 20 days now contemplated to three months or more to give educators time to comment.

Rob Busch  
Superintendent/PK-6 Principal  
Edgewood-Colesburg CSD  
409 East Street – P.O. Box 125  
Colesburg, Iowa 52035



I do not believe the changes proposed by the National Assessment Governing Board to the achievement benchmarks for the National Assessment of Educational Progress will produce achievement levels that are "reasonable, useful, and informative to the public," the goal defined in the first paragraph of the proposed policy statement.

These standards and benchmarks must be expressed in terms that parents and the public can understand. To that end, I point out the following: The original achievement levels were developed in a rushed process. Those levels have produced results that have confused educators, citizens, and policymakers. Instead of being reasonable, the benchmarks represent "wishful thinking" and defy "reason" and "common sense," according to knowledgeable experts. The latest research linking NAEP's benchmarks to international assessments reveals that the majority of students in most nations cannot clear NAEP's proficiency bar. Finally, the proposed modifications to the policy definitions for NAEP achievement levels are unresponsive to the history of criticisms of the achievement levels and to the latest research.

NAGB should rewrite the policy definitions for its standards and extend the comment period from the completely inadequate 20 days now contemplated to three months or more to give educators time to comment.

**Rob Terrill**  
District Assessment and Elementary Curriculum Director  
Highland Local Schools  
6506 SR 229  
Marengo, Ohio 43334

**To whom it may concern:**

I do not believe the changes proposed by the National Assessment Governing Board to the achievement benchmarks for the National Assessment of Educational Progress will produce achievement levels that are "reasonable, useful, and informative to the public," the goal defined in the first paragraph of the proposed policy statement.

These standards and benchmarks must be expressed in terms that parents and the public can understand. To that end, I point out the following: The original achievement levels were developed in a rushed process. Those levels have produced results that have confused educators, citizens, and policymakers. Instead of being reasonable, the benchmarks represent "wishful thinking" and defy "reason" and "common sense," according to knowledgeable experts. The latest research linking NAEP's benchmarks to international assessments reveals that the majority of students in most nations cannot clear NAEP's proficiency bar. Finally, the proposed modifications to the policy definitions for NAEP achievement levels are unresponsive to the history of criticisms of the achievement levels and to the latest research.

NAGB should rewrite the policy definitions for its standards and extend the comment period from the completely inadequate 20 days now contemplated to three months or more to give educators time to comment.

**Sincerely,**

--

Robert L. Cullison, Jr.  
Superintendent  
Prospect Mountain High School SAU #301  
242 Suncook Valley Road  
Alton, NH 03809

To Whom It May Concern:

I do not believe the changes proposed by the National Assessment Governing Board to the achievement benchmarks for the National Assessment of Educational Progress will produce achievement levels that are "reasonable, useful, and informative to the public," the goal defined in the first paragraph of the proposed policy statement.

These standards and benchmarks must be expressed in terms that parents and the public can understand. To that end, I point out the following: The original achievement levels were developed in a rushed process. Those levels have produced results that have confused educators, citizens, and policymakers. Instead of being reasonable, the benchmarks represent "wishful thinking" and defy "reason" and "common sense," according to knowledgeable experts. The latest research linking NAEP's benchmarks to international assessments reveals that the majority of students in most nations cannot clear NAEP's proficiency bar. And in many of those nations, they educate only their best and brightest.

NAGB should rewrite the policy definitions for its standards and extend the comment period from the completely inadequate 20 days now contemplated to three months or more to give educators time to comment.

**Shannon Peterson**  
**Director**



Main Campus  
246 11th Ave SE  
Forest Lake MN 55025 USA  
651-464-0771

LILA Annex  
121 11th Ave SE  
Forest Lake MN 55025 USA  
651-464-8989

Headwaters Campus  
19850 Fenway Ave N  
Forest Lake MN 55025 USA  
651-464-8989

I do not believe the changes proposed by the National Assessment Governing Board to the achievement benchmarks for the National Assessment of Educational Progress will produce achievement levels that are "reasonable, useful, and informative to the public," the goal defined in the first paragraph of the proposed policy statement.

These standards and benchmarks must be expressed in terms that parents and the public can understand. To that end, I point out the following: The original achievement levels were developed in a rushed process. Those levels have produced results that have confused educators, citizens, and policymakers. Instead of being reasonable, the benchmarks represent "wishful thinking" and defy "reason" and "common sense," according to knowledgeable experts. The latest research linking NAEP's benchmarks to international assessments reveals that the majority of students in most nations cannot clear NAEP's proficiency bar. Finally, the proposed modifications to the policy definitions for NAEP achievement levels are unresponsive to the history of criticisms of the achievement levels and to the latest research.

