Assessment Development Committee

May 7, 2021
5:30 – 7:30 pm, Eastern Time
Zoom Meeting

AGENDA

5:30 – 5:50 pm	**NAEP Reading Framework Update** *Dana Boyd, Chair* *Mark Miller, Vice Chair*	Attachment A
5:50 – 6:50 pm	**NAEP Mathematics and Reading Framework Processes** *Dana Boyd* *Mark Miller* *Michelle Blair, Assistant Director for Assessment Development* *Mark Loveland, Mathematics and Reading Framework Update, Project Co-Director*	Attachment B
6:50 – 7:05 pm	**Other Framework Processes: Initial Discussion** *Dana Boyd and Mark Miller*	Attachment C
7:05 – 7:30 pm	**NAEP Science Framework** *Dana Boyd and Mark Miller*	Attachment D
Information Item	**Strategic Vision 2025 Update**	See Executive Committee Material

NAEP Reading Framework Update

After over 2 years of Committee review and deliberation and after careful consideration of issues raised in Committee and full Board discussions, the Assessment Development Committee (ADC) had reached consensus on all issues relevant to the NAEP Reading Framework update.

Following the April 30, 2021 informational webinar on the Reading Framework, Board member Russ Whitehurst submitted to Board Chair Haley Barbour a set of proposed revisions to the draft 2026 NAEP Reading Framework. During the Executive Committee meeting on May 5, 2021, these proposed edits were referred to the Assessment Development Committee (as the committee of jurisdiction) for review.

Accordingly, three documents are attached:

1. Board member Russ Whitehurst's proposed edits to the draft 2026 NAEP Reading Framework
2. Board member Russ Whitehurst's preamble to his proposed edits
3. Summary of and notes on the proposed edits from Project Officer Michelle Blair

For the latest draft of the reading framework update and other related documents, see the NAEP Reading Framework Plenary session materials.

At the May 7 ADC meeting, the ADC will:

1. Review the guidance provided to the Development Panel by the Framework's Technical Advisory Committee after the March Board meeting (a copy of that guidance is attached);
2. Review the edits proposed by Board member Whitehurst (these edits revise the April 2021 draft framework); and
3. Review the edits from the Framework Development Panel in the April 2021 draft framework (a copy of these edits is in the reading plenary session materials).

NAEP READING FRAMEWORK UPDATE

TECHNICAL ADVISORY COMMITTEE GUIDANCE FOR THE DEVELOPMENT PANEL

OVERVIEW

The NAEP Reading Framework Technical Advisory Committee (TAC) is a group of eight experts in psychometrics and large-scale assessment. The TAC's role in the NAEP Reading Framework update process is to support the Development Panel (DP), addressing measurement and assessment questions as they surface. Two members of the TAC attend each Development Panel meeting. After Development Panel meetings, the full TAC convenes virtually to address specific questions from the previous Panel meeting, and to provide guidance for the subsequent Panel meeting. The TAC met for the seventh time on March 26, 2021. The objective of the meeting was to discuss the full Framework document and to offer guidance to support fine-tuning the Framework in advance of submission to the Governing Board. The TAC discussion focused on primarily universal design elements (UDEs) and topical knowledge. TAC members offered more general thoughts on the 2026 Framework as well. The TAC's March 2021 feedback and recommendations are summarized below.

VALIDITY RESEARCH ON UNIVERSAL DESIGN ELEMENTS

The TAC discussion began with a focus on evidentiary standards for UDEs. In and of themselves, UDEs are neither valid nor invalid. Rather, assessment developers examine the extent to which these features minimize construct-irrelevant variance (i.e., when factors unrelated to the intended subject of the test influence performance on the test). Similarly, UDEs should not inadvertently *create* bias by providing an advantage to particular student groups.

In the assessment accommodations literature, statistical examinations for an accommodation's impact is often carried out via multiple regression (e.g., where test scores or item responses are regressed on [1] presence of a disability such as visual impairment, [2] use of an accommodation such as Braille, and [3] the interaction of [1] and [2].) From a validity standpoint, a positive interaction effect is good: it indicates that on average, the accommodation increases scores, but only for the students who are supposed to receive it.

The TAC agreed, however, that UDEs in the 2026 Reading Framework are not accommodations; None are intended to help one group of students over another. Therefore, in a multiple regression analysis focused on the 2026 Reading Framework's UDEs, looking for main effects – not interaction effects – would be the first order of business.[1]

[1] Note that multiple regression is one among many tools test developers use to determine whether certain features of an assessment are doing the job they were intended to do for the populations they are intended to support. Other techniques, such as cognitive interviews and classroom tryouts (carried out for NAEP routinely during item development) generate different types of evidence, equally important to the overall validity argument.

KNOWLEDGE-BASED UNIVERSAL DESIGN ELEMENTS

The TAC spent roughly half of the meeting discussing knowledge-based UDEs (e.g., a short introduction to a potentially unfamiliar topic, available to all students), including how they affect validity and fairness, how often they are used in large-scale assessments, and how their common pitfalls can be avoided.

The reading comprehension construct in the 2026 Reading Framework does not assume prior disciplinary knowledge ("items should not ask readers to draw upon text-independent domain knowledge") and the only two knowledge-based UDEs under consideration are glossaries and short introductions, both of which are standard features of large-scale summative assessment. The Framework provides reassurance that appropriate safeguards for the recommended UDEs are in place, and that the more ambitious, potentially problematic UDEs are not slated for the 2026 assessment and will instead be the subject of further validation research.

The TAC believes it will be useful for the DP to further reinforce these points in the narrative. That is, in each instance that knowledge-based UDEs are recommended in the Framework, the DP should clarify that they are based on substantial precedent, represent best practice, and, in fact, are uncontroversial. Examples will help, whether in the Framework, in the Assessment and Item Specifications, or in both documents. Examples are especially useful for the disciplinary contexts, such as reading in science, where background knowledge arguably presents the clearest potential threat to unbiased measurement of reading comprehension.

TOPICAL KNOWLEDGE AND GENERAL IMPRESSIONS OF THE REVISED FRAMEWORK

A related issue to UDEs is the relative importance of topical knowledge in reading comprehension and, by extension, the appropriate emphasis on topical knowledge in a reading comprehension assessment. The TAC reflected on the DP's treatment of topical knowledge, and the discussion served as a springboard to more general reactions to the revised draft Framework. Both discussions are summarized briefly here.

TOPICAL KNOWLEDGE

The Framework conceptualizes topical knowledge as separate from reading comprehension. So, for the same reasons knowledge-based UDEs are encouraged, items that draw upon topical knowledge are discouraged. Specifically,

> *…items should not assess knowledge sources irrelevant to the items and associated comprehension targets in a given block. For example, items should not ask readers to draw upon text-independent domain knowledge, topic knowledge, knowledge of technical vocabulary or idiomatic expressions, or conceptual or domain knowledge in particular subject areas.*

The DP wants to deliver a Framework that gives NAEP the best possible chance of measuring reading comprehension as it is defined above – untethered to topical knowledge. To that end, the DP sought to confirm with the TAC that (1) the Framework's stance on topical knowledge is in keeping with modern assessment practice, and that (2) the associated rationales provided in the Framework are well aligned with modern validity theory.

