

National Assessment Governing Board

Committee on Standards, Design and Methodology

Friday, May 18, 2018

10:30 am – 1:00 pm

AGENDA

10:30 am – 10:35 am Welcome and Review of Agenda

Andrew Ho, COSDAM Chair

10:35 am – 11:50 am Discussion of Revised Draft Policy on Achievement Levels Attachment A
Setting (SV #5)

Andrew Ho
Sharyn Rosenberg, Assistant Director for Psychometrics

11:50 am – 12:00 pm Break

12:00 – 1:00 pm Joint Session with Reporting & Dissemination Committee: Attachment B
Communication and Interpretation of Achievement Levels
(SV #3)

Andrew Ho
Rebecca Gagnon, R&D Chair

Information Item

Update on Implementing the Strategic Vision (SV#2-10) Attachment C

Discussion of Revised Draft Policy on Achievement Level Setting (SV #5)

Background

Over the past year, COSDAM members discussed the need to revise the 1995 Governing Board policy 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. 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 is also a need to revisit the policy more generally to ensure that it reflects current best practices in standard setting. COSDAM members have acknowledged the need to seek input from multiple stakeholders throughout the process of revising the policy. To get an initial sense of the potential scope of recommended revisions to the policy, Assistant Director for Psychometrics Sharyn Rosenberg conducted informal conversations with several standard setting experts in spring 2017. Feedback from those conversations was shared with COSDAM in May 2017 and informed the additional work that has been performed since then.

As part of its Technical Support contract, the Governing Board requested that the Human Resources Research Organization (HumRRO) undertake several activities to inform the revision of the Board policy on setting achievement levels for NAEP. These activities were discussed during the March 2018 COSDAM meeting, and the Committee report is excerpted below:

Excerpt of March 2018 COSDAM Report

Best Practices in Achievement Levels Setting

Thanos Patelis of the Human Resources Research Organization (HumRRO), the project director for the Technical Support contract, described several recommendations from a 2-day, in-person meeting of experts in achievement levels setting. Mr. Patelis highlighted the following suggestions: 1) the policy should be a statement of high level guidance, and procedural details should be relegated to a new "processes and procedures" manual; 2) NAEP should develop publicly accessible interpretative guides for using achievement level results; 3) the NAEP policy definitions for Basic, Proficient, and Advanced should not be changed; and 4) the Board should provide clarification on the meaning of "general public" standard setting panelists and consider reducing their number. Mr. Patelis noted that this is not an exhaustive list of feedback, and the

report from the expert panel meeting, available in the COSDAM materials, contains several additional recommendations for the committee's consideration.

Next, Karla Egan of EdMetric presented her review of the literature on achievement level descriptions (ALDs). She suggested that the Board consider: 1) developing multiple types of ALDs for different purposes, including item writing, standard setting, and reporting; 2) using "can" statements for reporting instead of the "should" statements that are in the current NAEP ALDs; and 3) writing descriptors for the lowest category (below Basic) which does not currently have a policy definition.

Finally, Art Thacker of HumRRO provided a brief description of his technical memo about developing a validity argument for the NAEP achievement levels, with the following suggested steps: 1) make intended inferences explicit; 2) investigate how stakeholders typically use the achievement levels; 3) create an interpretative guide; 4) craft claims necessary to support expected inferences; and 5) organize evidence for each claim into a validity argument.

Identifying Revision Goals for the Board Policy on Achievement Levels Setting

During and following the short presentations, COSDAM members engaged in a rich discussion about various aspects of the policy. There was general agreement with the expert panel that there is too much detail in the current policy. It was suggested that the Board balance stability and flexibility in the policy revision. That is, aspects of the standard setting that are most important, such as representation of panelists, should not vary too much across different standard settings. On the other hand, it is possible to make a statement to that effect rather than specifying exact percentages or numbers of different types of panelists. As the committee moves forward with the policy revision, it will be important to determine what should be codified for stability while allowing for flexibility to incorporate new developments in standard setting.

In terms of the standard setting participants, there was extensive discussion about the category of "general public" panelists. The current policy specifies, "one-third will represent the public, non-educator sector, for example, scholars, employers, parents, and professionals in occupations related to the content area" (p. 6). In practice, subject-matter expertise has been a requirement and these panelists have had some professional experience in the content area. COSDAM members discussed whether there should be general public panelists who do not possess subject matter expertise, and whether the number of general public panelists should be reduced. Some raised concerns that non-educators may have trouble making informed judgments about the knowledge and skills necessary for performance in a content area domain. Others noted that the general public was already represented on the Board itself, which is ultimately responsible for setting achievement levels. However, members acknowledged that it could be problematic to reduce perceived or actual general public participation in the development of the Nation's Report Card. Sharyn Rosenberg, the Assistant Director for Psychometrics, will consult documentation from previous NAEP standard settings to provide additional information about how the

background of general public panelists may have affected their participation in the process. COSDAM members generally agreed that the policy should better clarify what is meant by “general public,” and that if the current practice is maintained, there may not be a compelling reason to reduce the number of panelists in this category.

COSDAM members discussed whether or not the Board should consider developing a policy definition and content ALDs for performance below the Basic achievement level. In the current policy, only Basic, Proficient, and Advanced are considered achievement levels. The percentage of students whose performance is below the Basic level is reported but not described. COSDAM members did not see a compelling reason to develop a description for below Basic. They noted that it is difficult to develop an informative description when the bottom of the category starts at zero; any statements would need to be in terms of what students sometimes or may be able to do. The NAEP item maps do include items below Basic and therefore provide some information about performance in this range.

Some COSDAM members raised questions about whether it is appropriate for the NAEP ALDs to be written as what students “should” do rather than what they “can” do. Prior to conducting the standard setting, “should” statements indicate the performance that is expected at each level; but after the cut scores are established, “should” statements indicate that not every student has demonstrated every skill in a given category. It may be more informative to develop separate reporting ALDs that are written in terms of “can” statements, but this would need to be based on data produced following a standard setting. That is, panels of content experts could use an item mapping approach to summarize the knowledge and skills that are typical of students in each achievement level and produce reporting ALDs for NAEP.

COSDAM members agreed that it would be helpful to develop interpretative guides and noted that strong communication and clear reporting are keys to guarding against unintended inferences. There was a suggestion to provide both examples and non-examples; that is, both appropriate and inappropriate examples of interpreting NAEP achievement levels. It is important to explain what evidence there is to justify a given use, rather than only focusing on score meaning and interpretation. A research study that is currently underway by HumRRO as part of the Technical Support contract should help to inform this effort. It would be helpful to engage in discussion with the Reporting and Dissemination (R&D) Committee on some of these issues.

Finally, COSDAM members agreed that the policy should not require gathering public comment on the cut scores, since it is not feasible to release those data prior to the official release of the Nation’s Report Card. Public comment on the ALDs and the Design Document should be sufficient throughout the standard setting process. Multiple stakeholders are engaged in the process through the standard setting panels and by representation on the Governing Board, which has the ultimate responsibility for establishing NAEP achievement levels.

May 2018 COSDAM discussion

Sharyn Rosenberg worked with COSDAM Chair Andrew Ho and COSDAM Vice Chair Joe Willhoft to produce a draft revised policy for COSDAM discussion (attached). Compared to the current policy on Developing Student Performance Levels for NAEP, the attached version reflects:

- Reorganization of principles, streamlining of language, minimization of redundancies
- Removal of details on implementation directed to staff and contractors
- 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”)
- The possibility of using 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*
- 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
- 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

During the May 2018 Committee meeting, COSDAM members will discuss the draft policy and provide feedback on the revision.

