Achievement Level Description Review for the National Assessment of Educational Progress Mathematics and Reading Assessments

Submitted: November 2022

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

This study was funded by the National 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 Mathematics and Reading Assessments

Final Process Documentation and Technical Report

November 2022

Eric L. Moyer
Jennifer Galindo
Pearson

National Assessment Governing Board

BOARD MEMBERSHIP (2021-2022)

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
Yazoo City, Mississippi

Dana K. Boyd
Principal
Parkland Elementary School
El Paso, Texas

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

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

Eric Hanushek
Hanna Senior Fellow
Hoover Institution
Stanford, 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

Paul Gasparini
Secondary School Principal
Christian Brothers Academy
Syracuse, New York

Honorable Gary R. Herbert
Former Governor of Utah
Salt Lake City, Utah

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

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

Tonya Matthews
Chief Executive Officer
International African American Museum
Charleston, South Carolina

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

Mark Miller
Eighth Grade Mathematics Teacher and Department Chair
Cheyenne Mountain Junior High
Colorado Springs, Colorado

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

Ron Reynolds
Executive Director
California Association of Private School Organizations
Van Nuys, California

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

Martin R. West
Massachusetts Board of Elementary and Secondary Education
Professor of Education
Harvard Graduate School of Education
Cambridge, Massachusetts

Representative Mark White
Tennessee House of Representatives
Nashville, Tennessee

Grover J. "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

Table of Tables

Table of Figures

Executive Summary

The National Assessment Governing Board contracted with Pearson to design and implement a review of the achievement level descriptions (ALDs) for National Assessment of Educational Progress (NAEP) Reading and Mathematics assessments for grades 4, 8, and 12. This document describes the procedural and technical aspects and outcomes of the operational ALD Review study.

In particular, this report addresses the Board's updated achievement levels policy that called for the development of Reporting ALDs to state how the assessment content (defined by the Reporting ALDs) aligns with the existing content ALDs and achievement level policy definitions.

Background

The National Assessment Governing Board has a legislatively mandated responsibility to develop NAEP achievement levels. The Board 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 consistent across all NAEP assessments. 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 71 of the current Mathematics 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 response to the trial status stipulation. 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. This study used the 2019 Reading and Mathematics NAEP assessments at grades 4, 8, and 12. This work is based on the existing Mathematics and Reading frameworks, not the updated versions that will be used for the 2026 administration and beyond. Once the updated frameworks are operationalized, new ALD studies will be conducted.

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) and the 2009 NAEP Mathematics ALDs for grade 12 (Pitoniak, Dion, & Garber, 2010). The Governing 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 Governing 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. Sharyn Rosenberg, the Governing Board's Assistant Director for Assessment Development and Contracting Officer's Representative (COR) for this project, and 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 Governing 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 reading or mathematics 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 October 2021 to test all aspects of the logistical design of the ALD review procedures, including methods associated with a virtual meeting. The operational study was held in February 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 student in grades 4, 8, or 12 in either mathematics or reading. 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 in one of the NAEP grade levels and with the relevant subject area, mathematics or reading. 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 balanced by geographical region, gender, and race/ethnicity.

A multiphase process was used, that 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 32 panelists for the pilot study and 42 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 and ethnicity distribution for each panel in the operational study

	Mathematics			Reading		
	Grade 4	Grade 8	Grade 12	Grade 4	Grade 8	Grade 12
Gender						
Female	5	4	6	7	8	4
Male	1	2	0	1	1	3
Other	0	0	0	0	0	1
Race						
Asian	0	0	1	0	0	0
Asian/White	0	0	0	0	1	0
Black/African American	1	1	0	3	1	2
White	4	5	5	5	7	6
No Response	1	0	0	0	0	0
Ethnicity						
Hispanic/ Latin/ Spanish	2	1	1	0	2	0
No Response	4	5	5	8	7	8

Creation of Anchor Item Sets

For the ALD review pilot study, Pearson used a model-based approach for reviewing the alignment of the ALDs for NAEP Mathematics and Reading. The model-based approach included three stages. 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 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 mathematics, there were a total of 322 items for grade 4, 305 items for grade 8, and 238 items for

grade 12. For reading, there were a total of 163 items for grade 4, 212 items for grade 8, and 211 items for grade 12.

ALD Review Model-Based Approach Methodology

Panelists' activities during both the pilot and operational studies followed the same general process based on the model-based approach 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 about the knowledge and skills needed to respond 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 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, 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 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 cross-grade review to ensure consistency in format and language within the draft Reporting ALDs across the grades. An internal review of the draft Reporting ALDs was also completed by NCES content staff, to ensure that the statements in the draft Reporting ALDs did not conflict with the framework expectations, with adjustment made to the statements to address any identified issues. A final external review was completed by a selected set of potential users of NAEP data and the draft Reporting ALDs, to ensure clarity and usability of the statements, with some final adjustments to the statements.

Study Outcomes

The final results of the operational ALD Review study were presented at the August 2022 meeting of the Governing Board. Tables 2 and 3 present the results of the alignment judgment round 3 survey for mathematics and reading, respectively. The results presented provide evidence that the panelists observed 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.

Table 2. Round 3 mathematics alignment judgment agreement with content ALDs

Subject	Grade	NAEP Level	Alignment Judgment			
			Minimal	Weak	Moderate	Strong
Mathematics	4	Basic	0%	0%	25%	75%
		Proficient	0%	0%	0%	100%
		Advanced	0%	0%	0%	100%
	8	Basic	0%	0%	29%	71%
		Proficient	0%	0%	57%	43%
		Advanced	0%	0%	57%	43%
	12	Basic	0%	13%	13%	75%
		Proficient	13%	0%	50%	38%
		Advanced	13%	38%	50%	0%

Table 3. Round 3 reading alignment judgment agreement with content ALDs

Subject	Grade	NAEP Level	Alignment Judgment			
			Minimal	Weak	Moderate	Strong
Reading	4	Basic	0	0	50%	50%
		Proficient	0	0	17%	83%
		Advanced	0	0	17%	83%
	8	Basic	0	0	50%	50%
		Proficient	0	0	83%	17%
		Advanced	0	0	17%	83%
	12	Basic	0	0	33%	66%
		Proficient	0	0	0	100%
		Advanced	0	0	0	100%

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 developing the ALD Review study. During the operational study, there were two TAC members to observe the process and report the fidelity of the used process to the process used during the

study. The observations from the TAC members 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 study. The results of these process evaluations will be discussed further in the study.

For internal validity evidence, the panelist 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 Governing Board) is responsible for the development of achievement levels for NAEP. To help define the meaning of the achievement levels for NAEP, the Governing 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 Governing 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 Governing 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 Governing 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. This study would generate new Reporting ALDs that comply with the 2018 Governing Board policy statement.

Background on Current Project

The Governing Board awarded a contract to Pearson on September 24, 2020, to design and implement a study procedure to conduct anchoring studies using the 2019

NAEP data to review the mathematics and reading ALDs for grades 4, 8, and 12. 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 results of the 2021 and 2023 NAEP Mathematics and Reading assessments in grades 4, 8, and 12.
2. Review the alignment between the 2019 NAEP Mathematics and Reading 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 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 Governing 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 Governing Board to consider.

The methodology used for the NAEP ALD Review study had to conform to the Governing 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, so that the Reporting ALDs would comply with the Board policy statement.

Due to concerns related to the COVID pandemic, the ALD study was convened as virtual meetings. Pearson sent each panelist a Chromebook to join the virtual meeting, along with accessing documents and materials through the Pearson Standard Setting website. Panelists were also provided access to the NAEP Integrated Management System (IMS), where they were able to view and engage with items within the NAEP

administration system. This allowed panelists to work with the various types of items on the NAEP assessments, including passage-based items and situation-based tasks (SBT). 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 the online platform and accessed feedback using the online interface.

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 the Mathematics and Reading assessments at grades 4, 8, and 12 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 Governing Board actions and Pearson's recommendations for future studies.

Technical Advice

The project for reviewing the NAEP ALDs for mathematics and reading 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 serve on a TAC for the NAEP ALD Review study. The TAC convened for several meetings which were held virtually 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 Governing Board's COSDAM during quarterly Board meetings and scheduled virtual meetings between November 2020 and August 2022. Besides the members of the TAC, Dr. Sharyn Rosenberg, the Governing Board's Assistant Director for Assessment Development and Contracting Officer's Representative (COR) for this project, or Dr. Rebecca Norman Dvorak, the Governing 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 Governing Board and Assistant Director for Psychometrics at U.S. Department of Education
- Dr. Marianne Perie, Director of Assessment Research and Innovation, WestEd
- Dr. Mark Reckase, University of Distinguished Professor Emeritus, Michigan State University
- Dr. Lauress 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 were heavily involved in earlier NAEP standard setting 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 mathematics and reading 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 were responsible for logistics. The lead content facilitators for the project include Victoria Young for reading and Rick Wilmeth for mathematics. 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 32 panelists for the pilot study and 42 panelists for the operational study. Panelists were divided based on their education and experience into six panel groups that focused on completing the ALD review process for a specific subject and grade.

- Grade 4 Reading
- Grade 8 Reading
- Grade 12 Reading
- Grade 4 Mathematics
- Grade 8 Mathematics
- Grade 12 Mathematics

The target maximum composition of the study was a total of 48 panelists, with eight panelists for each subject and grade 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 which 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. For each panel, the recruitment effort had a goal of securing a panel where at least half of the panel were classroom teachers and at least two of the panelists being non-classroom educators.

NAEP studies have used a split panel design since studies were conducted in the early 1990's. As was typical for past studies, the project staff assigned the individuals from each panel into two replicate groups that were as equal as possible with respect to panelist demographics and panel recruitment targets. Each replicate group was selected to be as equivalent as possible with respect to demographic characteristics and to include at least one non-classroom educator. Each group worked through the same process using the same set of items.

Pearson identified the panelists through an iterative multi-phase 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 and panelists.

Identifying Nominees

Panelist nominators were recruited using multiple sources. One source for nominators were professional organizations that have strong backgrounds in providing professional development in mathematics education or reading and literacy education. The following organizations were among those targeted for recruiting panelists:

- Association of Literacy Educators and Researchers
- Conference on English Leadership
- National Council of Teachers of Mathematics
- National Council of Supervisors of Mathematics
- Teach Plus

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 which are currently teaching the related subject in a classroom, where 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 questionnaire regarding individuals which 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, panelist type, or grade level. 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 questionnaire 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 teacher courses in the specific subject area.

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 the attempt to create panels that were representative of educators across the country. While the goal of the selection process was to have 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 were 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.

Preparation of Panelists

Panelists were provided support prior to the meeting, to assist them in preparing to engage in the meeting activities, as a virtual meeting. Panelists were sent a computer for the process, so they were each using the same technology. Reading panelists were sent an additional PC laptop to assist with viewing secure items that required a special locked down browser, which was provided by NCES. Since the panelists had access to secure information and NAEP items during the ALD review meeting, the panelists 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 pilot and operational ALD review meetings, multiple technical check-in meetings were scheduled on different days and various times to ensure that there was an available time that each panelists could attend. During the check-in meeting, the panelists accessed the meeting materials and activities through the Pearson standard setting website and accessed items through the NAEP integrated management system (IMS). The pre-meeting technical check-in meetings were facilitated virtually using Zoom.