This topic generated relatively little discussion. The TAC was unanimous in its support both of the DP's decision to exclude topical knowledge from the NAEP Reading construct and of the convincing rationales presented in the Framework. The DP thought it was important to be clearer about topical knowledge in the NAEP Reading Framework update. As a result, the revised framework document addresses the issue head-on. This is rare in large-scale testing; only a few states even address the issue of topical knowledge in their definition of reading comprehension, and none argue that topical knowledge should be measured as a component of reading comprehension.

GENERAL IMPRESSIONS OF THE REVISED FRAMEWORK.

The topical knowledge discussion provided the TAC an opportunity to voice its support for the revised 2026 Reading Framework. Although very few states currently address potentially controversial issues such as topical knowledge, bias, and responsible reporting, the TAC agreed that there are strong arguments that NAEP has a responsibility to be clear on these issues.

LIST OF TAC MEMBERS

Derek C. Briggs, University of Colorado, Boulder

Howard Everson, SRI International

Joan Herman, National Center for Research on Evaluation, Standards, and Student Testing (CRESST)

Kristen L. Huff, Curriculum Associates

Michael Kolen, University of Iowa

Scott Marion, The National Center for the Improvement of Educational Assessment

Jennifer Randall, Center for Educational Assessment, University of Massachusetts, Amherst

Guillermo Solano-Flores, Stanford University

Reading Framework for the 2026 National Assessment of Educational Progress

*** *April 21, 2021 Draft* ***
May 2, 2021Edits

National Assessment Governing Board
U.S. Department of Education

Developed for the National Assessment Governing Board under contract number 91995918C0001 by WestEd, with a subcontract to the Council of Chief State School Officers.

TABLE OF CONTENTS

Deleted: APPENDIX A: ADDITIONAL ASSESSMENT DESIGN FEATURES 62¶
APPENDIX B: ACHIEVEMENT LEVEL DESCRIPTIONS 62¶
NAEP Policy Definitions 62¶
Range ALDs 62¶
Organizational Features and Structures of the Reading Construct: Contexts, Purposes, Comprehension Targets, and Text Complexity 62¶
Comprehension Targets and Text Complexity 62¶
Broad and Specific Reading Purposes 62¶
NAEP Reading Achievement Levels: Grade 4 62¶
NAEP Basic 62¶
NAEP Proficient 62¶
NAEP Advanced 62¶
NAEP Reading Achievement Levels: Grade 8 62¶
NAEP Basic 62¶
NAEP Proficient 62¶
NAEP Advanced 62¶
NAEP Reading Achievement Levels: Grade 12 62¶ ... [1]

Deleted: ¶ ... [2]
Deleted: ¶ ... [3]
Deleted: ¶ ... [4]
Deleted: ¶ ... [5]
Deleted: ¶ ... [6]
Deleted: ¶ ... [7]
Deleted: ¶ ... [8]
Deleted: ¶ ... [9]
Deleted: ¶ ... [10]
Deleted: ¶ ... [11]
Deleted: ¶ ... [12]
Deleted: ¶ ... [13]
Deleted: ¶ ... [14]
Deleted: ¶ ... [15]
Deleted: ¶ ... [16]
Deleted: ¶ ... [17]
Deleted: ¶ ... [18]
Deleted: ¶ ... [19]
Deleted: ¶ ... [20]
Deleted: ¶ ... [21]
Deleted: ¶ ... [22]
Deleted: ¶ ... [23]
Deleted: ¶ ... [24]
Deleted: ¶ ... [25]
Deleted: ¶ ... [26]
Deleted: ¶ ... [27]
Deleted: ¶ ... [28]
Deleted: ¶ ... [29]

Visioning Panel

[* indicates the subgroup who drafted this framework as part of the Development Panel]

Peter Afflerbach*
Professor, Reading
University of Maryland
Silver Spring, MD

Carolyn Aguirre
Middle School Teacher / Department Head
New Haven Unified School District
San Leandro, CA

Sarah Aguirre*
Field Education Specialist
University of Texas, San Antonio
San Antonio, TX

Minerva Anaya St John
President
A-SJ Properties, Inc.
McAllen, TX

Nancy Brynelson*
Co-Director
California State University Chancellor's Office, Center for the Advancement of Reading and Writing
Gold River, CA

Jinghong Cai
Senior Research Analyst
National School Boards Association (NSBA)
Center for Public Education
Arlington, VA

Gina Cervetti*
Associate Professor, Education
University of Michigan
Ann Arbor, MI

Byeong-Young Cho*
Associate Professor, Korean Language Education
Hanyang University
Seoul, Republic of Korea

Julie Coiro*
Professor, Education
University of Rhode Island
Quaker Hill, CT

Carol Connor*
President, Society for the Scientific Study of Reading /
Chancellor's Professor, University of California, Irvine
Irvine, CA

Elena Forzani*
Assistant Professor, Literacy Education
Boston University
Boston, MA

Josephine Franklin
Associate Director
National Association of Secondary School Principals (NASSP)
Reston, VA

John Guthrie*
Jean Mullan Professor Emeritus, Human Development and Quantitative Methodology
University of Maryland, College Park
Chestertown, MD

Bonnie Hain*
Senior Director, Learning and Professional Services
American College Testing (ACT)

Woodstock, MD

Robin Hall
Director, Language Arts and Literacy
Council of the Great City Schools (CGCS)
Fairburn, GA

Kathleen Hinchman*
Professor, Childhood and Adolescent Literacy
Syracuse University
Syracuse, NY

Christy Howard
Associate Professor, Content Area Literacy
East Carolina University
Raleigh, NC

Panayiota Kendeou
Guy Bond Chair in Reading / Professor
University of Minnesota
Minneapolis, MN

Emily Kirkpatrick
Executive Director
National Council of Teachers of English (NCTE)
Louisville, KY

Carol Lee*
Edwina S. Tarry Professor, Education and Social Policy
Northwestern University
Country Club Hills, IL

Karen Malone
Curriculum, Instruction and Assessment Education Specialist
Bureau of Indian Education, Navajo District
Gallup, NM

Marina Pacheco*
Associate Professor, Curriculum and Instruction
University of Wisconsin, Madison
Madison, WI

Cindy Parker
English Language Arts Collaborative Advisor
Council of Chief State School Officers (CCSSO)
Danville, KY

Jim Patterson
Executive Director
The College Board
Coralville, IA

P. David Pearson, Panel Chair*
Professor Emeritus, Education
University of California, Berkeley
Berkeley, CA

Sue Pimentel
Founding Partner
Student Achievement Partners
Fort Myers, FL

Alicia Ross*
Teacher / Educational Consultant
Blue Ridge High School
Throop, PA

Robert Rothman*
Senior Editor
National Center on Education and the Economy (NCEE)
Washington, DC

Allison Skerrett*
Professor, Curriculum and Instruction
University of Texas, Austin
Austin, TX

Eric Turman
Principal
Reading High School
Reading, PA

Paola Uccelli*
Professor, Education
Harvard University
Belmont, MA

Paul Wenger
President-Elect
National Association of Elementary School Principals (NAESP)
West Des Moines, IA