Additional planned next steps are described below:

June 2018: COSDAM teleconference to discuss revised draft policy
 August 2018: Revised policy statement for full Board discussion
 September 2018: Seek external feedback and public comment
 October 2018: Full Board (optional) call to discuss revised draft policy
 November 2018: Board action on revised policy statement

**Adopted: TBD**

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). Achievement levels consist of general policy definitions for the *Basic*, *Proficient*, and *Advanced* levels, specific achievement level descriptions (ALDs) for each subject and grade, 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, appropriate, 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 and guidelines of the Governing Board Student Achievement Levels 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 appropriate 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.

Under provisions of the National Assessment of Educational Progress Authorization Act of 2002 (P.L. 107-279), Congress authorized the Governing Board to continue its mandate for developing appropriate student achievement levels for NAEP, consistent with relevant widely accepted professional assessment standards, based on the

appropriate level of subject matter knowledge for grade levels assessed, and using a national consensus approach.

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.

To develop student achievement levels for Board adoption, the Governing Board engages multiple stakeholders throughout the 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 complies with the National Assessment of Educational Progress Authorization Act of 2002 (P.L. 107-279) and 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. A procedures manual shall provide additional details about how this policy is 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.

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 (4th ed.). (2006). R.L. Brennan (Ed.). Westport, CT: Praeger.

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 consist of general policy definitions for the Basic, Proficient, and Advanced levels, specific achievement level descriptions (ALDs) for each subject and grade, cut scores that demarcate adjacent levels, and exemplar items or tasks that illustrate performance at each level.

Guidelines

- a) The following policy definitions will be applied to all subject areas and grades in which achievement levels are set. It is the Board's view that the level of performance referred to in the policy definitions is what students *should know and be able to do*, not simply the current academic achievement of students or that which today's U.S. schools expect.

Proficient. 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.

Basic. This level denotes partial mastery of prerequisite knowledge and skills that are fundamental for Proficient work at each grade.

Advanced. This level signifies superior performance beyond Proficient.

- b) Content achievement level descriptions (ALDs) translate the general 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 the expected knowledge, skills, or abilities of students performing at a particular 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 subject and grade. During reporting, content ALDs communicate the specific knowledge and skills represented by *Basic*, *Proficient*, and *Advanced* for a given subject and grade.
- c) 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.
- d) Exemplar items and 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 can do.

Principle 2: Development of Achievement Level Recommendations

The Governing Board shall develop appropriate student achievement levels for NAEP, consistent with relevant widely accepted professional assessment standards, based on the appropriate level of subject matter knowledge for grade levels assessed, and using a national consensus approach.

Guidelines

- a) A Design Document shall be developed at the beginning of the achievement level setting process, to describe in detail all planned materials, procedures, and analyses 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) will 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 possibly 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 will be articulated in terms of what students *should know and be able to do*. There will be no content ALDs developed for performance below the *Basic* level.
- c) An achievement-level setting panel of subject matter experts shall be convened to recommend achievement level cut scores and exemplars.
 - i. 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. 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.
 - ii. This panel shall include both educators and non-educators who are considered outstanding in their field. The educator group shall include both teachers and other educators (e.g., curriculum directors, academic coaches, principals). Teachers shall comprise the majority of the panel, with non-teacher educators accounting for no more than half the number of teachers. The remaining panelists shall be non-educators who represent the perspectives of additional stakeholders, including parents, researchers, employers, and other members of the general public.
 - iii. The size of the panels should be responsive to what current research demonstrates is best practice and operationally feasible, but should be

large enough to allow for split panels. Most NAEP achievement level settings have included approximately 20-30 panelists per grade, divided into two comparable groups with a subset of shared items.

- iv. The size and specific composition of the panels may be adjusted within these general guidelines if professional standards in the field evolve.
- 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. Training must include: 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 *Basic*, *Proficient*, and *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 may differ depending upon the specific assessment. The method must have a solid research base and be appropriate for the content area, item types, number of items, scoring rubrics, and mode, 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 appropriate and 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 an opportunity to test 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 Guideline a apply to panelists of both meetings.

- i) The Governing Board or its contractor shall convene a Technical Advisory Committee on Standard Setting (TACSS) 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 the National Center for Education Statistics (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 will be made available to the Board 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.
- k) The exemplars chosen from the pool of released items for the current NAEP assessment shall reflect performance in the *Basic*, *Proficient*, and *Advanced* regions of the scale. The use of exemplars is intended to help the public better understand what students who are in each achievement levels actually know and are able to do 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 (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 appropriate validity evidence for the intended uses and interpretations, are reasonable, and are informative to the public.

Guidelines

- a) Professional testing standards define validity as the degree to which evidence supports intended interpretations and uses of test scores. The validity of

achievement level results is a property of their intended interpretations and uses. Standard setting is necessarily judgmental. There are no “true” or “correct” cut scores. Instead, there is a legitimizing process that results in an authoritative consensus. In making a policy judgment to set achievement levels, the Board will examine and consider available evidence about due process and the reasonableness of results, in order to support intended uses and interpretations.

- 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 subjects and grades. 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) To facilitate valid uses of ALDs for 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 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. These revised content ALDs shall be written in terms of what students *do* know and empirically *can* do rather than what they *should* know and *should* be able to do.
- d) The Board will examine and consider all evidence related to reliability and validity of the achievement level setting activities. These data shall include but need 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.
- e) In describing student performance using the achievement levels, terms such as students performing at the *Basic* level or students performing at the *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.
- f) In reporting the results of NAEP, the three achievement levels of *Basic*, *Proficient*, and *Advanced* refer to the three regions of the NAEP scale at and above each respective cut score. The remaining region that falls below the *Basic* cut score will be identified as “below *Basic*” when a descriptor is necessary.
- g) 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. *Proficient* on NAEP may convey a different meaning from other uses of the term

“proficient” in common terminology or in reference to other assessments.

- h) When interpreting student performance using achievement levels, it is important to discourage incorrect comparisons and interpretations. For example, a *Proficient* cut score of 235 in reading should not be interpreted to have the same meaning as a *Proficient* cut score of 235 in U.S. history.
- i) 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 student performance.

Guidelines

- 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 past and recent administrations of the assessment. In its review, COSDAM (in consultation with ADC) shall solicit input from technical and subject matter experts to determine whether changes to the content ALDs and/or cut scores are warranted, 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; 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 automatically, except in circumstances where scale score trends are maintained. In this latter instance, COSDAM will determine how to revise the ALDs and review the cut scores to ensure that they remain appropriate and meaningful.
- c) If there are major updates to a NAEP framework, the ALDs will 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 may decide to 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 redone.

Principle 5: Stakeholder Input

The process of developing student achievement levels is a widely inclusive activity. There are many opportunities to engage multiple stakeholders throughout the achievement level setting process.

Guidelines

- a) The content achievement level descriptions are developed through the framework development process, using a panel that represents all major constituents in the various NAEP audiences, as listed in the introduction above. If it is necessary to revise the ALDs for use in achievement level setting and/or reporting, a similar group of content experts will be convened, and public comment will be sought on the resulting achievement level descriptions.
- b) The process of seeking nominations for the achievement level setting panels shall include extensive outreach to multiple constituencies, such as: state and local educators; curriculum specialists; business representatives; and professional associations in a given content area.
- c) As noted in Principle 2, Guideline a, 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.
- d) As noted in Principle 2, Guideline 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.
- e) As noted in Principle 2, Guideline i, 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.
- f) Ongoing input and coordination with staff and contractors from the NCES will 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.
- g) The Governing Board may ask its standing groups representing various constituencies to provide input on the achievement level setting process.

