
Committee on Standards, Design and Methodology

August 5, 2022

8:30 am – 10:30 am ET



Hyatt House Charleston Historic District
560 King Street
Charleston, South Carolina
(843) 414-4900

AGENDA

8:30 – 8:45 am	Welcome and Updates <i>Suzanne Lane, Chair</i>	
8:45 – 9:20 am	Update: Achievement Level Descriptor Review Study for U.S. History, Civics, and Science <i>Eric Moyer</i> <i>Suzanne Lane</i>	Attachment A
9:20 – 10:30 am	Discussion: Achievement Levels Work Plan <i>Suzanne Lane</i>	Attachment B

Design Document for Achievement Level Descriptor Study for Grade 8 Science, U.S. History, and Civics

In May 2022, the National Assessment Governing Board executed a contract modification with NCS Pearson, Inc. to conduct a study of the achievement level descriptors (ALDs) for grade 8 Science, U.S. History, and Civics. The study includes two intended outcomes: a) a set of revised ALDs for use as reporting ALDs, describing what students demonstrate they know and can do based on assessment data, and b) judgments of the alignment between newly developed reporting ALDs and the Board's achievement level policy descriptions and framework content ALDs. This contract modification was an option included with a 2020 contract conducting the same work for Math and Reading. The contract modification was based on the Board's response to two recommendations (recommendations 1 and 3) included in an [evaluation of the NAEP achievement levels](#) conducted by the National Academies of Sciences, Engineering, and Medicine in November 2016. The Board's full response to the Academies report is outlined in its [Achievement Levels Work Plan](#). COSDAM approved the two studies prior to the award of the original 2020 contract.

The following pages include a design document developed by Pearson outlining the plans and timeline for carrying out the ALD study for Science, U.S. History, and Civics. The document incorporates lessons learned from the Math and Reading study and feedback from the project's Technical Advisory Committee (TAC).

National Assessment Governing Board

Developing Reporting Achievement Level Descriptions for Grade 8 Science, U.S. History, and Civics

Submitted: June 2022

NAEP ALD Development Design Document

Submitted to:
National Assessment Governing Board
800 North Capitol Street, NW, Suite 825
Washington, DC 20002-4233

This study was funded by the
National Assessment Governing Board under
Contract 91995920R0004.

Submitted by:
NCS Pearson, Inc.
2510 N. Dodge Street
Iowa City, IA 52245-9945
Phone: 319.354.9200

Table of Contents

Table of Contents	2
Executive Summary	4
Section 1: Achievement Level Descriptions Review Panels	8
Overview of Panelist Recruiting Process	9
Identification of Panelist Nominators.....	10
Selection of Panelists.....	11
Section 2: Briefing Materials	14
Section 3: Pilot Study	15
Section 4: ALD Alignment Review Tasks and Procedures	17
Computer Use and Software.....	20
Preparing for the ALD Alignment Review Meeting.....	22
Development of Anchor Item Sets.	22
Division of Panelists into Replicate Panels.	24
Provision of Advanced Briefing Materials.....	25
Training of Facilitators.....	25
Preparation of Pearson Laptops.....	26
Preparing Panelists for the ALD Alignment Review Process.....	26
Provide an Orientation.....	26
Provide an Overview of the Alignment Review and Revision Process.	26
Individual Item Review and Summary Descriptions.....	27
Item Review.....	27
Summary Descriptions.	28
Alignment Judgment Rounds with Panel Discussions	29

Alignment Judgment Rounds.29

Process Evaluations.30

Section 5: External Feedback.....31

Content Review.31

Internal Review.32

External Review.32

References34

Executive Summary

The National Assessment of Educational Progress (NAEP), known as the “The Nation’s Report Card,” provides information on what students in the United States know and can do in various subject areas. As part of its legislative mandate for overseeing and setting policy for NAEP, the National Assessment Governing Board (Governing Board) develops achievement levels that further define expectations of what students should know and be able to do.

Achievement on all NAEP assessments is reported using the following achievement levels, in accordance with the Board policy on Developing Student Achievement Levels for NAEP, which are defined as follows:

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

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

NAEP Advanced – This level signifies superior performance beyond NAEP Proficient.

The achievement level policy descriptions apply to all NAEP assessments, regardless of subject and grade. In addition to these achievement levels, content-specific achievement level descriptions (ALDs) are developed to define the expected knowledge and skills for student at each achievement level.

The National Academies of Sciences, Engineering, and Medicine completed an evaluation of NAEP achievement levels in November 2016, which included seven recommendations. Two of the recommendations from the report – recommendations #1 and #3 – were related to NAEP ALDs.

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

Recommendation #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.

In response to this evaluation of the NAEP achievement levels, the Governing Board updated its policy in November 2018 to include reporting ALDs for all NAEP assessments and developed and approved an Achievement Levels Work Plan in March 2020. To carry out the work on the ALDs, the Governing Board issued a contract to Pearson in September 2020 to develop and conduct anchoring studies using NAEP 2019 data to review the ALDs and create reporting ALDs for grades 4, 8, and 12 reading and mathematics. The contract included an option for conducting similar studies for grade 8 science, U.S. history and civics; this work began in May 2022.

On behalf of the Governing Board, Pearson has developed this Design Document, which describes in detail the activities for the anchoring studies for grade 8 NAEP science, U.S. history, and civics. This document is intended to provide the foundation for all ALD alignment review and revision activities. The Design Document will guide all aspects of the ALD review and revision process, including (1) a set of judgments about the alignment between the policy descriptions and current ALDs and expectations of what students should know and be able to do,

based on the items that anchor to each achievement level for each assessment and (2) a set of revised ALDs that could be used as reporting ALDs, describing what students demonstrate they know and can do.

Pearson plans to use a model-based approach for conducting the study of the ALDs for NAEP science, U.S. history, and civics. The Board's Achievement Levels Work Plan indicated that the methodology for these studies should be similar to previous ALD development and review meetings held in 2009 (Donahue, Pitoniak, & Beaulieu, 2010; Pitoniak, Dion, & Garber, 2010) to reduce the potential for possible inconsistencies from the use of different methods. This was also the process used to complete the companion study for the NAEP ALD review focusing on grades 4, 8, and 12 NAEP mathematics and reading, which was completed in 2022.