Victoria Young
Director, Reading, Writing and Social Studies Assessments
Texas Education Agency
Austin, TX

Technical Advisory Committee

Derek C. Briggs
Professor, Research and Evaluation Methodology
University of Colorado, Boulder
Boulder, CO

Howard Everson
Senior Principal Research Scientist
SRI International
New York, NY

Joan Herman
Senior Research Scientist, University of California, Los Angeles /
Co-Director Emeritus, National Center for Research on Evaluation, Standards, and Student Testing (CRESST)
Los Angeles, CA

Kristen L. Huff
Vice President
Curriculum Associates
North Billerica, MA

Michael Kolen
Professor Emeritus, Educational Measurement
University of Iowa
Estes Park, CO

Scott Marion
Executive Director
The National Center for the Improvement of Educational Assessment (NCIEA)
Dover, NH

Jennifer Randall
Associate Professor and Director of Evaluation for the Center for Educational Assessment, Education
University of Massachusetts, Amherst
Amherst, MA

Guillermo Solano-Flores
Professor, Education
Stanford University
Stanford, CA

WestEd Staff

Matthew Gaertner
Measurement Specialist
Director of Research, Assessment Research and Innovation
WestEd
Austin, TX

Georgia Earnest García
Reading Content Specialist
Professor Emerita
University of Illinois, Urbana-Champaign
Napa, CA / Champaign, IL

Cynthia Greenleaf
Reading Content Specialist
Senior Research Scientist
WestEd
Albany, CA

Mira-Lisa Katz
Reading Content Specialist
Associate Director in Learning and Technology
WestEd
San Francisco, CA

Mark Loveland
Deputy Project Director
Senior Research Associate
WestEd
Redwood City, CA

Matthew Rudoff
Assessment Specialist
Manager, English Language Arts Assessment
WestEd
San Francisco, CA

Megan Schneider
Content Team Coordinator
Program Associate
WestEd
Redwood City, CA

Steven Schneider
Project Director
Senior Program Director, Science, Technology, Engineering, and Mathematics Research and Entrepreneurship
WestEd
Redwood City, CA

Sarah Warner
Project Coordinator
Research Associate
WestEd
Nashville, TN

Kamilah Wilson
Administrative Assistant
WestEd
Washington, DC

Council of Chief State School Officers (CCSSO) Staff

Fen Chou
Program Director, Standards, Assessment, and Accountability

Scott Norton
Deputy Executive Director, Programs

National Assessment Governing Board Staff

Michelle Blair
Project Officer
Assistant Director for Assessment Development

Sharyn Rosenberg
Assistant Director for Psychometrics

The National Assessment of Educational Progress (NAEP), often called The Nation's Report Card, is the largest nationally representative and continuing assessment of what students in public and private schools in the United States know and are able to do in various subjects. Since 1969, NAEP has been a common measure of student achievement across the country in mathematics, reading, science, and other subjects. The Nation's Report Card provides national, state, and some district-level results, as well as results for different demographic groups. NAEP is a congressionally mandated project of the National Center for Education Statistics (NCES), located within the U.S. Department of Education's Institute of Education Sciences. By law and by design, NAEP does not produce results for individual students or schools. The National Assessment Governing Board (Governing Board), an independent, bipartisan organization made up of governors, state school superintendents, teachers, researchers, and representatives of the general public, sets policy for NAEP.

The 2026 NAEP Reading Framework describes the content and design of the 2026 NAEP Reading Assessment; it is intended for a general audience. A second document, the *Assessment and Item Specifications for the 2026 NAEP Reading Framework*, serves as the "test blueprint" with information about passage selection, item development and other aspects of test development; it is intended for a more technical audience, including NCES and the contractors that will develop the NAEP Reading Assessment. In accordance with Governing Board policy, the 2026 NAEP Reading Framework focuses on "important, measurable indicators of student achievement to inform the nation about what students know and are able to do without endorsing or advocating a particular instructional approach."

The Education Sciences Reform Act of 2002 (P.L. 107-279) is the governing statute of NAEP. This law stipulates that NCES develops and administers NAEP and reports NAEP results. Under the law, the Governing Board is given responsibility for setting the assessment schedule, developing the frameworks that provide the blueprints for the content and design of the assessments, and setting achievement levels. The NAEP Reading Assessment is given in English every two years to students in grades 4 and 8, and every four years to students in grade 12. The assessment measures reading comprehension by asking students to read grade-appropriate materials and answer questions based on what they have read.

Development of the 2026 NAEP Reading Framework

In 2018, the Governing Board conducted a review of the current NAEP Reading Framework. In accordance with the Board policy, the review included commissioned papers and discussions with an array of reading educators and experts. Based on the review, at its March 2019 meeting, the Governing Board determined that the Reading Framework needed updating to better align with changes in what students in the second quarter of the 21st centery need to know and do to read proficiently. The process of updating the 2026 NAEP Reading Framework was guided by Governing Board policies that specify that the work be undertaken by a Visioning Panel of educators; experts in reading, learning and development, and assessment; and other key stakeholders in education. From this group, a subset of members continued as the Development Panel to finalize a document to recommend to the Governing Board for approval. In 2019, the

Deleted: Current NAEP Reading Assessment in a Digital Environment ¶

The NAEP Reading Assessment has been administered on a digital platform since 2017. NAEP's move to dynamic and innovative technologies provides an opportunity for an engaging assessment experience for students and more meaningful data about students' skills and knowledge for educators. With digitally based assessments, students are asked to receive, gather, and report information just as they do in many aspects of their everyday lives. These assessments also are constructed to reflect the principles of Universal Design of Assessments (UDA) (National Center on Educational Outcomes, 2016). The principles of UDA are intended to increase assessment validity and accessibility and to provide a more accurate understanding of what students know and can do (Thompson, Johnstone, & Thurlow, 2002; Thompson, Thurlow, & Malouf, 2004). Examples of three of the seven UDA principles include precisely defined constructs, accessible, non-biased items, and maximum readability and comprehensibility. ¶

The current NAEP Reading Assessment is organized according to assessment blocks. These feature either discrete items (stand-alone text passages and related questions) or scenario-based tasks (simulated settings in which students read passages while following various steps to accomplish a particular purpose or solve a problem). Scenario-based tasks (SBTs) can include many innovative features, such as: ¶
Task characters (avatars acting as simulated task partners)¶
Increased guidance enabling students to navigate more complex items¶
Item resetting in which students, after locking in answers, receive information about the correct response, so they can avoid carrying misconceptions into the next portion of the task¶

Schools and students participating in NAEP assessments are supported in various ways so they can successfully engage with the digitally based assessment. The digital platform provides students with support features that are intended to replicate the types of support provided during reading instruction and practice in school and at home or the workplace. For both discrete and SBT assessment blocks, tools available to all students include annotation via an on-screen pencil or highlighter, selection of color themes, and zoom-in. In addition, a text-to-speech capability is available on the Directions and Help screens (but not available for the reading passages or questions). Texts or questions may include hyperlinks, such as pop-up notes to click for more information (typically a definition of a selected word), a look-back button that takes students back to the relevant sentence or location in the text, multi-part response frames, and more. Not all support features are available in every block, but all blocks include some support features.¶

At the beginning of the assessment session, students interact with a tutorial that presents all the information needed to take the assessment on the digital platform; the tutorial explains how to progress through the reading passage and how to indicate or provide answers to questions, as well as how to use the tools. Students try out the tools and then enter and edit responses in a brief practice session. A ... [30]

Formatted: Superscript

Deleted: address advances in research in reading

Board charged the Visioning and Development Panels with developing recommendations for updating the framework as follows:

> The Visioning and Development Panels will recommend to the Board necessary changes in the NAEP Reading Framework at grades 4, 8, and 12 that maximize the value of NAEP to the nation. The panels are also tasked with considering opportunities to extend the depth of measurement and reporting given the affordances of digital based assessment. The update process shall result in three documents: a recommended framework, assessment and item specifications, and recommendations for contextual variables that relate to student achievement in reading.

