



Achievement Level Description Review for the National Assessment of Educational Progress Grade 8 Science, U.S. History, and Civics Assessments

Submitted: June 2023

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

Achievement Level Description Review for the National Assessment of Educational Progress Grade 8 Science, U.S. History, and Civics Assessments

Final Process Documentation and Technical Report

June 2023

Eric L. Moyer
Jennifer Galindo
Pearson



National Assessment Governing Board

BOARD MEMBERSHIP (2022–2023)

Honorable Beverly Perdue, Chair
Former Governor of North Carolina
New Bern, North Carolina

Representative Alice Peisch, Vice Chair
Massachusetts House of Representatives
Wellesley, Massachusetts

Honorable Haley Barbour
BGR Group Founding Partner
General Public Representative

Alberto Carvalho
Superintendent
Miami-Dade County Public Schools
Miami, Florida

Tyler Cramer
CEO and Executive Manager
Remarc Associates LLC
San Diego, California

Christine Cunningham
Professor of Education and Engineering
College of Education
The Pennsylvania State University
University Park, Pennsylvania

Frank Edelblut
Commissioner
New Hampshire Department of
Education
Concord, New Hampshire

Viola Garcia
Local School Board Member
Aldine Independent School District
Houston, Texas

Eric (Rick) Hanushek
Hanna Senior Fellow
Hoover Institution
Stanford, California

Patrick Kelly
Director of Governmental Affairs
Palmetto State Teachers Association
Columbia, South Carolina

Anna King
General Public Representative
Oklahoma City, OK

Suzanne Lane
Professor of Research Methodology
University of Pittsburgh
Pittsburgh, Pennsylvania

Scott Marion

President and Executive Director
The National Center for the
Improvement of Educational
Assessment, Inc.
Dover, New Hampshire

Reginald McGregor

Manager, Engineering Employee
Development & STEM Outreach
Rolls Royce Corporation
Indianapolis, Indiana

Jon T. Pickinpaugh

Secondary School Principal
UMÓⁿHOⁿ Nation Public School
Macy, Nebraska

Michael A. Pope

Eighth-Grade Teacher
Zama American Middle School (DODEA)
Camp Zama, Japan

Julia Rafal-Baer

Managing Partner & Co-Founder
ILO Group
Cranston, Rhode Island

Ron Reynolds

Executive Director
California Association of Private
Organizations
Van Nuys, California

Nardi Routten

Fourth-Grade Teacher
Creekside Elementary School
New Bern, North Carolina

Dilhani Uswatte

Elementary School Principal
Rocky Ridge Elementary School
Hoover, Alabama

Martin R. West

Massachusetts Board of Elementary
and Secondary Education
Professor of Education
Harvard Graduate School of Education

Representative Mark White

Tennessee House of Representatives
Nashville, Tennessee

Grover (Russ) Whitehurst

Professor Emeritus
Stony Brook University
Fort Myers, Florida

Carey M. Wright

State Superintendent
Mississippi Department of Education
Jackson, Mississippi

Ex-officio Member**Mark Schneider**

Director, Institute of Education Sciences

Table of Contents

Executive Summary	1
Background.....	1
Technical Advice.....	3
Study Process.....	4
Panelist Recruitment and Selection	4
Creation of Anchor Item Sets	5
ALD Review Model-Based Approach Methodology.....	6
Study Outcomes	7
Validity Evidence	8
Introduction.....	9
Background on NAEP Achievement Level Descriptions	9
Background on Current Project.....	10
Purpose and Organization of the Document.....	12
Technical Advice.....	12
Project Staff	13
General Procedures Applied to the Pilot Study and the Operational ALD Review Study.....	14
Recruitment and Selection of Study Panelists	14
Identifying Nominees	15
Selection of Panelists.....	16

Preparation of Panelists	17
Development of Anchor Item Sets.....	18
ALD Review Study Process.....	21
Orientation to the ALD Review Study	21
Individual Assessment Review.....	22
Summary Statement Development	22
ALD Alignment Review Judgments	25
Process Evaluations	28
Pilot ALD Review Study	29
Panelists.....	29
Facilitators	32
Observers.....	33
ALD Review Pilot Study.....	33
Results.....	34
Process Evaluations	34
Lessons Learned.....	36
Operational ALD Review Study.....	39
Panelists.....	39
Facilitators	43
Observers.....	44
Adjustments to the Process for the ALD Review Operational Study	44
Independent Item Review	45

Group and Panel Summary Statement Development	45
Alignment Judgments	45
Results	46
Process Evaluations	59
Validity Evidence	63
Recommendations for Future Studies	65

Table of Tables

Table 1. Gender, race, and ethnicity distribution for each operational panel.....	5
Table 2. Round 3 grade 8 alignment judgment agreement with content ALDs	8
Table 3. Achievement level counts for Grade 8 Science.....	19
Table 4. Achievement level counts for Grade 8 U.S. History	20
Table 5. Achievement level counts for Grade 8 Civics.....	20
Table 6. Panelist type distribution for each panel	30
Table 7. Gender, race, and ethnicity distribution for each panel	30
Table 8. Distribution for each panel across geographic regions	31
Table 9. Experience of panel members with student populations.....	31
Table 10. Facilitators and technical assistants for each panel	32
Table 11. Panelists' response to the statement: "I understood the steps to follow as I completed the independent item review activity."	34
Table 12. Panelists' response to the statement: "I understood the steps to follow as I completed the individual alignment judgment activity."	35
Table 13. Panelists' opinion regarding their confidence in the summary statements by the panel	35
Table 14. Panelists' opinion regarding their confidence in the ALD alignment judgments by the panel.....	36
Table 15. Pilot study lessons learned and plans of action.....	37
Table 16. Panelist type description for each panel	39
Table 17. Gender, race, and ethnicity distribution for each panel	41
Table 18. Distribution for each panel across geographic regions.....	41
Table 19. Experience of panel members with student populations	42
Table 20. School/District Location.....	42
Table 21. School/District Size	43
Table 22. School/District Student Populations	43
Table 23. Facilitators and technical assistants for each panel	44
Table 24. Science Grade 8 – Reporting ALDs.....	48
Table 25. U.S. History Grade 8 – Reporting ALDs	51
Table 26. Civics Grade 8 – Reporting ALDs.....	55
Table 27. Round 3 alignment judgment agreement with achievement level policy definitions.....	58
Table 28. Round 3 alignment judgment agreement with content ALDs.....	58
Table 29. Purpose of the ALD alignment judgment activity was clearly explained	59
Table 30. Explanation of the ALD alignment judgment options.....	60

Table 31. Rationale for misalignment judgment.....	60
Table 32. Steps to follow to complete the alignment judgment activity	60
Table 33. Use of feedback data between judgment rounds	61
Table 34. Creation of panel summary descriptions.....	61
Table 35. Group discussions	61
Table 36. Amount of time provided for ALD alignment judgments.....	62
Table 37. Confidence in the panel's achievement level summary descriptions.....	62
Table 38. Confidence in the panel's alignment judgments.....	63
Table 39. Validity evidence for NAEP ALD Review study	64

Table of Figures

Figure 1. Sample Item Review Spreadsheet	23
Figure 2. ALD Alignment Judgment Training Slide.....	26

Executive Summary

The National Assessment Governing Board (the Board) contracted with Pearson to design and implement a review of the achievement level descriptions (ALDs) for National Assessment of Educational Progress (NAEP) Grade 8 assessments in Science, U.S. History, and Civics. This document describes the procedural and technical aspects and outcomes of the operational ALD Review study. This study commenced after Pearson had completed an ALD Review study for Mathematics and Reading in 2022, which itself was undertaken on behalf of the Board for the reasons noted in the Background section of this report.

In particular, this report addresses the Board's updated achievement levels policy that called for the development of Reporting ALDs, which state what student performance associated with each NAEP achievement level likely can demonstrate related to the assessment content, and how these align with the existing content ALDs included in the frameworks and achievement level policy definitions.

Background

The Board has a legislatively mandated responsibility to develop NAEP achievement levels. The Board's [Policy Statement on Developing Student Achievement Levels for the National Assessment of Educational Progress](#) provides policy definitions of *NAEP Basic*, *NAEP Proficient*, and *NAEP Advanced* and describes the principles for setting achievement levels. The policy definitions are general, high-level expectations of what students should know and be able to do and are consistent across all NAEP assessments and grade levels assessed. The same general policy definitions apply to all NAEP assessments, regardless of subject and grade. Content achievement level descriptions (ALDs) are specific descriptions of what students at each level should know and be able to do for each individual assessment as specified in the frameworks (for example, see page 128 of the current [Science Framework](#)). The Achievement Levels Procedures Manual further describes details for implementing the Board policy.

The Board first established the achievement levels policy in 1990 with the expectation that, in addition to scale scores, reporting should include the percentage of test takers at each defined level and those falling below the *NAEP Basic* level. As part of the NAEP

reauthorization in 1994, Congress stipulated the achievement levels be designated as trial until the NCES Commissioner determines, as the result of an evaluation, that the achievement levels are reasonable, reliable, valid, and informative to the public.

In 2016, the National Academies of Sciences, Engineering, and Medicine (NASEM) conducted the [Evaluation of the Achievement Levels for Mathematics and Reading on NAEP](#). In it, they acknowledged the value of the NAEP achievement levels: “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.” They made recommendations to enhance their utility, including:

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.

In response to the recommendations presented by the National Academies and updated guidance on achievement level setting, the Board updated its achievement level policy in November 2018 with guidance to develop new achievement level descriptions of what students can do based on student NAEP performance. These descriptions would be used in reporting to help increase the utility of NAEP data and are referred to as Reporting ALDs. The Board approved an Achievement Levels Work Plan in 2020 to address the recommendations in the evaluation.

In September 2020, the Board awarded a contract to Pearson to address the National Academies recommendation and updated Board policy for mathematics and reading. That study used the 2019 Reading and Mathematics NAEP assessments at grades 4, 8, and 12.

For the second phase, the Board contracted with Pearson to study the ALDs for three subjects in grade 8: Science, U.S. History, and Civics.

The primary outcomes of this study were a) the development of Reporting ALDs based on assessment items and data, and b) comparison of the Reporting ALDs to the content ALDs as validity evidence. The methodology used was specified in the Board's Achievement Levels Work Plan and was similar to what was done to evaluate the alignment and revise the 2009 NAEP Reading ALDs for grades 4, 8, and 12 (Donahue, Pitoniak, & Beaulieu, 2010), the 2009 NAEP Mathematics ALDs for grade 12 (Pitoniak, Dion, & Garber, 2010), and the most recent study by Pearson for Mathematics and Reading ALDs, completed in Spring of 2022. The Board's Committee on Standards, Design and Methodology (COSDAM) oversaw this work from the onset, and a Technical Advisory Committee (TAC) with six experts in achievement levels and ALDs participated in about 100 hours of discussions to provide technical guidance on all phases of the work.

Technical Advice

The Board policy on developing and reviewing achievement level descriptions for NAEP requires appointment of a committee of technical advisors who have expertise in achievement level descriptions and psychometrics in general, as well as issues specific to NAEP. These advisors served on a TAC that was convened for several meetings throughout the project to provide advice. They provided feedback on plans and materials before activities were implemented and reviewed results of the process and analyses.

In addition to the members of the TAC, Dr. Rebecca Norman Dvorak, Assistant Director for Psychometrics and Technical Point of Contact (TPOC), provided technical advice to Pearson throughout the project, participated in all TAC meetings, and attended all panel meetings. Plans for the studies and all results were presented to the Board's Committee on Standards, Design and Methodology (COSDAM) during each quarterly Board meeting and through conference calls.

Study Process

The study involved convening panels of teachers and non-teacher educators with content expertise in U.S. history, civics, or science to review items, develop summary statements indicating what students know and can do as evidenced by correctly responding to the items, and then compare the statements to the existing content ALDs and provide alignment judgments. A pilot study was conducted in September 2022 to test all aspects of the logistical design of the ALD review procedures. The operational study was held in December 2022, with a different set of panelists than the pilot. The operational study resulted in two sets of outcomes – draft Reporting ALDs and final alignment judgments of Reporting ALDs to the achievement level policy definitions and to the content ALDs.

Panelist Recruitment and Selection

Panelist recruitment involved multiple steps, designed to obtain broadly representative, well-qualified panelists familiar with the knowledge, skills, and abilities needed by students in grade 8, in science, U.S. history, or civics. The panels for both the pilot and operational ALD Review studies were recruited from across the nation.

All panelists were required to have direct experience with students at the eighth-grade level and with the relevant subject area, science, U.S. history, or civics. Both current classroom educators and non-classroom educators, such as curriculum coordinators and instructional coaches, were recruited for participation in this study. Recruitment also focused on the requirement that panelists' demographics should be representative by geographical region, gender, and race/ethnicity.