The model-based approach includes three stages. The first stage involves conducting statistical analysis to determine the items from the subject and grade that are anchored to each achievement level. The second stage relies on panels of content experts for each individual assessment. Each panelist will review the items that are anchored to each achievement level to identify the knowledge and skills needed to respond to the items associated with a level. The panelists will then work together to develop common summary descriptions of the knowledge and skills that students in each level know and can do, based on the review of all items anchored to an achievement level. In the third stage, the panelists complete several alignment judgment rounds, comparing the current ALDs for the respective assessment with the summary descriptions. During the alignment judgment rounds, the panelists rate the degree of alignment and take notes regarding the similarities and differences between the ALDs and summary descriptions. The goal of this process is for the panel to make a recommendation regarding whether the current ALDs accurately describe what students in each level are expected to know

and be able to demonstrate or if revisions to the current ALDs are needed to more accurately represent what students within each achievement level demonstrate they know and are able to do.

To facilitate this process, Pearson will use computers during both the pilot and operational ALD review meetings. Using computers along with the online interface in the Pearson Standard Setting website will increase the efficiency of the activities the panelists will need to complete for this study.

Below is a summary of what each section in the Design Document includes. There are additional details about each section in this process beginning on page 23.

Section 1: *Achievement Level Descriptions Review Panels* describes the panelist identification and recruitment plan designed to obtain broadly representative and well-qualified panelist groups for all studies.

Section 2: *Briefing Materials* describes the briefing materials sent to panelist prior to each panel study in the ALD review process (pilot study and operational ALD review meeting).

Section 3: *Pilot Study* describes the pilot study designed to incorporate the exact procedures planned for the operational ALD review meeting.

Section 4: *Achievement Level Descriptions Review Tasks and Procedures* describes the ALD review tasks, the nature of the tasks, and the procedures to be implemented prior to and as a part of the operational ALD review meeting—including how panelists are trained and supported in implementing all activities.

Section 5: *External Feedback* describes the process of obtaining feedback on the results and recommendations from the operational ALD review meeting.

Section 1: Achievement Level Descriptions Review Panels

Pearson will implement a multi-step panelist recruitment plan for the pilot study and operational ALD review meeting. The objective of the recruitment plan is to produce well-qualified panels with a high level of content expertise, consisting of classroom teachers and content experts who will be in the best position to provide the necessary judgments for the respective subject and grade. Panelists for the pilot study and operational meeting will be recruited from across the nation. For each meeting, there will be three committees convened, with each committee focused on a single subject and grade.

Grade 8 Science

Grade 8 U.S. History

Grade 8 Civics

A maximum of 24 panelists will be recruited for the pilot study, with up to eight panelists recruited for each committee. Additionally, a maximum of 24 panelists with similar background distribution will be recruited for the operational study, with up to eight panelists recruited for each committee. Panelists in each committee will be assigned to one of two replicate groups, allowing for a comparison of results across groups.

Committees for the ALD alignment review and revision operational and pilot studies will reflect an overall balance of gender, race/ethnicity, geographic location, and urbanicity, with the intent for no more than 75% homogeneous for any of these demographic categories. Each committee will consist of classroom teachers currently engaged in student instruction for grade 8 and the subject area and non-classroom educators with curriculum experience within the respective subject, such as state or local curriculum coordinators or higher education faculty teaching education courses associated with the respective subject and

level. To make ensure that each committee has a correct balance of experience for the activities, at least half of the panelists for each committee will consist of classroom teachers, and at least two of the panelists recruited will be non-classroom educators.

Our goal for this study is to obtain panelists with a high level of content expertise (though not direct experience working on NAEP ALD, item, or framework development) that would enable them to engage in the ALD review process and provide meaningful recommendations and judgments. Pearson will work with staff from the Governing Board along with allied organizations in recruiting panelists.

Overview of Panelist Recruiting Process

A multi-phase process will be used to identify panelists:

Phase 1: Identify nominators through allied organizations, state departments of education, and interested individuals. Nominators will be contacted and asked to nominate outstanding classroom and non-classroom educators using an online nomination form (e.g., name, contact information, and basic qualifications). Nominators will be asked to provide information for up to six nominees, to allow nominations for a classroom teacher and non-classroom educator for each committee. For each nominee, the nominator will be asked to briefly describe the rationale for the recommendation.

Phase 2: Notify nominees of selection and request additional information. Résumés and completed panelist forms (e.g., background in instruction with respective subject and grade, professional achievements, experience with students) will be requested and collected from each nominee. Nominees will be asked to specify training and experience that makes them an outstanding candidate for panel selection.

Phase 3: Evaluate and select panelists based on their background and experience. The most qualified panelists for each committee will be selected and assigned to replicate groups balanced with respect to gender, race/ ethnicity, geographic location, instruction experience, type of institutional affiliation, and urbanicity.

Prior to finalizing the selection of candidates to participate in the pilot and operational studies, the list of prospective panelists will be prepared and presented to the Governing Board for review and approval.

To optimize recruitment, Pearson project management will work with the Pearson meetings team to run the nomination and recruitment process. The meetings team will use email and phone calls to communicate with prospective panelists. Once selected to participate in a study, the meetings team will work with the panelists to organize travel and lodging for the meeting. An honorarium will be paid to panelists for the ALD alignment review and revision operational and pilot studies. Substitute teacher costs will be reimbursed directly to schools based on actual school costs for substitute teacher payments.

Identification of Panelist Nominators

Panelist nominators will be recruited using multiple sources. One source for nominators will be professional organizations that have a strong background in providing professional development in science education or history, civics, or social studies education. Focused recruitment will also be placed on professional organizations for minorities. Professional education organizations expected to be used to recruit science panelists should include:

- National Science Teaching Association (NSTA)
- Association of Science Teacher Education
- National Middle Level Science Teachers

Professional history, civics, and social studies education organizations expected to be used to recruit panelists include:

- National Council for the Social Studies
- Center for Civics Education

General professional education organizations expected to be used to recruit panelists should include:

- Teach for America
- Teach Plus
- National Alliance of Black School Educators
- Association of Mexican American Educators

In addition to these organizations, state education organizations in science and history and civics, state superintendents and departments of education, school board presidents, and district and school administrators of public and private education entities will be contacted in the four NAEP regions to propose qualified nominators across both panelist types (classroom teacher and non-classroom teacher educator). Panelists from the operational study from the grade 8 panels for NAEP ALD review study for mathematics and reading will also be contacted for nominations, since they understand the expectations for the study and may be aware qualified candidates for the study.