To undertake this charge the Visioning Panel reviewed the considerable developments in reading research, literacy standards, and assessment that have taken place since the Board adopted the 2009–2019 NAEP Reading Framework in 2004. The Visioning Panel also considered input from a special panel of state literacy leaders as well as a paper, commissioned by NCES and authored by the NAEP Validity Studies (NVS) Panel, that examined the degree to which NAEP's assessments in mathematics, reading, and writing reflected both the content standards and the assessments implemented by states. In this report, the NVS Panel recommended that NAEP "should continue to develop and implement reading blocks that use new formats similar to scenario-based tasks or other alternatives that prioritize purpose-driven, performance-oriented, multisource tasks" (Valencia, Wixson, Kitmitto & Blankenship, 2019).

The Visioning Panel thus wanted to ensure that updates to the 2009–2019 framework would enable students to draw on their accumulated knowledge and experiences to complete assessment tasks. To that end, the Visioning Panel asked the Development Panel to update the framework in a manner that would enhance the assessment's validity while minimizing bias. The Panel also called for assessment texts and tasks to be broadly representative of the knowledge and experiences of the nation's students and the many ways in which they engage with reading in today's world.

To address the Visioning Panel recommendations, the Development Panel considered frameworks for other large-scale literacy assessments, such as the Programme for International Student Assessment (PISA) and the Progress in International Reading Literacy Study (PIRLS). The Development Panel attended to educational and societal developments, including advances in technology and new types of texts (digital and multimodal), and they incorporated findings from new research in three areas: disciplinary literacy; the role of affect, motivation, and agency in shaping readers' performance; and the role of social and cultural experiences in human development and learning, particularly in reading comprehension. The Panel augmented its attention to principles of Universal Design of Assessments to address the experiences of the nation's increasingly diverse students in more inclusive ways, many states' recent adoption of new standards and assessments, and innovations in digitally based assessments. These broad developments in research, policy, and practice guided the drafting of this framework update for the 2026 administration of the NAEP Reading Assessment.

The Updated NAEP Reading Framework

Formatted: No bullets or numbering

Deleted: Accordingly, the Visioning Panel set guidelines for drafting an updated NAEP Reading Framework that would:¶
Expand the construct of reading;¶
Expand the definition of text;¶
Extend the range of comprehension tasks that require knowledge application;¶
Augment and expand the cognitive targets and the approaches to reporting performance on them;¶
Expand how language structures and vocabulary are defined and measured; and ¶
Include, measure, and report on the role of engagement in reading performance. ¶
At the heart of the Visioning Panel's guidelines was a commitment to equity, guided by two priorities in accordance with the most recent standards of fairness and equity in large-scale assessment to accomplish the following:¶
Measure disparities in students' reading achievement in a way that minimizes test bias to the maximum extent (American Educational Research Association, American Psychological Association, and National Council of Measurement in Education, 2014; International Testing Commission, 2019; Task Force on Assessment of the International Reading Association, 2010); and¶
Describe disparities in "access to resources and opportunities, including the structural aspects of school systems that may impact opportunity and exacerbate existing disparities in family and community contexts and contribute to unequal outcomes" in reading (the National Academies of Sciences, Engineering, and Medicine, 2019, p. 3).

Deleted: and fairness

Deleted: guidelines

Formatted: Indent: First line: 0"

Deleted: This updated framework for the 2026 NAEP Reading Assessment addresses reading comprehension within a sociocultural context. This framing is the natural outgrowth of recent understandings about the social and cultural nature of all learning and human development. The 2002 report of the RAND Reading Study Group identified three key components of reading comprehension—reader, text, and activity—and situated them in sociocultural contexts. The term sociocultural refers to the social and cultural features and practices of contexts, such as schools, homes, and communities, where students learn to read and engage in reading (Lee, 2020; Pacheco, 2015, 2018; Skerrett, 2020). This sociocultural perspective is important to reading comprehension assessment because it acknowledges that these practices influence how readers approach, engage with, and make meaning from texts (Mislevy, 2016; 2019). ¶
Since the RAND report, an even broader consensus has emerged across the multiple disciplines of the learning sciences—including psychology, developmental studies, anthropology, linguistics, cognitive science, and even biology—recognizing the central role of culture in lifelong learning (National Academy of Sciences, 2018). In this emerging consensus, learning—and reading—are still, at their cores, cognitive processes. However, cognitive acts, including reading, are influenced by the particular contexts in which texts are written and in which reading takes place.

Drawing from previous frameworks and newer understandings, this updated NAEP Reading Framework attends to four key features of reading comprehension—contexts, readers, texts, and activities. At the heart of the 2026 NAEP Reading Framework is the definition of reading comprehension:

> **Reading comprehension is making meaning with text,** a complex cognitive process shaped by the environments in which students live, including family, community, and school. To comprehend, readers:
>
> - Engage with text in print and multimodal forms;
> - Employ personal resources that include foundational reading skills, language, knowledge, and motivations; and
> - Extract, construct, integrate, critique, and apply meaning in activities across a range of contexts.

Readers draw on a range of resources to make sense from text:

- What readers know about a topic;
- What readers know about texts and how they work;
- Internal processes, or foundational skills, needed to render text sensible, including phonemic awareness, letter-sound knowledge, and word- and sentence-reading skills;
- Higher order cognitive processes, such as attention, working memory, language comprehension, inferential reasoning, and comprehension monitoring; and
- Socially and culturally situated knowledge and practices from home, community, and school.

Advances in measurement and in digitally administered assessment of reading comprehension, already initiated by NAEP in 2017, allow for a large-scale assessment that is more accessible to a greater number of individuals (National Center on Educational Outcomes, 2016). These advances have also allowed the assessment design to gather more information on environmental factors that influence the cognitive processes underlying reading comprehension. Enacting the definition of reading comprehension in the 2026 NAEP Reading Assessment—described in this and subsequent chapters of the updated Framework—will enable NAEP to:

- Develop assessments with greater ecological validity (e.g., reading with purpose, applying what one learns from reading to a new task, benefiting from the presence of Universal Design elements that are typically available when reading outside of an assessment context);
- Draw on a greater range of texts and tasks representative of students' diverse experiences;
- Report on a broader array of the resources that students bring to bear in the act of reading (knowledge, language, opportunities to learn); and
- Increase the quantity and quality of information that is available to users of NAEP data to make inferences about student reading achievement in the U.S.