A multiphase process was used, which was focused on identifying and contacting qualified nominators, collecting and reviewing nominees, notifying nominees and collecting nominee information, and selecting and recruiting the sample of nominees to serve as panelists. Overall, there were a total of 16 panelists for the pilot study and 23 panelists for the operational ALD Review study. Table 1 summarizes the demographic information for panel members who participated in the operational study.

Table 1. Gender, race, and ethnicity distribution for each operational panel

Characteristic	Science	U.S. History	Civics
Gender			
Female	7	6	4
Male	2	2	1
Other	0	0	1
Race			
Asian	0	0	0
Black/African American	1	0	1
American Indian, Alaska Native	0	0	0
White	8	8	3
No Response	0	0	2
Ethnicity			
Hispanic/Latino	0	0	0
No Response	0	0	2

Creation of Anchor Item Sets

For the ALD review pilot study, Pearson used a three-stage model-based approach for reviewing the alignment of the ALDs for NAEP Science, U.S. History, and Civics. The first stage involved conducting statistical analysis to determine the items from the subject and grade that are anchored to each achievement level. These anchored item sets were developed using the item banks for each NAEP assessment from the 2019 administration for NAEP Science and the 2018 administration for NAEP U.S. History and NAEP Civics. The anchored item sets provided panelists with information about which items and related content or processes students with ability associated with a NAEP achievement level likely could demonstrate. Additional information about how the anchored items sets is provided later in the report.

For each subject the items were anchored to one of the NAEP achievement levels, *NAEP Basic*, *NAEP Proficient*, and *NAEP Advanced*, or below *NAEP Basic* or Does Not Anchor. The items were organized into relevant domain categories and then ordered within categories from the easiest item to the most difficult item. For NAEP Science, there were a total of 290 items; for NAEP U.S. History, there were 255 items; and for NAEP Civics, there were a total of 197 items.

ALD Review Model -Based Approach Methodology

The method used for this study was similar to the method used for other anchoring studies, as requested by the study design. Panelists' activities during both the pilot and operational studies followed the same general process for ALD review. Initially the panelists received training on the general ALD review process and the NAEP framework for their respective subject and grade. The panelists then received specific training on the process for conducting the individual item review, including a modeling activity to assist panelists in using this procedure. After the training, panelists conducted the independent item review for the items in each content domain. The independent item review process included reviewing each individual item and providing a description of the knowledge and skills represented by student performance in response to the item. At the end of the independent item review activity, panelists were asked to write a summary description of what students at each achievement level know and can do, based student performance on the items anchored to the achievement level.

Following the review of the items for a single domain category, panelists met in groups to draft group summary descriptions. The group summary descriptions were initially created in separate replicate groups, so there were two separate sets of summary descriptions for each domain category. Later in the process, the replicate group summary descriptions were combined into a single set of summary descriptions that all panelists agreed represented the demonstrated achievement for each level. The panel summary descriptions were considered the initial draft Reporting ALDs.

Finally, panelists provided judgments regarding the alignment between what panelists determined students could demonstrate in relation to the NAEP assessment, as defined by the panel summary descriptions, and what students should know and be able to do, as defined by the content ALDs in the NAEP frameworks. Panelist alignment judgments were restricted to the following options:

- 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.

Panelists were also asked to provide rationale for their alignment judgments. Panelists completed three rounds of alignment judgments with the opportunity to discuss their rationale between rounds and refine their summary statements to clarify the level of achievement demonstrated at each level.

Process evaluations were completed throughout the ALD review process. The evaluations included both selected-response and open-ended questions that addressed several aspects of the ALD review process.

- Clarity of the overview and purpose of the ALD Review study
- Understanding of the NAEP assessment
- Clarity of training and instructions in the ALD review process and tasks
- Confidence in the process and results

After the operational study, the draft Reporting ALDs went through additional reviews. The content facilitators from the operational ALD Review study conducted a review to ensure consistency in format and language within the draft Reporting ALDs across the achievement levels. An internal review of the draft Reporting ALDs was also completed by NCES content staff, with the primary purpose to ensure that the statements in the draft Reporting ALDs did not compromise item security. Simultaneously, Board staff and Board communications contractors reviewed the draft Reporting ALDs to offer feedback on clarity for use with a wide audience. Also, a review was completed by two content experts per subject area who were familiar with the NAEP frameworks to offer feedback on language, clarity, and sensitivity concerns. These reviews resulted in some final adjustments to the statements. The modified Reporting ALDs were sent to panelists to ensure that the modifications did not change the intended meaning.

Study Outcomes

The final results of the operational ALD Review study were presented at the March 2023 meeting of the Board. Table 2 presents the results of the alignment judgment round 3 survey for the three grade 8 subjects. The results presented provide evidence that the panelists observed some alignment between the knowledge and skills students demonstrated in an achievement level, described by the summary statements, and the expected knowledge and skills for an achievement level, described by the content ALDs in the framework, based on the majority of judgments indicating either Moderate or Strong alignment.

Table 2. Round 3 grade 8 alignment judgment agreement with content ALDs

Subject	NAEP Level	Alignment Judgment			
		Minimal	Weak	Moderate	Strong
Science	Basic	0%	33%	33%	33%
	Proficient	0%	11%	77%	11%
	Advanced	0%	11%	44%	44%
U.S. History	Basic	0%	0%	75%	25%
	Proficient	0%	0%	62%	38%
	Advanced	0%	0%	100%	0%
Civics	Basic	0%	0%	83%	17%
	Proficient	0%	0%	17%	83%
	Advanced	0%	0%	50%	50%

The final Reporting ALDs are provided in later sections of the report.

Validity Evidence

For procedural validity evidence of the study, a design document was constructed to describe the procedures and process of the ALD Review study. This document, reviewed by various individuals, served as the guide for implementing the ALD Review study. During the operational study, there were two TAC members to observe the process and report the fidelity of the study design to the process used during the study. The observations from the TAC members stated that the process followed the design document, and any deviations were minor to the overall validity of the study. Additionally, process evaluations were used throughout the ALD study to provide procedural evidence for the validity of the study. The results of these process evaluations will be discussed further in the report.

For internal validity evidence, the panelist level of agreement with the draft Reporting ALDs and alignment judgment provide evidence for internal consistency. Also, the change in alignment judgments between judgment rounds provides additional evidence for internal consistency of the results from the ALD Review study.

Introduction

Background on NAEP Achievement Level Descriptions

The National Assessment of Educational Progress (NAEP) assesses and reports the educational achievement for student groups in terms of both numerical scale scores and the percentages of students at or above the NAEP achievement levels. The National Center for Educational Statistics (NCES) develops the numerical scale scores for each NAEP subject. These scale scores communicate the degree to which students have mastered the content assessed by NAEP, with higher scores indicating greater levels of mastery.

The National Assessment Governing Board (referred to hereafter as the Board) is responsible for the development of achievement levels for NAEP. To help define the meaning of the achievement levels for NAEP, the Board has established general policy definitions for *NAEP Basic*, *NAEP Proficient*, and *NAEP Advanced*. The same general policy definitions apply to all NAEP assessments, regardless of subject and grade. The specific achievement level descriptions for each subject and grade assessed by NAEP are found in the NAEP assessment frameworks and reports.

As part of the Board's policy on Developing Student Achievement Levels for NAEP, Principle 1a states: "Content achievement level descriptions translate the policy definitions into specific expectations about student knowledge and skills in a particular content area, at each achievement level, for each subject and grade. Content Achievement Level Descriptions (ALDs) provide descriptions of specific expected knowledge, skills, or abilities of students performing at each achievement level. Content ALDs reflect the range of performance that items and tasks should measure. During the achievement level setting process, the purpose of content ALDs is to provide consistency and specificity for the panelists' interpretations of policy definitions for a given assessment."

The final report from the evaluation of the NAEP achievement levels for mathematics and reading by the National Academies of Sciences, Engineering, and Medicine included seven recommendations. The first recommendation from the report was to review the alignment among the achievement level descriptions and the cut scores for the achievement levels.

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.

Additionally, the third recommendation from the report was also related to the NAEP ALDs.

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 the recommendations, the Board adopted a comprehensive [Achievement Levels Work Plan](#) in 2020. The purpose of this plan was to provide details concerning how each of the seven recommendations from the evaluation would be addressed.

In response to the first recommendation, the Board issued a procurement for conducting studies to ensure that the current NAEP mathematics and reading ALDs at all three grade levels align with the knowledge and skills of students in each achievement level category as measured by the assessment items. That study was completed early in 2022 and generated new Reporting ALDs that comply with the 2018 Board policy statement. The current study is a follow-on to that study, focusing on NAEP grade 8 assessments in science, U.S. history, and civics.

Background on Current Project

The Board signed a contract with Pearson in 2022 to design and implement a study procedure to conduct anchoring studies using the most recent NAEP data to review the grade 8 Science (2019), U.S. History (2018), and Civics ALDs (2018). There were two primary goals of the study.

1. Create Reporting ALDs that describe what students performing at each achievement level know and can demonstrate, reflecting empirical evidence of the knowledge, skills, and abilities demonstrated within each achievement level. The Reporting ALDs will be used to report grade 8 results of the 2022 NAEP U.S. History and Civics assessments and the 2024 Science assessment.
2. Review the alignment between these assessment items that anchor to each achievement level range and the ALDs that describe what students should know and be able to demonstrate for each achievement level.

Pearson staff designed and implemented studies to test and refine procedures for the NAEP ALD studies, including the anchored item response theory (IRT) approach used to establish the association between the items in the item bank and the achievement levels. Throughout the process, Pearson staff worked with a Technical Advisory Committee (TAC) composed of testing and measurement experts to help ensure that the procedures were psychometrically sound and could be implemented with a representative set of panelists from a variety of backgrounds. In addition to guidance from the TAC, Pearson staff provided briefings and updates to the Board's Committee on Standards, Design and Methodology (COSDAM). Throughout the process of designing and implementing the ALD review procedures and preparing reports, COSDAM monitored activities and provided general guidance and direction regarding the conduct of the work and offered recommendations for the full Board to consider.

The methodology used for the NAEP ALD Review study had to conform to the Board's policy on Developing Student Achievement Levels for NAEP, especially as it applies to the review and revision of ALDs and development of Reporting ALDs, and similar to what was done to evaluate the alignment and revise the 2009 NAEP Reading ALDs for grades 4, 8, and 12. For the NAEP ALD Review Study for Reading and Mathematics, conducted in 2022, a methodology was selected so that the resulting Reporting ALDs would comply with the Board policy statement. For the current study for NAEP Civics, U.S. History, and Science for grade 8, the same methodology was used, to ensure consistency between the studies.

This ALD study was convened in Austin, Texas. Each panelist accessed documents and materials through the Pearson Standard Setting website. This allowed panelists to gain an understanding of the assessment and the students' experiences. Through each step of the ALD review process, the panelists accessed the items, documents, and activities through physical copies as well as the online platform.

Purpose and Organization of the Document

This document provides a detailed description of the ALD review process implemented by Pearson to develop Reporting ALDs for three grade 8 NAEP assessments and results from the alignment review. This will serve as the primary source of all information for all components of that process and all outcomes.

This document is organized to first provide context for the ALD review process. It will give details about the operations and procedures used by describing the activities that were part of the design and development of the ALD review procedures and provide information on support provided by the TAC for the ALD review process. These parts are followed by sections providing in-depth information on each of the studies conducted as part of the overall ALD review, including the pilot study and the operational ALD Review study. Descriptions of the ALD studies are followed by detailed information on the outcomes of the ALD review process and on technical procedures conducted. The document closes with information on Board actions and Pearson's recommendations for future studies.

Technical Advice

The project for reviewing the grade 8 NAEP ALDs for Science, U.S. History, and Civics required the appointment of a committee of technical advisors who have expertise in achievement level descriptions and psychometrics in general, as well as issues specific to NAEP. These advisors served on a TAC for the NAEP ALD Review study. The TAC convened for several virtual meetings to provide advice at every key point in the process. They provided feedback on plans and materials before activities were implemented and reviewed results of the process and analyses. The discussions with the TAC were summarized for each meeting and recommendations were noted.

Plans for the NAEP ALD Review study and all results were presented to the Board's COSDAM during quarterly Board meetings and scheduled virtual meetings between August 2022 and March 2023. Besides the members of the TAC, Dr. Rebecca Norman Dvorak, the Board's Assistant Director for Psychometrics and Technical Point of Contact (TPOC), provided technical advice to Pearson throughout the project, participated in all TAC meetings, and attended all panel meetings.

The names of the experts in standard setting who served on the TAC are shown below.