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 Governing Board-adopted achievement level descriptions, cut scores, and exemplars comply with all principles and guidelines of the Governing Board Student Achievement Levels policy.

Guidelines

- 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 will 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.
- 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) A COSDAM member may elect to attend any achievement level setting panel meeting(s) as an observer at the discretion of the COSDAM Chair.
- f) 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 assessments.
- g) 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.
- h) Consistent with Principle 4 above, COSDAM shall periodically review existing achievement levels to determine whether new achievement level descriptions and/or cut scores are needed to continue valid and reliable measurement of student performance, while recognizing the value of stability and the value that is accrued by using achievement levels over time.

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.

Communication and Interpretation of Achievement Levels

At the May 2018 Governing Board meeting in Montgomery, Alabama, COSDAM will hold a joint meeting with the Reporting and Dissemination Committee (R&D). The purpose of this joint meeting is to discuss the intersection of the two committees' work on achievement levels for the National Assessment of Educational Progress (NAEP).

Background

From 2014 to 2016, the National Academies of Sciences, Engineering, and Medicine evaluated the NAEP achievement levels in mathematics and reading, which are the responsibility of the Governing Board. In their evaluation, the National Academies noted eight common uses of NAEP achievement levels, specifically:

- Trends or comparisons of successive cohorts, e.g., the percentage of students at or above Proficient in reading has increased over time;
- Comparison to a state assessment;
- Point-in-time comparisons across states, districts, or population groups, e.g., more students in state A who are at or above Proficient in reading compared to state B;
- Rank ordering states or districts;
- Comparison across population groups to examine performance gaps;
- Comparison across subject areas, e.g., more students perform at or above Proficient on mathematics than in reading;
- Comparison of before and after an action or policy implementation; and
- Relationships among achievement results and contextual data.

The evaluation recognized the usefulness and value of the achievement levels but made several important recommendations, most of which focus on the work of COSDAM as well as two that also address the work of the R&D Committee:

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

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

Since the release of these recommendations in November 2016, Governing Board staff and COSDAM members have started working to fulfill these recommendations. The draft revision of the Board policy on developing student achievement levels (scheduled for full Board discussion in August 2018 and action in November 2018) establishes an

“interpretative guide [which] shall accompany NAEP reports, including specific examples of appropriate and inappropriate interpretations and uses of the results” (Principle 3i).

COSDAM will develop the content of this interpretative guide, but the responsibility to include and disseminate such a guide in reporting will fall to the R&D Committee and NCES. This joint meeting between R&D and COSDAM will focus, in part, on the development and use of an interpretative guide to facilitate the understanding of achievement levels.

As part of the Governing Board’s contract on Technical Support in Psychometrics, Assessment Development, and Preparedness for Postsecondary Endeavors, the Human Resources Research Organization (HumRRO) is conducting research to understand the various actual uses of NAEP data, including achievement levels. Information taken from published documents and interviews will guide development of a Board policy statement on appropriate uses of NAEP and development of an interpretative guide.

Within this task, HumRRO started work by providing advice on building a validity argument for the NAEP achievement levels. An excerpt of that memo, which focuses on how various audiences use NAEP achievement levels, is attached.

Finally, with the April release of the 2017 *Nation’s Report Card* in Mathematics and Reading, issues in understanding achievement levels re-emerged. During pre-release briefings with media, a reporter asked how the *Proficient* level on NAEP differs from what proficient means on a given state assessment. Material presented at the same time as the data release explicated what achievement levels mean (see attached) in hope of avoiding confusion, but misuses still appeared. During this joint meeting, R&D will seek a more concise and more comprehensible way of explaining the achievement levels and of distinguishing them from other uses of the term proficient.

Guiding Questions

With this background, the members of both committees will address the following questions in the course of the hour-long discussion:

- Does the revised achievement levels policy (Principle 3 in particular) capture the components critical to communicating the achievement levels effectively?

- How and to whom should an interpretative guide to the inappropriate and appropriate uses of NAEP achievement levels be presented and disseminated? Knowing the intended outcome and audience will inform the content development.
- How should the Governing Board highlight exemplary uses of NAEP achievement levels and address misuses of NAEP achievement levels?
- How can the Governing Board clearly and concisely explain achievement levels accurately? How can these explanations most effectively avoid misinterpretation? How can these explanations cleanly distinguish what NAEP means from what states mean by terms such as *Basic* and *Proficient* and *Advanced*?

Excerpt of Technical Memo: Uses of NAEP Achievement Levels¹

Arthur A. Thacker, Ph.D.
Tonya Longabach, Ph.D.
Human Resources Research Organization (HumRRO)

Introduction

The National Assessment Governing Board's (Governing Board) recent Strategic Vision² identifies policymakers, educators, researchers and business leaders, the media, and the general public as stakeholders who are expected to use National Assessment of Educational Progress (NAEP) results. The Strategic Vision is not so specific as to describe how each group is expected to use NAEP results, but it does indicate that they should be informed "about what America's students know and can do in various subject areas and compare achievement data over time and among student demographic groups." The Strategic Vision also states that NAEP should "inform education policy and practice."

The Governing Board is working towards developing a statement of intended and appropriate uses for both scale scores and achievement levels. HumRRO is currently conducting a research study to determine how various audiences have used and interpreted NAEP results. However, the current lack of specificity in the inferences each group might make represents a substantial challenge for validation. We will seek out inferences the identified groups have actually made from NAEP results.

Note that this memorandum is not comprehensive. Our goal is to provide guidance on how NAEP achievement levels might be validated for making specific inferences. The number of potential inferences that might be made and the amount of documentation available to potentially support those inferences is well beyond the scope of this memorandum. The examples we include in this memorandum, while important, do not necessarily represent the most important validation issues or interpretations of NAEP achievement levels rather, they were chosen to be illustrative of the range of inferences. Where possible, we summarize the literature related to common claims, but these summaries do not represent an exhaustive literature review.

Inferences from Various Stakeholders

Policymakers

For purposes of this memorandum, we define policymakers as national and state legislators, board and committee members at the federal, state, and district level who make policy and/or recommendations for policy in education, and other individuals who make or influence educational policy (e.g., congressional staffers, lobbyists). These individuals are responsible for policy across educational institutions and have considerable power to influence curriculum,

¹ This is an excerpt of Technical Memorandum #1 (HumRRO Report 2017 NO. 089), developed under contract #ED-NAG-17-C-0002, Technical Support in Psychometrics, Assessment Development, and Preparedness for Postsecondary Endeavors.

² See <https://www.nagb.gov/content/nagb/assets/documents/newsroom/press-releases/2016/nagb-strategic-vision.pdf>.

instruction, assessment, teacher professional development, and other factors. They must address information regarding what students know and can do, and whether students are prepared for their next experiences, as policymakers strive to improve the state of American education.

Policymakers use NAEP scores and performance level descriptors for the following purposes:

- making comparisons to other districts, states, and the nation;
- making within-state subgroup comparisons;
- analyzing state achievement trends;
- suggesting changes to state assessments and to aid in defining levels of student performance;
- validating state standards and building the case for educational reform and change in their states (Zenisky, Hambleton, & Sireci, 2009); and
- building arguments for new or amended legislation and for requesting funding related to education (Edley & Koenig, 2017).