Selection of Panelists

Nominees identified by the nominators will be asked to complete an online questionnaire regarding their qualifications and experiences for serving on the panel. Candidates that present the credentials required will be contacted by phone to collect any missing information, verify the information provided, and confirm their willingness to serve on the panel, if selected. The goal is

to select the most qualified panelists who are knowledgeable about the related subject at the appropriate level, while maintaining the goal of recruiting a mix of classroom teachers and non-classroom teacher educators for each panel with diverse demographics.

Panelists nominated in each panel must meet the following minimal qualifications.

Classroom Teacher Qualifications:

The nominee must meet the following qualifications:

- At least five years of overall teaching experience
- At least two years of experience teaching the related subject and grade
- Judged to be “outstanding” in their professional performance by a nominator

Non-Classroom Teacher Educator Qualifications:

The nominee must meet one of the following qualifications:

- Non-teacher educational staff within a school or district with education experience in the related subject and grade
- Curriculum director or content specialist serving schools at the related level or state department of education with education experience in the related subject and grade
- Postsecondary teacher education faculty teaching courses in the related subject and grade

The credentials of panelists will be evaluated based on the number and importance of the credentials that are presented, by giving each panelist a score that takes into account distinguishing credentials, professional activities at the state and national, years of experience teaching at the specific grade and subject, and other qualifications specific to the panelist type (classroom teacher or non-teacher educator). Nominees with the highest scores are given top priority by being considered the best-qualified candidates and being placed at the beginning of the candidate list. The selection process then identifies panelists to reach the targets for representation listed above, with nominees having the highest qualifications being the first

selected each time. All panels will be selected to have approximately equal proportions of males and females and equal proportions of representation from each of the four NAEP geographic regions. Every attempt will be made to create panels in which at least 25 percent of the panelists self-identify as a minority.

Each panelist for the pilot study and operational ALD review and revision meeting will be offered an honorarium. Pearson acknowledges that the funds available to offer panelists are not commensurate with their contribution. They will emphasize that panelists' participation in the ALD review and revision project represents an exceptional contribution to education in the United States. Additionally, all expenses related to the travel and lodging for the committee will be covered by Pearson, in accordance with federal travel regulations.

Section 2: Briefing Materials

Pearson will send access to a set of briefing materials to each confirmed panelist for their review and familiarization prior to the relevant panel meeting (pilot study/operational). The Pearson Standard Setting website, customized specifically for the project, will provide panelists with secure online access. The first time the panelist logs in to the website, he or she must read and electronically sign a nondisclosure agreement. Once signed, the panelist will be guided through a brief online training for using the website before having access to the non-secure advanced materials designated for the particular panel through links on the website. Panelists will use an online checklist to guide them through the online materials and will be able to check off each document after it has been reviewed. Documents will include the following:

- Confidentiality agreement
- Purpose and overview of the meeting
- Meeting roles and responsibilities
- Meeting agenda
- NAEP framework for the relevant assessment
- Other materials identified as appropriate

Communication with panelists will encourage them to review the briefing materials as those materials are intended to serve as a foundation for successfully carrying out the process designed for each panel. Pearson staff will be able to monitor panelist activities on the website to determine which materials a panelist has accessed.

Additionally, if needed, panelists will work with the Pearson team and NCES staff to complete the security requirements to access items through the NAEP Item Management

System (IMS). This will need to be completed prior to the first day of the meeting so panelists will be able to engage in the item review activity on the first day. Pearson project management will also work with the Governing Board Technical Point of Contact (TPOC) to collect the required non-disclosure agreements from panelists for access to review NAEP items for the associated grade and course.

Section 3: Pilot Study

From September 19 to 23, 2022, Pearson will conduct a pilot study over five days to implement the exact meeting procedure planned for the operational ALD alignment review and revision meeting. Conducting the pilot study at this time offers an opportunity to preview, revise, and resolve issues prior to the operational ALD alignment review and revision meeting planned for December 2022. To maintain uniformity of conditions, the pilot study and operational ALD alignment review and revision meetings will be held using the same process, including the same agenda, location, and procedures as those planned for the operational ALD alignment review and revision meeting. The pilot will inform whether any modifications are needed for the operational study.

The Technical Advisory Committee (TAC) will be consulted for advice and recommendations regarding details of the design prior to and after the pilot study. Up to two members of the TAC will be asked to observe the pilot study. All pilot and operational study materials will be based on data from the 2019 operational administration of the NAEP science assessment and the 2018 administrations of the U.S. History, and civics assessments (all of which are the most recent administrations). The pilot study has the following goals:

- Determine whether modifications for training, instructions, materials, timing, and logistics will be needed for the operational ALD alignment review and revision study.
- Provide an opportunity for facilitators to practice the process before moving to the operational ALD alignment review and revision meeting.

The initial process for the pilot meeting will be based on the one used during the operational meeting conducted for the ALD review meeting for NAEP mathematics and reading conducted in February 2022.

Given that all steps in the pilot study represent those planned for the operational ALD alignment review meeting, details on the process for both the pilot study and the operational meeting are provided under Section 4 of this document which clearly describes the operational ALD alignment review meeting.

A pilot study report will be prepared no later than October 2022, for presentation to COSDAM during a webinar. The timing of the pilot study allows for preparation of the report and review of the report by the TPOC and COR, COSDAM and TAC so that improvements to the process can be made in advance of the operational ALD review meeting.

Section 4: ALD Alignment Review Tasks and Procedures

Pearson will conduct an operational ALD alignment review meeting from December 5 to 9, 2022. To maintain uniformity of conditions, the operational ALD alignment review meeting will be held with the same agenda and procedures used for the pilot study, with the exception of adjustments made based on lessons learned from the pilot study process. This section describes the ALD alignment review procedures and tasks that Pearson will implement during both the pilot study and operational ALD alignment review.