- Dr. Karla Egan, Founder, EdMetric, LLC
- Dr. Ellen Forte, CEO & Chief Scientist, edCount, LLC
- Dr. Susan Loomis, Former Technical Consultant for the Board and Assistant Director for Psychometrics for the National Assessment Governing Board
- Dr. Marianne Perie, President, Measurement in Practice, LLC
- Dr. Mark Reckase, University of Distinguished Professor Emeritus, Michigan State University and former member of the National Assessment Governing Board
- Dr. Laurren Wise, Retired Principal Scientist, HumRRO

Note that Dr. Egan, Dr. Forte, and Dr. Perie have extensive experience designing and conducting standard setting and alignment workshops. Dr. Loomis and Dr. Reckase have been heavily involved in NAEP standard setting and alignment work, and Dr. Egan and Dr. Wise also served on the National Academies of Science, Engineering, and Medicine committee that conducted the most recent evaluation of the NAEP achievement levels.

Project Staff

Dr. Eric L. Moyer, Principal Research Scientist at Pearson, served as the project director for the NAEP ALD review for the U.S. history, science, and civics project. The assistant director for the project was Dr. Jennifer Galindo, Senior Research Scientist at Pearson. Other members of the leadership team for the project included Kevin Baker as the program manager and Julie Downey as the senior project manager, both of whom were responsible for logistics. The lead content facilitators for the project were Kadie Patterson for civics, Rachel Williams for U.S. history, and Michaela Viering for science. The lead content facilitators acted as content experts and ensured that the ALD review process was implemented as described. The specific meeting facilitators and moderators during the pilot and operational studies will be provided in the descriptions for each meeting.

General Procedures Applied to the Pilot Study and the Operational ALD Review Study

The sections below provide descriptions of each component of the NAEP ALD Review study process Pearson applied across both the pilot and operational studies. Variations from the general descriptions are addressed under each study description.

Recruitment and Selection of Study Panelists

Pearson implemented a multi-step panelist recruitment plan, which resulted in 16 panelists for the pilot study and 23 panelists for the operational study. Panelists were divided based on their education and experience into six replicate groups across three panels, that each focused on completing the ALD review process for a specific subject.

- Grade 8 Science
- Grade 8 U.S. History
- Grade 8 Civics

The target maximum composition of the study was a total of 24 panelists, with eight panelists for each subject-specific panel.

The recruitment plan followed the same process for both the pilot and operational studies with a focus on securing broadly representative, well-qualified panelist groups that reflected an overall balance of gender, race/ethnicity, geographic location, and classroom experience. Recruitment efforts were undertaken with the goal of securing a panel composed of both classroom teachers and non-classroom educators (e.g., state or district curriculum coordinator). For each panel, the recruitment effort had a goal of securing a panel for which at least half of the panelists were classroom teachers and at least two of the panelists were non-classroom educators.

Pearson identified the panelists through an iterative multiphase process focused on identifying and contacting qualified nominators; collecting and reviewing nominees;

notifying nominees and collecting nominee information; and selecting and recruiting the sample of nominees to serve as panelists.

Identifying Nominees

Panelist nominators were recruited using multiple sources. One source for nominators was professional organizations that have strong backgrounds in providing professional development in education for the related subjects and supporting educators. The following organizations were among those targeted for recruiting panelists:

- National Science Teaching Association
- Association of Science Teacher Education Regional Groups
- National Middle Level Science Teachers
- National Council for the Social Studies
- National Association of Independent Schools
- National Alliance for Public Charter Schools
- Teach Plus
- National Alliance of Black School Educators
- Association of Mexican American Educators
- Asian Educators Alliance
- Teach for America

In addition to these organizations, staff from state departments of education, teacher organizations, and other education entities were contacted in the four NAEP regions to propose qualified nominees. Nominating individuals and organizations were asked to provide nominations for two panelist types: classroom educators and non-classroom educators. The specific qualifications for each panelist type will be described later, but classroom educators are individuals who are currently teaching the related subject in a classroom, whereas a non-classroom educator is an individual that is engaged with the related subject in a non-classroom role (e.g., curriculum specialist).

Nominators were asked to complete an online form regarding individuals they believed met the qualifications for participating in the ALD Review study. Each nominator was allowed to nominate multiple individuals so they could nominate qualified individuals for each subject or panelist type. For each nominated individual, the nominator provided information concerning which panel the individual was nominated for and their rationale for the nomination.

Selection of Panelists

Nominees were asked to complete an online form regarding their qualifications and experiences for serving on the panel. Additionally, nominees indicated their availability to participate in either the pilot or operational study. The goal was to select the most qualified candidates for each panel, while maintaining a combination of classroom teachers and non-classroom educators and maximizing the representativeness of each panel.

Nominees recruited for each panel met the following qualifications:

Classroom Teacher:

- At least five years of overall teaching experience; and
- At least two years of recent experience teaching the respective subject at the specific grade level.

Non-Classroom Educator:

- Non-teacher school staff with education and/or experience in the respective subject area at the specific grade level; or
- Curriculum director or content specialist serving school or state department of education with education and/or experience in the respective subject area at the specific grade level; or
- Postsecondary teacher education faculty who teach courses in the specific subject area with focus on appropriate grade level.

Pearson project staff evaluated potential panelists based on the number and importance of their professional credentials presented in each panelist's informational survey. For each pilot and operational meeting, the selection process then chose candidates in an attempt to create panels that were representative of educators across the country. While the goal of the selection process was to have an approximately equal proportion of males and females and representation from each NAEP region in each panel, this was not possible given the distribution of qualified panelists.

For the pilot and operational ALD review studies, panelists were provided an honorarium of \$500 each. School districts were reimbursed for the cost of substitute teachers. Pearson acknowledged that the honorarium provided to panelists was not commensurate with their contribution and emphasized to panelists that their participation in the NAEP ALD Review study represented an exceptional contribution to

education in the United States. Also, all travel expenses the panelists were covered according to federal government guidelines.

Preparation of Panelists

Because the panelists had access to secure information and NAEP items during the ALD review meeting, they were required to complete several non-disclosure agreements. Panelists were also provided with the necessary credentials to access the various systems to participate in the meeting.

Prior to the meeting in Austin, panelists were provided access to the standard setting website and instructed to complete some pre-meeting work to prepare them for the meeting. The pre-meeting work included the following:

- An overview video that provided information about the purpose and process of the NAEP ALD Review study
- An agenda of activities for the 5 days of the meeting
- The applicable NAEP framework for their respective subject for the most recent administration

Panelists were sent emails during the week prior to the meeting to encourage them to complete the pre-meeting work and to remind them of the logistics of the meeting in Austin, Texas.

Development of Anchor Item Sets

Pearson used a three-step model-based approach for reviewing the alignment of the ALDs for the three grade 8 NAEP subjects. The Board's Achievement Levels Work Plan indicated that the methodology for these studies should be similar to previous ALD development and review studies held in 2009 (Donahue, Pitoniak, & Beaulieu, 2010; Pitoniak, Dion, & Garber, 2010) to reduce the potential for possible inconsistencies from the use of different methods.

The approach included three stages. The first stage involved conducting statistical analysis to determine the items from each subject and grade assessment that were anchored to each achievement level. It was recommended by the TAC that the methodology from the previous NAEP ALD Review studies for mathematics and reading completed in 2022 would be the most defensible approach.

The development of the anchor item sets starts by grouping performances of 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 performance is based on the average of their NAEP plausible values.¹ A student's performance was classified into either *NAEP Basic*, *NAEP Proficient*, or *NAEP Advanced* if their mean plausible value was greater than or equal to the cut score for the respective achievement level and less than the cut score for the next achievement level. Student performance was also classified into the region below *NAEP Basic* when their average plausible value was below the cut score for *NAEP Basic*. This approach used all students in the NAEP sample from the most recent administration to ensure that there were sufficient student responses associated with each achievement level for the analysis to determine each anchor item set.

After performance indicators for students were assigned to an achievement level, the conditional p -value, or probability of each student in that achievement level answering each item correctly, was calculated using the IRT statistics from the most recent administration of the assessments. The conditional p -value for students across a given level was averaged to derive the anchoring probability for that item or score point for

¹ Plausible values are proficiency estimates for an individual NAEP respondent, drawn at random from a conditional distribution of potential scale scores for all students in the sample who have similar characteristics and identical patterns of item responses.

multi-point items. Each item or score point was 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 three NAEP achievement levels and below *NAEP Basic*. Items were anchored to the lowest achievement level for which the average conditional p -values for the achievement level were greater than or equal to 0.67. Items that did not anchor to any achievement level, because their average conditional p -value for any achievement level did not meet or exceed the 0.67 criteria, were classified as Does Not Anchor. An item discrimination criterion was not used to anchor items to achievement levels, based on recommendations from the TAC.

Based on the anchoring criteria, items were classified into one of five categories: (1) below *NAEP Basic* level, (2) *NAEP Basic* level, (3) *NAEP Proficient* level, (4) *NAEP Advanced* level, or (5) Does Not Anchor. For items with a score point greater than 1, each possible non-zero score value was anchored to one of the five categories, so the item would appear in the item list one time for each possible non-zero score value. The items in the anchor item sets for the respective assessment were grouped by domain for each subject. By reviewing the items within a content area or passage type, across all achievement levels, the panelists were able to maintain a consistent focus on the knowledge and skills associated with the content area. Tables 3 through 5 present the number of items anchored to each achievement level.

Table 3. Achievement level counts for Grade 8 Science

Achievement Level	Science Content Domains			Total
	Earth and Space	Life	Physical	
Does Not Anchor	22	18	7	47
<i>NAEP Advanced</i>	28	19	32	79
<i>NAEP Proficient</i>	44	30	32	106
<i>NAEP Basic</i>	9	14	7	30
Below <i>NAEP Basic</i>	11	10	7	28
Total	114	91	85	290

Table 4. Achievement level counts for Grade 8 U.S. History

Achievement Level	U.S. History Content Domains				Total
	Peoples, Cultures, and Ideas	American Democracy	Economic and Technological Changes	Role of America in the World	
Does Not Anchor	14	12	8	4	38
<i>NAEP Advanced</i>	12	22	11	8	53
<i>NAEP Proficient</i>	38	27	19	22	106
<i>NAEP Basic</i>	12	9	13	7	41
Below <i>NAEP Basic</i>	5	4	6	2	17
Total	81	74	57	43	255

Table 5. Achievement level counts for Grade 8 Civics

Achievement Levels	Civics Content Domains					Total
	Constitution and American Government	Foundations of American Political System	Civic Life, Politics, and Government	Roles of U.S. Citizens	U.S. and World Affairs	
Does Not Anchor	8	6	2	4	3	23
<i>NAEP Advanced</i>	12	12	8	9	7	48
<i>NAEP Proficient</i>	16	18	13	14	9	70
<i>NAEP Basic</i>	7	9	5	7	5	33
Below <i>NAEP Basic</i>	6	4	5	4	4	38
Total	59	49	33	39	28	305

For each subject, the items associated with a content area were 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. The items classified as Does Not Anchor were included so the panelists could determine distinctions of what students with performance at the *NAEP Advanced* level would be able to demonstrate. Within an achievement level, the items were in decreasing order of conditional p -value, so the easiest item associated with the achievement level was first and the most difficult item was last. In this way, panelists would 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, there was one set of items from the Earth and Space domain that were presented within a scenario-based task. This type of item set was presented within a context or scenario and stimuli were presented progressively throughout the set of items. When panelists reviewed the items for the Earth and Space domain, they completed a review of items associated with the scenario-based task, where the items were presented in the administration order.

ALD Review Study Process

The three-stage model-based approach, mentioned previously, was implemented for both the pilot and operational studies. The first stage, which was described above, involved conducting statistical analysis to determine the items from the subject and grade that are anchored to each achievement level. The following sections provide a general description of the training, ALD Review study activities, and feedback that was used for the pilot and operational studies.

Orientation to the ALD Review Study

The meeting began with a general session, where Dr. Rebecca Norman Dvorak, a staff member of the Board, presented to all the panelists an overview of the NAEP Assessment program and Dr. Eric Moyer presented an overview of the ALD review along with security and meeting ground rules.

Following the general session presentation, the panelists were then split into subject-specific breakout groups. Each of the subject-specific breakout groups had a lead content facilitator and an assistant content facilitator. The role of the lead facilitator was to provide training to the overall panel and lead a single group. The role of the assistant facilitator was to assist the lead facilitator during the panel activities and to lead the second group in the group summary discussion. The lead content facilitators for each subject provided training on the key components of the NAEP framework for the relevant assessment. This involved a discussion of the specific item types, and the content included within each subscale.

Individual Assessment Review

Following the overview of the NAEP framework, panelists were provided an opportunity to review a set of items that a student might see during a NAEP administration. The purpose of the assessment review activity was to give the panelists the opportunity to acquaint themselves with a sample of items that was assessed during the most recent administration of NAEP for that subject at grade 8.