Once panelists completed the technical check and were able to access the standard setting website, they were 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 NAEP framework for their respective subject, reading or mathematics, from the 2019 administration.
- Access to download and install Zoom for the virtual meeting.

Panelists were sent emails during the week prior to the meeting to encourage them to complete the pre-meeting work and remind them of the starting date and time for the meeting.

Development of Anchor Item Sets

For the ALD review and revision pilot study, Pearson used a model-based approach for reviewing the alignment of the ALDs for NAEP mathematics and reading. The Governing 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 model-based 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. Prior to selecting the item anchoring approach used for the pilot study, several variations were investigated. The anchoring approaches that were investigated included:

- The anchoring method used during the 2009 NAEP anchoring studies,
- Item response theory (IRT) item mapping approaches,
- The anchoring method used for the Trends in International Mathematics and Science Study (TIMSS),
- The method from the 2009 NAEP anchoring studies using the criteria from the TIMSS study.

The investigated methods and the data were shared and discussed with the TAC. Each method resulted in different anchor item sets, which resulted from different conceptualizations of the set of students used to calculate the conditional probability utilized to anchor the item to an achievement level. It was recommended by the TAC that the methodology from the previous NAEP alignment studies for mathematics and reading in 2009 would be the most defensible approach. This is because of the need to make consistent inferences about what the ALDs represent across the years by using the same anchoring methodology.

The development of the anchor item sets starts by grouping performances of individual students from the most recent (2019) administration of the grade 4, grade 8, and grade 12 NAEP Mathematics and Reading assessments into achievement levels.

The achievement level classification for each student performance is based on the average of their NAEP "plausible values". 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 classified into the region just below *NAEP Basic* when their average plausible value was below the cut score for *NAEP Basic* and the region between the mean plausible value and two standard errors above the mean includes the cut score for *NAEP Basic*. This approach used all students in the NAEP sample from the most recent administration in 2019 to ensure that there are 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 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 four achievement levels. Items were anchored to the first achievement level where 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 content area for mathematics or passage type for reading. 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 1 through 6 present the number of items anchored to each achievement level.

Table 4. Achievement level counts for mathematics grade 4

Achievement Level	Mathematics Content Domains					Total
	Number Properties and Operations	Algebra	Geometry	Measurement	Data and Statistics	
Does Not Anchor	22	19	17	10	13	81
NAEP Advanced	23	12	12	13	9	69
NAEP Proficient	28	10	7	14	6	65
NAEP Basic	20	10	7	7	4	48
Below *NAEP Basic*	16	6	6	7	4	39
Total	109	57	49	51	36	302

Table 5. Achievement level counts for mathematics grade 8

Achievement Level	Mathematics Content Domains					Total
	Number Properties and Operations	Algebra	Geometry	Measurement	Data and Statistics	
Does Not Anchor	17	14	11	20	10	72
NAEP Advanced	13	25	7	16	13	74
NAEP Proficient	25	26	18	10	7	86
NAEP Basic	7	12	7	4	5	35
Below *NAEP Basic*	8	8	6	6	10	38
Total	70	85	49	56	45	305

Table 6. Achievement level counts for mathematics grade 12

Achievement Levels	Mathematics Content Domains				Total
	Number Properties and Operations	Algebra	Geometry and Measurement	Data, Statistics, and Probability	
Does Not Anchor	2	17	16	12	47
NAEP Advanced	6	21	22	17	66
NAEP Proficient	10	28	18	22	78
NAEP Basic	4	6	6	10	26
Below *NAEP Basic*	4	4	7	6	21
Total	26	76	69	67	238

Table 7. Achievement level counts for reading grade 4

Achievement Level	Reading Passage Types		Total
	Literary Passages	Informational Passages	
Does Not Anchor	16	22	38
NAEP Advanced	17	16	33
NAEP Proficient	20	28	48
NAEP Basic	21	12	33
Below *NAEP Basic*	8	3	11
Total	82	81	163

Table 8. Achievement level counts for reading grade 8

Achievement Level	Reading Passage Types		Total
	Literary Passages	Informational Passages	
Does Not Anchor	16	13	29
NAEP Advanced	16	19	35
NAEP Proficient	24	27	51
NAEP Basic	24	24	48
Below *NAEP Basic*	25	24	49
Total	105	107	212

Table 9. Achievement level counts for reading grade 12

Achievement Level	Reading Passage Types		Total
	Literary Passages	Informational Passages	
Does Not Anchor	8	10	18
NAEP Advanced	17	29	46
NAEP Proficient	19	50	69
NAEP Basic	8	32	40
Below *NAEP Basic*	16	22	38
Total	68	143	211

For mathematics, 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 provide distinction of what students with performance in the *NAEP Advanced* 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 reading assessment, the items within a content area were presented as sets with their associated passage. The passages were ordered by average *p*-value. The items associated with a typical passage set were also ordered by achievement level

and by average conditional *p*-value. For items associated with Situation Based Tasks (SBT), the items were presented in administration order.

ALD Review Study Process

The 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 were used for the pilot and operational studies.

Orientation to the ALD Review Study

On the first day of the meeting, panelist accessed the Zoom meeting link through the Pearson website. There was a unique Zoom link for reading and one for mathematics. When panelists joined the meeting, attendance was taken. The meeting began with an introduction followed by an overview of the NAEP Assessment program by a staff member from the Governing Board. An overview of the ALD review along with security and meeting ground-rules was presented by Dr. Eric Moyer, for the mathematics group, or Dr. Jennifer Galindo, for the reading group. The lead content facilitators for each subject provided training on the key components of the NAEP framework for the relevant assessment. For reading, this involved the different item types, passage types/genres for the respective grade level, and a discussion on text complexity. For math, this involved a discussion of the specific item types, the content included within each subscale, and the use of calculators.

Following the general session presentation, the panelists were then split into grade-level breakout groups on Zoom. Each of the grade-level breakout groups had a lead content facilitator, assistant content facilitator, and a moderator. 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 role of the moderator was to monitor the Zoom chat and assist with technology issues.

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 presented during the 2019 administration of NAEP.

The mathematics panelists accessed a link on the Chromebook that took them out to the IMS system where they were able to navigate through the set of items. For the reading group, a NAEP lockdown browser was installed on the separate PC and the panelists were instructed on how to access the NAEP IMS through the lockdown browser. Panelists were then able to review the passages with the related set of items. The reading group was required to view the items through the separate NAEP lockdown browser because when reading items with the passages are viewed in the current version of Chrome the item does not render correctly.

Due to the use of the IMS system, some panelists struggled to access the items. This was a result of lost passwords, Statements of Behavior agreements that weren't received by NCES, and PC issues. In most cases, the issues were resolved, and the panelists were able to conduct this activity individually. In some groups, the lead facilitator resolved the issues by sharing the screen and navigating through the items as a group rather than individually.

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

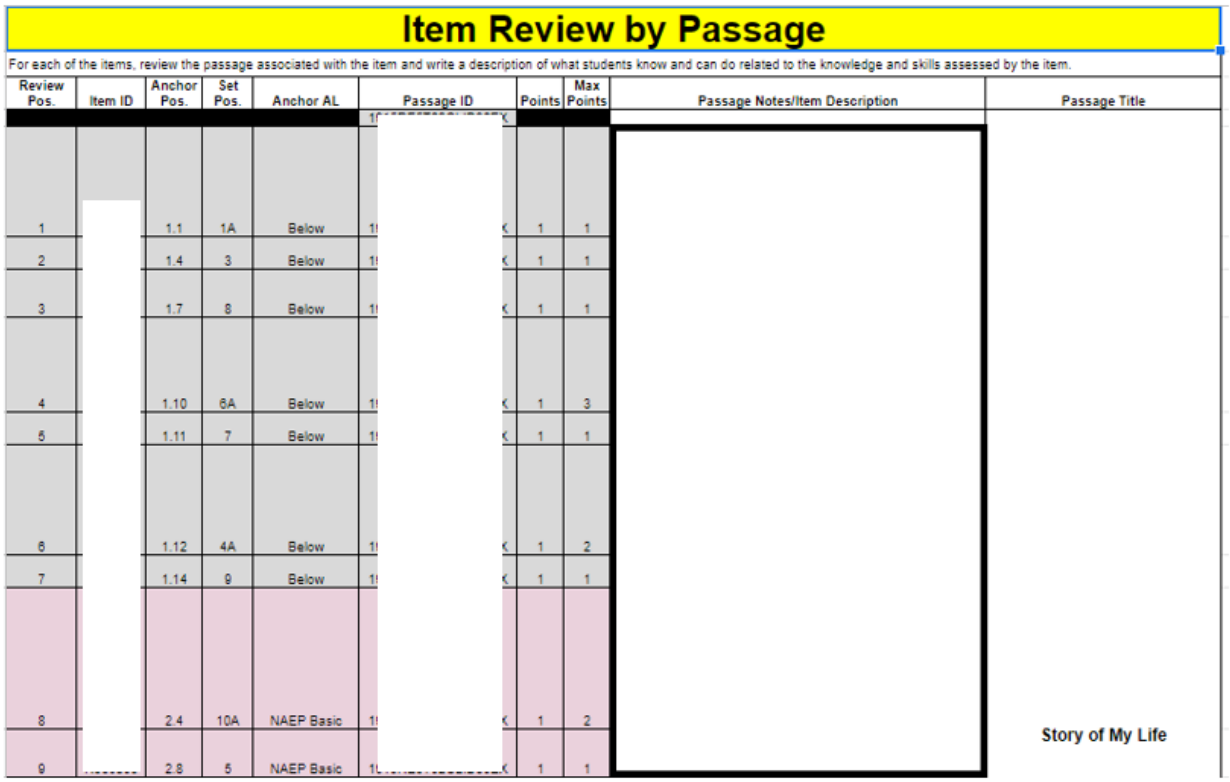
The lead facilitator described the process that was used to create the anchor item sets and an orientation to the information that is provided for each item. The panelists were also trained in how the items are scored with the item key or scoring rubrics. Panelists were oriented in the structure of the 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 they access the individual item review spreadsheets, assess the passages for reading only, enter item descriptions, and the items for review. Figures 1 and 2 contain 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 (Mathematics)



Item Review by Passage

For each of the items, review the passage associated with the item and write a description of what students know and can do related to the knowledge and skills assessed by the item.