The current plan is to conduct both the pilot and operational meetings in-person. Pearson is working with the Governing Board to identify a location that meets contractual requirements and the needs of the meeting. The purpose for conducting both meetings in the same location is to test the procedures and technology in a consistent setting.

The operational meeting will involve 24 panelists, with approximately eight per subject/grade. The panelists within each subject/grade will be assigned to two groups with approximately equal demographic representation, which will be used as replicate panels during the meeting. Each replicate panel will complete the same process, and receive training and modeling from the same lead facilitator; however, the group facilitation will be separate and conducted by different facilitators. Up to two members of the TAC will be asked to observe the meeting.

As previously indicated, Pearson will implement a model-based anchor approach and use the Pearson Standard Setting website platform to facilitate key aspects of the ALD alignment review process, including panelist training, review of ALDs, housing and reviewing of selected materials, recording of panelists' feedback and ratings during each round of review, provision of feedback, and evaluation of the ALD alignment review process. Access to viewing items in the

NAEP IMS for the NAEP Grade 8 Science assessment is also controlled through the Pearson Standard Setting website. The Pearson Standard Setting website will be set up to guide panelists through the steps of the ALD review process, with facilitators having the ability to restrict or provide panelists with access to sections of the site or activities, as needed.

To assist with panelists accessing the Pearson Standard Setting website during the meeting, the panelists will be assigned to specific computers at the meeting with their login information already entered. The night before the first day of the pilot study and the operational meeting, Pearson will reset the panelists' passwords to a common meeting password. Access to the sites is controlled by Pearson administrators. The sites are deactivated overnight and during any extended breaks during the day.

The model-based approach used in this study includes three stages. The first stage involves anchoring each item to a NAEP achievement level (*NAEP Basic*, *NAEP Proficient*, and *NAEP Advanced*). Additionally, an anchor set will be developed for items that map to the higher end of the region below *NAEP Basic*. Items that fall above *NAEP Advanced* will be classified as Does Not Anchor. This step occurs prior to the pilot ALD alignment review meeting. Pearson will conduct statistical analyses to determine the items that anchor to each achievement level. Images of the items associated with each achievement level in the anchor set will be uploaded to the Pearson standard setting website, along with links to access the items in the NAEP IMS for the NAEP Science assessment for grade 8. All operational materials will be based on data from the most recent administrations of the grade 8 NAEP science, U.S. history, and civics assessments.

The second stage relies on panels of content experts recruited to participate for each individual assessment. The panelists in each committee individually review the items that are

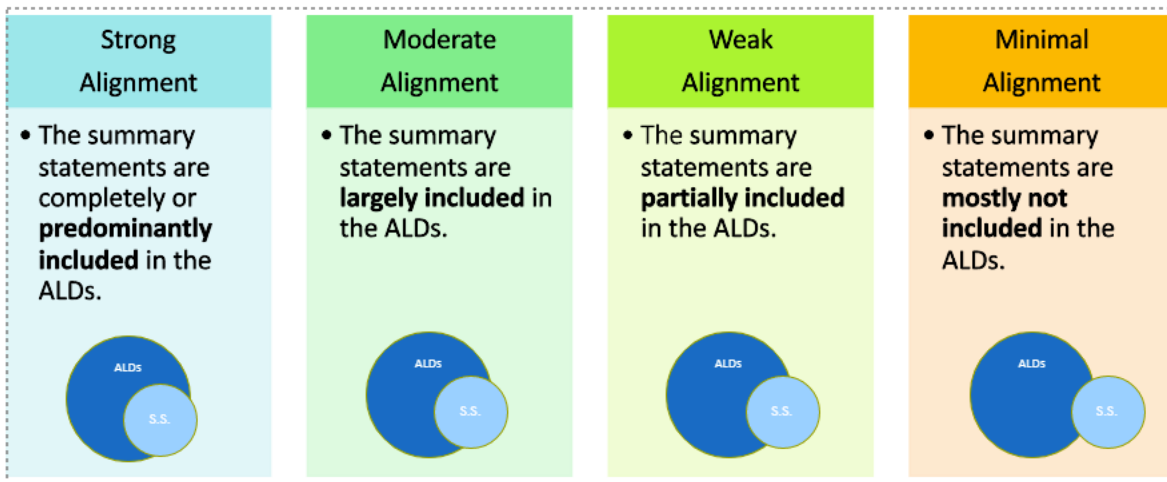
anchored to each performance level. While reviewing the items, they will be instructed to write a description of the knowledge/skills required to answer the item correctly or, on items worth multiple points, to receive a specific score point. The panelists will record their notes from their item reviews in individual panelist spreadsheets, accessed through the Pearson standard setting website. At the end of the individual item reviews for each achievement level, panelists will answer two question that will encourage them to think about the development of summary statements.

- Now that you have reviewed all the items for the achievement level anchor set, write a short summary of what students falling in the achievement level know and can do.
- How does what students know and can do in this achievement level differ from what students know and can do at the lower achievement level?

Next, group level summary descriptions are developed in each replicate group based on the individual item descriptions. The facilitator will guide the panelists within the replicate group they are working with to develop summary descriptions of what student in each achievement level know and can do, based on the individual item descriptions. Later in the process, the two replicate groups will work with the lead facilitator to combine the summary descriptions into final draft summary statements.

In the final stage, the panelists compare the current ALDs for the respective assessment with the summary descriptions, noting the similarities and differences, to make a recommendation regarding whether the current ALDs describe what students in each level know and can do. During the alignment judgment rounds, the panelists will classify the alignment of the ALDs for each achievement level into one of four classifications.

- Strong alignment: The summary statements are completely or predominantly included in the ALDs.
- Moderate alignment: The summary statements are largely included in the ALDs.
- Weak alignment: The summary statements are partially included in the ALDs.
- Minimal alignment: The summary statements are mostly not included in the ALDs.



The panelists will complete three judgment rounds with opportunities to discuss and revise their alignment ratings and the summary statements between rounds. Summary statements may be revised to improve clarity of the statements. The final version of the summary statement at the end of the process will establish draft reporting ALDs for each achievement level.