NAEP is well-structured in many ways for policymakers, who tend to be most interested in aggregate reports of student performance rather than individual student scores. NAEP is designed to generate comparable results across states and demographic groups. NAEP maintains a scale across years and allows for tracking of trends. However, when policymakers use NAEP to justify changes to state assessments or state performance definitions, build a case for educational reforms, or for requesting funding, they must support those uses based on their own understanding of NAEP and their judgements about NAEP's suitability for those purposes.

Educators

For purposes of this memorandum, we define educators as those persons who work most directly with students. They are responsible for instruction and for implementing curriculum and assessments. Educators include teachers, teachers' support personnel, content area specialists, academic coaches, etc. We also include school principals in this category, although there is some overlap with policymakers, since principals greatly influence policy within their particular schools.

Because NAEP does not produce results for individual students or at the school level, score interpretations are of limited use for educators. The achievement level descriptions (ALDs) and the frameworks, however, may provide considerable useful information. The frameworks indicate the content that students are expected to know in specific subjects at specific grades. The ALDs indicate how students will be categorized based on the level of their knowledge and skill related to that content. The ALDs help educators better understand how student performance is differentiated.

Educators receive their information about NAEP from various sources, including three main NAEP websites. They receive much of their information from their state education agency's website and the media. NCES also supports a NAEP state coordinator in each state who serves as a liaison between the state department of education and the NAEP programs. They are available to assist in the interpretation of NAEP results. We reviewed a sample of state websites as part of preparing this memorandum. We selected websites to reflect either high or low performance on NAEP to highlight any qualitative differences in the information presented to educators.

The three lowest performing states on NAEP 4th and 8th grade reading and mathematics and the three highest performing states based on 2015 results³ are shown in Table 1. The state Department of Education (DOE) websites and state education agency websites were searched to determine whether and how the states use NAEP data. We specifically searched for information on using NAEP for standard setting purposes.

Table 1. Highest and Lowest Performing States on 2015 NAEP Reading and Mathematics, Grades 4 and 8

Subject/Grade	High Performing	Low Performing
Mathematics		
Grade 4	MA MN NH	AL NM MS
Grade 8	MA MN NH	AL CA MS
Reading		
Grade 4	MA NH VT	NM CA AK MS
Grade 8	NH MA VT	MS NM LA

There were both differences and similarities in how the low and high performing states referred to the available NAEP data. The low performing states provided much less information about participating in NAEP and the purposes of NAEP, in general, compared to the high performing states. High performing states, on the other hand, were more likely to provide details about student performance and participation on NAEP. Many state DOE websites include links to the state NAEP results on the Nation's Report Card website. Some state websites made a statement that comparisons can be made of how students from different states performed on NAEP, or reference studies that linked state standards to the NAEP standards. However, both low and high performing states provided little information about the explicit uses of the NAEP data for the purposes of creating state level ALDs and informing the determination of cut scores at the state level.

The websites did not include any explicit reference to whether or how NAEP standards may inform state performance standards, or how NAEP data may serve as impact data in state standard settings. The most explicit statement of the connection between state assessment and NAEP was found on the MA DOE website: "...NAEP has taken on a greater prominence under the No Child Left Behind Act and serves to externally confirm results of state assessments, such as the Massachusetts Comprehensive Assessment System (MCAS)" (National Assessment of Educational Progress Frequently Asked Questions, 2017)." The state of Vermont makes another explicit comparison between the structure of its own state science test and the NAEP science assessment standards: "The tests were designed to measure different standards, or frameworks, on separate scoring scales, but both assessments address similar skills and content areas. These assessments provide a way to reference national, state and local science achievement" (*Vermont Students Score among Best in the Nation on the National Assessment of Educational Progress*, 2016). The state also points out some similarities in the pattern of scores on both the state assessment and NAEP.

³For more information see the website

<https://www.nationsreportcard.gov/profiles/stateprofile?chort=2&sub=RED&sj=AL&sfj=NP&st=MN&year=2015R3>.

Among the state websites studied, most high performing states reported:

- trends or comparisons of successive cohorts;
- comparison of the percentage of students at or above Proficient on NAEP to the percentage of students at or above Proficient on a state test;
- point-in-time comparisons across states, districts, or population groups (e.g., Vermont included information showing an increase in the performance of students of low SES);
- performance on subscales (e.g. algebra, vocabulary, etc.)
- rank ordering of states or districts;
- comparisons across population groups to examine performance gaps; and
- comparisons across subject areas.

Lower performing states tended to mention NAEP reports less often. However, we did find some information in the comments of school administrators to the media that NAEP results were used as an indication that the current state education system was in need of reform. For example, in 2013 the superintendent of Louisiana, [John White](#), “used the [NAEP state achievement] report to reiterate his push for the [Common Core](#) national education standards. ‘The growth this year was moderate. If we want to see something beyond incremental growth, we’ve got to raise our standards, and the Common Core standards is the best way to do that,’ he said” (Bacon-Blood, 2013).

Researchers and Business Leaders

For purposes of this memorandum, researchers and business leaders include persons conducting educational research and individuals from private industry with an interest in elementary and secondary student performance. Currently, NAEP data use and interpretation research by these stakeholders may take the following directions (Edley & Koenig, 2017):

- track trends in and compare the performance of successive cohorts,
- make point-in-time comparisons across states and school districts,
- compare the performance of population groups within and across states (performance gaps),
- rank order the performance of states and compare state to national performance;
- compare performance across tested subject areas,
- examine relationships among student performance and selected student/school/family variables, and
- compare states’ standards for proficient performance in reading and mathematics by placing them on a common scale defined by NAEP scores (“mapping studies”).

Beginning with NAEP results from 2003, NCES conducted a series of studies that mapped each state’s grade 4 and 8 reading and mathematics proficiency levels to the NAEP scale. This mapping was designed as a mechanism to evaluate the extent to which state standards reflected the same rigor as NAEP standards, and it was used as a policy lever to encourage states to set challenging standards for their students (Edley et al., 2017). In the mapping study report by Bandeira de Mello, Bohrnstedt, Blankenship, & Sherman (2015), the NAEP score that corresponds to a state’s standard (i.e., the NAEP scale equivalent score) is determined by a

direct application of equipercentile mapping. For a given subject and grade, the percentage of students reported in the state assessment to be meeting the standard in each NAEP school is matched to the point on the NAEP achievement scale corresponding to that percentage. The percentage of students passing the state standard was mapped onto the NAEP scores. The results are then aggregated over all of the NAEP schools in a state to provide an estimate of the NAEP scale equivalent of the state's threshold for its standard (Bandeira de Mello et al., 2015).

Peterson and Ackerman (2015) took a different approach to the comparison of state achievement scores and NAEP scores. They calculated the difference between the percentage of students considered “proficient” by both the state and NAEP assessments. The magnitude of the difference was considered to indicate how rigorous the state standards are as compared with NAEP standards.

These examples indicate that some researchers and policymakers do consider NAEP achievement levels to be a standard that states should strive toward. At the same time, some researchers caution against using NAEP as an infallible measure of state educational achievement due to fundamental differences between the state and NAEP frameworks and standards (e.g., Ho & Haertel, 2007). It is important to remember that determining the score equivalency between NAEP scale and state scale does not say anything about the equivalency or lack thereof in knowledge and skills associated with the score. The NAEP and state assessments may or may not measure the same knowledge and skills. An alignment study would need to be conducted to assess the extent to which the two assessments measured the same construct.