NAGB should rewrite the policy definitions for its standards and extend the comment period from the completely inadequate 20 days now contemplated to three months or more to give educators time to comment.

Thank you and I am able to give more information if someone would like.

Shawn

--

Shawn M Chabot  
Assistant Superintendent/Chief Academic Officer  
Lewiston Public Schools

I do not believe the changes proposed by the National Assessment Governing Board to the achievement benchmarks for the National Assessment of Educational Progress will produce achievement levels that are "reasonable, useful, and informative to the public," the goal defined in the first paragraph of the proposed policy statement.

These standards and benchmarks must be expressed in terms that parents and the public can understand. To that end, I point out the following: The original achievement levels were developed in a rushed process. Those levels have produced results that have confused educators, citizens, and policymakers. Instead of being reasonable, the benchmarks represent "wishful thinking" and defy "reason" and "common sense," according to knowledgeable experts. The latest research linking NAEP's benchmarks to international assessments reveals that the majority of students in most nations cannot clear NAEP's proficiency bar. Finally, the proposed modifications to the policy definitions for NAEP achievement levels are unresponsive to the history of criticisms of the achievement levels and to the latest research.

NAGB should rewrite the policy definitions for its standards and extend the comment period from the completely inadequate 20 days now contemplated to three months or more to give educators time to comment.

Sincerely,

Steven D. Parker, Ed.D.  
Superintendent  
Lancaster County Public Schools  
P.O. Box 2000  
Kilmarnock, VA 22482



I do not believe the changes proposed by the National Assessment Governing Board to the achievement benchmarks for the National Assessment of Educational Progress will produce achievement levels that are "reasonable, useful, and informative to the public," the goal defined in the first paragraph of the proposed policy statement.

These standards and benchmarks must be expressed in terms that parents and the public can understand. To that end, I point out the following: The original achievement levels were developed in a rushed process. Those levels have produced results that have confused educators, citizens, and policymakers. Instead of being reasonable, the benchmarks represent "wishful thinking" and defy "reason" and "common sense," according to knowledgeable experts. The latest research linking NAEP's benchmarks to international assessments reveals that the majority of students in most nations cannot clear NAEP's proficiency bar. Finally, the proposed modifications to the policy definitions for NAEP achievement levels are unresponsive to the history of criticisms of the achievement levels and to the latest research.

NAGB should rewrite the policy definitions for its standards and extend the comment period from the completely inadequate 20 days now contemplated to three months or more to give educators time to comment.

Sincerely,  
Susette Bollard

--

Susette L. Bollard  
Superintendent of Schools  
Central Vermont Supervisory Union  
Serving the Towns of Northfield, Orange, Washington, and Williamstown

## To Whom It May Concern:

I do not believe the changes proposed by the National Assessment Governing Board to the achievement benchmarks for the National Assessment of Educational Progress will produce achievement levels that are "reasonable, useful, and informative to the public," the goal defined in the first paragraph of the proposed policy statement.

These standards and benchmarks must be expressed in terms that parents and the public can understand. To that end, I point out the following: The original achievement levels were developed in a rushed process. Those levels have produced results that have confused educators, citizens, and policymakers. Instead of being reasonable, the benchmarks represent "wishful thinking" and defy "reason" and "common sense," according to knowledgeable experts. The latest research linking NAEP's benchmarks to international assessments reveals that the majority of students in most nations cannot clear NAEP's proficiency bar. Finally, the proposed modifications to the policy definitions for NAEP achievement levels are unresponsive to the history of criticisms of the achievement levels and to the latest research.

NAGB should rewrite the policy definitions for its standards and extend the comment period from the completely inadequate 20 days now contemplated to three months or more to give educators time to comment.

Dr. Theresa R Rouse

Superintendent

Joliet Public Schools District 86

420 N. Raynor Ave.

Joliet, IL 60435

District Website: <https://www.joliet86.org/>

Superintendent's Blog: <https://www.joliet86.org//superintendent-blog/>

To whom it may concern:

I do not believe the changes proposed by the National Assessment Governing Board to the achievement benchmarks for the National Assessment of Educational Progress will produce achievement levels that are "reasonable, useful, and informative to the public," the goal defined in the first paragraph of the proposed policy statement.

These standards and benchmarks must be expressed in terms that parents and the public can understand. To that end, I wish to note the following:

The original achievement levels were developed in a rushed process. Those levels have produced results that have confused educators, citizens, and policymakers. Instead of being reasonable, the benchmarks represent "wishful thinking" and defy "reason" and "common sense," according to knowledgeable experts.

The latest research linking NAEP's benchmarks to international assessments reveals that the majority of students in most nations cannot clear NAEP's proficiency bar.