Computer Use and Software

Pearson will use computers during the pilot study and the operational ALD review meeting. The computer will be a laptop (provided by Pearson) used by panelists to access the online Pearson standard setting website interface. Panelists will also be provided an external monitor, which will provide them more screen area during the meeting to review materials and complete activities. As a part of supporting both efficiency and accuracy of the ALD review process, Pearson will monitor panelist progress and provide required access of materials

through to the Pearson standard setting website. The website can be accessed only through a user identification code (ID) assigned by Pearson. Permissions will be set up for each user ID so each panelist can access only the materials he/she will be using.

The use of the website will ensure the security of the materials during the standard setting meeting, since the system will be used throughout the ALD review process and requires a secure login by all users. Pearson designed the interface for the NAEP ALD review process to have the following features:

- Simultaneous access by multiple users, with each individual user assigned a profile which defines their level of access to the site, including the materials the user can access and the specific functionalities available. Facilitators and site administrators can use conditional access features based on date, activity completion, or custom settings to define access to materials and functionalities of the site.
- User access prior to the ALD review meeting, so users can interact with materials and experience training that will prepare them for the meeting and the ALD review process.
- Consistent user interface throughout the entire ALD review process, from pre-meeting work to post-meeting feedback.
- Management of storage and access to materials through the website, as a single access point for the meeting.
- Assignment of panelists to materials and items, based on panelists grouping.
- Embedded quality control features through conditional settings within activity that limit the range of responses to valid entries only and ensure no blank entries are accepted.
- Export of panelists' item review notes, as a group or individually. Facilitators will be able to view the panelists' notes within the website, to verify panelists' responses.

- Secure storage of all panelist information and judgments within the website, with access restricted to facilitators and site administrators.

Preparing for the ALD Alignment Review Meeting

This section provides further information related to the key activities Pearson will complete before the pilot and operational ALD alignment review meetings that contribute to the success of the meetings.

Development of Anchor Item Sets.

Prior to selecting the proposed anchoring approach outlined in this section, several variations were investigated to determine the effect of using different approaches as part of the development of the process used for the mathematics and reading study, including the anchoring method used during the 2009 NAEP anchoring studies, item response theory item mapping approaches, and the anchoring method used for the Trends in International Math and Science Study (TIMSS). The investigated methods and the data were shared and discussed with the TAC at the onset of the mathematics and reading study. Each method resulted in different anchor item sets, which resulted from different conceptualizations of the set of students used to calculate the conditional probability utilized to anchor the item to an achievement level. It was recommended by the TAC that the methodology from the previous alignment studies for mathematics and reading in 2009 (and specified by the Governing Board in the Achievement Levels Work Plan) would be the most defensible approach. This is because of the need to make consistent inferences about what the ALDs represent across the years by using the same anchoring methodology. That is, the anchor sets across years are created in the same manner and therefore the descriptions of the achievement levels represent the same range of students. This process will also be used for the current ALD review study for NAEP assessments in grade 8 science, civics, and U. S. History.

The development of the anchor item sets starts by grouping performances representing individual students from the most recent administration of the grade 8 NAEP science, U.S. History, and civics assessments into achievement levels. The achievement level classification for each student is based on the average of their NAEP “plausible values” and the relationship to the boundaries of the achievement levels for the respective assessment. A student will be classified into either *NAEP Basic*, *NAEP Proficient*, or *NAEP Advanced* if their mean plausible value is at greater than or equal to the cut score for the respective achievement level. A student is classified into the region just below *NAEP Basic* when their average plausible value is below the cut score for *NAEP Basic* and the region between the mean plausible value and two standard errors above the mean includes the cut score for *NAEP Basic*. This approach will use all students in the NAEP samples from the most recent administration of each assessment to ensure that there are sufficient students associated with each achievement level for the analysis to determine each anchor item set. This is an approach has been utilized in previous NAEP anchor studies.

After performance indicators for students are assigned to an achievement level, the conditional *p*-value, or probability of each student in that achievement level answering each item correctly, will be calculated using the IRT statistics from the most recent administration of the assessments. The conditional *p*-value for students across a given level will be averaged to derive the anchoring probability for that item or score point for multi-point items. Each item or score point will be assigned four conditional *p*-values, one each for below *NAEP Basic*, *NAEP Basic*, *NAEP Proficient*, and *NAEP Advanced*, which represent the average performance on the item of the typical student within the four achievement levels. Items will be anchored to the first achievement level where the conditional *p*-values for the achievement level are greater than or equal to 0.67.

Based on the anchoring criteria, items will be classified into one of five categories: (1) just below the *NAEP Basic* level, (2) *NAEP Basic* level, (3) *NAEP Proficient* level, (4) *NAEP Advanced* level, or (5) does not anchor. The items in the anchor item sets for the respective assessment will be grouped by content area. By reviewing the items within a content area, across all achievement levels, the panelists will be able to maintain a consistent focus on the knowledge and skills associated with the content area. For the civics U.S. history assessments, the items associated with a category will be ordered by achievement level from below the *NAEP Basic* level to the *NAEP Basic* level, to the *NAEP Proficient* level, and then finally the *NAEP Advanced* level. Within an achievement level, the items will be presented in decreasing order of conditional *p*-value, so the easiest item associated with the achievement level is first and the most difficult item is last. In this way, panelists will see a progression in what students know and are able to demonstrate while working through the items that anchor to that achievement level. For the science assessment, the items within a content area will be presented as sets with their associated stimuli. The item sets will be ordered by average *p*-value. The items within an item set will also be ordered by achievement level and by average conditional *p*-value.

Division of Panelists into Replicate Panels. In order to assess the degree of internal validity of the process, two replicate panels will be created within a panel. Approximately three to four panelists will be in each replicate panel. Pearson will assign panelists to a replicate panel with the intent of creating panels that are as equivalent as possible to one another. The purpose of the replicate panel is to be able to assess the degree of internal validity. In addition, a qualitative analysis will be conducted to assess the degree to which the summary descriptions created by each replicate panel vary.