Review Pos.	Item ID	Anchor Pos.	Set Pos.	Anchor AL	Passage ID	Points	Max Points	Passage Notes/Item Description	Passage Title
1		1.1	1A	Below		1	1		
2		1.4	3	Below		1	1		
3		1.7	8	Below		1	1		
4		1.10	6A	Below		1	3		
5		1.11	7	Below		1	1		
6		1.12	4A	Below		1	2		
7		1.14	9	Below		1	1		
8		2.4	10A	NAEP Basic		1	2		
9		2.8	5	NAEP Basic		1	1		Story of My Life

Figure 2. Sample Item Review Spreadsheet (Reading)

Following the training, panelists conducted the independent item review. This activity was completed one subscale at a time. For the reading panels, the two subscales were based on the types of passages included as part of the assessment: literary passages and informational passages. For the mathematics panel, the subscales were based on the mathematics content domains: number properties and operations, algebra, geometry, measurement, and data, statistics, and probability. To complete the activity, panelists accessed the anchor item sets for the specific subscale via the Pearson website. For mathematics, the sets were organized by achievement level and item difficulty. For reading, in most cases the items were arranged by passage set and then by item difficulty. For the scenario-based tasks, the items were presented in administration order. 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 skill necessary to answer each of the items correctly or to receive the indicated score point. For reading, there was also a place for notes to be taken about the passages.

In the website, panelists were provided the items as screenshots, which were provided 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 mathematics panelists were able to access the NAEP IMS on their Chromebook, where the reading panelists needed to access the NAEP IMS through the lockdown browser on the separate PC. 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 questions was to help facilitate the thinking of a summary description individually prior to the group discussion. Approximately half-way 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.

The summary statements were to indicate the knowledge and skills that students with achievement 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 ensured 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 took additional time to assist the groups in creating 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 can likely demonstrate in relation to the NAEP assessments. Individual groups within each subject and grade 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 student 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 3 displays the PowerPoint slide that was used to discuss the alignment categories with the panelists.

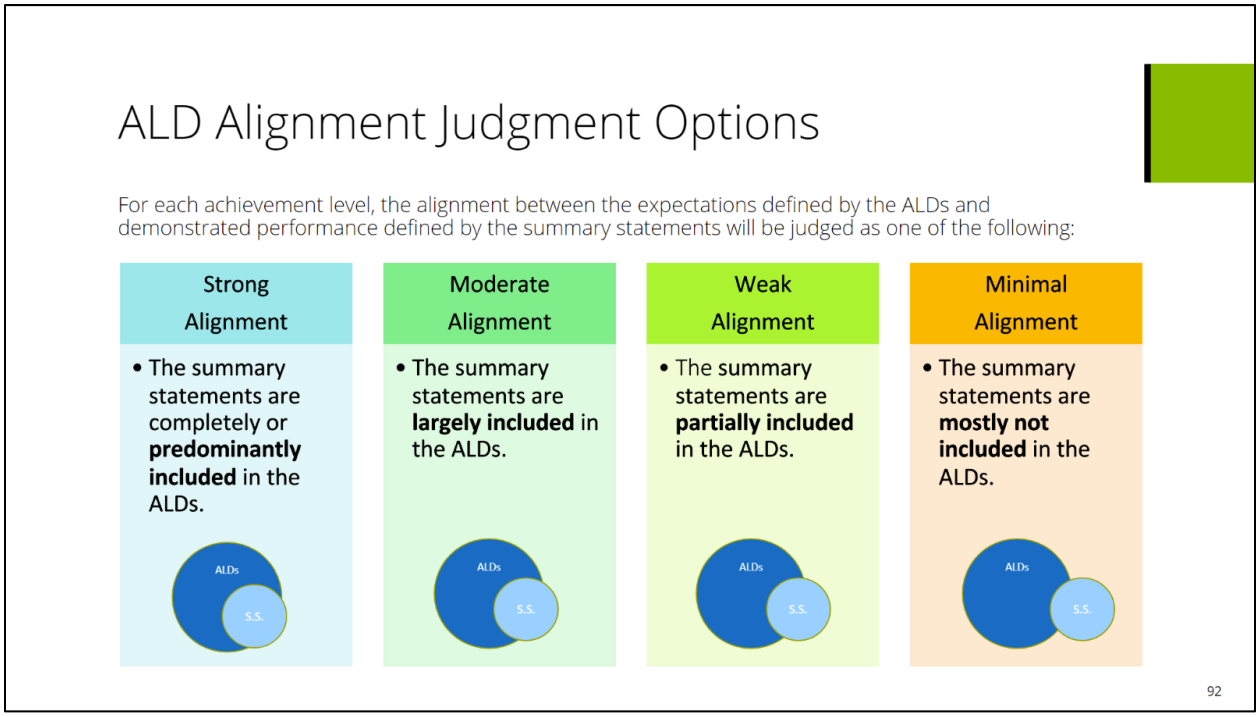


Figure 3. ALD Alignment Judgment Training Slide

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 the basis for their recommendation 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 the 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 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

Of the individuals nominated to participate in the ALD Review study, 32 individuals indicated that they were available to participate in the pilot study. Since the 32 panelists that indicated their availability for the pilot study was less than the maximum of 48 panelists, we reviewed the qualifications of each of the possible panelists with the COR. 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 32 panelists to participate in the pilot study.

There were six panels that focused on completing the ALD review process for a specific subject and grade level 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. 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 expected. This was attributed to the difficulty of teachers taking off for a week, since it was challenging to find substitutes during the COVID-19 pandemic. Table 10 summarizes information about the panelist type for panel members who participated in the pilot study.

Table 10. Panelist type distribution for each panel

Type	Mathematics			Reading		
	Grade 4	Grade 8	Grade 12	Grade 4	Grade 8	Grade 12
Classroom Teacher	2	2	2	2	3	2
Non-Classroom Educator	4	4	3	3	2	2
Total	6	6	5	5	5	4

Table 11 summarizes the gender and ethnicity distributions for the panel members who participated in the pilot study.

Table 11. Gender and ethnicity distribution for each panel

	Mathematics			Reading		
	Grade 4	Grade 8	Grade 12	Grade 4	Grade 8	Grade 12
Gender						
Female	6	5	2	4	5	3
Female/Other	0	0	0	0	0	1
Male	0	1	3	1	0	0
No Response	0	0	0	0	0	0
Race						
Black, African-American	1	1	0	0	4	0
White, American Indian, Alaska Native	0	0	0	1	0	0
White	4	5	5	3	2	4
No Response	1		0	1	0	0
Ethnicity						
Hispanic/ Latin/ Spanish	0	0	0	1	0	0
No Response	0	0	0	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 12 summarizes the distribution of the panel members who participated in the pilot study across the geographic regions.

Table 12. Distribution for each panel across geographic regions

	Mathematics			Reading		
Region	Grade 4	Grade 8	Grade 12	Grade 4	Grade 8	Grade 12
Northeast	1	1	0	0	0	0
Midwest	0	1	1	2	0	0
South	5	4	2	2	4	3
West	0	0	2	1	1	1

Table 13 summarizes the education experience with students from special populations for the panel members who participated in the pilot study.

Table 13. Experience of panel members with special populations

	Mathematics			Reading		
Experience	Grade 4	Grade 8	Grade 12	Grade 4	Grade 8	Grade 12
English Language Learners	6	4	4	3	5	3
Mainstream Special Education	6	5	5	5	4	4
Self-contained Special Education	3	0	2	1	0	0
Gifted and Talented Education	5	5	4	4	4	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 committee 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 moderator helped with technology issues and monitored the Zoom chat. The facilitators and moderators for each panel are shown in Table 14.

[redacted]

Content Area	Grade	Lead Facilitator	Assistant Facilitator	Moderator
Reading	4	Melia Franklin	Noemi Nolter	Mark Robeck
	8	Victoria Young	Lillian Moore	Brad Ungurait
	12	Carol Jago	Tammy Visco	Brian Wrobel
Mathematics	4	Lois Yoder	Mandy Speights	Sarah Esparza
	8	Rick Wilmeth	Mandip Gill	Grant Smith
	12	Kate Brian	Michael Morony	Imelda Martinez

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 Zoom platform – The facilitators were provided an overview of the online meeting platform.
- 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. The website provided a framework for facilitating the ALD review process that each of the facilitators followed. Specific guidelines for modeling the website and 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 on 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 walkthrough 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.

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 one staff member from the Governing Board, Dr. Sharyn Rosenberg.

ALD Review Pilot Study Process

The model-based ALD review process that was described in the design document was successfully implemented for the pilot study. The process was implemented using a virtual process using Pearson Chromebooks and PC laptops that were sent to the panelists, as approved by the TAC and the Governing Board.

The schedule for the pilot study was modified from the agenda that was originally developed for the pilot study. Some activities took more time than initially anticipated, requiring for the decrease of allowed time for some activities, specifically the individual alignment judgement activity. Additionally, the cross-grade draft reporting ALD review that was scheduled as part of the pilot meeting was completely removed from the schedule, after discussion with the TAC observers and Dr. Sharyn Rosenberg. The agenda for the operational ALD Review study was adjusted based on information collected during the pilot study.

As a virtual process, the panelists participating in the pilot study mostly worked effectively individually, as table groups, and as a whole group. When required by the process, they 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. There were some circumstances where panelists were not exclusively in secure settings while completing the tasks. When this was identified, Pearson study staff addressed this with the individual to ensure that they could move to a secure location for the rest of the meeting. This increased the information related to the need for a secure location for the operational meeting.

Panelists did not report any challenges working virtually with the Zoom interface or the Pearson standard setting website. There were some initial challenges reported with getting the information required to access the NAEP IMS prior to the meeting. The Pearson staff worked with panelists that had difficulty accessing IMS, with the support of NCES staff, to resolve all these issues. These issues informed the development of refined technology check-in meetings for the operational ALD Review study.

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. There were also two intended outcomes from the pilot study.

1. Draft Reporting ALDs that describe what students within each achievement level actually can 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 from the NAEP frameworks that describe what students at each achievement level should know and be able to do.

The panelists were able to work together after the round 1 alignment judgment feedback discussion to develop panel-level summary statements. Some panelists stated that creating the panel summary statement was a more challenging task than the initial creation of the summary statements. One of the challenges identified by the panelists was that they were combining group summary descriptions that were created one or two days previously and were not as familiar with the specific items that supported their creation.

When reviewing the final summary statement from the pilot ALD review meeting there was some inconsistency in the level of specificity in the statements from the different replicate groups. This was especially demonstrated with grade 12 mathematics panel, where panelists did not seem to keep to the process guidelines of making summary descriptions within an achievement level only when at least two items were available. This was due to the large number of concepts covered by the grade 12 Mathematics NAEP assessment, which resulted in having single items addressing specific content within an achievement level. Another possible limitation to the process of creating summary descriptions for the grade mathematics panel was the limited number of items available to cover the wide range of content objectives for the subject. More training in the process for the panelists was added to the operational study with the addition of a modeling activity to determine a more cohesive level of specificity for the summary descriptions.

When reviewing the rationale from the panelists for the different alignment judgments, there seemed to be various perspectives in what led to a judgment indicating a lack of alignment. There were some panelists that indicated a lack of alignment between the content ALDs and summary descriptions because a topic was found in the content ALDs, but not in the summary descriptions, but this could be from a sampling issue, where some content may not have been included in the test.