The panelists accessed a link on the Chromebook provided for their work in this study that took them out to the NAEP Item Management System (IMS) system, where they were able to navigate through the set of items and view them in the digital format administered to students.

Summary Statement Development

The second stage of the model-based approach, defined earlier, results in a set of draft Reporting ALDs that describe the knowledge and skills likely exhibited in performance associated within each achievement level. The process of creating these draft Reporting ALDs included the following steps:

- Independent item review
- Group summary statements
- Panel-level summary statements

The following sections describe each of these activities.

Independent Item Review

The lead facilitator described the process that was used to create the anchor item sets and an orientation to the information that is provided for each item. The panelists were also trained in how the items are scored with the item key or scoring rubrics. Panelists were oriented in 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. The facilitator modeled the item review process showing how panelists could access the individual item review spreadsheets, assess the passages for reading only, enter item descriptions, and the summary statements for review. Figure 1 contains sample item review spreadsheets.

1 Below

For each of the items in the below **NAEP BASIC** anchor set, write a description of what students know and can do related to the knowledge and skill assessed by the item.

Anchor Pos.	Item ID	Points	Max Points	Item Description
1.1		1	2	
1.2		2	2	
1.3		1	1	
1.4		1	2	
1.5		1	1	
1.6		1	1	

Figure 1. Sample Item Review Spreadsheet

Following the training, panelists conducted the independent item review. This activity was completed one subscale at a time. The subscales were based on the content domains for each subject as defined in the NAEP framework for the subject assessment. To complete the activity, panelists accessed the anchor item sets for the specific subscale via the Pearson website. The sets were organized by achievement level and item difficulty. The Pearson website contained screenshots of each item; in addition, a link to access the item in the NAEP IMS system was provided. Panelists accessed a spreadsheet that contained the item identifiers and subsequent metadata. In the spreadsheet, they were asked to make notes about the knowledge and skills necessary to answer each of the items correctly or to receive the indicated score point.

On the website, panelists were provided the items as screenshots, which were made available by NCES. Along with the item, panelists were provided either the item key or access to the scoring guide for the item, and information to access the item in the NAEP IMS. As with the independent item review, the panelists were able to access the NAEP IMS on their Chromebook. For each item anchored to a NAEP achievement level, panelists were asked to provide a description of the knowledge and skills required for a student to provide a correct response or earn the specific score point, for multi-point items. The items anchored to below *NAEP Basic* and Does Not Anchor were provided for panelists to review to ensure accurate descriptions for the NAEP achievement levels.

For each NAEP achievement level, panelists were asked to write a summary description of what students at that achievement level know and can do based on the items they reviewed. They were also asked to specify what students know and can do at that level and how that differs from the lower adjacent achievement level. The purpose of these

descriptions was to help facilitate the development of a summary description individually prior to the group discussion. Approximately halfway through their review of the items for the first domain, panelists were given training about completing this individual summary activity for each achievement level.

Group-level Summary Statement Development

Following the review of the items for a single subscale or passage type, panelists met in their replicate groups to draft group summary descriptions. The lead and assistant facilitators led their own groups through a discussion about what students know and can do based on the items they reviewed, as well as the summary descriptions they wrote for each achievement level. The facilitators captured the group summary descriptions in a spreadsheet. A set of summary descriptions was developed for each group. The group summary descriptions were used to make round 1 alignment judgments, which will be described later in this report.

The summary statements were to indicate the knowledge and skills that students with performance associated with the achievement level would likely be able to demonstrate. So that the summary statements were based on sufficient evidence, the group was instructed to make summary statements when a minimum of two or more items represented similar content within the domain. The two items used as evidence for the summary statement could either be within the same achievement level or across achievement levels, to differentiate demonstrated abilities across levels. The requirement of having two supporting items aimed to ensure that the summary statements were not overly influenced by idiosyncrasies that could affect item difficulty. The facilitators supported the panelists as they grouped items representing knowledge and skills from similar content and generalized summary statements. Summary statements that were supported by only single items did not contribute to the overall summaries. The creation of the summary statement for the first subdomain required additional time to assist the groups in developing the skill to create summary statements at an appropriate level of generalization.

Panel-level Summary Statement Development

A major outcome of the ALD Review study was draft summary statements that could be used as draft Reporting ALDs to describe the range of knowledge and skills that students at each achievement level likely can demonstrate in relation to the NAEP

assessments. Individual groups within each subject panel created summary statements, which were based on the results of their review process and group discussions. After the round 1 alignment judgments, the lead facilitator led the panel through a discussion to combine the summary statements developed by the two groups into a single set of summary statements for the panel at each grade and subject. The panel worked together to review the group summary statements and worked to create a single set of summary statements that the entire panel agreed represented the demonstrated achievement for the level, within the subscale. Between alignment judgment rounds, which will be discussed next, panels were provided the opportunity to review and provide edits to the summary statements that clarify what achievement is expected within a level.

ALD Alignment Review Judgments

In the third stage of the model-based approach used during the ALD Review study, the panelists completed three alignment judgment rounds, comparing the current content ALDs from the NAEP framework document for the respective assessment with the drafted summary descriptions. The purpose of this alignment review was to evaluate the alignment between what the panelists determined students could demonstrate in relation to the NAEP assessment, as defined by the panel summary descriptions, and what students should know and be able to do, as defined by the ALDs in the NAEP framework. The policy definitions for the NAEP Achievement Levels (*NAEP Basic*, *NAEP Proficient*, and *NAEP Advanced*) provide high-level expectations of what students should know and be able to do. The content ALDs included as part of the NAEP frameworks for each subject area communicate descriptions of specific expected knowledge, skills, and abilities of students performing at each achievement level.

Prior to starting their individual alignment judgments, the panelists were provided training in how they would record their judgments and the criteria for each rating. Their alignment judgments were restricted to the following options:

- 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.

Figure 2 displays the PowerPoint slide that was used to discuss the alignment categories with the panelists.

ALD Alignment Judgment Options

For each achievement level, the alignment between the expectations defined by the ALDs and demonstrated performance defined by the summary statements will be judged as one of the following:

Strong Alignment	Moderate Alignment	Weak Alignment	Minimal Alignment
<ul style="list-style-type: none"> • The summary statements are completely or predominantly included in the ALDs. 	<ul style="list-style-type: none"> • The summary statements are largely included in the ALDs. 	<ul style="list-style-type: none"> • The summary statements are partially included in the ALDs. 	<ul style="list-style-type: none"> • The summary statements are mostly not included in the ALDs.

92

Figure 2. ALD Alignment Judgment Training Slide

The panelists were instructed that they would complete three individual judgment rounds, with opportunity for discussion between the judgment rounds. The focus of the discussion between rounds was not to improve the alignment judgment ratings but to improve the coherence of their understanding of the alignment categories and the rationale for their judgments. After each judgment round, panelists are provided feedback data based on the judgment agreement for the round and the opportunity to discuss their judgments and rationale before making their alignment judgments during the next round.

Round 1

The first round of judgments focused on the alignment between the group summary descriptions and the content ALDs. The round 1 judgments were using the group summary descriptions, since that was what they had been working on up to this point in the process. Panelists were first asked to rate the level of alignment between the group summary descriptions and the policy definitions by achievement level. They

were then asked to rate the level of alignment between the group summary descriptions and the content ALDs by achievement level. They were also asked to provide written rationale for their alignment judgment.

After round 1 judgments, the panelists reviewed the feedback data and discussed their alignment judgments in their groups. This discussion provided the panelists the opportunity to discuss their perspectives of the alignment classifications and summary descriptions. During the judgment round, the groups had the opportunity to revise and clarify their group summary descriptions to improve their understanding and use of them.

After the group discussions, the panelists came together as a single panel and reviewed the judgment agreement between the groups and the summary descriptions between the two groups. The facilitator led the panelists through a discussion to combine the group summary descriptions into a single panel set of summary descriptions.

Round 2

The second round of judgments focused on the alignment between the panel summary descriptions with the policy definitions and the content ALDs. As with the first judgment round, panelists were first asked to rate the level of alignment between the panel summary descriptions and the policy definitions by achievement level. They were then asked to rate the level of alignment between the panel summary descriptions and the content ALDs by achievement level. They were also asked to provide written rationale for their alignment judgment.

After round 2 judgments, the panelists reviewed the feedback data and discussed their alignment judgments in the panel. This discussion provided the panelists the opportunity to discuss their perspectives of the alignment classifications and the panel summary descriptions. The groups then had the opportunity to further revise and clarify their panel summary descriptions to improve their understanding and use during the judgment round.

Round 3

The third round of judgments again focused on the alignment between the panel summary descriptions and the content ALDs. During the preparations for this round, the facilitator reminded the panelists that this would be the final judgment round and would be presented to the Board. The panelists made their third-round judgments following the same process used during the second judgment round. They were also asked to provide written rationale for their alignment judgment.

After this round of judgments, the panelists were presented with the final alignment judgments by the panel, but there was no specific discussion about the rationale for their judgments since there were no additional judgment rounds. The panelists were also provided the opportunity to complete a final review of the panel summary descriptions and recommend any final adjustments before they were recommended as the draft Reporting ALDs from the panel.

Process Evaluations

The validity of the outcomes of the ALD Review study depends, in part, on the evidence of the procedural validity of the process implemented. One source of evidence of procedural validity for the ALD Review study results from process evaluations given to panelists at key points in the process. The questionnaires included both selected-response and open-ended questions that addressed the panelists' understanding and evaluation of the instructions, tasks, and materials, as well as their comfort level with the process and confidence in the results. The process evaluations were completed using the website interface.

Pilot ALD Review Study

This section of the report provides general results for the pilot study and describes only those features of the pilot study that differ from the procedures described under the section on general procedures. Pearson designed the pilot study to test all tools, processes, and procedures planned for the operational ALD Review study.

Panelists

Though the goal was to include 24 panelists, only 16 individuals nominated indicated that they were available to participate in the pilot study. We reviewed the qualifications of each of the possible panelists with the TPOC. The review of the panelists' experience and qualification was to determine if there were any possible conflicts with other NAEP work and that their qualifications and experiences would match the needs of the study. The decision was to invite all 16 panelists to participate in the pilot study.

There were five or six panelists per subject assessment. The participants in each panel were divided into two replicate groups, to verify that the process worked for separate groups and to observe any differences in outcomes between groups, because each group reviewed the same set of items. Each replicate group was selected to have both teachers and non-classroom educators, along with representation across other classifications.

The representation of panelist type for the pilot study was more distributed to the non-classroom educators than planned. This was attributed to the difficulty of teachers taking a week off, since it was challenging to find substitutes for schools, a problem that began with the COVID-19 pandemic and that has continued into the 2022–2023 school year. Additionally, the timing of the pilot study was close to the beginning of the school year, which was also a deterrent for panelists participating in the meeting. Table 6 summarizes information about the panelist type for panel members who participated in the pilot study.

Table 6. Panelist type distribution for each panel

Type	Current Role in Education		
	Science	U.S. History	Civics
Classroom Teacher	3	3	4
Non-Classroom Educator	2	3	1
Total	5	6	5

Table 7 summarizes the gender and ethnicity distributions for the panel members who participated in the pilot study.

Table 7. Gender, race, and ethnicity distribution for each panel

	Gender and Ethnicity		
	Science	U.S. History	Civics
Gender			
Female	3	3	3
Male	2	3	2
No Response	0	0	0
Race			
Asian	0	1	0
Black, African American	1	1	0
American Indian, Alaska Native	0	0	0
White	4	3	3
Two or More Races	0	0	1
No Response	0	1	1
Ethnicity			
Hispanic/Latino	0	0	0
No Response	0	0	0

There was a desire to have representation across the different NAEP regions. Although that representation did not exist in each individual panel, across all panels there was

representation from each region. Table 8 summarizes the distribution of the panel members who participated in the pilot study across the geographic regions.

Table 8. Distribution for each panel across geographic regions

Region	Geographic Region		
	Science	U.S. History	Civics
Northeast	2	1	0
Midwest	1	0	1
South	1	0	0
West	1	5	4

Table 9 summarizes the education experience with students from special populations for the panel members who participated in the pilot study.

Table 9. Experience of panel members with student populations

Experience with Student Population	Panel		
	Science	U.S. History	Civics
General Education	5	6	5
English Language Learners	3	4	3
Mainstream Special Education	5	4	4
Self-contained Special Education	1	2	3
Gifted and Talented Education	3	6	4

Facilitators

Each of the grade and subject groups was facilitated by a lead content expert and a supporting content expert. Both content experts for a panel had multiple years of experience with the respective content. The lead facilitator was responsible for conducting all panel-level training, and both facilitators ensured that appropriate processes were followed. The technical assistants helped with technology issues. The content facilitators and technical assistant for each panel are shown in Table 10.