Provision of Advanced Briefing Materials. As discussed in Section 3 Briefing Materials, panelists will have access to advance materials through the Pearson Standard Setting website. Pearson will send each confirmed panelist access to the materials for their review and familiarization prior to the operational meeting, including information to log into the website and change the assigned password. The first time the panelist logs in to the website, the panelist will be guided through a brief online training for using the website before having access to the non-secure advanced materials. Panelists will then use an online checklist to progress through the review and to ensure that each document is reviewed. Panelists will also have access to an orientation activity that includes an overview video as well as multiple engagement check-ins to gauge interaction and preparedness.

Training of Facilitators. The ALD alignment review and revision study will involve two content facilitators as well as a process facilitator per panel. There are three panels for each study, one panel for grade 8 science, one panel for grade 8 U.S. history, and grade 8 civics. The content facilitators are selected for their expertise and experience in instruction in the related subject and grade as the framework they bring. The process facilitator is selected for their expertise and experience conducting meetings. Recognizing that facilitators may introduce individual difference that can result in slightly different instructions, content and process facilitators will be properly trained to implement the process uniformly and as intended. Pearson will prepare the PowerPoint presentations that facilitators will use during the meeting. Facilitators will attend multiple virtual training sessions prior to both the pilot study and the operational meeting. The project director overseeing the activities will lead the training. In addition, the facilitators and project director will do a final review of the entire meeting the day before the pilot study and the day before the operational meeting.

Preparation of Pearson Laptops. Pearson will be providing each of the panelists a Pearson laptop that has been configured to have the appropriate software needed to access the NAEP items as well as the Pearson Standard Setting website. The laptops will be shipped to the meeting site and stored in a secure location when not in use. Additionally, each panelist will be provided an external monitor to ensure that they have the screen area to complete their work with minimal need to switch tabs or screens during the process.

Preparing Panelists for the ALD Alignment Review Process

Provide an Orientation. The pilot and operational meetings will have up to 24 panelists, with approximately eight panelists assigned to each panel. The three panels, along with the process and content facilitators, will begin the meeting with introductions and a description of the panelist recruitment process.

Provide an Overview of the Alignment Review and Revision Process. Facilitators will provide an overview of the purpose of the ALD alignment review and revision in general and describe of the process that will be used. Panelists will receive training in the key components of the NAEP framework for the related assessment and the rationale supporting these components. The administration and sampling used for NAEP administration will also be presented as background information. The process facilitator will describe the process that was used to create the anchor item set and an orientation to the information that is provided for each item. The panelists will also be trained in how the items are scored with the item key or scoring rubrics. We will also orient the panelists to the structure of the alignment judgment process, including the anchoring of items to specific achievement levels and the ordering of the items within achievement levels.

Individual Item Review and Summary Descriptions

Item Review. During the individual item review process, panelists will review all items within the anchor item set. Prior to panelists beginning the item review process, the facilitator will model how to use the website to review each item, view item information, and collect individual panelists' notes about each item using a practice anchor set. Additionally, the content facilitator will model to the panel how to approach the items in the anchor set and the process of developing item descriptions, with the focus of establishing a level of specificity in the item notes that will be useful during the creation of summary descriptions. The facilitator will also model for panelists how to convert the item description into a summary statement about what students know and can do for an achievement level.

The panelists will work independently to review each item and create a description of the knowledge and skills demonstrated by students who answered the item correctly or who received a particular score for constructed-response items. During this step, panelists will review items within the same content area. For civics and U.S. History, items will be grouped by increasing achievement level (including the items anchored to just below the *NAEP Basic* level) and conditional *p*-value, from easiest to hardest. For science, the item sets will be ordered by increasing difficulty (from easiest to most difficult, based on average conditional *p*-value) and all items within an item set will be presented to panelists before they review the next item set. Items within an item set will be ordered by increasing achievement level and conditional *p*-value. After the item sets, stand-alone items will be ordered by achievement level and from easiest to most difficult, within achievement level.

When panelists have completed their review of the anchor items for an achievement level, they will be asked to write a single set of summary statement about what students within the

three NAEP achievement levels (*NAEP Basic*, *NAEP Proficient*, and *NAEP Advanced*) know and can do, based on the item descriptions. Once panelists have completed their review of items across all achievement levels, they will be asked to consider what students in one achievement level can do compared to the adjacent achievement level(s). Panelists will complete the entire process for a content area before progressing to the next grouping of items by content. For science, after creating the summary statement for the science content areas, they will work to create performance level summary statements for the science practices.

Summary Descriptions. After individually writing descriptions for each item for a content area, the panelists will convene in their replicate panel to review the collected set of knowledge and skills associated with the items within an achievement level and construct a summary of what students performing in that content area and achievement level know and can do. (For science, the panelist's item descriptions will be sorted by achievement level and conditional *p*-value so that they can reference the items during the discussions). Prior to beginning the replicate panel development of summary descriptions, the facilitator will model the process of using the item descriptions to develop summary descriptions by achievement level. This process will start with the *NAEP Basic* level, then the *NAEP Proficient* level, then the *NAEP Advanced* level. They will create their summary descriptions for each achievement level, within a content area, on the website using a shared document. For Science, the panelists will create summary descriptions for each content area and the science practices. The summary descriptions will be captured by the content facilitator assigned to that replicate panel.

After each content area of individual review, they will meet in the replicate panels to create the summary descriptions for that content area. Once the replicate panel has completed the summary descriptions for a content area, they will then begin individual work for the next content area. When all content areas are completed, the panelists will progress to the alignment

judgment rounds with replicate panel discussions, whole panel discussions, and ALDs revisions, if needed. After the first round of individual alignment judgments and feedback discussion, the panels will work together to develop committee-level summary statements.

Alignment Judgment Rounds with Panel Discussions

The process of creating and reviewing ALDs is an iterative process, where individuals provide individual alignment judgments, followed by replicate panel discussions and whole panel discussions that are informed by results from the alignment judgments, resulting in more informed judgments in subsequent rounds. There will be three alignment judgment rounds. Prior to the first alignment judgment round, the content facilitator will provide the panelists training about the meaning of alignment between the summary statements and the policy ALDs as well as the current content ALDs.