Overview of the Updated NAEP Reading Framework's Key Components

Deleted: The understanding of reading comprehension informing the 2026 NAEP Reading Framework is an outgrowth of earlier and current cognitively oriented work in reading comprehension (Anderson & Pearson, 1984; Kintsch, 1998; RAND Reading Study Group, 2002; Pearson, et al., 2020). Descriptions of the cognitive activities involved in constructing meaning have increasingly implicated social and cultural dimensions over time, dimensions that were also foreshadowed in NAEP reading frameworks adopted by the Governing Board in 1992 and 2004. Research evidence has highlighted that, like all human learning, reading comprehension is a meaning-making activity that involves socially and culturally specific characteristics and practices (Bronfenbrenner & Morris, 2006; Lee, 2016b, 2020; National Academy of Sciences, 2018; Zelazo, 2013). ¶

Deleted: these

Deleted: research

Deleted: The cognitive processes involved in reading are shaped by social interaction and mediated by many aspects of cultural practice, including the traditions and modes of speaking, that are part of students' daily lives (Nasir & Hand, 2006).

Deleted: '

Deleted: social and cultural influences. To

Formatted: Indent: Left: 0.25", Outline numbered + Level: 1 + Numbering Style: Bullet + Aligned at: 0.5" + Indent at: 0.75", Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Deleted: contexts

Deleted: ¶
The definition of reading comprehension included in the 2026 NAEP Reading Framework acknowledges and incorporates the cognitive roots of previous reading frameworks. Also, the definition illustrates how what readers know, do, and understand from reading is tied to the variations in knowledge, skills, and experiences they bring to their reading from experiences at home, in their communities, and in school. It embraces the understanding that social and cultural practices also influence texts, including who reads and writes them and under what circumstances, how they are generated, how they appear, and how they are used. And finally, the definition emphasizes the integration of reading with other ... [31]

Deleted: address

Deleted: the

Deleted: sociocultural

Deleted: aspects

Deleted: of

Deleted: known

Deleted: as

Deleted: motivations, prior experiences, agency,

Deleted: precision

Deleted: of

The new framework maintains many aspects of the 2009–2019 NAEP Reading Framework. It also introduces some changes in the assessment design that are based on current research in human development and learning, including reading comprehension. The advent of digitally based assessments in 2017 has allowed NAEP to provide an engaging assessment experience for students and explore new testing methods and question types. Framework updates also reflect trends in international reading comprehension assessments, such as the Programme for International Student Assessment (PISA) and the Progress in International Reading Literacy Study (PIRLS).

Comprehension Targets

Like its predecessors, the 2026 NAEP Reading Assessment engages students in reading texts and responding to questions that assess their comprehension of these texts. Comprehension Targets are used to generate test items that assess four important dimensions of reading comprehension. Three of these—Locate and Recall, Integrate and Interpret, and Analyze and Evaluate—are similar to the cognitive targets used in the 2009–2019 Framework. One new target—Use and Apply—reflects a frequent and authentic purpose in disciplinary and workplace reading. Assessment of students' comprehension of vocabulary and language structures is systematically woven throughout the comprehension items.

Other Key Components

Disciplinary contexts for reading have taken on an expanded role in the 2026 NAEP Reading Framework to mirror the increased focus in schools on reading comprehension within disciplines, as well as in state standards and large-scale reading comprehension assessments. Two broad purposes for reading comprehension—reading to develop understanding and reading to solve a problem—will be delineated to systematically sample students' reading performance in literature, science, and social studies. Texts, too, are sampled to address purposes within disciplines, affordances offered by digital and multimodal formats, and text complexity criteria for each tested grade.

Reporting 2026 NAEP Reading Assessment Results

Results of the NAEP Reading Assessment are reported in terms of average scores for groups of students on the NAEP 0–500 scale and as percentages of students who attain each of the three achievement levels (*NAEP Basic*, *NAEP Proficient*, and *NAEP Advanced*). They are reported in the aggregate for the nation, states, and select large urban districts participating in the NAEP Trial Urban District Assessment; they are not reported for individual students, classrooms, or schools.

The 2026 NAEP Reading Framework updates the reporting system. The aim is to provide more nuanced reporting and useful data to key stakeholders across the nation. Currently, results of the NAEP Reading Assessment are disaggregated by gender, race/ethnicity, socioeconomic status, English learner status, state, region, type of community, public and nonpublic school, and literary and informational texts. Building on this system, the 2026 Framework proposes to disaggregate results by disciplinary contexts—literature, social studies, and science—rather than literature and informational texts. In addition, reporting categories are expanded to include former English (ELs) learners in addition to current ELs and non-ELs, in order to describe student performance in more precise and detailed ways.

Deleted: scientific

Deleted: A continuing commitment to equity, non-biased and valid assessments, and the principles of Universal Design of Assessments were central to the updates in the 2026 NAEP Reading Framework.

Deleted: contexts

Deleted: Finally, task-based, motivational, and knowledge-based Universal Design Elements are included as appropriate to support precise measurement of students' reading comprehension in ecologically valid ways.

Deleted: to emphasize equity, rigor, precision, and validity.

Deleted: (1) socioeconomic status within race/ethnicity and

Deleted: (2)

Comparison of the 2009–2019 NAEP Reading Framework and the 2026 NAEP Reading Framework

The framework for the 2026 NAEP Reading Assessment updates the framework developed and used for the 2009–2019 assessments. Building from this previous framework and on digital innovations, updates include:

- Expansion of the definition of reading comprehension. Reading comprehension is defined as making meaning with text and four key features are highlighted—contexts, readers, texts, and activities.
- Emphasis on two additional, research-based factors: how reading varies across disciplines; and the increasing use of digital and multimodal texts.

Key similarities and differences between the two frameworks are presented in exhibit 1.1. While updated, the continuity between the current framework and assessment and the 2026 NAEP Reading Framework is substantial.

Exhibit 1.1. Similarities and Differences Between the 2009–2019 and 2026 NAEP Reading Frameworks

	Current Framework and Assessment	2026 Framework Update
Comprehension Targets	Locate and Recall Integrate and Interpret Critique and Evaluate	Locate and Recall Integrate and Interpret Analyze and Evaluate Use and Apply
Disciplinary Contexts	Literary Text Informational Text	Literature Contexts Social Studies Contexts Science Contexts
Purposes	Specific purposes communicated to students for scenario-based tasks in digitally based assessment as of 2017	Broad Purposes ● Reading to Develop Understanding ● Reading to Solve Problems Specific purposes for all assessment tasks are communicated to students
Text Types	Literary Texts Informational Texts	Literature Texts Social Studies Texts Science Texts
Text Source	Authentic	Authentic except in rare instances
Text Format	Digital texts as of 2017 ● Static – non-moving print, graphics, or images on screen ● Dynamic – navigation across modes (print, video, other) or nonlinear locations (hypertext link)	Digital texts ● Static – non-moving print, graphics, or images on screen ● Expanded use of dynamic formats – navigation across modes (print, video,

Deleted: The framework also proposes to measure an expanded set of contextual variables via questionnaires and the increased use of digital process data to provide more information on student performance. The contextual variables are clustered by two sets of reader characteristics: (1) cognition and metacognition and (2) engagement and motivation; and by two sets of environmental characteristics: (1) perceptions of school and community resources and (2) perceptions of teacher, instructional, and classroom supports. Ultimately, the framework envisions a reporting system that has enhanced explanatory capacity to assist educators in accessing, interpreting, and acting on the valuable information provided in NAEP reports and databases.¶

Deleted: to explicitly acknowledge the sociocognitive processes of reading.