Table 10. Facilitators and technical assistants for each panel

Content Area	Lead Facilitator	Assistant Facilitator	Technical Assistant
Science	Michaela Viering	Ashtyn Greenstein	Brian Wrobel
U.S. History	Rachel Williams	Meredith Hohe	Ebony Gaines
Civics	Lydia Mantis	Judy Goss	Grant Smith

Weekly 90-minute facilitator trainings for the pilot study were held over the course of six weeks. All of these meetings were conducted virtually. The extensive trainings included:

- Use of the Pearson meeting website – because the Pearson website was used as a facilitation tool during the meeting, facilitators needed to become familiar with the use of the platform. Specific guidelines for providing access to the panelists were discussed as well as how panelists would interact and access the anchor item sets.
- Use of the NCES IMS system – Facilitators were trained in how to access the IMS system as well as the NAEP lockdown browser.
- NAEP assessment overview and NAEP framework – Facilitators were provided with an overview of the NAEP assessment program and detailed information about the NAEP frameworks.
- ALD review meeting process – The facilitators participated in a walk-through of the ALD review meeting agenda, with a focus on specific issues such as time management, the use of the online platform, capturing panelist discussions, and communicating feedback information.
- Presentation slides and script – As part of the walkthrough of the ALD review process, the facilitators also reviewed the slides. The script provided along with the presentation slides offered facilitators guidance throughout the presentation, including when specific language was to be used during the panelist training and use of the Pearson website. Although the presentations

were based on the presentations for the ALD Review Study for mathematics and reading, these were adjusted for the specific subjects for this study.

Observers

In addition to the facilitators, there were three observers of the pilot study. They included two members of the TAC, Dr. Karla Egan and Dr. Susan Loomis. There was also the Assistant Director for Psychometrics and TPOC for the Board, Dr. Rebecca Norman Dvorak.

ALD Review Pilot Study

The model-based ALD review process that was described in the design document was successfully implemented for the pilot study.

The agenda that was originally developed for the pilot study required modification. Some activities took more time than initially anticipated, especially for science, requiring a decrease in allotted time for some activities, specifically the individual alignment judgment activity. The agenda for the operational ALD Review study was adjusted based on information collected during the pilot study.

At the Austin meeting site, the panelists engaged in meaningful discussions throughout the ALD review process and worked collectively to address areas of disagreement and come to a working agreement on critical tasks, especially when working on the development of the draft Reporting ALDs. During the pilot meeting, the panelists in the Civics and U.S. History panel engaged in some discussions related to the age of the NAEP frameworks for these content areas and how instructional practices have changed since they were developed.

Results

The purpose of the pilot study was to implement the exact meeting procedures for the planned operational ALD Review study, to offer an opportunity to preview, revise, and resolve any issues prior to the operational meeting related to the collection of the two intended outcomes of the study:

1. Draft Reporting ALDs that describe what students within each achievement level demonstrate based on evidence from an actual administration of the assessment
2. Judgments of the alignment between the draft Reporting ALDs generated by the panelists about what students can do and the policy definitions and current content ALDs based on the NAEP frameworks that describe what students at each achievement level should know and be able to do

Process Evaluations

As stated earlier, process evaluations were administered during different parts of the process. Panelists' responses to these questions suggest that most panelists understood the process for reviewing the items, creating the summary descriptions, and completing the alignment judgment. This is indicated by the distribution of panelists choosing "Agree" or "Strongly Agree" in the evaluation survey, as seen in Tables 11–12.

Table 11. Panelists' response to the statement: "I understood the steps to follow as I completed the independent item review activity."

Subject	Strongly Disagree	Disagree	Slightly Disagree	Slightly Agree	Agree	Strongly Agree
Science	0	0	1 (20%)	0	2 (40%)	2 (40%)
U.S. History	0	0	1 (17%)	2 (33%)	3 (50%)	0
Civics	0	0	1 (20%)	0	2 (40%)	2 (40%)

Table 12. Panelists' response to the statement: "I understood the steps to follow as I completed the individual alignment judgment activity."

Subject	Strongly Disagree	Disagree	Slightly Disagree	Slightly Agree	Agree	Strongly Agree
Science	0	0	1 (20%)	0	2 (40%)	2 (40%)
U.S. History	0	0	1 (17%)	2 (33%)	3 (50%)	0
Civics	0	0	1 (20%)	0	2 (40%)	2 (40%)

Most panelists also indicated that they were at least somewhat confident in the panel-level summary descriptions and ALD alignment judgments, as shown in Tables 13 and 14.

Table 13. Panelists' opinion regarding their confidence in the summary statements by the panel

	Responses	Not Confident at All	Slightly Confident	Somewhat Confident	Completely Confident
Science	NAEP Basic	0	1 (20%)	3 (60%)	1 (20%)
	NAEP Proficient	0	1 (20%)	3 (60%)	1 (20%)
	NAEP Advanced	0	0	4 (80%)	1 (20%)
U.S. History	NAEP Basic	0	0	5 (83%)	1 (17%)
	NAEP Proficient	0	1 (17%)	4 (67%)	1 (17%)
	NAEP Advanced	1 (17%)	1 (17%)	3 (50%)	1 (17%)
Civics	NAEP Basic	0	0	2 (40%)	3 (60%)
	NAEP Proficient	0	0	1 (20%)	4 (80%)
	NAEP Advanced	0	0	2 (40%)	3 (60%)

Table 14. Panelists' opinion regarding their confidence in the ALD alignment judgments by the panel

	Responses	Not Confident at All	Slightly Confident	Somewhat Confident	Completely Confident
Science	NAEP Basic	0	0	3 (60%)	2 (40%)
	NAEP Proficient	0	0	3 (60%)	2 (40%)
	NAEP Advanced	0	0	3 (60%)	2 (40%)
U.S. History	NAEP Basic	0	1 (17%)	4 (67%)	1 (17%)
	NAEP Proficient	0	2 (33%)	3 (50%)	1 (17%)
	NAEP Advanced	1 (17%)	2 (33%)	2 (33%)	1 (17%)
Civics	NAEP Basic	0	1 (20%)	2 (40%)	2 (40%)
	NAEP Proficient	0	0	3 (60%)	2 (40%)
	NAEP Advanced	0	0	3 (60%)	2 (40%)

Lessons Learned

The purpose of the pilot study was to try out all tools and procedures developed for the operational meeting scheduled for December 5–9, 2022. As noted, there were no insurmountable issues encountered, but several opportunities for improvement were identified and revisions were made in preparation for the operational meeting. These are presented in Table 15.

Table 15. Pilot study lessons learned and plans of action

Meeting Segment	Lesson Learned	Plan of Action
Pre-meeting	We were able to identify 24 qualified panelists who were willing and able to take part; however, not all were ultimately able to attend.	<p>Communicate with panelists who could not participate in pilot meeting to recruit for operational meeting.</p> <p>Complete recruiting process by November 14, so panels can be selected and locked in time to send schedule travel by November 30.</p> <p>NAGB has given approval to increase the number of attendees per panel to 10, with the hope that at least 8 will attend the meeting in Austin.</p>
	Some panelists in science had difficulty seeing the scenario based task (SBT) items in IMS.	Panelists in science will continue to have access to IMS; however, the facilitator will lead the panelists through a group review of the two SBTs.
	There were some inconsistencies in how facilitators led panelists through some discussions.	Facilitator training prior to the meeting will be used to model specific activities to ensure consistency in procedures utilized.
General Activity Agenda	The time allotted for completing several of the activities was inadequate and will need to be modified.	The agenda will be revised to make time for the item review activity and the group and panel summary statement creations.
General Session and Orientation	Panelists indicated they would have benefited from greater clarity up front of the final deliverables produced by the meeting.	Presentation of general session materials will be modified to add reiteration of the meeting goals.
Item Review Activity	Panelists provided item descriptions at various levels of specificity.	Panelist training on the item review activity will be improved by sharing item descriptions from the pilot meeting found to be on target. The term “item descriptions” will also be modified to “item notes” to clarify the purpose of the individual review activity.
	Panelists in science did not have enough time to complete individual review activity.	Agenda will be adjusted to provide panelists instruction on the permitted time per item and help them focus their time effectively.

Meeting Segment	Lesson Learned	Plan of Action
	U.S. History and Civics panelists expressed frustration with content gaps and vocabulary use of the NAEP assessments.	The general session will include added information about the age of the current frameworks and plans for future updates.
Group Summary Descriptions	Panelists expressed frustration at not knowing the level of specificity expected of the group summary statements.	Final Reporting ALDs from the grade 8 subjects will be shared with the panelists prior to the group summary descriptions as examples. Facilitators will indicate which content may be most beneficial for their content area.
	Panelists in science needed additional time to complete group summary descriptions.	Agenda will be revised to provide additional time for groups to complete creation of summary descriptions.
Panel Summary Descriptions	Panel summary descriptions were created at different levels of specificity.	<p>Panelists will create panel summary descriptions that are statements that summarize the bullets from the groups, with a focus on creating statements with specificity similar to the content ALDs, but with some additional detail.</p> <p>Facilitators will remind panelists of the step in creating the panel summary descriptions throughout the process.</p>
Alignment Judgments	Some U.S. History panelists judged the alignment to be weak, based on their perspective of what should be covered by the assessment, not due to inconsistencies between the summary statements and content ALDs in the framework.	Reiterate the purpose for this activity and the process for making alignment judgments. Remind panelists that the upcoming framework revisions will be the opportunity to address gaps in content. Review justifications between rounds to ensure they are appropriate.

Operational ALD Review Study

This section of the report provides general results for the operational study and describes only those features of the operational study that differ from the procedures described under the section on general procedures. Pearson designed the operational study to apply the lessons learned from the pilot study.

Panelists

During the nomination process, over 90 individuals were nominated as possible participants for the study. The week of the meeting, December 5–9, 2022, a total of 23 panelists participated in the operational study, including 10 teachers and 13 non-teacher educators. The target composition of the study was a total of 24 panelists, eight panelists for each subject panel, with at least half of the panelists composed of teachers and at least two panelists being non-classroom educators. The minimum number of panelists for a subject area, informed by recommendations by the TAC, was six. Table 16 summarizes information about the panelist type for panel members who participated in and completed the operational study.

Table 16. Panelist type description for each panel

Type	Current Role in Education		
	Science	U.S. History	Civics
Classroom Teacher	3	3	4
Non-Classroom Educator	6	5	2
Total	9	8	6

A total of 13 teachers and 10 non-teacher educators (23 panelists total) participated in the operational study. The primary goal was to obtain panelists with the level of content expertise that would enable them to fully engage in the content-focused review process utilized during the meeting and to provide meaningful judgments. To achieve this goal, a multi-step recruitment process was implemented, starting with collecting nominations for qualified individuals from state and district education departments, and education organizations, then following up with nominated

individuals to collect information about experience and interest. To assist with recruiting panelists that reflected an overall representation, specific effort was spent in collecting nominations for individuals with diverse characteristics and experiences and from a range of demographic locations. This proved challenging as substitute teacher shortages were common at the time the study was conducted. The resulting panels included current classroom teachers and non-classroom educators (i.e., prior classroom teachers now working in state or district setting), educators from varied demographic location and urbanicity, and educators with experience with various student populations. Table 17 summarizes the gender and ethnicity distributions for the panel members who participated in the operational study.

Despite a targeted effort to increase the racial/ethnic and gender diversity of the panels, most participants were white females (two Black or African American panelists expected to participate were ultimately unable to). The TAC and COSDAM members discussed the implications of this on the study outcomes. Because the tasks involved reviewing assessment items and identifying the skills and knowledge required to respond to the items correctly rather than judging what students should know and be able to do, having strong content expertise was considered the most important panelist qualification. In addition, COSDAM members requested Pearson collect information about the student populations the panelists had experience with. The data indicated educators represented a diverse set of districts with varied percentages of minorities, income levels, size, and urbanicity. Information about the final set of panelists can be found in Tables 19 through 22.

The participants in each subject panel were divided into two replicate groups, to verify that the process worked for separate groups and to observe any differences in outcomes between groups. Each replicate group was selected to have both teachers and non-classroom educators, along with equal representation across other classifications.

Table 17. Gender, race, and ethnicity distribution for each panel

	Gender and Ethnicity		
	Science	U.S. History	Civics
Gender			
Female	7	6	4
Male	2	2	1
Other	0	0	1
Race			
Asian	0	0	0
Black/African American	1	0	1
American Indian, Alaska Native	0	0	0
White	8	8	3
No Response	0	0	2
Ethnicity			
Hispanic/Latino	0	0	0
No Response	0	0	2

There was a desire to have representation across the different NAEP regions. Although that representation did not exist in each individual panel, across all panels there was representation from each region. Table 18 summarizes the distribution of the panel members who participated in the pilot study across the geographic regions.