Process Evaluations

As stated earlier, process evaluations were administered at different parts of the process. Panelists' responses to these questions suggest that most panelists were comfortable with 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 15–18,

Table 15. Panelists' response to the statement: "I understand the steps to follow as I completed the independent item review activity."

Subject	Grade	Strongly Disagree	Disagree	Slightly Disagree	Slightly Agree	Agree	Strongly Agree
Mathematics	4	0%	0%	0%	0%	17%	83%
	8	0%	0%	0%	0%	0%	100%
	12	0%	0%	0%	0%	0%	100%
Reading	4	0%	0%	0%	0%	25%	75%
	8	0%	0%	0%	0%	80%	20%
	12	0%	0%	0%	0%	20%	80%

Table 16. Panelists' response to the statement: "I understood the steps to follow as I completed the individual alignment judgment activity."

Subject	Grade	Strongly Disagree	Disagree	Slightly Disagree	Slightly Agree	Agree	Strongly Agree
Mathematics	4	0%	0%	0%	33%	50%	17%
	8	0%	0%	0%	0%	50%	50%
	12	0%	0%	0%	0%	40%	60%
Reading	4	0%	0%	0%	0%	20%	80%
	8	0%	0%	0%	0%	100%	0%
	12	0%	0%	0%	0%	25%	75%

The panelists also indicated that they were confident in the panel-level summary descriptions and ALD alignment judgments, as shown in Tables 22 and 23.

Table 17. Panelists' opinion regarding their confidence in the summary statements by the panel

Subject	Grade	ALD	Strongly Disagree	Disagree	Slightly Disagree	Slightly Agree	Agree	Strongly Agree
Mathematics	4	*NAEP Basic*	0%	0%	0%	0%	33%	67%
		NAEP Proficient	0%	0%	0%	0%	17%	83%
		NAEP Advanced	0%	0%	0%	0%	33%	67%
	8	*NAEP Basic*	0%	0%	0%	0%	0%	100%

Subject	Grade	ALD	Strongly Disagree	Disagree	Slightly Disagree	Slightly Agree	Agree	Strongly Agree
		NAEP Proficient	0%	0%	0%	0%	0%	100%
		NAEP Advanced	0%	0%	0%	0%	0%	100%
	12	*NAEP Basic*	0%	0%	0%	0%	60%	40%
		NAEP Proficient	0%	0%	0%	0%	60%	40%
		NAEP Advanced	0%	0%	0%	0%	60%	40%
Reading	4	*NAEP Basic*	0%	0%	0%	0%	20%	80%
		NAEP Proficient	0%	0%	0%	0%	20%	80%
		NAEP Advanced	0%	0%	0%	0%	20%	80%
	8	*NAEP Basic*	0%	0%	0%	0%	67%	33%
		NAEP Proficient	0%	0%	0%	0%	67%	33%
		NAEP Advanced	0%	0%	0%	0%	67%	33%
	12	*NAEP Basic*	25%	0%	0%	0%	0%	75%
		NAEP Proficient	25%	0%	0%	0%	0%	75%
		NAEP Advanced	25%	0%	0%	0%	0%	75%

Table 18. Panelists' opinion regarding their confidence in the ALD alignment judgments by the panel.

Subject	Grade	ALD	Strongly Disagree	Disagree	Slightly Disagree	Slightly Agree	Agree	Strongly Agree
Mathematics	4	*NAEP Basic*	0%	0%	0%	17%	33%	50%
		NAEP Proficient	0%	0%	0%	17%	33%	50%
		NAEP Advanced	0%	0%	0%	17%	83%	0%
	8	*NAEP Basic*	0%	0%	0%	0%	0%	100%
		NAEP Proficient	0%	0%	0%	0%	0%	100%
		NAEP Advanced	0%	0%	0%	0%	25%	75%
	12	*NAEP Basic*	0%	0%	0%	0%	60%	40%
		NAEP Proficient	0%	0%	0%	0%	60%	40%
		NAEP Advanced	0%	0%	0%	0%	60%	40%
Reading	4	*NAEP Basic*	0%	0%	0%	0%	20%	80%
		NAEP Proficient	0%	0%	0%	0%	20%	80%
		NAEP Advanced	0%	0%	0%	0%	20%	80%
	8	*NAEP Basic*	0%	0%	0%	0%	67%	33%
		NAEP Proficient	0%	0%	0%	0%	67%	33%
		NAEP Advanced	0%	0%	0%	0%	67%	33%
	12	*NAEP Basic*	25%	0%	0%	0%	50%	25%
		NAEP Proficient	25%	0%	0%	0%	25%	50%
		NAEP Advanced	25%	0%	0%	0%	50%	25%

Lessons Learned

Feedback from the pilot study provided information in multiple areas the supported adjustments prior to the operational ALD review meeting. All adjustments described in this section were based on discussion and approval by the TAC and the Governing Board. Table 16 presents the places in the process where improvements could be made and what revisions would be made.

Table 19. Pilot study lessons learned and plans of action

Meeting Segment	Lesson Learned	Plan of Action
Pre-Meeting	We were only able to identify 32 qualified panelists who were willing and able to participate.	Communicate with panelists who could not participate in pilot meeting to recruit for operational meeting. Continue nomination process, starting process earlier than for pilot meeting, completed by December 10. Complete recruiting process by January 14, so panels can be selected and locked in time to send panelist materials and equipment by February 7.
	Some panelists did not receive all computers, monitors, log-ins, etc., in advance of the meeting starting.	Equipment will be sent to panelists by February 7 to ensure that all panelists receive equipment. Panelists will participate in a technology "check-in" meeting to ensure that all panelists have set up equipment and have access to the meeting websites and tools.

Meeting Segment	Lesson Learned	Plan of Action
		Communication plan is adjusted to streamline communication with panelists and ensure panelists receive all required information prior to meeting.
	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	General session and orientation included duplicate instructions about NAEP.	Presentation of general session materials will be modified to streamline presentation of NAEP information.
Item Review Activity	Panelists provided item descriptions at various levels of specificity.	Panelist training in item review activity process will use items to model level of specificity in item descriptions.
	Panelists did not have enough time to complete individual review activity.	Agenda will be adjusted to provide panelists the time needed to complete individual reviews.

Meeting Segment	Lesson Learned	Plan of Action
Group Summary Descriptions	Groups provided summary descriptions at varying levels of specificity.	Facilitators will model the creation of summary descriptions. As part of the modeling process, the panelists will be shown example summary descriptions at the appropriate level of specificity, close to the content ALDs with some additional detail. Panelists will be trained to write summary descriptions for groups using bullets. Panelists will create group summary descriptions based on at least two items, either within achievement level or across achievement levels, with a focus on the "big ideas" covered by the framework for the specific grade and subject
	Panelists 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.	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. Facilitators will remind panelists of the step in creating the panel summary descriptions throughout process.

Meeting Segment	Lesson Learned	Plan of Action
	Panelists needed additional time to complete panel summary descriptions.	Agenda will be revised to provide additional time for the panel to complete the creation of the panel summary descriptions.
Alignment Judgment Survey	Panelists could have different reasons for selecting alignment judgments.	The training on the alignment judgments will include discussions about various reasons for the different classifications, including that the summary descriptions describe statements that are not contained within the achievement level or the summary descriptions describe content not included in the achievement level. The discussion will also include discussions about what does not a source of misalignment. The alignment judgment survey process will include questions to capture specifics about the panelist judgment rationale. As part of the feedback discussion after round 3 alignment judgments, if there are any alignment classifications of "Weak" or "Minimal" the facilitators will lead a discussion to determine the source of any concerns related to alignment.
Cross-Grade Review	Cross-grade review was not completed because of the need for additional time in other activities.	The work to create summary descriptions at similar levels of specificity will assist with alignment across grades.

Meeting Segment	Lesson Learned	Plan of Action
		The cross-grade review could be conducted by content experts following the operational meeting but prior to public comment. Additional information about coherence across grades could be collected through the public comment process.

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. Due to current challenges in education including the Omicron surge of COVID-19, which was peaking during the final recruitment stages, several individuals confirmed that they were interested in participating but were unable to due to shortages of substitute teachers and lack of administrative support for being absent for five days. The week of the meeting, February 21-25, 2022, included Presidents' Day, which could be a vacation time for some schools, which also impacted panelists' commitment. A total of 43 panelists started the operational study, including 21 teachers and 22 non-teacher educators. During the meeting, one panelist dropped from the study, resulting in 42 panelists completing the study. The target composition of the study was a total of 48 panelists, eight panelists for each subject and grade specific panels, with at least half of the panelists composed of teachers and at least two panelists being non-classroom educators. Table 20 summarizes information about the panelist type for panel members who participated in and completed the pilot study.

Table 20. Panelist type description for each panel

	Mathematics			Reading		
Type	Grade 4	Grade 8	Grade 12	Grade 4	Grade 8	Grade 12
Classroom Teacher	4	4	4	2	4	4
Non-Classroom Educator	4	4	4	4	2	2
Total	8	8	8	6	6	6

There were six panels that focused on completing the ALD review process for a specific subject and grade level 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. Each replicate group was selected to have both teachers and non-classroom educators, along with representation across other classifications.

The composition of the panels for the operational study was to include panelists who reflected an overall balance of gender, race/ethnicity, geographic location and student experience. The ultimate 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 provide meaningful judgments. Table 21 summarizes the gender and ethnicity distributions for the panel members who participated in the pilot study.

Table 21. Gender and ethnicity distribution for each panel

	Mathematics			Reading		
	Grade 4	Grade 8	Grade 12	Grade 4	Grade 8	Grade 12
Gender						
Female	7	7	4	5	4	6
Male	1	1	3	1	2	0
Other	0	0	1	0	0	0
Race						
Asian	0	0	0	0	0	1
Asian/White	0	1	0	0	0	0
Black/African American	3	1	2	1	1	0
White	5	6	6	4	5	5
No Response	0	0	0	1	0	0
Ethnicity						
Hispanic/ Latin/ Spanish	0	1	0	0	1	1
No Response	0	0	0	0	0	0

There is 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 22 summarizes the distribution of the panel members who participated in the pilot study across the geographic regions.

Table 22. Distribution for each panel across geographic regions

Region	Mathematics			Reading		
	Grade 4	Grade 8	Grade 12	Grade 4	Grade 8	Grade 12
Northeast	1	0	1	0	0	0
Midwest	2	1	1	2	1	1
South	4	5	3	4	4	4
West	1	2	3	0	1	1

Table 23 summarizes the education experience with students from special populations for the panel members who participated in the pilot study.

Table 23. Experience of panel members with special populations

Experience	Mathematics			Reading		
	Grade 4	Grade 8	Grade 12	Grade 4	Grade 8	Grade 12
English Language Learners	4	6	5	3	5	5
Mainstream Special Education	5	7	8	6	6	6
Self-contained Special Education	2	2	0	4	2	0
Gifted and Talented Education	4	6	7	3	5	5

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 moderator helped with technology issues and monitored the Zoom chat. The facilitators and moderators for each panel are shown in Table 24.