Finally, the proposed modifications to the policy definitions for NAEP achievement levels are unresponsive to the history of criticisms of the achievement levels and to the latest research.

NAGB should rewrite the policy definitions for its standards and extend the comment period from the completely inadequate 20 days now contemplated to three months or more to give educators time to comment.

Tim O. Mains, Superintendent  
Pine Bush Central School Dist.

twitter: @mainsPBsuper

*Knowing is not enough; we must apply.*

*Willing is not enough; we must do.*

~ Johann Wolfgang von Goethe

CONFIDENTIALITY NOTICE: This message is from the Superintendent's Office of the Pine Bush Central School District. The message and any attachments may be confidential or privileged and are intended only for the individual or entity identified above as the addressee. If you are not the addressee, or if this message has been addressed to you in error, you are not authorized to read, copy, or distribute this message or any attachments. Please delete this message and any attachments and notify the sender by return email or call the Superintendent's office at (845)744-4009.



I do not believe the changes proposed by the National Assessment Governing Board to the achievement benchmarks for the National Assessment of Educational Progress will produce achievement levels that are "reasonable, useful, and informative to the public," the goal defined in the first paragraph of the proposed policy statement.

These standards and benchmarks must be expressed in terms that parents and the public can understand. To that end, I point out the following: The original achievement levels were developed in a rushed process. Those levels have produced results that have confused educators, citizens, and policymakers. Instead of being reasonable, the benchmarks represent "wishful thinking" and defy "reason" and "common sense," according to knowledgeable experts. The latest research linking NAEP's benchmarks to international assessments reveals that the majority of students in most nations cannot clear NAEP's proficiency bar. Finally, the proposed modifications to the policy definitions for NAEP achievement levels are unresponsive to the history of criticisms of the achievement levels and to the latest research.

NAGB should rewrite the policy definitions for its standards and extend the comment period from the completely inadequate 20 days now contemplated to three months or more to give educators time to comment.

--



**Trina Evans**

**Liberty Middle School Principal  
High School Mathematics Department**

I do not believe the changes proposed by the National Assessment Governing Board to the achievement benchmarks for the National Assessment of Educational Progress will produce achievement levels that are "reasonable, useful, and informative to the public," the goal defined in the first paragraph of the proposed policy statement.

These standards and benchmarks must be expressed in terms that parents and the public can understand. To that end, I point out the following: The original achievement levels were developed in a rushed process. Those levels have produced results that have confused educators, citizens, and policymakers. Instead of being reasonable, the benchmarks represent "wishful thinking" and defy "reason" and "common sense," according to knowledgeable experts. The latest research linking NAEP's benchmarks to international assessments reveals that the majority of students in most nations cannot clear NAEP's proficiency bar. Finally, the proposed modifications to the policy definitions for NAEP achievement levels are unresponsive to the history of criticisms of the achievement levels and to the latest research.

NAGB should rewrite the policy definitions for its standards and extend the comment period from the completely inadequate 20 days now contemplated to three months or more to give educators time to comment.

--

**Troy Parton, Superintendent**  
**Munday CISD**  
**PO Box 300**  
**Munday, TX 76371**  
**Home of the Moguls**

Good morning-

I do not believe the changes proposed by the National Assessment Governing Board to the achievement benchmarks for the National Assessment of Educational Progress will produce achievement levels that are "reasonable, useful, and informative to the public," the goal defined in the first paragraph of the proposed policy statement.

These standards and benchmarks must be expressed in terms that parents and the public can understand. To that end, I point out the following: The original achievement levels were developed in a rushed process. Those levels have produced results that have confused educators, citizens, and policymakers. Instead of being reasonable, the benchmarks represent "wishful thinking" and defy "reason" and "common sense," according to knowledgeable experts.

The latest research linking NAEP's benchmarks to international assessments reveals that the majority of students in most nations cannot clear NAEP's "proficiency" bar. The proposed modifications to the policy definitions for NAEP achievement levels are unresponsive to the history of criticisms of the achievement levels and to the latest research.

NAGB should rewrite the policy definitions for its standards and extend the comment period from the completely inadequate 20 days, now contemplated, to three months or more to give educators time to comment.

Thank you,

William M. Ward, Ed.D.  
Superintendent of Schools  
Pawling Central School District  
Pawling, New York

I do not believe the changes proposed by the National Assessment Governing Board to the achievement benchmarks for the National Assessment of Educational Progress will produce achievement levels that are "reasonable, useful, and informative to the public," the goal defined in the first paragraph of the proposed policy statement.