Table 18. Distribution for each panel across geographic regions

Region	Panel		
	Science	U.S. History	Civics
Northeast	2	2	2
Midwest	2	3	1
South	1	3	2
West	4	0	1

Table 19 summarizes the education experience with students from special populations. Tables 20–22 summarize the different educational environments for the panel members who participated in the pilot study.

Table 19. Experience of panel members with student populations

Experience with Student Population	Panel		
	Science	U.S. History	Civics
General Education	9	8	6
English Language Learners	6	6	4
Mainstream Special Education	7	7	5
Self-contained Special Education	2	0	1
Gifted and Talented Education	6	6	3

Table 20. School/District Location

Experience	Panel		
	Science	U.S. History	Civics
City	9	2	2
Suburb	2	2	2
Town	1	3	0
Rural	2	1	1
No Response	2	0	1

Table 21. School/District Size

Experience	Panel		
	Science	U.S. History	Civics
Large	3	2	4
Medium	0	1	0
Small	2	2	0
Distant	1	2	0
Remote/Fringe	1	1	1
N/A	2	0	1

Table 22. School/District Student Populations

Experience	Panel		
	Science	U.S. History	Civics
Percent Minorities			
High (More than 60%)	1	0	2
Medium (30% to 59%)	4	2	2
Low (Less than 30%)	2	6	1
Family Income			
Above National Average	3	3	2
Below National Average	4	5	3
N/A	2	0	1

Facilitators

Each of the grade and subject groups was facilitated by a lead content expert and a supporting content expert. The lead facilitator was responsible for conducting all panel-level training, and both facilitators ensured that appropriate processes were followed. The technical assistants helped with technology issues. The content facilitators and technical assistant for each panel are shown in Table 23.

Table 23. Facilitators and technical assistants for each panel

Content Area	Lead Facilitator	Assistant Facilitator	Technical Assistant
Science	Michaela Viering	Ashtyn Greenstein	Brian Wrobel
U.S. History	Rachel Williams	Meredith Hohe	Ebony Gaines
Civics	Kadie Patterson	Judy Goss	Grant Smith

All but one content facilitator for the operational study were facilitators during the pilot study for the same subject. All facilitators participated in two three-hour training meetings that focused on reviewing the process for the NAEP ALD Review study and adjustments that were made for the operational study, based on results from the pilot study. The facilitator that was new to the study received additional training from the previous facilitators and the project director to ensure that she was prepared to lead the meeting for their groups. All training was conducted virtually.

Observers

In addition to the content facilitators, there were four observers of the operational study. They included two members of the TAC, Dr. Karla Egan and Dr. Susan Loomis. There were also staff members from the Board, the Assistant Director for Psychometrics and Technical Point of Contact, Dr. Rebecca Norman Dvorak, and Assistant Director for Assessment Development, Dr. Sharyn Rosenberg.

Adjustments to the Process for the ALD Review Operational Study

The model-based ALD review process that was described in the design document and used during the pilot study was successfully implemented for the operational study. This process was also successfully implemented previously during the NAEP ALD Review Study for mathematics and reading. The activities and materials for the study were standardized across subjects, as much as possible. The following sections provide descriptions of the activities each panel completed.

Independent Item Review

After the training by the facilitators in completing the individual item review, the facilitators conducted a modeling activity in which the facilitator showed examples of item descriptions at various levels of detail. The panelists were asked to draft item descriptions for several items and have a discussion about the strength of the descriptions. The focus of this activity was to assist panelists in understanding the item review activity and the level of detail required for the next activity.

After the training on the item review process and the modeling activity, the rest of the independent item review activities were completed using the same process used during the pilot meeting. Based on their experiences during the pilot meeting, the facilitators in each of the panels adjusted the time and order of the item reviews to ensure that there was enough time for panelists to complete the review activities for each content area.

Group and Panel Summary Statement Development

Based on the lessons learned from the pilot meeting, the group summary statement spreadsheets were updated so that the individual summary statements from the individual item review activity were automatically completed in the group spreadsheets. To assist with achieving the correct level of specificity in the summary statements, training was provided by the facilitators to understand the level of supporting evidence needed to create a usable summary statement. The panelists were instructed that the evidence needed to create a summary statement should be derived either from at least two items at the same achievement level or from two items from adjacent achievement levels that highlighted the difference between what is achieved at the two levels.

Alignment Judgments

Prior to panelists making their individual alignment judgments, the panels were brought together to provide training on the alignment judgment process. This alignment judgment training was provided by one of the project directors, Eric Moyer or Jennifer Galindo. The training focused on the different rationale for a lack of alignments and the possible judgments for the level of alignment. The rationale for indicating a lack of alignment included:

- Knowledge and skills associated with one achievement level on the content ALD are associated with a different achievement level in the summary statement.
- Knowledge and skills associated with an achievement level in the summary statement are not represented on the content ALDs.

Additionally, it was emphasized that if knowledge and skills are associated with an achievement level in the content ALDs, but not represented in the Reporting ALDs, this is not a reason for misalignment, because this could be a sampling issue of the assessment. In other words, one assessment administration may not address all content specified by the framework at the degree necessary for inclusion in the Reporting ALDs (i.e., 2 items at an achievement level). The panelists were also asked to provide a written rationale for their judgment.

Results

After the operational NAEP ALD review study, the lead facilitators, who were content experts in the specific subject, completed a review of the draft Reporting ALDs produced by the panels. During the review, the facilitators from each content area reviewed the language and format of the draft Reporting ALDs across the achievement levels. Minor adjustments were made to improve the cohesiveness of the statements across the achievement levels. The facilitators were mindful that the edits should not impact the performance statements communicated by the committees. The draft Reporting ALDs resulting from the content review were created as bulleted lists so the most useful format for the Reporting ALDs could be determined.

The draft Reporting ALDs from the content review, as bulleted lists, went through various internal reviews.

- Internal review by National Center for Educational Statistics (NCES) and NAEP contractors to ensure that there were no issues from an operational perspective (e.g., item security concerns, inconsistent language with frameworks)
- External review by two content experts per subject, who were highly familiar with NAEP frameworks. These reviews focused on bias and sensitivity concerns, consistency with NAEP frameworks, and clarity. The subject area experts were also asked to identify skills or knowledge that showed progression across achievement levels.

- Review by Board communications staff and contractors. These reviews focused on general readability to consider aspects that may be unclear to a general audience.

The internal reviews were designed to improve the overall clarity and utility of the Reporting ALDs. The feedback from the internal review was reviewed by the lead content facilitators and discussed with NCES during an articulation meeting to determine which changes should be made to address issues, while maintaining the substance of the statements from the operational ALD Review study.

Following the internal review, Pearson emailed the revised Reporting ALDs to panelists and included the original statements developed during the study. The participants who responded approved the modifications and did not have concerns with the revised statements.

The Board reviewed and took action on the final Reporting ALDs for Grade 8 NAEP assessments during the March 3, 2023, Board meeting. They were unanimously approved. The final Reporting ALDs for Grade 8 Science, U.S. History, and Civics are presented in Tables 24 through 26.

Table 24. Science Grade 8 – Reporting ALDs

Achievement Level	Reporting ALDs
<i>NAEP Basic</i>	<p>Regarding the content for <i>Earth and space sciences</i>, students performing at the <i>NAEP Basic</i> achievement level likely can</p> <ul style="list-style-type: none"> • apply knowledge that sedimentary rock relates to fossil formation and how fossils are evidence of past environments • recall information about the solar system with an emphasis on the Sun and Earth • relate an aspect of the water cycle to weather formation. <p>Regarding the content for <i>life science</i>, students performing at the <i>NAEP Basic</i> achievement level likely can</p> <ul style="list-style-type: none"> • identify simple relationships between organisms within an ecosystem (mutualism, competition) • determine simple relationships within food webs • identify the effect of resource availability on a population provided a specific context • demonstrate understanding that reproduction is an essential part of population survival • recall that plants need sunlight to grow and reproduce. <p>Regarding the content for <i>physical science</i>, students performing at the <i>NAEP Basic</i> achievement level likely can</p> <ul style="list-style-type: none"> • identify that matter has unique chemical and physical properties • recognize kinetic energy and that it can be converted to a different form. <p>Regarding the <i>science practices</i>, students performing at the <i>NAEP Basic</i> achievement level likely can</p> <ul style="list-style-type: none"> • describe trends in data displayed on graphs • identify simple information from a graph or data table • summarize the interactions between components of a model • identify simple relationships in a scientific context • identify foundational science principles • describe 1–2 steps of an experimental design.
<i>NAEP Proficient</i>	<p>Regarding the content for <i>Earth and space sciences</i>, students performing at the <i>NAEP Proficient</i> achievement level likely can</p> <ul style="list-style-type: none"> • identify the components and causal processes of the water cycle • describe how human activity can impact the environment provided a specific context • identify the impact of ocean currents on air temperature when provided geographic locations

- use knowledge of the Earth's structure including tectonic plate movement to explain physical characteristics of the Earth's surface
- identify how Earth's materials are made and broken down)
- interpret evidence from fossils and rock layers to determine past environments
- describe the relative age of rocks in a diagram
- identify the effect of tilt and rotation on the impact of solar radiation on different locations on Earth
- support an argument about orbital motion in the solar system from evidence provided in a specific context
- recognize that weather changes when two air masses meet.

Regarding the content for *life science*, students performing at the *NAEP Proficient* achievement level likely can

- analyze how changes to living and non-living components of the environment impact food webs
- predict the effect of changing resource availability on a population when provided a specific context
- recognize the role of decomposition within a food web
- identify advantages of asexual reproduction
- describe the function of body systems (circulatory, respiratory, digestive)
- identify the foundational process and components of photosynthesis (inputs and outputs)
- recognize the relationship between structure and function in organisms
- classify organisms based on their characteristics
- identify adaptations that impact the survival of a species when provided a specific context.

Regarding the content for *physical science*, students performing at the *NAEP Proficient* achievement level likely can

- recognize and/or explain the characteristics (speed and spacing of molecules, effect of heat) of the states of matter
- interpret a position vs time graph to describe the motion of an object
- identify how kinetic and/or potential energy changes when an object is in motion
- demonstrate a cause and effect understanding of energy transfer on an object when provided a specific context
- apply knowledge of a substance's density to determine the properties of an unknown substance
- identify a substance's physical and chemical properties, and those properties make the substance suitable for use in certain applications
- interpret a model to identify the direction of gravity within the context of the model
- apply knowledge that atoms of elements can be combined to form different substances with their own unique properties.

Regarding the *science practices*, students performing at the *NAEP Proficient* achievement level likely can

- set up an experiment including identifying specific variables
- identify the data to collect in an experiment in order to support or test a claim

	<ul style="list-style-type: none"> • explain how to use a tool when provided a specific context • analyze figures, including graphs, to use data as evidence to support a claim • interpret maps to draw conclusions • explain how components of a model interact with one another • recognize connection between structure and function • describe individual parts of a system or process.
<p><i>NAEP Advanced</i></p>	<p>Regarding the content for <i>Earth and space sciences</i>, students performing at the <i>NAEP Advanced</i> achievement level likely can</p> <ul style="list-style-type: none"> • use knowledge of the Earth's structure and movement of tectonic plates to explain geological phenomena • identify that convection currents in water and/or air are caused by uneven heating • use models of the Sun-Earth system to infer the effect of Earth's tilt, rotation, and differences in solar radiation • use evidence from a given specific context to support an explanation of how human activity can impact the environment • use knowledge of geologic processes to explain how materials of the Earth are made and broken down. <p>Regarding the content for <i>life science</i>, students performing at the <i>NAEP Advanced</i> achievement level likely can</p> <ul style="list-style-type: none"> • explain relationships among different trophic levels of a food web • use evidence from a specific context to support an explanation of how resource availability affects population dynamics • use the structures of organisms to identify specific adaptations of organisms and infer how these adaptations help organisms to survive • explain the function of the organs that make up the body systems • identify advantages or disadvantages of asexual reproduction • demonstrate an understanding of the interaction between organisms and environmental factors in nutrient cycling. <p>Regarding the content for <i>physical science</i>, students performing at the <i>NAEP Advanced</i> achievement level likely can</p> <ul style="list-style-type: none"> • use evidence to support a claim about the effect of temperature on pressure within a given specific context • know and describe the properties of water (including the effects of temperature and pressure) • explain how the density of a gas affects its behavior for a given specific context • describe the effect of gravitational force on objects presented in a diagram • apply knowledge of energy to identify energy transformations within a complex or unique specific context • describe the role of conservation of mass in a chemical reaction • describe the relationship between kinetic and potential energy within a system • describe the relationship between solar energy and energy production in plants • apply knowledge of the properties of matter to select the appropriate material to perform a defined function.