Table 24. Facilitators and moderators for each panel

Content Area	Grade	Lead Facilitator	Assistant Facilitator	Moderator
Reading	4	Melia Franklin	Noemi Nolter	Mark Robeck
	8	Victoria Young	Lillian Moore	Grant Smith
	12	Carol Jago	Cristina Everett	Brian Wrobel
Mathematics	4	Lois Yoder	Mandy Speights	Sarah Esparza
	8	Mandip Gill	Bobby Morgan	Brad Ungurait
	12	Kate Brian	Michael Morony	Imelda Martinez

All but two facilitators for the operational study were facilitators during the pilot study for the same subject and grade. 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 two facilitators that were new to the study received additional training from the previous facilitators and the project director to ensure that they were prepared to lead the meeting for their groups. All trainings were conducted virtually.

Observers

In addition to the facilitators, there were four observers of the operational study. They included two members of the TAC, Dr. Karla Egan and Dr. Susan Loomis. There was also two staff members from the Governing Board, Dr. Sharyn Rosenberg and Dr. Rebecca Norman Dvorak.

ALD Review Operational Study Process

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. The activities and materials for the study were standardized across subjects and grades, as much as possible. The following sections provide descriptions of the activities each panel completed.

Pre-Meeting Technical Check

The process was implemented using a virtual process using Pearson Chromebooks and PC laptops that were sent to the panelists, as approved by the TAC and the Governing Board. As a virtual meeting there were additional pre-meeting preparations for the panelists to complete. Panelists were sent technology so that all panelists were using the same computers for the process. As defined based on the lessons learned from the pilot study, prior to the ALD Review study, multiple technical check-in meetings were scheduled on different days and at various times to ensure that there was an available time that each panelist could attend. The technical check-in meetings were scheduled during the week prior to the review study. All panelists attended at least one of the check-in meetings. Prior to the check-in meeting, panelists were supposed to complete the process of collecting all information to log-in to the different websites that would be used during the meeting, including the Pearson standard setting website and the IMS system. During the technology check-in meeting, the panelists signed onto the Zoom meeting and were asked to confirm that each panelist had accessed the websites. If they had any issues or questions, the leader of the check-in meeting would try to assist the panelist in resolving their issues.

Independent Item Review

Based on the recommendations after the pilot study, after the training from the facilitators in completing the individual item review, the facilitators conducted a modeling activity. The modeling activity was conducted 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. During the pilot meeting, the amount of time scheduled to complete the independent item review was found to be not enough. So, for the operational study more time was provided in the agenda for the item review activity and the creation of the summary descriptions.

Group and Panel Summary Statement Development

During the pilot study, it was found that the groups needed more time to both create the initial summary statements in their replicate groups and to combine the group-level summary statements into panel summary descriptions. To address this issue during the operational study, the agenda provided more time to complete the summary statement development activity for each subset of items and to create the panel-level summary descriptions.

To assist with achieving the correct level of specificity in the summary statements, additional training was provided for the panelists to understand the amount of supporting evidence needed to create a summary statement. The evidence for a summary statement needed to consist of two items from the same achievement level that supported the statement or two items from adjacent achievement levels that provided differentiation between what is demonstrated at the different levels.

Alignment Judgments

A result of the pilot study was that there seemed to be some confusion about what constituted a lack of alignment. To address this issue, additional training was added to the alignment judgment process training that discussed the different rationale for a lack of alignment.

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

Additionally, it was emphasized that if knowledge and skills is associated with an achievement level on the content ALDs, but not represented in the Reporting ALDs,

that this is not a reason for misalignment, because this could be a sampling issue of the assessment.

To emphasize this representation of alignment and misalignment, the alignment judgment survey was adjusted. After the initial questions from the pilot study, which asked panelists about their perception of the strength of the alignment between the content ALDs and the Reporting ALDs, the panelists were asked about their rationale for judgment. When selecting an alignment rating other than "Strong," panelists were asked to indicate one of two possible rationales. The rationale options are:

- Judgment Rationale #1 – the summary statements (SS) for a particular achievement level include knowledge, skills, and processes that are associated with a different achievement level in the content ALDs.
- Judgment Rationale #2 **–** the summary statements (SS) for an achievement level include knowledge, skills, and processes that are not associated with any achievement level in the content ALDs.

Additionally, the panelists were asked to provide a written explanation for their judgment and rationale.

Results

After the operational NAEP ALD review study, the lead facilitators, who were content experts in the specific subject and grade, conducted a cross-grade review of the draft Reporting ALDs drafted by the committees. During the cross-grade review, the lead facilitators from each content area reviewed the language and format of the draft Reporting ALDs across the grades and achievement levels. Minor adjustments were made to improve the cohesiveness of the statements across the grade levels. The facilitators were mindful that the edits did not impact the expectations communicated by the committees. The draft Reporting ALDs resulting from the cross-grade review were created as both statements that paralleled the format of the content ALDs from the NAEP framework and bulleted lists so the most useful format for the Reporting ALDs could be determined.

During the COSDAM meeting held April 2022, the draft Reporting ALDs from the cross-grade review were presented for discussion. The recommendation from COSDAM, which was agreed to by the Governing Board, was to present the draft Reporting ALDs

as bulleted lists with the probability language of "likely" to differentiate the Reporting ALDs from the content ALDs.

The draft Reporting ALDs from the cross-grade review, as bulleted lists, went through both internal and external reviews. The internal and external reviews were designed to improve the overall clarity and utility of the Reporting ALDs. For the internal review NCES and NAEP contractors provided written feedback related to the level of specificity, vocabulary, and the order in which the content was presented. The feedback from the internal review was reviewed by the lead content facilitators and discussed with NCES 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 workshop participants and included the original statements developed during the workshop. The participants who responded did not have concerns with the revisions.

The external review was conducted to ensure that the Reporting ALDs were clear to the individuals that would most likely access and use them. The external review yielded 22 responses with representation across 13 states. There were 11 reviewers who provided input on the mathematics Reporting ALDs (five had no concerns, six provided feedback), and 12 for reading (six had no concerns, six provided feedback). Pearson content leads reviewed all feedback and made adjustments when a) modifications would lead to increased clarity, and/or b) suggestions were supported by the assessment data and did not result in substantive changes against what the workshop participants intended.

The Governing Board reviewed and took action on the final Reporting ALDs for Mathematics and Reading during the August 2022 meeting. The final Reporting ALDs for Mathematics and Reading are presented in Tables 25 through 30.

Table 25. Mathematics Grade 4 - Reporting ALDs

Achievement Level	Reporting ALDs
NAEP Basic	Students performing at the NAEP Basic achievement level can likely • determine place value of whole numbers up to hundred thousands • locate whole numbers on a number line • read, write, compose, and decompose multi-digit whole numbers in a variety of forms based on place value • identify even and odd numbers and understand factors • add and subtract multi-digit whole numbers with single-step and/or regrouping • add and subtract decimals to the hundredths place • understand inverse operations and their properties and apply concepts of multiplication Students performing at the NAEP Basic achievement level can likely • identify appropriate measurement tools in real-world scenarios • measure or estimate lengths of objects in standard and non-standard units • find the perimeter of polygons given a visual aid Students performing at the NAEP Basic achievement level can likely • identify lines of symmetry • identify attributes of polygons as well as 3D shapes • compare these attributes with the support of visual aids Students performing at the NAEP Basic achievement level can likely • correlate information between tables and data displays • read and interpret tables and scaled graphs Students performing at the NAEP Basic achievement level can likely • identify the rule for a pattern and extend, complete, or determine missing numbers in patterns • complete input/output tables • locate points on a map/grid system with whole number or letter coordinates
NAEP Proficient	Students performing at the NAEP Proficient achievement level can likely • demonstrate an understanding of the relationships between the four operations (addition, subtraction, multiplication, and division) • add and subtract multi-digit whole numbers, fractions, and decimals in single and multi-step problems • apply basic properties of operations to solve problems

Achievement Level	Reporting ALDs
	• divide with whole numbers using one-digit divisors and understand remainders • solve problems with real-world contexts involving fractions with like denominators • identify and explain factors and multiples • compare and order whole numbers • identify, understand, and sort even and odd numbers • identify and compare decimals, fractions, and whole numbers on a number line • identify reasonable estimates Students performing at the NAEP Proficient achievement level can likely • identify appropriate units or tools of measurement within the same system • convert measurements within the same system • measure lengths of objects to the nearest whole or ½ unit • solve or estimate problems involving area Students performing at the NAEP Proficient achievement level can likely • demonstrate knowledge of 2D shapes by identifying, comparing, contrasting, and analyzing their attributes and describe attributes of 3D shapes with support of visual aids • select the final image of translations and reflections with no dashed lines and create parallel lines Students performing at the NAEP Proficient achievement level can likely • interpret and analyze data with scales of 2 or greater to solve problems • identify possible outcomes in probability events • determine the probability of events using terms of likelihood Students performing at the NAEP Proficient achievement level can likely • select expressions and equations to represent real-world situations • solve one-step equations with whole numbers • determine and/or apply rules to write, identify, or extend values in input/output tables • locate and name points on a map/grid system with whole number or letter coordinates
NAEP Advanced	Students performing at the NAEP Advanced achievement level can likely • compare and order whole numbers, fractions, and decimals to hundredths • apply understanding of factors and multiples and the structure of all operations with whole numbers • understand and use inverse operations and use simple ratios • multiply and divide 2- and 3-digit whole numbers with no remainders

Achievement Level	Reporting ALDs
	Students performing at the NAEP Advanced achievement level can likely • select appropriate and reasonable measurements in real-world scenarios • know vocabulary and units associated with area, perimeter, and volume • solve one-step and multi-step problems involving area and/or perimeter Students performing at the NAEP Advanced achievement level can likely • identify and apply attributes of 2D and 3D shapes in more complex contexts • compose and decompose 2D shapes to create more complex shapes • identify, draw and/or describe parallel lines • apply and draw lines of symmetry • generalize and reason with attributes of symmetrical figures • identify a series of rotations Students performing at the NAEP Advanced achievement level can likely • identify or describe events based on general probability categories • interpret and analyze data from single or multiple line, bar, and circle graphs • determine and interpret probability of an event with more than one condition • create a visual representation of equivalent fractions in relation to probability Students performing at the NAEP Advanced achievement level can likely • select expressions and equations to represent real-world situations with unknowns in all positions • solve for unknowns in all positions with division of whole numbers • determine, apply, and/or write a rule for a given pattern or input/output table as well as extend patterns and input/output tables • locate and name points (*x*, *y*) on a coordinate grid with whole number or letter coordinates