These standards and benchmarks must be expressed in terms that parents and the public can understand. To that end, I point out the following: The original achievement levels were developed in a rushed process. Those levels have produced results that have confused educators, citizens, and policymakers. Instead of being reasonable, the benchmarks represent "wishful thinking" and defy "reason" and "common sense," according to knowledgeable experts. The latest research linking NAEP's benchmarks to international assessments reveals that the majority of students in most nations cannot clear NAEP's proficiency bar. Finally, the proposed modifications to the policy definitions for NAEP achievement levels are unresponsive to the history of criticisms of the achievement levels and to the latest research.

NAGB should rewrite the policy definitions for its standards and extend the comment period from the completely inadequate 20 days now contemplated to three months or more to give educators time to comment.

Sincerely,

Brian Hummingbird  
Principal  
Warner High School

I do not believe the changes proposed by the National Assessment Governing Board to the achievement benchmarks for the National Assessment of Educational Progress will produce achievement levels that are "reasonable, useful, and informative to the public," the goal defined in the first paragraph of the proposed policy statement.

These standards and benchmarks must be expressed in terms that parents and the public can understand. To that end, I point out the following: The original achievement levels were developed in a rushed process. Those levels have produced results that have confused educators, citizens, and policymakers. Instead of being reasonable, the benchmarks represent "wishful thinking" and defy "reason" and "common sense," according to knowledgeable experts. The latest research linking NAEP's benchmarks to international assessments reveals that the majority of students in most nations cannot clear NAEP's proficiency bar. Finally, the proposed modifications to the policy definitions for NAEP achievement levels are unresponsive to the history of criticisms of the achievement levels and to the latest research.

NAGB should rewrite the policy definitions for its standards and extend the comment period from the completely inadequate 20 days now contemplated to three months or more to give educators time to comment.

Chuck Benge  
601 E. Madison  
Fairfield, IA  
Curriculum Director

I do not believe the changes proposed by the National Assessment Governing Board to the achievement benchmarks for the National Assessment of Educational Progress will produce achievement levels that are "reasonable, useful, and informative to the public," the goal defined in the first paragraph of the proposed policy statement.

These standards and benchmarks must be expressed in terms that parents and the public can understand. To that end, I point out the following: The original achievement levels were developed in a rushed process. Those levels have produced results that have confused educators, citizens, and policymakers. Instead of being reasonable, the benchmarks represent "wishful thinking" and defy "reason" and "common sense," according to knowledgeable experts. The latest research linking NAEP's benchmarks to international assessments reveals that the majority of students in most nations cannot clear NAEP's proficiency bar. Finally, the proposed modifications to the policy definitions for NAEP achievement levels are unresponsive to the history of criticisms of the achievement levels and to the latest research.

NAGB should rewrite the policy definitions for its standards and extend the comment period from the completely inadequate 20 days now contemplated to three months or more to give educators time to comment.

Dr. Glen Fenter  
Superintendent  
Marion School District  
200 Manor Street  
Marion, AR 72364

Hello:

I do not believe the changes proposed by the National Assessment Governing Board to the achievement benchmarks for the National Assessment of Educational Progress will produce achievement levels that are "reasonable, useful, and informative to the public," the goal defined in the first paragraph of the proposed policy statement.

These standards and benchmarks must be expressed in terms that parents and the public can understand. To that end, I point out the following: The original achievement levels were developed in a rushed process. Those levels have produced results that have confused educators, citizens, and policymakers. Instead of being reasonable, the benchmarks represent "wishful thinking" and defy "reason" and "common sense," according to knowledgeable experts. The latest research linking NAEP's benchmarks to international assessments reveals that the majority of students in most nations cannot clear NAEP's proficiency bar. Finally, the proposed modifications to the policy definitions for NAEP achievement levels are unresponsive to the history of criticisms of the achievement levels and to the latest research.

NAGB should rewrite the policy definitions for its standards and extend the comment period from the completely inadequate 20 days now contemplated to three months or more to give educators time to comment.

Danielle Bolduc, Gilford Elementary Principal  
Gilford Elementary School  
76 Belknap Mountain Road  
Gilford, NH 03249

Ken Kunin, Superintendent of Schools  
South Portland Schools  
130 Wescott Road  
South Portland, ME 04016

Dear National Assessment Governing Board,

My name is Ken Kunin, Superintendent of Schools in South Portland, Maine. We are a small suburban district of 3,000 students. I am writing regarding proposed changes to the NAEP achievement benchmarks.

I do not believe the changes proposed by the National Assessment Governing Board to the achievement benchmarks for the National Assessment of Educational Progress will produce achievement levels that are "reasonable, useful, and informative to the public," the goal defined in the first paragraph of the proposed policy statement.