	<p>Regarding the <i>science practices</i>, students performing at the <i>NAEP Advanced</i> achievement level likely can</p> <ul style="list-style-type: none"> • interpret experimental data to draw conclusions • apply scientific knowledge to determine the appropriate tool to use in a given context • use evidence, including quantitative data from graphs, to support a scientific claim • design a controlled experiment • explain relationships between parts of a system or process.
--	--

Note: The content descriptions represented within the reporting ALD statements are intended to reflect the content defined within the framework.

Table 25. U.S. History Grade 8 – Reporting ALDs

Achievement Level	Reporting ALDs
<i>NAEP Basic</i>	<p>Students performing at the <i>NAEP Basic</i> achievement level in U.S. History likely can</p> <ul style="list-style-type: none"> • recall major historical terms and concepts • identify the context of major historical figures, places, ideas and events • identify simple historical concepts in primary or secondary sources • make simple conclusions based on primary or secondary sources. <p>Regarding the content for <i>change and continuity in American democracy</i>, students performing at the <i>NAEP Basic</i> achievement level likely can</p> <ul style="list-style-type: none"> • recall fundamental knowledge of the Civil War, including causes, key events, and outcomes • recall fundamental knowledge about the U.S. Constitution. <p>Regarding the content for <i>gathering and interactions of peoples, cultures, and ideas</i>, students performing at the <i>NAEP Basic</i> achievement level likely can</p> <ul style="list-style-type: none"> • recall knowledge of the experience of Black or African Americans (and enslaved Africans) through Reconstruction • identify major social and cultural characteristics in various time periods.

	<p>Regarding the content for <i>economic and technological changes</i>, students performing at the <i>NAEP Basic</i> achievement level likely can</p> <ul style="list-style-type: none"> • demonstrate knowledge of the impact of economic policies and technological innovations • identify major technological and economic developments • identify and describe the economic motivations of European colonization. <p>Regarding the content for <i>the changing role of America in the world</i>, students performing at the <i>NAEP Basic</i> achievement level likely can</p> <ul style="list-style-type: none"> • identify the relationships between different nations and groups of people • recall fundamental knowledge about major events related to foreign policy.
<p><i>NAEP Proficient</i></p>	<p>Students performing at the <i>NAEP Proficient</i> achievement level in U.S. History likely can</p> <ul style="list-style-type: none"> • read and interpret primary and secondary sources to make inferences and draw conclusions • recall knowledge of historical events without source material to provide context • demonstrate understanding and knowledge of change over time. <p>Regarding the content for <i>change and continuity in American democracy</i>, students performing at the <i>NAEP Proficient</i> achievement level likely can</p> <ul style="list-style-type: none"> • demonstrate understanding of the influences and content of major founding documents • identify principles in American founding documents • recall knowledge about political and social reform movements • make inferences or connections using primary sources • read and interpret maps to develop conclusions • identify inequities involving freedom and opportunity for women and Black or African Americans. <p>Regarding the content for <i>gathering and interactions of peoples, cultures, and ideas</i>, students performing at the <i>NAEP Proficient</i> achievement level likely can</p> <ul style="list-style-type: none"> • identify experiences of Black or African Americans from Reconstruction to the Great Migration • identify the nature and consequences of American Indian interactions with European explorers, colonists, and the United States government • identify the motivations and influence of abolitionism • demonstrate understanding of the perspectives or contributions of individuals and groups to the development of unique American culture • read and interpret graphs and maps to identify trends in migration to and within the United States

	<ul style="list-style-type: none"> • demonstrate understanding of the composition of the workforce and the impact of different groups in various time periods • make inferences using primary sources • use historical terms to answer a question. <p>Regarding the content for <i>economic and technological changes</i>, students performing at the <i>NAEP Proficient</i> achievement level likely can</p> <ul style="list-style-type: none"> • describe the relationship between government policies and the economy • apply historical knowledge to analyze a source • explain the effects of economic and technological change • demonstrate understanding of the economic consequences of the labor of enslaved and free Black or African Americans in the South • determine the perspective of individuals or groups regarding economic systems • determine or provide reasons for an effect of a historical event or process. <p>Regarding the content for <i>the changing role of America in the world</i>, students performing at the <i>NAEP Proficient</i> achievement level likely can</p> <ul style="list-style-type: none"> • identify U.S. foreign policy across various time periods • interpret an author's purpose or point of view in historical sources • use sources to infer the meaning or significance of historical events • understand two historical concepts and the connections between them to explain events • demonstrate understanding of the interaction between peoples and nations.
<p><i>NAEP Advanced</i></p>	<p>Students performing at the <i>NAEP Advanced</i> achievement level in U.S. History likely can</p> <ul style="list-style-type: none"> • demonstrate understanding of historical events and concepts through writing • analyze primary and secondary sources to contextualize and explain historical ideas and events. <p>Regarding the content for <i>change and continuity in American democracy</i>, students performing at the <i>NAEP Advanced</i> achievement level likely can</p> <ul style="list-style-type: none"> • analyze and interpret primary and secondary sources to explain in writing their impact or effect in specific time periods • explain motivations for westward migration and expansion by white and Black or African American settlers • understand the causes and effects of federal government policies and actions regarding slavery • recall detailed information about historical people or events without the assistance of a source

- make complex or detailed connections between concepts related to rights of individuals and groups.

Regarding the content for *gathering and interactions of peoples, cultures, and ideas*, students performing at the *NAEP Advanced* achievement level likely can

- place events within historical time periods
- analyze primary and secondary sources to determine purpose, supply evidence, reach conclusions, or draw inferences
- make comparisons between different roles and lifestyles within and across different historical time periods in American history (colonial, Reconstruction, modern day)
- provide multiple points of evidence for a historical claim
- demonstrate historical knowledge through written expression.

Regarding the content for *economic and technological changes*, students performing at the *NAEP Advanced* achievement level likely can

- demonstrate understanding of how changes in technology impacted economic growth
- use multiple pieces of evidence from historical sources to arrive at a conclusion
- use maps, charts, and graphs to analyze historical trends
- demonstrate understanding of the significance of major economic developments
- provide an explanation or justification for a historical claim.

Regarding the content for *the changing role of America in the world*, students performing at the *NAEP Advanced* achievement level likely can

- explain in writing the impact of government policies on different groups or nations
- make inferences or contextualize ideas from a source or time period
- describe the intended purpose or impact of foreign policy.

Note: The content descriptions represented within the reporting ALD statements are intended to reflect the content defined within the framework.

Table 26. Civics Grade 8 – Reporting ALDs

Achievement Level	Reporting ALDs
<i>NAEP Basic</i>	<p>Regarding the content for <i>civic life, politics, and government</i>, students performing at the <i>NAEP Basic</i> achievement level likely can</p> <ul style="list-style-type: none"> • describe the structure and function of government • identify the difference between civic and private life • interpret stimuli to identify governing documents and their purpose. <p>Regarding the content for <i>U.S. and world affairs</i>, students performing at the <i>NAEP Basic</i> achievement level likely can</p> <ul style="list-style-type: none"> • identify potential areas of conflict and cooperation between countries • recognize that the United States is part of an interconnected world. <p>Regarding the content for <i>the roles of U.S. citizens</i>, students performing at the <i>NAEP Basic</i> achievement level likely can</p> <ul style="list-style-type: none"> • identify restrictions to fundamental freedoms • identify ways in which citizens influence American society • describe the rights and responsibilities of U.S. citizens. <p>Regarding the content for <i>the Constitution and American government</i>, students performing at the <i>NAEP Basic</i> achievement level likely can</p> <ul style="list-style-type: none"> • identify and explain the sources and purposes of tax dollars • identify the purpose of each level of government: national, state, and local • identify ways in which the media and private citizens can express opinions and play a role in the political process. <p>Regarding the content for <i>foundations of the American political system</i>, students performing at the <i>NAEP Basic</i> achievement level likely can</p> <ul style="list-style-type: none"> • identify key democratic ideals, including equality and individual rights • identify equality under law, consent of the governed, and natural rights • describe how the U.S. has not always lived up to its founding ideals and principles.
<i>NAEP Proficient</i>	<p>Regarding the content for <i>civic life, politics, and government</i>, students performing at the <i>NAEP Proficient</i> achievement level likely can</p> <ul style="list-style-type: none"> • describe different types of government

- identify separation of powers and checks and balances
- analyze and infer the meaning from a variety of different civics-related sources
- analyze the responsibilities/purposes of government
- recognize the need to balance rights and responsibilities of citizens.

Regarding the content for *U.S. and world affairs*, students performing at the *NAEP Proficient* achievement level likely can

- examine ways in which the United States influences other countries
- evaluate global scenarios and determine the effect that these scenarios may have on the United States and its policies
- identify foreign policy issues.

Regarding the content for *the roles of U.S. citizens*, students performing at the *NAEP Proficient* achievement level likely can

- identify the process of becoming a U.S. citizen
- make inferences about media sources to gain political and civic information
- explain the differences between civic rights, civic responsibilities, and the duties of citizens
- describe how rights are protected and limited by the U.S. Constitution
- explain how rights have evolved in the U.S. Constitution.

Regarding the content for *the Constitution and American government*, students performing at the *NAEP Proficient* achievement level likely can

- describe the rights of citizens, including due process of law
- describe how the media can play a role in elections and the democratic process
- explain the purpose and functions of each level of government: national, state, and local
- describe ways in which citizens influence government.

Regarding the content for *foundations of the American political system*, students performing at the *NAEP Proficient* achievement level likely can

- describe the purpose of the Bill of Rights and apply the Bill of Rights to real-world scenarios
- explain that the United States is made up of diverse groups, whose ideas have contributed to the American political system
- draw conclusions from sources to describe the foundations of American democracy
- describe how U.S. constitutional democracy relies on an educated citizenry
- identify reasons why the United States can be viewed as the land of opportunity.

<p><i>NAEP Advanced</i></p>	<p>Regarding the content for <i>civic life, politics, and government</i>, students performing at the <i>NAEP Advanced</i> achievement level likely can</p> <ul style="list-style-type: none"> • evaluate sources to draw conclusions about early American political history • identify information from multimedia sources with differing points of view • analyze and apply the concept of federalism. <p>Regarding the content for <i>U.S. and world affairs</i>, students performing at the <i>NAEP Advanced</i> achievement level likely can</p> <ul style="list-style-type: none"> • identify the roles, purposes, and limitations of international organizations • explain global issues and develop potential solutions to global problems. <p>Regarding the content for <i>the roles of U.S. citizens</i>, students performing at the <i>NAEP Advanced</i> achievement level likely can</p> <ul style="list-style-type: none"> • explain how individuals participate in and influence the democratic process • evaluate the importance of civic responsibilities in a democracy. <p>Regarding the content for <i>the Constitution and American government</i>, students performing at the <i>NAEP Advanced</i> achievement level likely can</p> <ul style="list-style-type: none"> • explain the role of political parties and interest groups in the democratic process • explain how the media plays a role in elections and the democratic process • analyze the functions of the three branches of government. <p>Regarding the content for <i>foundations of the American political system</i>, students performing at the <i>NAEP Advanced</i> achievement level likely can</p> <ul style="list-style-type: none"> • describe the purpose of various founding documents, including the U.S. Constitution and <i>The Federalist Papers</i> • evaluate competing ideas within the U.S. political system • evaluate how the diversity of the United States has contributed to the development of the American political system • describe changes in American society and government.
---------------------------------	--

Note: The content descriptions represented within the reporting ALD statements are intended to reflect the content defined within the framework.

Tables 27 and 28 present the results from round 3 of the ALD alignment judgment activity from the operational study.

Table 27. Round 3 alignment judgment agreement with achievement level policy definitions

Subject	NAEP Level	Alignment Judgment			
		Minimal	Weak	Moderate	Strong
Science	Basic	0%	0%	55%	44%
	Proficient	0%	11%	44%	44%
	Advanced	11%	0%	44%	44%
U.S. History	Basic	0%	0%	17%	83%
	Proficient	0%	0%	0%	100%
	Advanced	0%	0%	33%	67%
Civics	Basic	0%	0%	37%	63%
	Proficient	0%	0%	63%	37%
	Advanced	0%	0%	50%	50%

Table 28. Round 3 alignment judgment agreement with content ALDs

Subject	NAEP Level	Alignment Judgment			
		Minimal	Weak	Moderate	Strong
Science	Basic	0%	33%	33%	33%
	Proficient	0%	11%	77%	11%
	Advanced	0%	11%	44%	44%
U.S. History	Basic	0%	0%	75%	25%
	Proficient	0%	0%	63%	37%
	Advanced	0%	0%	100%	0%
Civics	Basic	0%	0%	83%	17%
	Proficient	0%	0%	17%	83%
	Advanced	0%	0%	50%	50%

The alignment judgments for U.S. History and Civics comparing the summary statements to the policy definitions and the content ALDs were either moderate or strong. The alignment judgments for the policy definitions were higher than the judgments for the content ALDs. For U.S. History *NAEP Advanced*, all eight panelists listed the alignment to the content ALDs as moderate. The relatively positive alignment judgments indicated that the panelists believed that knowledge and skills likely

demonstrated by the students on the NAEP assessment represented the knowledge and skills described by the NAEP content ALDs.