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

Table 26. Mathematics Grade 8 - Reporting ALDs

Achievement Level	Reporting ALDs
NAEP Basic	Students performing at the NAEP Basic achievement level can likely • simplify expressions involving integers • use operations to solve real-world problems involving integers or fractions • use proportional relationships to find equivalent ratios and create fractions and fractional relationships, with or without models • demonstrate understanding of scientific notation Students performing at the NAEP Basic achievement level can likely • reason and determine measurements, including length, area, and volume, with descriptions, labeled diagrams, and units provided • apply proportional reasoning to solve problems in context using scale factor, distance, unit conversion and quantities • apply simple scale factor value to find unknown lengths of triangles and rectangles without setting up a proportion Students performing at the NAEP Basic achievement level can likely • find a missing angle in a triangle given two angles and understand that angles of a triangle add to 180 degrees • recognize quadrilaterals given a description of their shared attributes Students performing at the NAEP Basic achievement level can likely • interpret, create, and/or compare different data set representations to determine a specific set of values for mean, mode, and range while identifying errors and appropriateness Students performing at the NAEP Basic achievement level can likely • use a coordinate plane to identify and plot coordinate points precisely • find the distance between points • recognize and extend patterns within an arithmetic or geometric sequence of numbers in a list or table (arithmetic, geometric) to solve problems in context • identify, solve, and/or evaluate one- and two-step equations, and apply slope, given linear relationships
NAEP Proficient	Students performing at the NAEP Proficient achievement level can likely • demonstrate an understanding of using and creating ratios to solve problems mathematically or in context • calculate GCF and LCM • perform basic operations with rational numbers to solve problems in context while applying proper units and converting between fractions, decimals, and percent

Achievement Level	Reporting ALDs
	• compare and order rational numbers with rational or common irrational numbers with or without a number line • apply problem-solving strategies to solve square roots and ratio and proportions Students performing at the NAEP Proficient achievement level can likely • demonstrate an understanding of solving problems that relate to comparing measures of two or three dimensions of space • determining possible dimensions given area and volume as well as selecting appropriate units of measure and applying scale factor to area • reason abstractly using addition and subtraction in contextual situations • solve problems involving capacity, area, and weight • classify angle measurements using diagrams and protractors Students performing at the NAEP Proficient achievement level can likely • understand concepts of parallel and perpendicular lines • use angle relationships and/or measurements formed when parallel lines are cut by a transversal. • apply concepts of corresponding parts between similar and congruent figures with some containing composite shapes in context • apply problem-solving strategies to solve Pythagorean Theorem problems • solve problems in context by creating a figure in the coordinate plane that satisfies area and perimeter criteria • reflect a shape on the coordinate plane over the x- and y-axis and plot some of the corresponding points • determine unknown side lengths by decomposing a polygon using given constraints • determine coordinates of missing endpoints of vertical or horizontal line segments on a coordinate plane Students performing at the NAEP Proficient achievement level can likely • use problem solving skills to make calculations based on multiple representations of data sets in context to determine measures of central tendencies, theoretical probability, and basic probability concepts • estimate values along the line of best fit • identify sources of bias in a sample design • calculate the mean from tables of data in multiple sets of values Students performing at the NAEP Proficient achievement level can likely • create, model, identify, and solve one-step inequalities and multi-step equations with or without context and with or without constraints • evaluate and extend sequential and recursive patterns using tables, models, multiple steps or from translating a written description

Achievement Level	Reporting ALDs
	• graph and identify key features of linear and nonlinear functions • recognize the effects on a graph when the slope and y-intercept are changed
NAEP Advanced	Students performing at the NAEP Advanced achievement level can likely • solve mathematical problems and problems in context with rational numbers including absolute values and variables by interpreting, creating, and using diagrams • engage with abstract situations and apply properties such as even and/or odd numbers, divisibility rules, and prime and composite numbers in mathematic situations Students performing at the NAEP Advanced achievement level can likely • create a proportion to represent scale in context and analyze number lines with fractional intervals • use ratios and proportion to find and/or explain measurements in multi-step situations, including unit rate and speed/distance • solving problems involving area including composing and decomposing complex figures • reason abstractly using multiple steps of addition and subtraction in context • estimate length and measure using tools appropriately in context Students performing at the NAEP Advanced achievement level can likely • understand angle relationships formed when parallel lines are cut by a transversal(s) to solve complex problems • determine relative positions of points using the geometric ideas of midpoint involving directionality on the coordinate plane • classify geometric solids by their properties and recognize cross-sections of plane and solid figures predicting results of subdividing complex plane figures in a variety of ways • perform a sequence of transformations in the coordinate plane and identify corresponding parts and points • understand and apply relationships of circles and central angles • evaluate proportionality and similarity to find unknown values • justify the sum of interior angles of polygons • demonstrate knowledge of the faces of geometric solids • determine unknown sides of triangles using Pythagorean Theorem or similar triangles Students performing at the NAEP Advanced achievement level can likely • analyze and compare problems using problem-solving skills to make calculations and connections based on multiple representations of data sets to determine measures of central tendencies and their effect, theoretical and experimental probability, and basic probability concepts • use a line of best fit or line graph to make predictions and interpretations

Achievement Level	Reporting ALDs
	• analyze and critique graphical displays to justify appropriateness and solve problems Students performing at the NAEP Advanced achievement level can likely • interpret, analyze, apply, and justify mathematical or contextual linear and nonlinear relationships and their key features represented through equations, tables, and graphs of a line, including with scales other than 1 • evaluate, interpret, justify, solve, and write multiple-step equations and/or expressions and inequalities with fractions and/or multiple variables in contextual situations with or without constraints

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

Table 27. Mathematics Grade 12 - Reporting ALDs

Achievement Level	Reporting ALDs
NAEP Basic	Students performing at the NAEP Basic achievement level can likely • use operations with rational numbers • apply single-step percentages to solve real-world problems • apply proportional relationships to solve real-world problems Students performing at the NAEP Basic achievement level can likely • use the properties of operations to determine equivalent polynomial expressions • solve single-step radical equations • perform a single reflection of a parent function graphically and/or algebraically • identify a type of function given a verbal description, table of values, and graph • analyze graphs of linear functions to compare rates of change or slope Students performing at the NAEP Basic achievement level can likely • apply proportional reasoning to solve real-world problems using scale or unit rate • determine how rigid and non-rigid transformations affect an object and its measurements, including area • calculate vertical and horizontal distances given the coordinates of two points • compare areas of simple figures with or without a grid Students performing at the NAEP Basic achievement level can likely • determine probabilities of simple events from 2-way tables and verbal descriptions

Achievement Level	Reporting ALDs
	• determine the characteristics of a well-designed survey including valid sampling methods • identify the mean from the graph of a normal distribution • describe the effect changing the value of a data point has on the mean or median of the data set • analyze a scatterplot to identify a correct trend line
NAEP Proficient	Students performing at the NAEP Proficient achievement level can likely • demonstrate an understanding of real numbers and operations with real numbers • analyze information to solve real-world problems with proportional reasoning • write numbers in scientific notation and compute with scientific notation • simplify numeric expressions and perform operations that involve whole-number exponents • use common multiples to solve real-world problems Students performing at the NAEP Proficient achievement level can likely • use the properties of operations and exponents to determine equivalent polynomial expressions in a single step • perform a rigid transformation (reflection or translation) of a function graphically and/or algebraically • identify a type of function based on equations, verbal descriptions, tables, or graphs • translate between different representations of functions • analyze key features of functions (including slope, intercept, domain, and range) given coordinates of points, a table of values, a verbal description, an equation, or a graph • evaluate algebraic expressions or functions (including piecewise, linear, radical, quadratic, exponential, step) in the form of an equation or a graph for a given value • apply exponential relationships to solve problems or write expressions • determine the number of solutions for equations or inequalities • write inequalities from a verbal description • apply understanding of sequences to solve problems Students performing at the NAEP Proficient achievement level can likely • apply proportional relationships to solve problems about similar figures or represent scale relationships using diagrams • perform dilations, centered at the origin, with coordinates • compare areas of complex figures on a grid • reason about properties (angles, line segments, etc.) of diagrams based on perpendicular or parallel lines • use properties of plane figures to solve problems • demonstrate an understanding of 3-dimensional shapes by composing/decomposing them and taking cross sections

Achievement Level	Reporting ALDs
	• determine the type of measurement of a 3-dimensional figure for a given unit • complete an expression for a trigonometric ratio from a right triangle Students performing at the NAEP Proficient achievement level can likely • distinguish between, find, and compare experimental and theoretical probabilities • analyze the characteristics of experimental and survey designs and what can likely be inferred by each • describe the impact of increasing a data value on the mean or median • analyze data sets to determine or compare measures of center or spread • analyze trends in scatterplots to make predictions or determine when predictions are appropriate • analyze both appropriate and misleading data displays to draw conclusions
NAEP Advanced	Students performing at the NAEP Advanced achievement level can likely • reason with and apply the properties and operations of real numbers • solve multi-step real-world problems using percentages • compute and/or estimate the values of numeric expressions involving square roots and cube roots Students performing at the NAEP Advanced achievement level can likely • use the properties of operations and exponents to determine equivalent polynomial expressions in multiple steps • perform a series of rigid transformations (including translations and reflections) of a function graphically and/or algebraically • determine an appropriate family of functions to model a real-world problem given a diagram or a table of values • analyze multiple key features (including slope, intercept, domain, range, and vertex) of functions (including linear, quadratic, and logarithmic) given coordinates of points, a graph, or an equation • analyze real-world problems to determine the meaning of values within an equation • solve and/or graph compound, absolute value, and quadratic inequalities • determine and apply recursive rules for sequences and functions Students performing at the NAEP Advanced achievement level can likely • apply proportional reasoning to solve problems relating to area, similar figures, and converting between measurement systems • describe a series of transformations, when two or more types of transformations are used, for a figure on the coordinate plane • use coordinate geometry to find the midpoint of a segment • reason about the relationship between lines in parallel planes and lines that intersect parallel planes • apply properties and measurements of figures, including angles, perimeter, and area to solve problems

Achievement Level	Reporting ALDs
	• decompose 3-dimensional shapes to solve problems • apply the Pythagorean Theorem to find lengths in 3-dimensional figures • apply trigonometric ratios to solve problems involving a single right-triangle Students performing at the NAEP Advanced achievement level can likely • determine joint probabilities from 2-way tables or data sets • generalize results of an experiment to a population • describe or find the effects on summary statistics when the data set is changed • create a data set with a given median, mode, and range • compare correlation coefficients from scatterplots • understand that correlation does not imply causation • create data displays, including 2-way tables and boxplots, given data or summary statistics