Many studies focused on validity evidence based on relationships with external variables, that is, setting benchmarks on NAEP that are related to concurrent or future performance on measures external to NAEP. Examples are academic preparedness for college; international tests; state tests and their alignment with NAEP (Edley et al., 2017). The studies indicate that there is considerable correspondence between the percentages of students at NAEP achievement levels and the percentages on other assessments (Gattis et al., 2016; Jia et al., 2014; Lim & Sireci, 2017; Neidorf, Binkley, Gattis, & Nohara, 2006; Phillips, 2014a, 2014b; Poland & Plevyak, 2015; Provasnik, Lin, Darling, & Dodson, 2013). These studies show that the NAEP achievement-level results (the percentage of students at the advanced level) are generally consistent with the percentage of U.S. students scoring at the reading and mathematics benchmarks on the Programme for International Student Assessment (PISA), the mathematics benchmarks on Trends in International Mathematics and Science Study (TIMSS), and at the higher levels for College Board Advanced Placement (AP) exams. For example, a report by Fields (2014) states that the content of the 12th grade NAEP reading and mathematics assessments was found to be similar to widely recognized tests used for college admission and placement. A linking study by Moran, Freund, & Oranje (2012) determined that there is a higher correlation between NAEP and SAT mathematics scores than between NAEP and SAT reading scores. The SAT reading benchmark, however, was closer to the NAEP Proficient score than the SAT math benchmark.. Several studies investigated the relationship between NAEP Proficient and college and career readiness (Moran, Oranje, & Freund, n.d.; Schneider, Kitmitto, Muhusani, & Zhu, 2015), but the relationship was found to be fairly weak. Additional research in this area was proposed.

During the August 2016 Governing Board quarterly meeting, researchers provided the following recommendations regarding the use of NAEP data.

- Panelists urged the Governing Board to enable linkages from NAEP data to state-level or national-level to conduct research about the long-term effects of educational policies.

- All panelists agreed that while NAEP data describe trends in student achievement, the data do not support conclusions about the reasons for these trends. Additional research is needed to discover factors that can improve schools and student learning.
- It was suggested that the NAEP data be used to compare the performance of districts with similar demographic characteristics, such as poverty levels. NAEP data may be used to guide best practices on what works in the improvement of educational achievement.

The Media

While academic and research articles provide scientific, well-reasoned rationales for or against the specific interpretations of NAEP, articles by the media present a different side. They tell the story of those who are trying to use information under real-life conditions from the assessments that the academics are studying, and the real-world challenges and issues experienced by practitioners in the field.

Articles in publications like *Education Week* illustrate that there is a large degree of confusion accompanying the application and interpretation of NAEP standards. While many researchers and even state officials may assume the debate about the application of NAEP standards is resolved, magazine and newspaper articles question whether it is appropriate for states to incorporate NAEP standards into the standards of the state, and what the appropriate uses for NAEP scores are in general.

One point of argument is lack of clarity on the meaning of “proficient” and the application of that meaning to state standards. Not all media representatives consistently clarify for the public that NAEP Proficient is not grade-level proficiency and that NAEP Proficient is intended to be an aspirational standard. What makes this matter more complicated is that under the No Child Left Behind Act (NCLB), states had to create achievement levels that were grade-specific and most states chose to adopt the ALD title of “Proficient.” Reconciling these sets of standards causes additional conflict and confusion when states are trying to create their achievement levels and communicate them to the public. One suggestion to make the situation more understandable is for policymakers to explain to the stakeholders “what are good goals for educational purposes compared to what is appropriate for accountability when establishing cut scores on their state assessments” (Hull, 2008), why they may be different, and which performance levels are more appropriate for each specific purpose.

Many researchers are concerned that information from NAEP gets misinterpreted by the media and politicians, sometimes to serve the interests of specific groups. Various misinterpretations of NAEP results are frequently used by the politicians and media, giving rise to the term “misnaepery” (Sawchuk, 2013). One prominent example of this inappropriate interpretation includes tying an increase in state NAEP scores to some specific policy or intervention implemented by the state, and a decrease – to a policy that was proposed by an organization, but then not implemented. In practice, it is very challenging to make these causal connections. Organizations that are using NAEP scores to bolster claims about the effects of a specific policy are likely not interpreting the NAEP scores correctly (Chingos & Blagg, 2015).

A number of misinterpretations come from the misunderstanding of NAEP’s definition of “proficient”, with some reporters claiming that being below proficient means being “below grade level.” Yet another source of confusion comes from comparing state assessment scores with NAEP scores and arriving at opposing conclusions. Comparing the achievement of different student population groups is often fraught with misinterpretations as well (e.g., treating the

NAEP achievement scale as continuous between grades and comparing achievement of one population at a higher grade to the achievement of another population at a lower grade).

At least in part, these misinterpretations arise from a lack of readily available or accessible information on how the NAEP scores should be interpreted, what the appropriate uses of these scores are, and what conclusions are appropriate to make. Educational researchers call for using caution in deciphering which claims are appropriate, and discouraging the propagation of false claims about NAEP data interpretation (Polikoff, 2015a, 2015b).

The General Public

The general public may not have sufficient knowledge and training to deeply understand the intent and the meaning of state or national assessments is, and may have a difficult time interpreting and critically evaluating information coming from various, often conflicting, sources. The media may make the situation in education appear more critical or negative than it really is. For example, if a state performs as one of the best on NAEP, but there is no growth in scores, the general public may see headlines like “Public education test results are dismal. Schools are failing NH children” (Levell, 2016). In addition, as mentioned earlier, the information provided by the media may not be completely objective, and score interpretations may be promoting a specific political agenda.

There is some confusion among the general public regarding why their state may have high scores on the state assessments, but low scores on NAEP (Weiss, 2016; Dillon, 2005). This may occur if the state set standards lower than NAEP standards, or if the state simply has different content standards. There may also be conflicting information on exactly how the state standards compare to NAEP standards; this may cause one study to claim that a state has low standards, and another study – that the state is either lagging behind others, or low on scores from some other perspective. A study by Achieve⁴, describes several NAEP objectives at grade 4 contrasted with the grade those same objectives are introduced in several states’ standards documents. The objective “Use simple ratios to describe problem situations,” is typically introduced in grade 6 in many states. Discrepancies like this add complexity to potential comparisons between NAEP results and state testing results.

One potential goal would be for the general public to be able to use state and national assessments to make decisions about whether children are getting the best education in their particular state. It is likely impossible to make such inferences at the school or even classroom level from state and national assessments. The media, however, may make it sound like those conclusions are appropriate and necessary. The same article by Levell (2016) that proclaimed the failure of New Hampshire public education, for example, suggests that, based on the fact that there was little to no growth in the student scores on state assessments or NAEP, the parents should “[e]ngage your local school board and question why they are using College and Career Readiness Standards and tests that are not providing a better education for our children;” consider a transfer to a charter or private school; or refuse to have their child take a state assessment. It may be helpful for the general public to have access to a source of clear, easy to understand, reliable information on the kinds of inferences that can legitimately be made from state and national assessments.

⁴ See https://www.achieve.org/files/16-149_Achieve_NAEP%20math%20report.pdf.