Deleted: three

Deleted: concepts

Deleted: (1) how social and cultural experiences shape learning and development; (2)

Deleted: (3)

	Current Framework and Assessment	2026 Framework Update
		other) or nonlinear locations (hypertext link)
Text Complexity	Determined by: ● Expert judgment ● Passage length ● Two or more research-based readability measures	Determined by: ● Expert judgment ● Passage length ● Quantitative and qualitative research-based complexity measures
Language Structures and Vocabulary	Vocabulary assessed Potential for subscore	Language structures and vocabulary assessed No subscore
Universal Design Elements (UDE)	Digitally based assessment as of 2017 includes tools and support features: ● Highlighting and notetaking ● Text-to-speech on Directions and Help screens ● Zoom-in and selection of color schemes ● Sequential directions and transitions ● Look-back buttons to return to relevant section of text ● Graphic organizers ● Item foreshadowing ● Multi-part response frames ● Purpose statements ● Task characters (avatars that act as partners in simulated settings) ● Pop-up notes for definitions of vocabulary ● Resetting by providing correct response to answered questions ● Topic or passage introductions	Types of UDEs and possible examples: ● Task-based UDEs – Highlighting and notetaking – Text-to-speech on Directions and Help Screens – Zoom-in and selection of color schemes – Sequential directions and transitions for reading collection of texts – Look-back buttons to return to relevant section of text – Graphic organizers – Item foreshadowing – Multi-part response frames – Samples of student writing as examples ● Motivational UDEs – Explicit connections between broad and specific purposes – Task characters that provide oral or written directions, act as peers or experts, or serve as an audience ● Knowledge-based UDEs – Text providing brief topic previews – Pop-up notes for definitions of words or phrases that are rare and not part of the comprehension target being tested – Resetting by providing correct response to answered questions

	Current Framework and Assessment	2026 Framework Update
Reporting	Overall scale score and achievement levels (NAEP Basic, NAEP Proficient, NAEP Advanced) Disaggregation by gender, race/ethnicity, socioeconomic status, English learner status, state, region, type of community, public or nonpublic school, and literary and informational texts Data collected from student, teacher, and administrator questionnaires on contextual variables of interest Some data collected from students' test taking behaviors (process data) in digital administrations	Overall scale score and achievement levels (NAEP Basic, NAEP Proficient, NAEP Advanced) Disaggregation by all existing categories, adding ● Disciplinary contexts ● Socioeconomic status within race/ethnicity (subject to the availability of valid information on students' socioeconomic status) ● Former English learners (ELs) as well as current ELs and non-ELs Data collected from student, teacher, and administrator questionnaires on expanded set of contextual variables Data collected from students' test taking behaviors (process data) on expanded set of contextual variables

The remainder of the framework is organized to provide greater detail about the proposed content and design of the assessment and the reporting of results:

- Chapter 2 presents the **2026 NAEP Reading Assessment**, including the definition of reading comprehension and major assessment components.
- Chapter 3 describes the **Development of the 2026 NAEP Reading Assessment**, including specific design elements.
- Chapter 4 explains the **Reporting of NAEP 2026 Results**, including the expansion of reporting categories, contextual variables, and explanatory reporting capacity.

Deleted: 4

The 2026 NAEP Reading Framework recommends updates necessary to deliver assessments that are relevant and valid measures of student achievement in the U.S. The 2026 Framework builds on the current NAEP framework and operational assessment, especially the advances made possible by digitally-based assessment, by drawing on current understandings of reading comprehension and assessment. Chapter 2 provides a detailed description of the components that will be included in NAEP Reading assessments that students will take beginning in 2026. The chapter begins with the 2026 NAEP Definition of Reading Comprehension, presents the definition's origins in policy and scholarship on reading comprehension, and concludes with a description of the components of the assessment.

The NAEP Definition of Reading Comprehension

The 2026 NAEP Reading Framework attends to four key features involved in reading comprehension—contexts, readers, texts, and activities. The cognitive processes involved in reading are shaped by social interaction and mediated by many aspects of cultural practice, including the traditions and modes of speaking, that are part of students' daily lives (Nasir & Hand, 2006). At the core of the 2026 NAEP Reading Framework is the definition of reading comprehension:

> **Reading comprehension is making meaning with text,** a complex cognitive process shaped by the environments in which students live, including family, community, and school. To comprehend, readers:
>
> - Engage with texts in print and multimodal forms;
> - Employ personal resources that include foundational reading skills, language, knowledge, and motivation; and
> - Extract, construct, integrate, critique, and apply meaning in activities across a range of contexts.

Texts. Texts are generated by authors to communicate to readers. Texts take many forms, drawing on multiple genres and combinations of genres. They relay vastly different content to address many kinds of purposes. They draw on a wide array of modalities (e.g., static print, nonlinear hypertext, images, videos), sometimes combining modalities into multimodal forms (e.g., print with images or links to videos). They may be printed on paper or published in digital forms. They also differ in complexity, a term that usually refers to the density and nuance of texts' ideas and language structures.

Texts are composed according to conventions tied to cultural traditions and social practices. These traditions and practices are developed within and across such disciplines as literature, science, or history. Such conventions include genre traditions favored by disciplines and modalities that are selected because of the ways they communicate certain kinds of ideas. Texts also vary in terms of the people, points of view, and experiences that are or are not represented. This means that texts may be readily understood by readers who find the ideas familiar or compelling but more challenging to others.

Activities. Activities include all the things readers do as they comprehend text and communicate and apply their understanding after reading. For example, readers *read the lines*,

Deleted: , fair,

Deleted: shaped by

Deleted: students' social and cultural experiences.