These standards and benchmarks must be expressed in terms that parents and the public can understand. To that end, I point out the following: The original achievement levels were developed in a rushed process. Those levels have produced results that have confused educators, citizens, and policymakers. Instead of being reasonable, the benchmarks represent "wishful thinking" and defy "reason" and "common sense," according to knowledgeable experts. The latest research linking NAEP's benchmarks to international assessments reveals that the majority of students in most nations cannot clear NAEP's proficiency bar. Finally, the proposed modifications to the policy definitions for NAEP achievement levels are unresponsive to the history of criticisms of the achievement levels and to the latest research.

NAGB should rewrite the policy definitions for its standards and extend the comment period from the completely inadequate 20 days now contemplated to three months or more to give educators time to comment.

Sincerely,  
Ken Kunin



I do not believe the changes proposed by the National Assessment Governing Board to the achievement benchmarks for the National Assessment of Educational Progress will produce achievement levels that are "reasonable, useful, and informative to the public," the goal defined in the first paragraph of the proposed policy statement. The current levels have produced results that have confused educators, policymakers, and families. The benchmarks represent "wishful thinking" and are inconsistent with children's development. The latest research linking NAEP's benchmarks to international assessments reveals that the majority of students in most nations cannot clear NAEP's proficiency bar.

I recommend that NAGB rewrite the policy definitions for its standards and extend the comment period from the 20 days now contemplated to three months or more to give educators time to provide informed and detailed comments.

Sincerely,  
*Iris C. Rotberg*

Iris C. Rotberg, Ph.D.  
Research Professor of Education Policy  
Graduate School of Education and Human Development  
The George Washington University  
2134 G Street, NW  
Washington, D.C. 20052

To whom it may concern,

As the superintendent of one of the largest school districts in Washington State, I am actively involved in legislation concerning both graduation requirements and student assessment. As such, I am very concerned with the direction that NAEP is presently taking.

I do not believe the changes proposed by the National Assessment Governing Board to the achievement benchmarks for the National Assessment of Educational Progress will produce achievement levels that are "reasonable, useful, and informative to the public," the goal defined in the first paragraph of the proposed policy statement.

These standards and benchmarks must be expressed in terms that parents and the public can understand. To that end, I point out the following: The original achievement levels were developed in a rushed process. Those levels have produced results that have confused educators, citizens, and policymakers. Instead of being reasonable, the benchmarks represent "wishful thinking" and defy "reason" and "common sense," according to knowledgeable experts. The latest research linking NAEP's benchmarks to international assessments reveals that the majority of students in most nations cannot clear NAEP's proficiency bar. Finally, the proposed modifications to the policy definitions for NAEP achievement levels are unresponsive to the history of criticisms of the achievement levels and to the latest research.

NAGB should rewrite the policy definitions for its standards and extend the comment period from the completely inadequate 20 days now contemplated to three months or more to give educators time to comment.

Serving Evergreen's Future

John Steach EdD  
Superintendent  
13501 NE 28<sup>th</sup> St.  
Vancouver WA, 98682



*If we teach today's students as we taught yesterday's, we rob them of tomorrow.* John Dewey

The Department's proposal to amend the policy definitions for NAEP's benchmarks must be revised to make them more informative to the American public. Without such changes, most Americans will continue to believe that "Proficient" means performance at grade level.

Have a wonderful day!

MaryAnn Bragg

To Whom it May Concern:

I do not believe the changes proposed by the National Assessment Governing Board to the achievement benchmarks for the National Assessment of Educational Progress will produce achievement levels that are "reasonable, useful, and informative to the public," the goal defined in the first paragraph of the proposed policy statement.

These standards and benchmarks must be expressed in terms that parents and the public can understand. To that end, I point out the following: The original achievement levels were developed in a rushed process. Those levels have produced results that have confused educators, citizens, and policymakers. Instead of being reasonable, the benchmarks represent "wishful thinking" and defy "reason" and "common sense," according to knowledgeable experts. The latest research linking NAEP's benchmarks to international assessments reveals that the majority of students in most nations cannot clear NAEP's proficiency bar. Finally, the proposed modifications to the policy definitions for NAEP achievement levels are unresponsive to the history of criticisms of the achievement levels and to the latest research. I agree with the National Superintendent's Roundtable summary <http://files.constantcontact.com/d6ed868c001/b63f4624-49ad-40a1-afd8-660913afe0cd.pdf> and ask that feedback be considered.

NAGB should rewrite the policy definitions for its standards and extend the comment period from the completely inadequate 20 days now contemplated to three months or more to give educators time to comment.

Respectfully,

Steven T. Webb, Ed.D.  
Superintendent  
Vancouver Public Schools  
2901 Falk Road  
Vancouver, WA 98661

**Connect with me on Twitter @SuptVPS**

Dear Members of the National Assessment Governing Board,

I know that the national Assessment Governing Board is considering changes to the achievement benchmarks. I do not believe that the proposed changes for the National Assessment of Educational Progress will produce achievement levels that are "reasonable, useful, and informative to the public," which is the goal defined in the first paragraph of the proposed policy statement.