References

- American Educational Research Association, American Psychological Association, National Council on Measurement in Education, Joint Committee on Standards for Educational, & Psychological Testing (US). (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Angoff, W. H. (1971). Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.), *Educational Measurement* (2nd ed., pp. 508-597). Washington, DC: Council on Education.
- Angoff, W. H. (1988). Proposals for theoretical and applied development in measurement. *Applied Measurement in Education*, 1, 215-222.
- Bacon-Blood, L. (2013, November 7). Louisiana students score near bottom on national tests. Retrieved from http://www.nola.com/education/index.ssf/2013/11/louisiana_students_score_near.html
- Bandeira de Mello, V., Bohrnstedt, G., Blankenship, C., & Sherman, D. (2015). *Mapping State Proficiency Standards onto NAEP Scales: Results from the 2013 NAEP Reading and Mathematics Assessments* (NCES 2015-046). National Center for Education Statistics. Washington, DC: U.S. Department of Education.
- Bourque, M. L. (2009). A History of NAEP Achievement Levels: Issues, Implementation, and Impact 1989-2009. *National Assessment Governing Board*.
- Chingos, M., & Blagg, K. (2015, October 28). How do states really stack up on the 2015 NAEP? Retrieved from <https://www.urban.org/urban-wire/how-do-states-really-stack-2015-naep>
- Dillon, S. (2005, November 26). Students ace state tests, but earn D's from U.S. Retrieved from <http://www.nytimes.com/2005/11/26/education/students-ace-state-tests-but-earn-ds-from-us.html#story-continues-1>
- Egan, K.L., Schneider, M.C., & Ferrara, S. (2012). Performance level descriptors: History, practice and a proposed framework. In G.J. Cizek (Ed.), *Setting Performance Standards: Foundations, Methods, and Innovations* (pp. 79-106). New York: Routledge.
- Fields, R. (2014). Towards the National Assessment of Educational Progress (NAEP) as an Indicator of Academic Preparedness for College and Job Training. Washington, DC: National Assessment Governing Board. Retrieved from <http://ed.sc.gov/scdoe/assets/File/tests/middle/naep/NAGB-indicator-of-preparedness-report.pdf>
- Gattis, K., Kim, Y., Stephens, M., Dager, L., Fei, H., & Holmes, J. (2016). A Comparison Study of the Program for International Student Assessment (PISA) 2012 and the National Assessment of Educational Progress (NAEP) 2013 Mathematics Assessments. AIR-NAEP Working paper #02-2016.
- Hambleton, R., Pitoniak, M., & Copella, J. (2012). Essential steps in setting performance standards on educational tests and strategies for assessing the reliability of results. In G.J. Cizek (Ed.), *Setting Performance Standards: Foundations, Methods, and Innovations* (pp. 47-76). New York: Routledge.

- Ho, A. D., & Haertel, E. H. (2007). [Apples to apples? The underlying assumptions of state-NAEP comparisons](https://scholar.harvard.edu/files/andrewho/files/ho_haertel_apples_to_apples.pdf). Paper commissioned by the Council of Chief State School Officers. Retrieved from https://scholar.harvard.edu/files/andrewho/files/ho_haertel_apples_to_apples.pdf
- Hull, J. (2008, June 17). The proficiency debate. Retrieved from <http://www.centerforpubliceducation.org/Main-Menu/Evaluating-performance/The-proficiency-debate-At-a-glance/The-proficiency-debate-A-guide-to-NAEP-achievement-levels.html>
- Jia, Y., Phillips, G., Wise, L.L., Rahman, T., Xu, X., Wiley, C., & Diaz, T.E. (2014). 2011 NAEP-TIMSS Linking Study: Technical Report on the Linking Methodologies and Their Evaluations (NCES 2014-461). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, Washington, D.C.
- Kane, M. (2001). So much remains the same: conception and status of validation in setting standards. In G. J. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives* (pp.53-88). Mahwah, NJ: Lawrence Erlbaum.
- Kane, M. (2006). *Validation*. In R. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17–64). Westport, CT: American Council on Education and Praeger.
- Kane, M. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, v50(1), pp. 1-73.
- Levell, M. (2016, November 4). Public education test results are dismal. Schools are failing NH children. Retrieved from <http://nhpoliticalbuzz.org/public-education-test-results-are-dismal-schools-are-failing-nh-children/>
- Lim, H., & Sireci, S. G. (2017). Linking TIMSS and NAEP assessments to evaluate international trends in achievement. *Education Policy Analysis Archives/Archivos Analíticos de Políticas Educativas*, (25), 1-25.
- Loomis, S. (2012). Selecting and training standard setting participants. In G. J. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives* (pp.107-134). Mahwah, NJ: Lawrence Erlbaum.
- Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2003). A brief introduction to evidence-centered design. *ETS Research Report Series*, 2003(1).
- Moran, R., Freund, D., & Oranje, A. (2012). Analyses relating Florida students' performance on NAEP to preparedness indicators and postsecondary performance. Washington, DC: Author. Retrieved from https://www.nagb.org/content/nagb/assets/documents/what-we-do/preparedness-research/statistical-relationships/Florida_Statistical_Study.pdf
- Moran, R., Oranje, A., & Freund, D. (n.d.). NAEP 12th grade preparedness research: Establishing a statistical relationship between NAEP and SAT. Retrieved from http://www.nagb.org/content/nagb/assets/documents/what-we-do/preparednessresearch/statistical-relationships/SAT-NAEP_Linking_Study.pdf
- National Academies of Sciences, Engineering, and Medicine. (2017). *Evaluation of the Achievement Levels for Mathematics and Reading on the National Assessment of Educational Progress*. Washington, DC: The National Academies Press.

- National Assessment Governing Board. (2017). National Assessment Governing Board's Response to the National Academies of Sciences, Engineering, and Medicine 2016 Evaluation of NAEP Achievement Levels. Provided by the National Assessment Governing Board November 2017. Author.
- National Assessment of Educational Progress Frequently Asked Questions. (2017, October 11). Retrieved from <http://www.doe.mass.edu/mcas/natl-intl/naep/faq.html?section=overview>
- Neidorf, T., Binkley, M., Gattis, K., & Nohara, D. (2006). Comparing Mathematics Content in the National Assessment of Educational Progress (NAEP), Trends in International Mathematics and Science Study (TIMSS), and Program for International Student Assessment (PISA) 2003 Assessments. NCES 2006-029.
- Norcini, J.J., & Shea, J.A. (1992). The reproducibility of standards over groups and occasions. *Applied Measurement in Education*, 5, 63-72.
- Peterson, P., & Ackerman, M. (2015). States raise proficiency standards in math and reading. Retrieved from file:///C:/NAEP/ednext_XV_3_peterson%20ackerman%202015.pdf
- Phillips, G. W. (2007). Expressing International Educational Achievement in Terms of US Performance Standards: Linking NAEP Achievement Levels to TIMSS. *American Institutes for Research*.
- Phillips, G. W. (2014a). International Benchmarking: State and National Education Performance Standards. *American Institutes for Research*.
- Phillips, G. W. (2014b). Linking the 2011 National Assessment of Educational Progress (NAEP) in Reading to the 2011 Progress in International Reading Literacy Study (PIRLS). *American Institutes for Research*.
- Poland, S., & Plevyak, L. (2015). US student Performance in science: A review of the four major science assessments. *Problems of Education in the 21st Century*, 64.
- Polikoff, M. (2015a, October 6). Friends don't let friends misuse NAEP data. [Blog post]. Retrieved from <https://morganpolikoff.com/tag/naep/>
- Polikoff, M. (2015b, October 28). My quick thoughts on NAEP. [Blog post]. Retrieved from <https://morganpolikoff.com/tag/naep/>
- Provasnik, S., Lin, C., Darling, D., & Dodson, J. (2013). A comparison of the 2011 Trends in International Mathematics and Science Study (TIMSS) assessment items and the 2011 National Assessment of Educational Progress (NAEP) frameworks. *National Center for Education Statistics*.
- Sawchuk, S. (2013, July 24). When bad things happen to good NAEP data. Retrieved from <https://www.edweek.org/ew/articles/2013/07/24/37naep.h32.html>
- Schneider, M. C., Huff, K. L., Egan, K. L., Tully, M., & Ferrara, S. (2010). Aligning achievement level descriptors to mapped item demands to enhance valid interpretations of scale scores and inform item development. In *annual meeting of the American Educational Research Association, Denver, CO*.