Alignment Judgment Rounds. The replicate panels will be a significant part of the review and judgment process. To identify areas of alignment and lack of alignment, panelists will then work independently to make comparisons between the replicate panel summary descriptions with the policy descriptions and current ALDs. Panelists will be asked to assess the degree of alignment of the replicate panel summary descriptions by content area and achievement level to the policy descriptions and to the ALDs. For their alignment judgment to the ALDs, they will be prompted to provide suggestions/comments for their alignment judgment for each achievement level. The panelists will be asked to select a rationale for their alignment judgment, to check for the source of any lack of alignment.

- Summary statements and content ALDs were aligned (Only select for Strong judgment)
- Knowledge and skills in the summary statements associated with an achievement level are contained within a different achievement level in the NAEP content ALDs.

- Knowledge and skills in the summary statements are not included in the content areas

After the first round of individual alignment judgments, panelists will meet back in their replicate panels and discuss why they made their specific alignment judgments. If necessary, they can make adjustments to the summary descriptions to clarify or modify the language to better indicate the original intended meaning. The panelists will then proceed with another round of individual alignment judgments. They will then meet back as a whole panel and review summary statistics of the alignment judgments made and have a whole panel discussion similar to the replicate panel discussion. Where necessary, they will visit the summary descriptions to evaluate if the difference in the judgments relates to differences in the summary descriptions between the replicate panels. Panelists will be shown the summary descriptions from the two replicate panels side-by-side to evaluate the degree to which they are similar or different and develop a set of whole panel summary descriptions. The panelists will have one more opportunity to make individual alignment judgments based on the whole panel summary descriptions.

After the third round of individual alignment judgments, a whole panel discussion will take place. For the final whole panel discussion, panelists will be asked to review the final version of the summary statements for each achievement level and discuss any final revisions they would recommend to these statements. If there are any areas where they are more than 50% of alignment judgments that are minimal or weak, the facilitator will lead a discussion for the panelist to identify areas where the lack of alignment between the summary statements and content ALDs exists.

Process Evaluations. Procedural evidence refers to the appropriateness of the procedures and how well those procedures were implemented. Evidence for procedural validity may come

from a number of sources, including criteria for selecting panelists, the justification for the method, the quality of the implementation of the procedure, and the completeness of the documentation of the process. As another source of evidence of procedural validity, panelists will be asked to complete evaluation forms after each major activity of the process. Evaluations will include both selected-response and open-ended questions that address the panelists' understanding of the process and confidence in the results. No key-entry by staff is required because panelists will use the secure website interface to complete their evaluations. Panelist entries will be available for viewing using the facilitator login to the website. Facilitators will scan written responses for possible problems as they are collected during each day. Summary statistics will be computed for all ratings items and written responses. These analyses will be reviewed in real time throughout each day, and any sources of confusion will be identified for clarification with individual panelists or the panel as a whole. The operational meeting must be completed in time for the recommended ALDs to be used for reporting the next NAEP U.S. history and civics results at grade 8 in spring/summer 2023 and the next science results at grade 8 in 2025.

Section 5: External Feedback

Pearson understands the potential contribution of external feedback to the review and revision of ALDs for the NAEP science, U.S. History, and civics assessments. Pearson has designed a strategy for collecting and using feedback from content experts and NAEP stakeholders for the panel recommendations for revising the ALDs.

Content Review. After the meeting, the content facilitators will meet and complete a final review of the summary statements across achievement levels to address any areas where the consistency of language used could be improved, without changing the implied expectations

from the committee. The content facilitators from the meeting will be best suited to complete this review, since they know the related content along with the discussed expectations by the committee. This group will also format the summary statements as draft reporting ALDs in the format of a bulleted list. The initial statement in the list would be the same for each content area:

“Students in grade 8 performing at the NAEP <Achievement Level> when working within the content area can likely:”

Internal Review. Once the draft reporting ALDs have been created by the content facilitators, an internal review by NCES will be completed. The focus of this internal review is to ensure that there are no security issues with the reporting ALDs, where a statement would provide too much information about the assessment, and no consistency issues with the framework or assessment design. The comments from NCES will be focused on the following issues.

- Framework – notes an inconsistency with the NAEP content framework
- Assessment – notes an inconsistency with the skills measured in the assessment items
- Security – notes a potential security concern with the descriptions in the ALD
- Note – notes any other issues that could cause confusion with the interpretation of skills for external audiences

After the NCES review, content facilitators from the meeting will review the NCES feedback and make any revisions to the draft reporting ALDs, making sure that the implied expectations from the committee for each achievement level. There will be a reconciliation meeting with NCES content staff to discuss any final issues and determine any final edits.

External Review. Subsequent to the internal review, a focused external review will be completed. The purpose of the external review is to provide a subset of pre-identified individuals

with the appropriate experience the opportunity to review the draft reporting ALDs with regard to their clarity and usefulness. Participants in the external review will be provided the opportunity to complete an online survey after reviewing the draft reporting ALDs. For each grade and subject that they reviewed the draft reporting ALDs, they will answer two questions.

- Are the reporting ALDs clear?
- Can the achievement of students between the levels be distinguished from the reporting ALDs?

If the participant responds with a “No” for either question, they will provide responses to improve the reporting ALDs. The content facilitator will review the comments from the external review and make any necessary edits to the draft reporting ALDs.

Additionally, the panelists from the operational study will be provided the opportunity to review the draft reporting ALDs. The panelists will be sent the version of the draft reporting ALDs along with the final summary statement from the operational committee. The purpose of this review is to ensure that panelists do not perceive that the draft reporting ALDs have not changed from the committee summary statements significantly.