These standards and benchmarks must be expressed in terms that parents and the public can understand. To that end, I point out the following: The original achievement levels were developed in a rushed process. Those levels have produced results that have confused educators, citizens, and policymakers. Instead of being reasonable, the benchmarks represent wishful thinking and defy common sense, according to knowledgeable experts. The latest research linking NAEP's benchmarks to international assessments reveals that the majority of students in most nations cannot clear NAEP's proficiency bar. Finally, the proposed modifications to the policy definitions for NAEP achievement levels are unresponsive to the history of criticisms of the achievement levels and to the latest research.

NAGB should rewrite the policy definitions for its standards and extend the comment period from the completely inadequate 20 days now contemplated to three months or more to give educators time to comment.

Sincerely,

Ms. Fern Lox  
Assistant Principal  
Chancellor Livingston Elementary School  
Rhinebeck Central School District

Please revise the Department's proposal to amend the policy definitions for NAEP's benchmarks. These benchmarks should be changed to be more clear for parents and schools. Without such changes, most Americans will continue to believe that "Proficient" means performance at grade level.

Rob McEntarffer

To: U.S. Department of Education  
Re: Jack McKay, Executive Director  
Date: September 28, 2018

The Horace Mann League was founded a century ago to promote the benefits of public education for all in a democratic republic. In support of our mission, the League was a major contributor to the development of a 2018 report, *How High the Bar? How Would Other Nations Perform if their Students were judged by Common Core or NAEP Benchmarks?* In addition to the original research produced by that report, it contained a detailed literature review of the many expert criticisms of NAEP's performance benchmarks, dating from their introduction in 1990 to 2018.

Against that background, I want to state clearly that the League's board does not believe the changes to the NAEP achievement benchmarks proposed by the National Assessment Governing Board will produce achievement levels that are "reasonable, useful, and informative to the public," the goal defined in the first paragraph of the proposed policy statement.

These standards and benchmarks must be expressed in terms that parents and the public can understand. To that end, our board wants me to point out the following:

1. The original achievement levels were developed in a rushed process. Those levels have produced results that have confused educators, citizens, and policymakers. Instead of being reasonable, the benchmarks represent "wishful thinking" and defy "reason" and "common sense," according to knowledgeable experts.
2. The latest research linking NAEP's benchmarks to international assessments reveals that the majority of students in most nations cannot clear NAEP's proficiency bar.
3. The proposed modifications to the policy definitions for NAEP achievement levels are unresponsive to the history of criticisms of the achievement levels and to the latest research.
4. Benchmarks or proficiency standards appears to be arbitrary and are set to have winner and losers. These standards produce a zero sum game which seems to be contrary to the intent and mission of America's public school. Using benchmarks and standards is folly when comparing America's diverse population, limited social services, and associated degrees of poverty.

Sincerely,

Jack McKay

Jack McKay, Ed.D., Executive Director  
The Horace Mann League of the USA

Dear NAEP Achievement-Level-Setting Program,

As a former member of the National Assessment Governing Board, I am keenly interested in the improvement and credibility of the NAEP program.

I am writing to express my strong support for a complete rethinking of the NAEP "achievement levels." I urge the National Assessment Governing Board to abandon the achievement levels, because they are technically unsound and utterly confusing to the public and the media. They serve no purpose other than to mislead the public about the condition of American education.

The achievement levels were adopted in 1992 for political reasons: to make the schools look bad, to convey simplistically to the media and the public that "our schools are failing."

The public has never understood the levels. The media and prominent public figures regularly report that any proportion of students who score below "NAEP proficient" is failing, which is absurd. The two Common Core-aligned tests (PARCC and SBAC) adopted "NAEP Proficient" as their passing marks, and the majority of students in every state that use these tests have allegedly "failed," because the passing mark is out of reach, as it will always be.

The National Center for Educational Statistics (NCES) has stated clearly that "Proficient is not synonymous with grade level performance." Nonetheless, public figures like Michelle Rhee (who was chancellor of the DC public schools) and Campbell Brown (founder of the website "The 74") have publicly claimed that the proficiency standard of NAEP is the bar that ALL students should attain. They have publicly stated that American public education is a failure because there are many students who have not reached NAEP proficient.

In reality, there is only one state in the nation--Massachusetts--where as much as 50% of students have attained NAEP Proficient. No state has reached 100% proficient, and no state ever will.

When I served on NAGB for seven years, the board understood very well that proficient was a high bar, not a pass-fail mark. No member of the board or the staff expected that some day all students would attain "NAEP Proficient." Yet critics and newspaper consistently use NAEP proficient as an indicator that "all students" should one day reach. This misperception has been magnified by the No Child Left Behind Act, which declared in law that all students should be "proficient" by the year 2014.