Deleted: Key Terminology in the Definition¶
Each feature of the definition (contexts, readers, texts, activities) is important to understand how readers make meaning in the presence of texts. ¶
Contexts. A central principle of the 2026 NAEP Definition of Reading Comprehension is that, as a human meaning-making activity, reading comprehension is situated within, and shaped by, social and cultural contexts. Social contexts, the settings within which individuals interact with one another, are governed by particular norms and expectations for the roles that different participants take up (e.g., student and teacher; youngest and eldest sibling). Social contexts are also inherently cultural. Cultural socialization occurs in classrooms, families, communities, and many other social contexts. With repeated ways of acting, interacting, knowing, believing, and valuing being passed down across generations all social groups develop cultures (Nasir & Hand, 2006). ¶
Experiences students have in these contexts shape every aspect of reading comprehension: understanding of what to do, how to engage with text, and how to respond to and learn from reading. Contexts influence everything that readers bring to reading—including the language, knowledge, motivations, and cognition that are acquired and refined in home, community, and school settings. Contexts shape the texts readers read. Although there is a common thread to the cognition involved in reading across contexts, much of the process of comprehension is influenced by context (Scribner & Cole, 1981; Skerrett, 2020). ¶
Readers. Each reader is a distinctive human being who brings a unique and diverse repertoire of cultural, cognitive (including metacognitive), motivational, and linguistic resources to every encounter with text. These resources are developed through experiences in multiple settings and communities and applied as readers make sense of text. For instance, first graders will use their knowledge of the stories they have listened to at home and in daycare settings to understand the stories they now have to read on their own. Adolescents in the U.S. would face a challenge when reading an unfamiliar text about the game of cricket in India, using their knowledge of other sports to make sense of the text. Bilingual readers often use what they know about reading in one language to read in another language (August & Shanahan, 2006; García & Godina, 2017). Readers' motivations and purposes are also impacted by their previous experiences and by the particular contexts in which the reading is being performed. They read to enjoy and be carried away by stories, to appreciate an author's use of language, to learn about themselves and the natural and social worlds in which they live, or to gather information and insight to act on the world. They re ... [32]

Deleted: artifacts

Deleted: heir ideas

making sense of individual propositions in a text; they *read between the lines*, drawing inferences that connect ideas in one part of the text with ideas in another; and they *read beyond the lines*, using what they know to fill in gaps and draw more global meanings, such as themes and concepts. Evidence of comprehension-related activity comes from the things readers do to communicate and apply their understanding. For example, readers discuss their understanding of text and engage in activities in which they apply their understanding, such as preparing for a debate. They offer evaluations of texts, and they apply what they learn from their reading to solve problems and act in the world. They also use foundational skills, such as decoding, word recognition, and fluency (Vorstius, Radach, Mayer, & Lonigan, 2013). While these activities enable comprehension, they do not provide direct evidence of comprehension; thus, they are not directly assessed in the NAEP Reading Assessment.

The Specialized Role of Readers' Knowledge. Many different kinds of knowledge play important roles in reading comprehension (Willingham, 2006). The categories of knowledge include world knowledge, knowledge of the topics of texts readers encounter, knowledge of text genres and structures, and linguistic knowledge, including vocabulary and syntax. In the process of extracting meaning, readers use this knowledge to clarify potential sources of ambiguities, including use of pronouns, words with multiple meanings, and ambiguous syntax. These forms of knowledge enable readers to make connections between adjacent ideas in texts even when authors do not make these connections explicitly. In more transparently construction-oriented processes, readers use knowledge to fill in gaps left by the author. Readers also use knowledge related to key ideas or themes in the text to construct mental models of meaning.

Of all of the types of knowledge involved in reading comprehension, the role of topic knowledge is probably the best understood. Contemporary cognitive models of reading describe the essential role of topic knowledge in text comprehension (Graesser, Singer, & Trabasso, 1994; Kintsch, 1998; McCarthy & McNamara, 2021; van den Broek, Risden, Fletcher, & Thurlow, 1996). These models represent the relationship between knowledge and comprehension as one in which existing knowledge is continually activated and integrated with textual information as readers develop a propositional understanding and, ultimately, a coherent mental representation of the text. Moreover, a large body of research has documented the impact of readers' topic knowledge and domain knowledge on reading comprehension across grade levels and text genres (e.g., Pearson, Hansen, & Gordon, 1979; Taft & Leslie, 1985; Alexander, Kulikowich, & Schulze, 1994). These studies also explain that while topic knowledge often influences readers' ability to recall information from text and to answer text explicit comprehension questions, the most consistent impact of topic knowledge is on readers' abilities to respond to questions that require bridging inferences (connecting information within texts) and more global inferences (such as understanding concepts or themes). Readers may be generally skilled at such mental operations but not able to do so when texts focus on unfamiliar topics.

Updating the NAEP Reading Framework

The 2026 NAEP Reading Framework is updated to reflect two research-based developments that help to ensure that the NAEP Reading Assessment is a valid measure of reading achievement by students in the nation's schools. The first is how reading varies across disciplines. The second regards the use of digital and multimodal texts.

Formatted: Font: Not Italic

Formatted: Indent: First line: 0"

Deleted: Reading comprehension depends on who is doing the reading, what they are reading, why and where they are reading, how they have been prepared for the reading, with whom they are reading, and what schools and society will take as evidence of successful comprehension. Because all of these factors influence a complex process like reading comprehension, assessments must be sufficiently complex in their design and implementation (Mislevy, 2016). ¶
Roots of the Definition¶
The NAEP Definition of Reading Comprehension and the resulting assessment are grounded in important developments in reading comprehension theory, research, practice, and policy over the three decades since the first NAEP Reading Framework was published in 1992. This definition draws on robust features from earlier NAEP reading frameworks and research describing cognitive processes involved in reading comprehension. It also attends to recent sociocultural understandings of learning and development, to disciplinary reading, and to an expanding conceptualization of what counts as text in today's society.¶
NAEP's definitions of reading comprehension in both the 1992-2007 Reading Framework and the 2009-2019 Reading Framework reflected dominant cognitive models of their times. The construction integration (C-I) models proposed by theorists such as Kintsch (1998), Perfetti (1999), and van den Broek (van den Broek, Risden, Fletcher, Thurlow, Britton, & Graesser, 1996) are still regarded as the most valid and useful cognitive accounts of reading comprehension. These models emphasize the multiple levels of meaning readers create, including a representation of the surface form that reflects accurate decoding; a text-base that includes all of the key ideas in the text plus the text-based inferences that link ideas within texts; and a situation model that represents the integrative links readers make between ideas expressed in the text and the knowledge they bring to reading.¶
Although earlier NAEP Reading frameworks were grounded in cognitive models of comprehension, they also acknowledged the importance of readers' purposes and the contexts in which they read and learned to read. In the first Reading Framework published in 1992, reading comprehension was defined as "… a complex process that involves an interaction among the reader, the text, and the context in which something is read" (p. 6). Purpose was mentioned when describing characteristics of good readers, who "can read a variety of texts for different purposes" (p. 9). The 2002 RAND Model of Reading Comprehension, which was heavily influenced by C-I models, was ex[... [33]

Deleted: three

Deleted: precise, fair, and

Deleted: accurate

Deleted: comprehension

Deleted: is how sociocultural experiences shape learning and development, including the learning and development of reading comprehension and, consequently, its assessment. The second

Deleted: third

A first update in the 2026 NAEP Reading Framework draws on recent research demonstrating that reading and texts are shaped by disciplinary contexts. While a core set of academic literacy skills and strategies can be applied across areas of study, there are important differences in disciplinary reading practices. These include differences in the genres and discourse conventions and structures of texts, what counts as explanation, argument, and evidence, and the kinds of reasoning needed to formulate new understandings (Goldman, et al., 2016; Moje, 2007; Shanahan & Shanahan, 2008; Snow, 2010). These differences, which are related to the core activities in each discipline, require readers to employ different resources as they read and respond to text.