References

- Bourque, M. L., (2009). *A history of NAEP achievement levels: Issues, implementation, and impact, 1989-2009*. Washington, D.C.: National Assessment Governing Board.
- Donahue, P., Beaulieu, N., Freund, D., & Pitoniak, M. (2009). *Report on the 2009 Reading Achievement Level and Scale-Anchoring Study (Draft)*. Princeton, NJ: Educational Testing Service.
- Egan, K. (2021). *Reviewing achievement levels and cut scores*. Washington D.C.: National Assessment Governing Board.
- Loomis, S. (2016). *Anchor Studies for Analysis of NAEP Achievement Levels*. Prepared for presentation to an Expert Panel convened by HumRRO for the National Assessment Governing Board, Arlington, Virginia, on July 12–13, 2018.
- Mullis, I.V.S, Cotter, K.E, Centurino, V.A.S, Fisbein, B.G & Liu, J. (2016). Using Scale Anchoring to Interpret the TIMSS 2015 Achievement Scales. In M.O. Martin, I.V.S. Mullis, & M. Hooper (Eds.), *Methods and Procedures in TIMSS 2015* (pp. 14.1-14.47). Retrieved from Boston College, TIMSS & PIRLS International Study Center website: <http://timss.bc.edu/publications/timss/2015-methods/chapter-14.html>
- Pearson (2016). *Developing achievement levels on the National Assessment of Educational Progress in grade 8 technology and engineering literacy: Technical report*. Iowa City, IA: Author.
- Perie, M. (2018). *Validating achievement level descriptors*. Paper prepared for presentation to an Expert Panel convened by HumRRO for the National Assessment Governing Board, Arlington, Virginia, on July 12–13, 2018.

Achievement Levels Work Plan Status

The Board developed and adopted the [Achievement Levels Work Plan](#) in 2020 in response to the [2016 National Academies of Sciences, Engineering, and Medicine \(NASEM\) evaluation of NAEP achievement levels](#). In its evaluation, NASEM presented seven recommendations to inform future removal of the achievement level Trial status¹ and to increase evidence that the NAEP achievement levels are reasonable, valid, and informative.

The Achievement Levels Work Plan outlines proposed activities to address each NASEM recommendation. Most of these actions are presented as the responsibility of the Governing Board; though it also calls out activities under the purview of the National Center for Education Statistics (NCES) the Board identified as relevant to the recommendations. The Work Plan also indicates the proposed timeline based on the intended assessment schedule at the time of development; however, the 2021 assessments were delayed by one year due to the COVID-19 pandemic and, likewise, achievement levels activities were delayed. The documentation below presents the current progress of all activities identified. The original timeline can be found in the Work Plan document.

The first set of proposed activities in the Work Plan address recommendations 1 – 3. The table below presents the progress to date on those activities. The activities identified as Governing Board responsibilities have been conducted or are underway.

Table 1. Achievement Levels Work Plan activities associated with NASEM recommendations 1 – 3.

Proposed Activity	Responsibility	Progress
COSDAM approval of Achievement Levels Procedures Manual (described in policy statement)	COSDAM	Completed in June 2020. Available here .
Conduct studies to examine and/or document alignment between NAEP Math and Reading Frameworks and item pools for grades 4, 8, 12	NCES	For NCES' consideration
Conduct studies to review and revise Math and Reading ALDs at grades 4, 8, and 12.	NAGB	Math and Reading study concluding
Conduct studies to review and revise U.S. History, Civics, and Science ALDs at grade 8		U.S., History, Civics, and Science project started in May 2022 with expectation for completion in spring of 2023

¹ Removal of the Trial status of NAEP achievement levels is at the discretion of the National Center for Education Statistics (NCES) Commissioner. The Commissioner has indicated additional activities beyond those outlined in the NASEM evaluation will be required for the removal. COSDAM leadership and the Commissioner will discuss next steps for moving forward.

Full Board action on revised Reporting ALDs	NAGB	Action for Math and Reading set for August 5, 2022 Action for U.S. History, Civics, and Science expected in May, 2023
Conduct studies to examine and/or document alignment between NAEP U.S. History, Civics, Science, and TEL Frameworks and item pools for grade 8	NCES	For NCES' consideration

The second set of proposed activities address NASEM recommendation 4. These activities are presented in Table 2 and focus on the utility of linking studies to demonstrate NAEP's relationship with measures of student preparedness for various academic and life outcomes. These activities are the focus of the recently established Linking Studies Working group composed of COSDAM and Reporting and Dissemination Committee (R&D) members.

Table 2. Achievement Levels Work Plan activities associated with NASEM recommendation 4.

Proposed Activity	Responsibility	Progress
Review of technical memo on various ideas (including pros/cons) for synthesizing and representing findings about how other assessments and external indicators of student performance relate to NAEP (including a summary of existing linking studies) and what the findings mean for NAEP.	NAGB	Efforts are ongoing through the Linking Studies Working Group
As the Governing Board works to develop its next Strategic Vision, deliberations will take place as part of that effort to determine how to approach the goal of making NAEP more relevant by connecting NAEP results to important real world indicators of student achievement.	NAGB	The Strategic Vision 2025 includes linking studies as a priority Efforts are ongoing through the Linking Studies Working Group to consider how to make NAEP more relevant through linking
Decision on additional studies that should be pursued to connect NAEP to other assessments and external indicators of student performance	NAGB/NCES	This will be discussed in the Linking Studies Working Group

The final set of activities outlined in the Achievement Levels Work Plan are presented in Table 3. These activities address NASEM recommendations 5 and 6, with a focus on the need to better

understand intended and unintended uses of achievement levels and to communicate their proper use with the public. These activities have not yet started and are primarily the responsibility of the Governing Board. During the August COSDAM meeting, the Chair will facilitate a discussion regarding how to proceed with starting up these activities.

Table 3. Achievement Levels Work Plan activities associated with NASEM recommendations 5 - 6.

Proposed Activity	Responsibility	Progress
Convene ongoing advisory group to discuss and provide feedback on the development of materials for communicating NAEP achievement levels	NAGB/NCES	Not yet started
Collect information about current uses of NAEP achievement levels via focus groups and evaluate appropriateness of interpretations and uses that are not directly intended	NAGB	Not yet started Plan to begin with informal discussions with CCSSO and TUDA task force members in Fall of 2022
Adopt statement of intended purpose and meaning of NAEP	NAGB	Intended Meaning adopted by Board in March, 2020.
Improve communications of what NAEP frameworks and achievement levels represent	NAGB/NCES	Some initial efforts to develop infographics
Develop and finalize interpretative guide for NAEP achievement levels; iterative drafts will be discussed by COSDAM and R&D	NAGB	Not yet started
Collect and document validity evidence to support intended interpretations and uses of NAEP achievement levels • Collect and summarize validity evidence to support intended interpretations and uses of NAEP scale scores	NAGB NCES	Some validity evidence has been collected, may want to consider additional collection; have not yet formally summarized findings