Schneider, M., Kitmitto, S., Muhusani, H., & Zhu, B. (2015). Using the National Assessment of Educational Progress as an Indicator for College and Career Preparedness. Washington, DC: Author. Retrieved from <http://www.air.org/sites/default/files/downloads/report/Using-NAEP-as-an-Indicator-College-Career-Preparedness-Oct-2015.pdf>

Shepard, L., Glaser, R., Linn, R., & Bohrnstedt, G. (1993). Setting performance standards for student achievement: A report of the National Academy of Education Panel on the Evaluation of the 1992 Achievement Levels. Stanford, CA: National Academy of Education

Vermont students score among best in the nation on the National Assessment of Educational Progress. (2016, November 3). Retrieved from <http://education.vermont.gov/sites/aoe/files/documents/edu-press-release-naep-necap-scores.pdf>

Vinovskis, M. A. (1998). Overseeing the Nation's Report Card: The Creation and Evolution of the National Assessment Governing Board (NAGB).

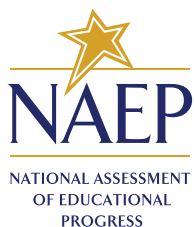
Weiss, J. (2016, January 27). Report says Texas school standards are worst in nation. Retrieved from <https://www.dallasnews.com/news/education/2016/01/27/report-says-texas-school-standards-are-worst-in-nation>

Zenisky, A., Hambleton, R.K., & Sireci, S.G. (2009). Getting the message out: An evaluation of NAEP score reporting practices with implications for disseminating test results. *Applied Measurement in Education*, 22(4), 359-375.



What Are NAEP Achievement Levels and How Are They Determined?

Every two years, the National Assessment of Educational Progress (NAEP) in reading and mathematics is administered and results are reported. The results are presented for the nation, for states and jurisdictions like the District of Columbia, and for 27 select urban school districts. The Nation's Report Card, as NAEP is known, provides more than just average scores, so that student performance can be understood more fully. Results can be examined by characteristics of schools, teachers, and students, such as urbanicity, teacher certification, and student demographics.



The National Assessment Governing Board was created by Congress in 1988 as an independent, nonpartisan board to set policy for NAEP. The Governing Board's duties include determining what subjects are assessed when and developing student achievement levels for NAEP. Achievement levels represent what students should know and be able to do, essentially answering the question: How good is good enough on NAEP?

The Governing Board defines the meaning of the achievement levels of *Basic*, *Proficient*, and *Advanced* through a careful and deliberate process. Achievement-level setting, also called standard setting, is not unique to NAEP or to educational testing. Medical boards and state bar organizations use cut scores to determine who may practice in their respective professions. More broadly, athletic competitions may set cut scores for what skills need to be demonstrated before moving to the next round. For The Nation's Report Card, the process translates content expectations to the NAEP scale, producing cut scores that separate each category of performance.

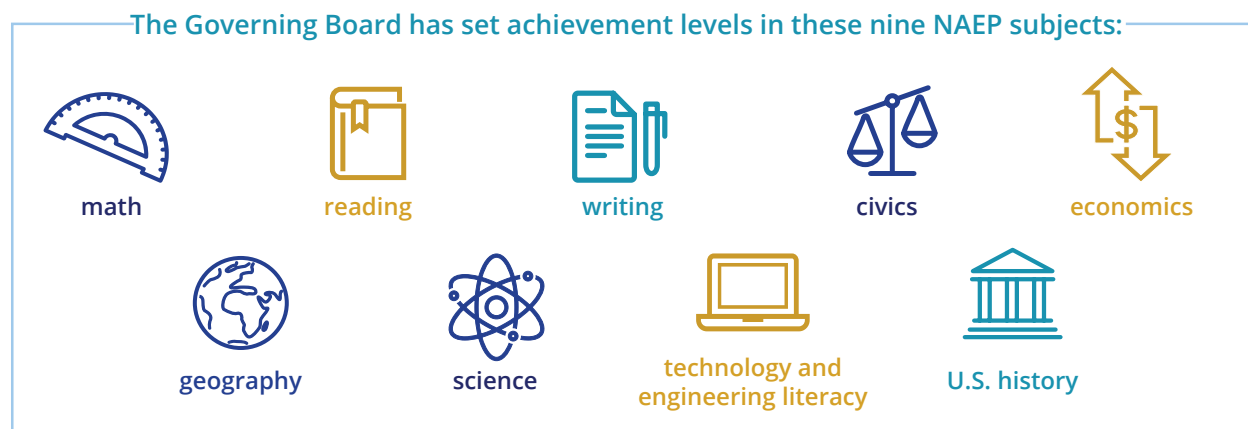
ACHIEVEMENT LEVELS



NAEP standard-setting involves approximately 20 to 30 content experts, typically teachers and other subject-matter experts, who meet for several days and receive in-depth training. Standard-setting experts provide technical guidance throughout the process and the content experts apply their knowledge and experience to recommend the achievement levels. Panelists also recommend exemplars—items or student responses that are representative examples of performance at the *Basic*, *Proficient*, and *Advanced* levels. Achievement Level Descriptions, specific to each assessment, accompany the recommendations to clarify what *Proficient* means, for example, on the NAEP grade 4 mathematics assessment.

NAEP defines *Proficient* differently from other uses of the term. For example, the Every Student Succeeds Act refers to student “proficiency.” State assessment systems may use the terms “proficient” and “proficiency,” but there is wide variation in how states define proficient., e.g., equivalent to grade-level performance or a description of what students already know. This variation in terminology is often a source of confusion when it comes to understanding the NAEP achievement levels. For NAEP, *Proficient* 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. Thus, *Proficient* represents an aspirational goal for what all students should know.

The National Assessment Governing Board is responsible for setting achievement levels based on the outcome of the standard-setting process. The Board has set achievement levels in nine NAEP subjects and always solicits public comment on the plans for the standard-setting process and on the content of the Achievement Level Descriptions. The Nation’s Report Card includes information about the [types of items or responses](#) that exemplify each achievement level, which helps to illustrate what performance at each achievement level looks like.



By presenting both average scores and achievement levels, the National Assessment Governing Board fulfills its Congressionally mandated obligation to improve the reporting of results on The Nation’s Report Card. For more information, please visit the Governing Board’s website: www.nagb.gov.

Strategic Vision Activities Led by COSDAM

During the November 2016 Board meeting, a [Strategic Vision](#) was formally adopted to guide the Board's work over the next several years. For each activity led by COSDAM, information is provided below to describe the current status and recent work, planned next steps, and the ultimate desired outcomes. Please note that many of the Strategic Vision activities require collaboration across committees and with NCES, but the specific opportunities for collaboration are not explicitly referenced in the table below. In addition, the activities that include contributions from COSDAM but are primarily assigned to another standing committee (e.g., framework update processes) or ad hoc committee (i.e., exploring new approaches to postsecondary preparedness) also have not been included below.