Schools have been closed, and teachers and principals have been fired and lost their careers and their reputations because their students were not on track to reach an impossible goal.

As you well know, panels of technical experts over the years have warned that the achievement levels were not technically sound, and that in fact, they are "fatally flawed." They continue to be "fatally flawed." They cannot be fixed because they are in fact arbitrary and capricious. The standards and the process for setting them have been criticized by the General Accounting Office, the National Academy of Sciences, and expert psychometricians.



Whether using the Angoff Method or the Bookmarking Method or any other method, there is no way to set achievement levels that are sound, valid, reliable, and reasonable. If the public knew that the standards are set by laypersons using their "best judgment," they would understand that the standards are arbitrary. It is time to admit that the standard-setting method lacks any scientific validity.

When they were instituted in 1992, their alleged purpose was to make NAEP results comprehensible to the general public. They have had the opposite effect. They have utterly confused the public and presented a false picture of the condition and progress of American education.

As you know, when Congress approved the achievement levels in 1992, they were considered experimental. They have never been approved by Congress, because of the many critiques of their validity by respected authorities.

My strong recommendation is that the board acknowledge the fatally flawed nature of achievement levels. They should be abolished as a failed experiment.

NAGB should use scale scores as the only valid means of conveying accurate information about the results of NAEP assessments.

Thank you for your consideration,

Diane Ravitch  
NAGB, 1997-2004  
Ph.D.  
New York University

Dear NAEP Policy Board:

I strongly support the views of Diane Ravitch expressed in her recent letter to you, advocating that "... the board acknowledge the fatally flawed nature of achievement levels. They should be abolished as a failed experiment.

NAGB should use scale scores as the only valid means of conveying accurate information about the results of NAEP assessments.

Sincerely,

Dan Drmacich, Coordinator; Rochester Coalition for Public Education

**TUTTLE PUBLIC SCHOOLS**  
515 E. MAIN STREET P.O. BOX 780  
TUTTLE, OK 73089-0780  
PHONE 405.381.2605 FAX 405.381.4008

**BOBBY WAITMAN, SUPERINTENDENT**

SUSAN DOMBEK, Elem. School Principal TWYLLA HESTER, High School Principal  
CHERYL WILLIAMS, Intermediate School Principal SCOTT MOORE, Athletic Director  
TARA NORVELL, Middle School Principal JEFF DEWITT, Director of Federal Programs  
SARA GARRETT, Treasurer

---

September 19, 2018

To whom it may concern,

**I do not believe the changes proposed by the National Assessment Governing Board to the achievement benchmarks for the National Assessment of Educational Progress will produce achievement levels that are "reasonable, useful, and informative to the public," the goal defined in the first paragraph of the proposed policy statement.**

**These standards and benchmarks must be expressed in terms that parents and the public can understand. To that end, I point out the following: The original achievement levels were developed in a rushed process. Those levels have produced results that have confused educators, citizens, and policymakers. Instead of being reasonable, the benchmarks represent "wishful thinking" and defy "reason" and "common sense," according to knowledgeable experts. The latest research linking NAEP's benchmarks to international assessments reveals that the majority of students in most nations cannot clear NAEP's proficiency bar. Finally, the proposed modifications to the policy definitions for NAEP achievement levels are unresponsive to the history of criticisms of the achievement levels and to the latest research.**

**NAGB should rewrite the policy definitions for its standards and extend the comment period from the completely inadequate 20 days now contemplated to three months or more to give educators time to comment.**

Respectfully,



**Bobby Waitman**

I do not believe the changes proposed by the National Assessment Governing Board to the achievement benchmarks for the National Assessment of Educational Progress will produce achievement levels that are "reasonable, useful, and informative to the public," the goal defined in the first paragraph of the proposed policy statement.

These standards and benchmarks must be expressed in terms that parents and the public can understand. To that end, I point out the following: The original achievement levels were developed in a rushed process. Those levels have produced results that have confused educators, citizens, and policymakers. Instead of being reasonable, the benchmarks represent "wishful thinking" and defy "reason" and "common sense," according to knowledgeable experts. The latest research linking NAEP's benchmarks to international assessments reveals that the majority of students in most nations cannot clear NAEP's proficiency bar. Finally, the proposed modifications to the policy definitions for NAEP achievement levels are unresponsive to the history of criticisms of the achievement levels and to the latest research.

NAGB should rewrite the policy definitions for its standards and extend the comment period from the completely inadequate 20 days now contemplated to three months or more to give educators time to comment.

Thanks for your consideration.