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

Table 28. Reading Grade 4 - Reporting ALDs

Achievement Level	Reporting ALDs
NAEP Basic	When reading literary texts such as fiction, poetry, and literary nonfiction, fourth-grade students performing at the NAEP Basic level can likely • determine the relevant meaning of familiar words using context within the same sentence or paragraph • identify a specific detail to make a simple inference about the characters' actions, motivations, or feelings, using a single point or multiple points in the text if they are in close proximity • sequence or categorize events from the story • make a general reference to an appropriate section of the text or provide some support for ideas related to the plot or characters • find meaning or provide evidence from one of the texts when making a comparison across texts • identify explicit details from the text • state an opinion with general support from one section of the text When reading informational texts such as articles and excerpts from books, fourth-grade students performing at the NAEP Basic level can likely • determine the relevant meaning of familiar words using context from a single section of the text • locate a specific detail from the text and make simple inferences from one section of the text • restate a problem or solution presented in a single section of the text • provide a description of a text feature or author's craft using a general reference to the text • provide an opinion using a general reference to the text
NAEP Proficient	When reading literary texts such as fiction, poetry, and literary nonfiction, fourth-grade students performing at the NAEP Proficient level can likely • determine the meaning of words using context from multiple sections of the text • provide a reason why a particular detail is important to the story • identify the key events to determine main idea and make complex inferences about the characters' actions, motivations, or feelings, using relevant evidence within or across texts • describe the impact of a character's actions or explain how characters influence others • recognize a text's structure and organization • draw conclusions from single or multiple locations across a text and provide limited support from the text • develop an opinion with relevant support from a text

Achievement Level	Reporting ALDs
	When reading informational texts such as articles and excerpts from books, fourth-grade students performing at the NAEP Proficient level can likely • determine the relevant meaning of words with multiple meanings • use a specific detail from the text to make inferences or provide a description or an explanation about text features • provide an opinion with relevant support from the text • restate a problem or solution presented in a single section of the text • describe, explain, or draw conclusions about text structures (e.g., compare and contrast, cause and effect, sequence and order) • integrate ideas across a text to determine purpose and main idea
NAEP Advanced	When reading literary texts such as fiction, poetry, and literary nonfiction, fourth-grade students performing at the NAEP Advanced level can likely • determine the meaning of nonliteral phrases • identify relevant details to support a detailed judgment about a character and infer a character's development from the beginning of the text to the end • interpret lines of poetry to determine meaning • make a judgment about the purpose of text structure with relevant evidence from the text • distinguish the theme of a text • provide support from across the text when selecting evidence • compare two texts to support an opinion • make inferences across texts and use relevant details for support When reading informational texts such as articles and excerpts from books, fourth-grade students performing at the NAEP Advanced level can likely • make complex inferences about words with multiple meanings or nonliteral phrases • select details to support a solution to a problem and provide relevant support for a given idea • provide a detailed opinion with relevant support, using details from the text • summarize ideas presented in a text and explain and/or interpret the purpose of a text feature • begin to evaluate text structures or an author's purpose

Table 29. Reading Grade 8 - Reporting ALDs

Achievement Level	Reporting ALDs
NAEP Basic	When reading literary texts such as fiction, poetry, and literary nonfiction, eighth-grade students performing at the NAEP Basic level can likely • use explicit context within the same sentence or paragraph to determine the meaning of essential words • make simple inferences about one part of the text using explicit details from that same part of the text • identify basic literary elements such as order of events, character traits and motivation, and main idea • formulate an opinion about the text and provide evidence that demonstrates limited or incomplete understanding using general information from or reference to the text When reading informational texts such as exposition and argumentation, eighth-grade students performing at the NAEP Basic level can likely • determine the meaning of words using context from one section of the text • locate and use explicit details to answer specific questions and make simple inferences about the text • determine the main idea or purpose of the text using explicit features from the text • demonstrate a general understanding of text features or graphics • demonstrate a general understanding of the concepts in the text but can support their understanding using only limited information from the text • formulate an opinion about a claim or argument and support this opinion using only limited information from the text
NAEP Proficient	When reading literary texts such as fiction, poetry, and literary nonfiction, eighth-grade students performing at the NAEP Proficient level can likely • use context explicitly and implicitly across the entire text to determine the meaning of words and nonliteral phrases • make inferences and draw conclusions about varied literary elements such as character interactions, comparison of characters, plot features, and theme • support ideas with relevant examples from the text and provide some explanation about the connection between the ideas and evidence • provide a reasonable opinion supported by some evidence from the text When reading informational texts such as exposition and argumentation, eighth-grade students performing at the NAEP Proficient level can likely • use context to determine the definition of multiple-meaning words

Achievement Level	Reporting ALDs
	• make inferences or judgments about text structures, features, and author's craft but can provide only partial explanations or text support • make connections between text features and graphics to explain how they support the primary text • identify one or both sides of an argument • offer an opinion about the evidence an author uses to support a claim or argument
NAEP Advanced	When reading literary texts such as fiction, poetry, and literary nonfiction, eighth-grade students performing at the NAEP Advanced level can likely • interpret descriptive or figurative language and how it impacts the meaning of the text • evaluate the relationships between literary elements such as setting, characterization, tone, structure, and how they impact the overall meaning of the text • construct an opinion and support it with relevant details and specific examples from the text • make specific connections within and across texts by using relevant evidence and providing a thorough explanation When reading informational texts such as exposition and argumentation, eighth-grade students performing at the NAEP Advanced level can likely • determine the meaning of words using information and ideas presented throughout the text • make connections within a text to determine similarities and differences, causes and effects, and problems and solutions • express an opinion that evaluates whether a text feature is critical to the overall understanding of the text • synthesize information across related texts and fully support their ideas with evidence from both texts • evaluate the effectiveness of an author's argument and support their evaluation with evidence • use appropriate text evidence from multiple sources to substantiate their own opinions or claims made by the author

Table 30. Reading Grade 12 - Reporting ALDs

Achievement Level	Reporting ALDs
NAEP Basic	When reading literary text such as fiction, poetry, and literary nonfiction, twelfth-grade students performing at the NAEP Basic level can likely • use context that is in close proximity to an unknown word or phrase to determine meaning • locate and identify relevant details in the text in order to support literal comprehension • make inferences that demonstrate a basic understanding of literary elements such as an author's purpose, character motivation, mood, and theme • draw general conclusions based on concepts that are presented with abundant and/or explicitly stated text evidence When reading informational text such as exposition and argumentation, twelfth-grade students performing at the NAEP Basic level can likely • use context, typically within close proximity, to identify the meaning of unknown words and phrases • identify and make judgments about key details within and across texts • use those details to draw simple inferences about author's purpose, tone, and word choice • provide opinions and sometimes support them with generalized text evidence • evaluate the effectiveness of an author's claim, organization, and evidence used • utilize text features and organizational structure to locate information and identify textually explicit details
NAEP Proficient	When reading literary texts such as fiction, poetry, and literary nonfiction, twelfth-grade students performing at the NAEP Proficient level can likely • infer the meaning of words from direct and indirect context and explain the impact of those words • locate and identify information and ideas from within a single text and across texts to build meaningful comparisons of ideas, characters, and author's craft • go beyond foundational comprehension skills and analyze complex themes and ideas in order to draw inferences • make effective judgments and critiques of the author's use of sensory/descriptive language, connotative language, figurative language, and other literary elements When reading informational texts such as exposition, argumentation, and documents, twelfth-grade students performing at the NAEP Proficient level can likely

Achievement Level	Reporting ALDs
	• interpret and evaluate a variety of organizational structures or types of media (e.g., diagrams, charts, videos, etc.) used in argumentative, persuasive, and informational texts and determine how those structures/text features enhance the effectiveness of the text • synthesize several documents and support one or more of their opinions using relevant and sufficient evidence from the text • connect key details within and across texts and use those details to draw complex inferences about author's purpose, tone, word choice, and related ideas • describe and evaluate the effectiveness of nuanced language use, specific details, and an author's stylistic, syntactical, and rhetorical choices based on certain sections of the passage or the passage as a whole • evaluate the effectiveness of the author's claims, organization, selection of ideas, and evidence used
NAEP Advanced	When reading literary texts such as fiction, poetry, and literary nonfiction, twelfth-grade students performing at the NAEP Advanced level can likely • evaluate literary elements and explain how they connect to the overall purpose of the text as well as how they develop over the course of the text • use complex strategies to navigate the literal, figurative, and implicit meaning of texts, including the analysis of author's choices, word choice, and language • analyze and interpret a variety of text types and devices. • synthesize ideas from sections of the text and multiple texts to generate new understandings and integrate new information When reading informational texts such as exposition, argumentation, and documents, twelfth-grade students performing at the NAEP Advanced level can likely • analyze and evaluate a variety of organizational structures or types of media (e.g., diagrams, charts, videos, etc.) used in argumentative, persuasive, and informational texts and critique how those structures/text features are used to enhance the effectiveness of the text • synthesize information within and across texts and use it to create and support their own arguments that can go beyond the literal interpretation of the text

Achievement Level	Reporting ALDs
	• critique the effectiveness of sophisticated language use and evaluate an author's stylistic, syntactical, and rhetorical choices based on certain sections of the passage or the passage as a whole, supporting their evaluation with precise text evidence • evaluate the effectiveness of an author's claims, organization, and selection of ideas and evidence used • develop opinions and support these with specific, relevant textual evidence • demonstrate a thorough understanding of the themes and ideas presented

Tables 31 through 34 present the results from round 3 of the ALD alignment judgment activity from the operational study.

Table 31. Round 3 mathematics alignment judgment agreement with achievement level policy definitions

Subject	Grade	NAEP Level	Alignment Judgment			
			Minimal	Weak	Moderate	Strong
Math	4	Basic	0%	0%	0%	100%
		Proficient	0%	0%	0%	100%
		Advanced	0%	0%	0%	100%
	8	Basic	0%	0%	0%	100%
		Proficient	0%	0%	0%	100%
		Advanced	0%	0%	29%	71%
	12	Basic	0%	0%	25%	75%
		Proficient	0%	0%	63%	37%
		Advanced	13%	0%	38%	50%

Table 32. Round 3 mathematics alignment judgment agreement with content ALDs

Subject	Grade	NAEP Level	Alignment Judgment			
			Minimal	Weak	Moderate	Strong
Math	4	Basic	0%	0%	25%	75%
		Proficient	0%	0%	0%	100%
		Advanced	0%	0%	0%	100%
	8	Basic	0%	0%	29%	71%
		Proficient	0%	0%	57%	43%
		Advanced	0%	0%	57%	43%
	12	Basic	0%	13%	13%	75%
		Proficient	13%	0%	50%	38%
		Advanced	13%	38%	50%	0%

Table 33. Round 3 reading alignment judgment agreement with achievement level policy definitions

Subject	Grade	NAEP Level	Alignment Judgment			
			Minimal	Weak	Moderate	Strong
Reading	4	Basic	0	0	0	100%
		Proficient	0	0	17%	83%
		Advanced	0	0	0	100%
	8	Basic	0	0	0	100%
		Proficient	0	0	0	100%
		Advanced	0	0	0	100%
	12	Basic	0	0	0	100%
		Proficient	0	0	0	100%
		Advanced	0	0	0	100%

Table 34. Round 3 reading alignment judgment agreement with content ALDs

Subject	Grade	NAEP Level	Alignment Judgment			
			Minimal	Weak	Moderate	Strong
Reading	4	Basic	0	0	50%	50%
		Proficient	0	0	17%	83%
		Advanced	0	0	17%	83%
	8	Basic	0	0	50%	50%
		Proficient	0	0	83%	17%
		Advanced	0	0	17%	83%
	12	Basic	0	0	33%	67%
		Proficient	0	0	0	100%
		Advanced	0	0	0	100%

The alignment judgments for reading comparing the summary statements to the policy definitions and content ALDs were either moderate or strong. The alignment judgments for the policy definitions were higher than the judgments for the content ALDs. 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.