The Governing Board's Assistant Director for Psychometrics, Sharyn Rosenberg, will answer any questions that COSDAM members have about ongoing or planned activities.

Strategic Vision Activity	Current Status and Recent Work	Planned Next Steps	Desired Outcome
<p>SV #2: Increase opportunities to connect NAEP to administrative data and state, national, and international student assessments</p> <p><i>Incorporate ongoing linking studies to external measures of current and future achievement in order to evaluate the NAEP scale and add meaning to the NAEP achievement levels in reporting. Consider how additional work could be pursued across multiple subject areas, grades, national and international assessments, and longitudinal outcomes</i></p>	<p>COSDAM discussions at May and August 2017 board meetings to examine how existing findings may be used to add meaning to scale scores and achievement levels, and what additional studies to take on</p> <p>Ongoing linking studies include: national NAEP-ACT linking study; longitudinal studies at grade 12 in MA, MI, TN; longitudinal studies at grade 8 in NC, TN; NAEP-TIMSS linking study; NAEP-HSLS linking study; NAEP Validity Studies (NVS) studies</p> <p>Informational update on current studies was provided in the March 2018 COSDAM materials</p>	<p>Complete ongoing studies</p> <p>Decide what new studies to take on</p> <p>Decide how to use and report existing and future results</p> <p>Complete additional studies</p>	<p>NAEP scale scores and achievement levels may be reported and are better understood in terms of how they relate to other important indicators of interest (i.e., other assessments and milestones)</p>

Strategic Vision Activity	Current Status and Recent Work	Planned Next Steps	Desired Outcome
<p>SV #3: Expand the availability, utility, and use of NAEP resources, in part by creating new resources to inform education policy and practice</p> <p><i>Research when and how NAEP results are currently used (both appropriately and inappropriately) by researchers, think tanks, and local, state and national education leaders, policymakers, business leaders, and others, with the intent to support the appropriate use of NAEP results (COSDAM with R&D and ADC)</i></p> <p><i>Develop a statement of the intended and unintended uses of NAEP data using an anticipated NAEP Validity Studies Panel (NVS) paper and the Governing Board's research as a resource (COSDAM with NCES)</i></p> <p><i>Disseminate information on technical best practices and NAEP methodologies, such as training item writers and setting achievement levels</i></p>	<p>Ina Mullis of the NVS panel spoke with COSDAM at the March 2017 board meeting and is working on a white paper about the history and uses of NAEP</p> <p>Technical Support contract specifies that the research study topic for year 1 will focus on how NAEP results are used by various stakeholders. The study is currently underway and is planned for COSDAM discussion during the August 2018 Board meeting.</p> <p>This idea was generated during the August 2017 COSDAM discussion of the Strategic Vision activities</p>	<p>Use research to draft short document of intended and appropriate uses for Board discussion (November 2018)</p> <p>NCES produces documentation of validity evidence for intended uses of NAEP scale scores</p> <p>Governing Board produces documentation of validity evidence for intended uses of NAEP achievement levels</p> <p>Work with NCES and R&D to refine list of technical topics for dissemination efforts</p>	<p>Board adopts formal statement or policy about intended uses of NAEP. The goal is to increase appropriate uses and decrease inappropriate uses (in conjunction with dissemination activities to promote awareness of the policy statement)</p> <p>Stakeholders benefit from NAEP technical expertise</p>

Strategic Vision Activity	Current Status and Recent Work	Planned Next Steps	Desired Outcome
<p>SV# 5: Develop new approaches to update NAEP subject area frameworks to support the Board's responsibility to measure evolving expectations for students, while maintaining rigorous methods that support reporting student achievement trends</p> <p><i>Consider new approaches to creating and updating the achievement level descriptors and update the Board policy on achievement levels</i></p>	<p>Panel of standard setting experts convened in January 2018 to discuss technical issues and recommendations for achievement levels policy</p> <p>Literature review on considerations for creating and updating achievement level descriptors (ALDs)</p> <p>Technical Memo on developing a validity argument for the NAEP achievement levels (February 2018)</p> <p>The efforts described above were discussed at the March 2018 COSDAM meeting</p>	<p>COSDAM discussion of draft revised policy statement to occur at this Board meeting (May 2018)</p> <p>COSDAM call to discuss revised draft policy (June 2018)</p> <p>Revised policy statement for full Board discussion (August 2018)</p> <p>Seek external feedback and public comment (September 2018)</p> <p>Full Board call to discuss revised draft policy (October 2018)</p> <p>Board action on revised policy statement (November 2018)</p>	<p>Board has updated policy on achievement levels that meets current best practices in standard setting and is useful for guiding the Board's achievement levels setting work</p>
<p>SV# 7: Research policy and technical implications related to the future of NAEP Long-Term Trend assessments in reading and mathematics</p> <p><i>Support development and publication of multiple papers exploring policy and technical issues related to NAEP Long-Term Trend. In addition to the papers, support symposia to engage researchers and policymakers to provide stakeholder input into the Board's recommendation</i></p>	<p>White papers commissioned, symposium held in Washington, DC (March 2017), and follow-up event held at American Educational Research Association (AERA) annual conference (April 2017)</p> <p>Full Board and Executive Committee discussions (March, May, and August 2017) and webinar on secure LTT items and p-values from 2012 administration (October 2017)</p>	<p>Per the discussion and next steps at the March 2018 Executive Committee meeting, COSDAM will discuss design considerations for the next administration of LTT. Additional information is expected to be provided by NCES at the August COSDAM meeting.</p>	<p>Determine whether changes to the NAEP LTT schedule, design and administration are needed (led by Executive Committee and NCES)</p>

Strategic Vision Activity	Current Status and Recent Work	Planned Next Steps	Desired Outcome
<p>SV# 9: Develop policy approaches to revise the NAEP assessment subjects and schedule based on the nation's evolving needs, the Board's priorities, and NAEP funding</p> <p><i>Pending outcomes of stakeholder input (ADC activity), evaluate the technical implications of combining assessments, including the impact on scaling and trends</i></p>	<p>COSDAM presentation and discussion on initial considerations for combining assessments (November 2017)</p> <p>Full Board presentation and discussion on efficiencies in what and how to measure student knowledge and skills (March 2018)</p>	<p>Plenary discussion of Assessment Schedule during May 2018 Board meeting</p> <p>Additional discussion planned for August 2018, with Board action tentatively scheduled for November 2018</p>	<p>Determine whether new assessment schedule should include any consolidated frameworks or coordinated administrations</p>
<p>SV# 10: Develop new approaches to measure the complex skills required for transition to postsecondary education and career</p> <p><i>Continue research to gather validity evidence for using 12th grade NAEP reading and math results to estimate the percentage of grade 12 students academically prepared for college</i></p>	<p>Several studies are ongoing (see activities under SV# 2)</p> <p>Per COSDAM discussion at August 2017 meeting, additional studies are on hold until at least November 2018 pending Board decision on how to move forward with findings from Ad hoc Committee on Measures of Postsecondary Preparedness</p>	<p>Decide whether additional research should be pursued at grade 8 to learn more about the percentage of students "on track" to being academically prepared for college by the end of high school</p> <p>Decide whether Board should make stronger statement and/or set "benchmarks" rather than current approach of "plausible estimates"</p> <p>Decide whether additional research should be conducted with more recent administrations of NAEP and other tests</p>	<p>Statements about using NAEP as an indicator of academic preparedness for college continue to be defensible and to have appropriate validity evidence</p>