--

Carl Bruner, Ed.D.  
Superintendent  
Mount Vernon School District No. 320  
[www.mountvernonschools.org](http://www.mountvernonschools.org)

Dear National Assessment Governing Board:

I do not believe the changes proposed by the National Assessment Governing Board to the achievement benchmarks for the National Assessment of Educational Progress will produce achievement levels that are "reasonable, useful, and informative to the public," the goal defined in the first paragraph of the proposed policy statement.

These standards and benchmarks must be expressed in terms that parents and the public can understand. To that end, I point out the following: The original achievement levels were developed in a rushed process. Those levels have produced results that have confused educators, citizens, and policymakers. Instead of being reasonable, the benchmarks represent "wishful thinking" and defy "reason" and "common sense," according to knowledgeable experts. The latest research linking NAEP's benchmarks to international assessments reveals that the majority of students in most nations cannot clear NAEP's proficiency bar. Finally, the proposed modifications to the policy definitions for NAEP achievement levels are unresponsive to the history of criticisms of the achievement levels and to the latest research.

NAGB should rewrite the policy definitions for its standards and extend the comment period from the completely inadequate 20 days now contemplated to three months or more to give educators time to comment.

Sincerely,

<b>Knoxville Panthers</b>	<b>Cassi Pearson</b> Superintendent Knoxville Community School District 309 W Main Street, Knoxville, IA 50138 o: (641) 842-6551 m: (641) 751-5891 e: <a href="mailto:cassi.pearson@kcsd.k12.ia.us">cassi.pearson@kcsd.k12.ia.us</a>
---------------------------	---



**JENNINGS PUBLIC SCHOOL**

475 North Oak Street – Jennings, OK 74038 – (918) 757-2536 – [www.jennings.k12.ok.us](http://www.jennings.k12.ok.us)  
Kurtis Lasater, President      Nathan Staley, Vice-President      Shawna Robinson, Clerk



Derrick Meador, Superintendent

Jay Reeves, Principal

Thursday, September 20, 2018

To Whom It May Concern:

I do not believe the changes proposed by the National Assessment Governing Board to the achievement benchmarks for the National Assessment of Educational Progress will produce achievement levels that are "reasonable, useful, and informative to the public," the goal defined in the first paragraph of the proposed policy statement.

These standards and benchmarks must be expressed in terms that parents and the public can understand. To that end, I point out the following: The original achievement levels were developed in a rushed process. Those levels have produced results that have confused educators, citizens, and policymakers. Instead of being reasonable, the benchmarks represent "wishful thinking" and defy "reason" and "common sense," according to knowledgeable experts. The latest research linking NAEP's benchmarks to international assessments reveals that the majority of students in most nations cannot clear NAEP's proficiency bar. Finally, the proposed modifications to the policy definitions for NAEP achievement levels are unresponsive to the history of criticisms of the achievement levels and to the latest research.

NAGB should rewrite the policy definitions for its standards and extend the comment period from the completely inadequate 20 days now contemplated to three months or more to give educators time to comment.

Sincerely,

A handwritten signature in black ink, appearing to read "Derrick Meador".

Derrick Meador, Superintendent  
Jennings Public Schools  
475 North Oak Street  
Jennings, OK 74038  
918-757-2536

The Department's proposal to amend the policy definitions for NAEP's benchmarks must be revised to make them more informative to the American public. Without such changes, most Americans will continue to believe that "Proficient" means performance at grade level. I support the findings of "How High the Bar," and I implore you to study that report further.

Martha Bruckner, Ph. D.

Executive Director, MOEC Collective Impact Initiative

Metropolitan Omaha Education Consortium

UNO: Barbara Weitz Community Engagement Center, 223 B

6001 Dodge Street, Omaha, NE 68182

I do not believe the changes proposed by the National Assessment Governing Board to the achievement benchmarks for the National Assessment of Educational Progress will produce achievement levels that are "reasonable, useful, and informative to the public," the goal defined in the first paragraph of the proposed policy statement.

These standards and benchmarks must be expressed in terms that parents and the public can understand. To that end, I point out the following: The original achievement levels were developed in a rushed process. Those levels have produced results that have confused educators, citizens, and policymakers. Instead of being reasonable, the benchmarks represent "wishful thinking" and defy "reason" and "common sense," according to knowledgeable experts. The latest research linking NAEP's benchmarks to international assessments reveals that the majority of students in most nations cannot clear NAEP's proficiency bar. Finally, the proposed modifications to the policy definitions for NAEP achievement levels are unresponsive to the history of criticisms of the achievement levels and to the latest research.

NAGB should rewrite the policy definitions for its standards and extend the comment period from the completely inadequate 20 days now contemplated to three months or more to give educators time to comment.

Jeff Morrison