For Science, there was one panelist who felt that the alignment to the policy statement was weak for the *NAEP Proficient* level and minimal for the *NAEP Advanced* level. For the alignment to the content ALDs, one-third (i.e., three) of the Science panelists said that the *NAEP Basic* level alignment was weak. A panelist rationale for this alignment judgment was that students at the *NAEP Basic* level “could not use evidence to support explanations, explain observations, or design experiments.” They felt aspects of the draft Reporting ALDs were reflective of a student not reaching *NAEP Basic* based on the content ALDs.

Process Evaluations

Panelists responded to a number of survey questions multiple times at different points in the operational NAEP ALD Review study. The responses to select questions have been placed into separate tables depending on the Likert scale used. These responses generally show a high rate of agreement to the question posed, though the level of agreement varies among the panelists. In Civics, there was one panelist who did not agree that the training provided was adequate for the study. The panelists indicated that the amount of time provided for the alignment judgment rounds and discussions was either exactly right or too much. Tables 29 through 38 present the results of select questions from the process evaluation surveys.

Table 29. Purpose of the ALD alignment judgment activity was clearly explained

Select the option that best reflects your opinion about the level of agreement for the statement. The purpose of the ALD alignment judgment activity was clearly explained.						
	Strongly Disagree	Disagree	Slightly Disagree	Slightly Agree	Agree	Strongly Agree
Science	0	0	0	0	7 (78%)	2 (22%)
U.S. History	0	0	0	1 (13%)	7 (88%)	0
Civics	0	1 (17%)	0	0	4 (67%)	1 (17%)

Table 30. Explanation of the ALD alignment judgment options

Select the option that best reflects your opinion about the level of agreement for the statement. The explanation of the ALD alignment judgment options (i.e., Strong, Moderate, etc.) was clear.						
	Strongly Disagree	Disagree	Slightly Disagree	Slightly Agree	Agree	Strongly Agree
Science	0	0	0	1 (11%)	5 (56%)	3 (33%)
U.S. History	0	0	0	2 (25%)	5 (63%)	1 (13%)
Civics	0	1 (17%)	1 (17%)	0	3 (50%)	1 (17%)

Table 31. Rationale for misalignment judgment

Select the option that best reflects your opinion about the level of agreement for the statement. I understood the possible rationale for misalignment between the summary descriptions and content ALDs.						
	Strongly Disagree	Disagree	Slightly Disagree	Slightly Agree	Agree	Strongly Agree
Science	0	0	0	1 (11%)	5 (56%)	3 (33%)
U.S. History	0	0	0	0	6 (75%)	2 (25%)
Civics	0	1 (17%)	0	0	4 (67%)	1 (17%)

Table 32. Steps to follow to complete the alignment judgment activity

Select the option that best reflects your opinion about the level of agreement for the statement. I understood the steps to follow as I completed the individual alignment judgment activity.						
	Strongly Disagree	Disagree	Slightly Disagree	Slightly Agree	Agree	Strongly Agree
Science	0	0	0	3 (33%)	4 (44%)	2 (22%)
U.S. History	0	0	0	3 (38%)	4 (50%)	1 (13%)
Civics	0	1 (17%)	0	2 (33%)	2 (33%)	1 (17%)

Table 33. Use of feedback data between judgment rounds

Select the option that best reflects your opinion about the level of agreement for the statement. I was able to use the feedback data during group discussions and judgment rounds.						
	Strongly Disagree	Disagree	Slightly Disagree	Slightly Agree	Agree	Strongly Agree
Science	0	0	0	1 (11%)	6 (67%)	2 (22%)
U.S. History	0	0	0	0	6 (75%)	2 (25%)
Civics	0	0	1 (17%)	0	3 (50%)	2 (33%)

Table 34. Creation of panel summary descriptions

Select the option that best reflects your opinion about the level of agreement for the statement. The creation of the panel summary descriptions was helpful in defining student abilities for each achievement level.						
	Strongly Disagree	Disagree	Slightly Disagree	Slightly Agree	Agree	Strongly Agree
Science	0	0	1 (11%)	1 (11%)	2 (22%)	5 (56%)
U.S. History	0	0	0	0	4 (50%)	4 (50%)
Civics	0	0	0	0	4 (67%)	2 (33%)

Table 35. Group discussions

Select the option that best reflects your opinion about the level of agreement for the statement. The group discussions helped me feel more confident with the work in subsequent judgment rounds.						
	Strongly Disagree	Disagree	Slightly Disagree	Slightly Agree	Agree	Strongly Agree
Science	0	0	0	2 (22%)	3 (33%)	4 (44%)
U.S. History	0	0	0	0	4 (50%)	4 (50%)
Civics	0	0	0	0	4 (67%)	2 (33%)

Table 36. Amount of time provided for ALD alignment judgments

The Amount of Time Training on the ALD Alignment Judgment Process and Discussion of Alignment Classifications			
Subject/Grade	Too Little Time	Exactly Right Time	Too Much Time
Science	1 (11%)	8 (89%)	0
U.S. History	0	7 (88%)	1 (13%)
Civics	1 (17%)	3 (50%)	2 (33%)

As part of the process evaluation, panelists were asked to rate their confidence that the final panel-level summary descriptions describe the knowledge and skills that students with achievement associated with the level are likely to demonstrate. The ratings of the panelists show that they were confident (Somewhat or Completely) that their summary statements accurately described student achievement based on the items reviewed during the process. Tables 37 and 38 present the results of these questions from the process evaluations.

Table 37. Confidence in the panel's achievement level summary descriptions

Select the option that best reflects your opinion about your confidence in the achievement level summary descriptions developed by the panel.					
	Performance Level	Not Confident at All	Slightly Confident	Somewhat Confident	Completely Confident
Science	<i>NAEP Basic</i>	0	2 (22%)	2 (22%)	5 (56%)
	<i>NAEP Proficient</i>	0	2 (22%)	2 (22%)	5 (56%)
	<i>NAEP Advanced</i>	0	2 (22%)	1 (11%)	6 (67%)
U.S. History	<i>NAEP Basic</i>	0	0	3 (38%)	5 (63%)
	<i>NAEP Proficient</i>	0	0	4 (50%)	4 (50%)
	<i>NAEP Advanced</i>	0	0	5 (63%)	3 (38%)
Civics	<i>NAEP Basic</i>	0	2 (33%)	1 (17%)	3 (50%)
	<i>NAEP Proficient</i>	0	0	2 (33%)	4 (67%)
	<i>NAEP Advanced</i>	1 (17%)	1 (17%)	0	4 (67%)

At the end of the alignment judgment process, panelists were asked to rate their confidence in their alignment judgment classification for each achievement level. For all subjects, greater than two-thirds of the panelists indicated that they were either

“Somewhat Confident” or “Completely Confident” in the overall alignment judgment by the committee for the achievement levels.

Table 38. Confidence in the panel's alignment judgments

Select the option that best reflects your opinion about your confidence in overall alignment judgment by the panel.					
	Performance Level	Not Confident at All	Slightly Confident	Somewhat Confident	Completely Confident
Science	<i>NAEP Basic</i>	0	3 (33%)	4 (44%)	2 (22%)
	<i>NAEP Proficient</i>	0	2 (22%)	3 (33%)	4 (44%)
	<i>NAEP Advanced</i>	0	3 (33%)	3 (33%)	3 (33%)
U.S. History	<i>NAEP Basic</i>	0	0	3 (38%)	5 (63%)
	<i>NAEP Proficient</i>	0	0	4 (50%)	4 (50%)
	<i>NAEP Advanced</i>	0	0	3 (38%)	5 (63%)
Civics	<i>NAEP Basic</i>	0	1 (17%)	2 (33%)	3 (50%)
	<i>NAEP Proficient</i>	0	0	2 (33%)	4 (67%)
	<i>NAEP Advanced</i>	0	1 (17%)	1 (17%)	4 (67%)

Validity Evidence

Two general categories of validity evidence were collected or used during the ALD Review study: procedural and internal. Procedural validity evidence refers to the appropriateness for the ALD Review study 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 ALD review methodology, the quality of the implementation of the procedure, and the completeness of the documentation of the process (Sireci et al., 2009). In general, it is the fidelity of the process implemented to the design document produced to guide the process. Internal validity evidence refers to the internal consistency of data generated within the ALD review meeting. Table 39 provides a list of the different types of validity evidence for the NAEP ALD Review study.

Table 39. Validity evidence for NAEP ALD Review study

Validity Evidence		
Type of Evidence	Evidence	Explanation
Procedural Validity Evidence	Design Document	A design document, which was fully reviewed by the TAC and COSDAM, provided the procedures and process intended to be carried out in the ALD Review study. The design document served as a guide for all aspects of the ALD Review study. The processes implemented and described in this report are consistent with those in the design document.
	Observations of Two TAC Members	Two of the TAC members observed all aspects of the NAEP ALD Review study to provide external evidence that the procedures described in the design document were implemented correctly. The resulting reports from the external reviewers clearly state that the intended procedures were followed.
	Process Evaluations	Panelists were asked to complete process evaluations after each major activity of the ALD review. The results of these evaluations, as previously described, provide strong evidence of the procedural validity of the study.
Internal Validity Evidence	Panelist Agreement with Draft Reporting ALDs	At the end of the ALD Review study, the panelists were asked about their level of agreement with the summary statements or draft Reporting ALDs, that they described the knowledge and skills that would be demonstrated by performance associated with each NAEP achievement level. The results of this evaluation provide strong evidence for the internal validity of the study results.
	Panelist Agreement with Final Round of Reporting ALDs	After the panelists' draft Reporting ALDs underwent several rounds of editorial reviews, the resulting final Reporting ALDs were shared with the panelists from the operational ALD Review study. The panelists were provided the opportunity to review the final Reporting ALDs and comment if there were any significant changes from the intention of the initial draft Reporting ALDs. The panelists' agreement with the final Reporting ALDs is additional evidence for the internal validity of the study results.
	Panelist Alignment Judgments	During the ALD Review study, panelists were provided the opportunity to complete three rounds of alignment judgments after panel discussions of the results and possible areas of variation. The change of the alignment judgment results across the rounds, which demonstrates increasing consistency in their judgments, provides evidence for the internal validity of the alignment judgment results.

Recommendations for Future Studies

Out of this ALD Review study, Pearson has several recommendations for future ALD review studies. As described in the achievement level procedures manual, ALD review studies are required to occur on a periodic basis, at least once every 10 years or three administrations of an assessment, whichever comes later. The goal of these ALD review meetings is defined in Principle 4(a).

At least once every 10 years or three administrations of an assessment, whichever comes later, the Board, through its Committee on Standards, Design and Methodology (COSDAM), shall review the alignment between the content ALDs and items, based on empirical data from recent administrations of NAEP assessments. In its review, COSDAM (in consultation with the Assessment Development Committee) shall solicit input from technical and subject matter experts to determine whether changes to the content ALDs are warranted or whether a new standard setting shall be conducted, making clear the potential risk of changing cut scores to trends and assessment of educational progress.

Since ALD review studies are planned to be conducted on a more periodic basis, it would be beneficial to explore methods that could either simplify the ALD review process used for this study or build off the results of this study. The method that was used for this study, resulting in Reporting ALDs and alignment judgments, is a time- and resource-intensive process. With the development of initial Reporting ALDs, methods could be explored to possibly build off the results of this study to complete future ALD reviews.

Additionally, the goal stated in Principle 4 for the ALD review is to determine if any changes to the content ALDs are warranted or whether a new standard setting is required. The results of an alignment study are judgments regarding the strength of the alignment between the content ALDs and the Reporting ALDs, or a comparison between what is expected that students should know and be able to do and what students demonstrate they actually know and can do. To support the decision required by the principle, research could be completed to assist in determining the criteria for the amount of misalignment that would lead to revising the ALDs or requiring a new standard setting. This research would incorporate the potential risk to changing the ALDs or changing the cut scores, to the trends and application of the NAEP achievement levels.

Finally, as reported, the results of the alignment study for science indicated various levels of misalignment for each of the NAEP achievement levels, but specifically *NAEP Basic*, where a third of the panelists indicated that there was weak alignment. The rationale provided by the panelists indicated that they believed that some content described in the Reporting ALDs, specifically associated with the science processes, was aligned with a different NAEP achievement level. There was much discussion with the TAC about various sources of this misalignment, but no clear determination was made. It would be beneficial to the current assessment and for future studies to examine the different sources that may lead to misalignment and to gather more qualitative information from the panelists to explain their thinking, which could be used to address questions regarding the decisions made using the ALD alignment judgment results.