The alignment judgments for mathematics comparing the summary statements to the policy definitions were generally moderate or strong, with one exception for grade 12 mathematics. For grades 4 and 8 mathematics, the alignment judgments for the

summary statements to the content ALDs were moderate or strong, indicating that the panelists believed that the knowledge and skills likely demonstrated by students represented the knowledge and skills described by the NAEP content ALDs. The alignment judgment for grade 12 mathematics also showed agreement except for the *NAEP Advanced* level. One panelist from the grade 12 mathematics panel indicated that the alignment for NAEP Proficient and *NAEP Advanced* was minimal, with three panelists selecting weak alignment for *NAEP Advanced*.

The comments from some of the panelists in the grade 12 mathematics panel indicated the students performing at the *NAEP Advanced* level are not likely demonstrating some of the more advanced knowledge and skills that are described in the content ALDs, so the *NAEP Advanced* summary statement may better align with the content ALDs at the *NAEP Proficient* level. This misalignment at the *NAEP Advanced* level could be due to various situations. Some possibilities include that the content ALDs for *NAEP Advanced* for grade 12 are very ambitious or that the cut score for *NAEP Advanced* is too low to represent the knowledge and skills defined by the content ALDs. Additionally, there were a significant number of items associated with Does Not Anchor, which could include some of the more advanced skills identified in the content ALDs. The focus of this study was to identify any areas of possible misalignment but cannot fully explain why that misalignment exists.

Process Evaluations

Panelists responded to a number of survey questions multiple times at different points of 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 grade 4 mathematics, 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 35 through 42 present the results of select questions from the process evaluation surveys.

Table 35. 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**	**Total**
Reading Grade 4	0	0	0	0	2 (33%)	4 (67%)	6
Reading Grade 8	0	0	0	0	3 (50%)	3 (50%)	6
Reading Grade 12	0	0	0	0	3 (50%)	3 (50%)	6
Mathematics Grade 4	1 (13%)	0	0	0	1 (13%)	6 (75%)	8
Mathematics Grade 8	0	0	0	1 (14%)	1 (14%)	5 (71%)	7
Mathematics Grade 12	0	0	0	1 (13%)	1 (13%)	6 (75%)	8

Table 36. 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**	**Total**
Reading Grade 4	0	0	0	0	2 (33%)	4 (67%)	6
Reading Grade 8	0	0	0	1 (17%)	3 (50%)	2 (33%)	6
Reading Grade 12	0	0	0	1 (17%)	2 (33%)	3 (50%)	6
Mathematics Grade 4	0	1 (13%)	0	0	2 (25%)	5 (63%)	8
Mathematics Grade 8	0	0	0	0	4 (57%)	3 (43%)	7
Mathematics Grade 12	0	0	0	1 (13%)	4 (50%)	3 (38%)	8

Table 37. 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	Total
Reading Grade 4	0	0	0	0	2 (33%)	4 (67%)	6
Reading Grade 8	0	0	0	2 (33%)	2 (33%)	2 (33%)	6
Reading Grade 12	0	0	0	2 (33%)	2 (33%)	2 (33%)	6
Mathematics Grade 4	0	1 (13%)	0	0	3 (38%)	4 (50%)	8
Mathematics Grade 8	0	0	0	0	4 (57%)	3 (43%)	7
Mathematics Grade 12	0	0	0	1 (13%)	3 (38%)	4 (50%)	8

Table 38. 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	Total
Reading Grade 4	0	0	0	0	2 (33%)	4 (67%)	6
Reading Grade 8	0	0	0	0	4 (67%)	2 (33%)	6
Reading Grade 12	0	0	0	0	3 (50%)	3 (50%)	6
Mathematics Grade 4	0	1 (13%)	0	0	2 (25%)	5 (63%)	8
Mathematics Grade 8	0	0	0	2 (29%)	3 (43%)	2 (29%)	7
Mathematics Grade 12	0	0	0	2 (25%)	1 (13%)	5 (63%)	8

Table 39. 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**	**Total**
Reading Grade 4	0	0	1 (17%)	0	1 (17%)	4 (67%)	6
Reading Grade 8	0	0	0	0	3 (50%)	3 (50%)	6
Reading Grade 12	0	0	0	0	3 (50%)	3 (50%)	6
Mathematics Grade 4	0	0	0	0	3 (38%)	5 (63%)	8
Mathematics Grade 8	0	0	0	1 (14%)	1 (14%)	5 (71%)	7
Mathematics Grade 12	0	0	0	0	2 (25%)	6 (75%)	8

Table 40. Creation of panel summary descriptions

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**	**Total**
Reading Grade 4	0	0	1 (17%)	0	1 (17%)	4 (67%)	6
Reading Grade 8	0	0	0	0	1 (17%)	5 (83%)	6
Reading Grade 12	0	0	0	1 (17%)	0	5 (83%)	6
Mathematics Grade 4	0	0	0	0	2 (25%)	6 (75%)	8
Mathematics Grade 8	0	0	1 (14%)	0	3 (43%)	3 (43%)	7
Mathematics Grade 12	0	0	0	0	3 (38%)	5 (63%)	8

Table 41. 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**	**Total**
Reading Grade 4	0	0	0	2 (33%)	3 (50%)	1 (17%)	6
Reading Grade 8	0	0	0	0	2 (33%)	4 (67%)	6
Reading Grade 12	0	0	0	0	0	6 (100%)	6
Mathematics Grade 4	0	0	0	0	3 (38%)	5 (63%)	8
Mathematics Grade 8	0	1 (14%)	0	0	1 (14%)	5 (71%)	7
Mathematics Grade 12	0	0	0	0	2 (25%)	6 (75%)	8

Table 42. 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**	**Total**
Reading Grade 4	0	4 (67%)	2 (33%)	6
Reading Grade 8	0	5 (83%)	1 (17%)	6
Reading Grade 12	0	6 (100%)	0	6
Mathematics Grade 4	0	8 (100%)	0	8
Mathematics Grade 8	0	5 (71%)	2 (29%)	7
Mathematics Grade 12	2 (25%)	6 (75%)	0	8
Total	2 (5%)	34 (83%)	5 (12%)	41

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 43 and 44 present the results of these questions from the process evaluations.

Table 43. Confidence in the panel's achievement level summary descriptions - reading

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**	**Total**
Grade 4	*NAEP Basic*	0	0	1 (17%)	5 (83%)	6
	NAEP Proficient	0	0	2 (33%)	1 (67%)	6
	NAEP Advanced	0	0	0	6 (100%)	6
Grade 8	*NAEP Basic*	0	0	3 (50%)	3 (50%)	6
	NAEP Proficient	0	0	2 (33%)	4 (67%)	6
	NAEP Advanced	0	0	2 (33%)	4 (67%)	6
Grade 12	*NAEP Basic*	0	0	1 (17%)	5 (83%)	6
	NAEP Proficient	0	0	0	6 (100%)	6
	NAEP Advanced	0	0	0	6 (100%)	6

Table 44. Confidence in the panel's achievement level summary descriptions - mathematics

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**	**Total**
Grade 4	*NAEP Basic*	0	0	1 (13%)	7 (88%)	8
	NAEP Proficient	0	0	1 (13%)	7 (88%)	8
	NAEP Advanced	0	0	2 (25%)	6 (75%)	8
Grade 8	*NAEP Basic*	0	0	0	7 (100%)	7
	NAEP Proficient	0	0	0	7 (100%)	7
	NAEP Advanced	0	1 (14%)	0	6 (86%)	7
Grade 12	*NAEP Basic*	0	0	1 (13%)	7 (88%)	8
	NAEP Proficient	0	0	3 (38%)	5 (63%)	8
	NAEP Advanced	0	0	3 (38%)	5 (63%)	8

At the end of the alignment judgment process, panelists were asked to rate their confidence in their alignment judgment classification for each achievement level. Overall, the panelists had confidence in their alignment judgment classifications. Tables 45 and 46 present the results of these questions from the process evaluations.

Table 45. Confidence in the panel's alignment judgments - reading

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**	**Total**
Grade 4	*NAEP Basic*	0	0	1 (17%)	5 (83%)	6
	NAEP Proficient	0	0	1 (17%)	5 (83%)	6
	NAEP Advanced	0	0	2 (34%)	4 (66%)	6
Grade 12	*NAEP Basic*	0	0	1 (17%)	5 (83%)	6
	NAEP Proficient	0	0	0	6 (100%)	6
	NAEP Advanced	0	0	0	6 (100%)	6

*Note: The Ratings for Grade 8 were missing.

Table 46. Confidence in the panel's achievement level summary descriptions - mathematics

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**	**Total**
Grade 4	*NAEP Basic*	0	0	1 (13%)	7 (88%)	8
	NAEP Proficient	0	0	0	8 (100%)	8
	NAEP Advanced	0	0	0	8 (100%)	8
Grade 8	*NAEP Basic*	0	1 (14%)	1 (14%)	5 (63%)	7
	NAEP Proficient	0	0	1 (14%)	6 (86%)	7
	NAEP Advanced	0	0	1 (14%)	6 (86%)	7
Grade 12	*NAEP Basic*	0	0	2 (25%)	6 (75%)	8
	NAEP Proficient	0	0	2 (25%)	6 (75%)	8
	NAEP Advanced	0	0	1 (13%)	7 (88%)	8

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). Internal validity evidence refers to the internal consistency of data generated within the ALD review meeting. Table 47 provides a list of the different types of validity evidence for the NAEP ALD Review study.

Table 47. 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 process implemented and described in this report are consistent with those in the design document.
	Observations of two TAC member	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 ALD 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 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 Governing Board, through its Committee on Standards, Design and Methodology (COSDAM), shall review the alignment between the content ALDs and items, based on empirical data from recent administrations of NAEP assessments. In its review, COSDAM (in consultation with the Assessment Development Committee) shall solicit input from technical and subject matter experts to determine whether changes to the content ALDs are warranted or whether a new standard setting shall be conducted, making clear the potential risk of changing cut scores to trends and assessment of educational progress.

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 of the results from 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, research could possibly find or develop a method that could possibly build off of the results from 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 ALD and the Reporting ALDs, or a comparison between what is expected that students should know and be able to do and what students actually can know and demonstrate. To support the decision required by the principle, research could be completed to assist in determining the criteria the amount of misalignment which 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 grade 12 mathematics indicated various levels of misalignment for each of the NAEP achievement levels, but specifically *NAEP Advanced*. There was much discussion with the TAC and COSDAM members 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, which could be used to address previous questions regarding the decisions made using the ALD alignment judgment results.