Also newly explicit in the 2026 Framework is recognition of the multimodal nature of texts used across all aspects of society. The widespread presence and rapid evolution of computers, smart devices, and software platforms have changed society's ideas about what counts as text and its uses. Students read digital/multimodal texts in and out of school. Even though there is a common thread to reading in print and multimodal texts, there are also substantial differences, particularly around navigation (Coiro, 2020; Hartman, Morsink, & Zheng, 2010; Serafini & Gee, 2017). The implication is that the NAEP Reading Assessment must sample multiple modes of text.

These updates allow the 2026 NAEP Reading Framework to better describe how well U.S. students comprehend what they read in texts and situations that more closely approximate reading practices in today's schools and society as a whole. By building on past frameworks and research traditions while embracing more recent developments in assessment, NAEP will continue to both lead and reflect reading assessment in the nation.

The NAEP 2026 Reading Assessment and the Definition of Reading Comprehension

The NAEP Definition of Reading Comprehension provides the foundation for how NAEP will assess reading comprehension. Each of the four aspects of the NAEP Definition of Reading Comprehension—contexts, readers, texts, and activities—is reflected throughout the 2026 NAEP Reading Assessment. The remainder of this chapter describes and explains key components of the NAEP Reading Assessment as well as their relationship to the definition. (See Exhibit 2.1.)

Components. The section begins with the core component of the assessment, the reading comprehension assessment items. After describing the items, the chapter takes on the challenge posed by Cronbach (1990) and Mislevy (2019), which is to address the variability inherent in complex domains of learning, including reading comprehension. To that end, five additional updated components are also presented: disciplinary contexts, purposes, texts, and contextual variables. Taken together, these components ensure that NAEP will assess students' reading comprehension in ways that reflect the NAEP Definition of Reading Comprehension and the natural variation that readers encounter in reading in home, school, community, and workplace settings. In this way, NAEP captures information on a wide range of factors that may influence reading comprehension.

Comprehension Items: The Role of Comprehension Targets

As in previous NAEP assessments, the 2026 NAEP Reading Assessment will engage students in reading sets of texts and responding to questions that assess their comprehension of

Deleted: Literacy scholarship has documented that cognitive actions associated with reading comprehension reflect the language and literacy practices (broadly, any activities through which students make and communicate meaning) of schools and communities (Frankel, Becker, Rowe, & Pearson, 2016; Heath, 1982; Lee, 2017; Scribner & Cole, 1981; Smagorinsky, 2001; Street, 1984), including disciplinary communities (Goldman, et al, 2016; Moje, 2007). This insight mirrors the broad consensus that has emerged across the learning sciences that learning is sociocultural in nature (Brown, Collins, & Duguid, 1989; Nasir & Hand, 2006). This finding is reflected in a 2018 report of the National Academies of Sciences, Engineering, and Medicine [NASEM]. The report explains that "each learner develops a unique array of knowledge and cognitive resources in the course of life that are molded by the interplay of that learner's cultural, social, cognitive, and biological contexts" (NASEM, p. 33). ¶
This NASEM finding is also reflected in other large-scale assessments. PIRLS, the international assessment of reading for fourth grade students, notes that "social interactions about reading in one or more communities of readers can be instrumental in helping young students gain an understanding and appreciation of texts and other sources of information" (Mullis & Marten, 2021, p. 7). PISA, an international assessment for many subjects for 15-year-olds, similarly states that reading "is viewed as an expanding set of knowledge, skills, and strategies that individuals build on throughout life in various contexts, through interaction with their peers and the wider community" (OECD, 2019, p. 27). ¶
Scholars who study assessment closely (Greeno, 1998; Mislevy, 2016, 2019; Pellegrino, 2013) also note the importance of attending to contextual factors that shape student performance in any domain of expertise or learning. Measurement scholar Mislevy's (2019) summary of the implications of recognizing these factors for educational assessment is far-reaching:¶
Situative, sociocognitive (SC) psychology is forcing a reconception of educational assessment. The SC perspective emphasizes the interplay between across-person linguistic, cultural, and substantive patterns that human activity is organized around and within-person cognitive resources that individuals develop to participate in activities. Rather than seeing assessment primarily as measurement, we are increasingly seeing it as an evidentiary argument, situated in social contexts, shaped by purposes, and centered on students' developing capabilities for valued activities... Implications follow for current challenges such as assessing higher order skills, performance in digital environments, and diverse student populations. (p. 164)¶ ... [34]

Deleted: second

Deleted: is the recognition of

Deleted: account more precisely for

Deleted: , universal design elements

Deleted: accounts

Deleted: for

Deleted: and mitigates potential bias that might result from a narrower operationalization of reading comprehension

these texts. Comprehension Targets are used in NAEP to generate the questions, i.e., the assessment items, that students respond to as they take the test. Students' answers to these questions provide the observable data that NAEP uses to represent how effectively students engage in important comprehension processes, such as recalling texts and forming connections among ideas within and across texts, when reading various kinds of texts. Three of the four targets— ***Locate and Recall***, ***Integrate and Interpret***, ***Analyze and Evaluate***— are closely aligned with those in the 2009-2019 NAEP Reading Framework. An additional target, ***Use and Apply,*** has been added to reflect the importance of applying comprehension to new situations.

Each comprehension target involves inferences that readers tend to find more or less challenging in general. Items based on each target will range in difficulty, depending on the particulars of the questions in relation to the texts they are designed to probe. Building on the attention to vocabulary in the 2009-2019 Framework, the 2026 assessment also attends to structures of language within each comprehension target.

Locate and Recall. The first Comprehension Target is Locate and Recall. In order to comprehend, readers need to identify important information and form connections among ideas in the text as they move through it. In addition, readers often need to locate information to fulfill a particular purpose, aid recall, and repair understanding. These kinds of processing help readers build a literal understanding of what the text "says".

Items assessing the Locate and Recall target typically focus on information stated directly in a single location in a text, such as a sentence, a paragraph, adjacent paragraphs, or a single graphic. However, in some cases, readers may need to navigate across different pages or documents, including hyperlinked and multimodal texts, to find additional information that is relevant to the test item. Test items might ask readers to recall or locate specific information about characters or settings in a story; or to locate a specific piece of information from a table in an expository text. Locate and Recall items can also require readers to form connections across text segments that are near one another in the text, such as fairly straightforward inferences about the relationships between ideas presented in adjacent sentences (e.g., A caused B or A occurred before B). Finally, readers may be asked to infer the meanings of unfamiliar words using information in the sentences immediately surrounding that word.

Integrate and Interpret. The second Comprehension Target describes what students do as they Integrate and Interpret information from one or more texts. These processes can involve making connections across sentences, paragraphs, or sections within or across texts to synthesize ideas under a common theme (e.g., justice or loss) or idea (e.g., how food goes from the farm to tables in people's houses). In making these connections, readers rely on their understanding of the ideas in the texts, their disciplinary knowledge, their knowledge of text genres, and even their knowledge of how language works to communicate ideas. In order to engage in these processes, readers may be required to navigate complex hyperlinks or multimodal elements, such as video or interactive graphics.

Test items that gauge readers' ability to Integrate and Interpret may ask readers to compare and contrast characters and settings, examine causal and chronological relations across aspects of text, or formulate explanations for events or information in texts. For example, items may ask readers to explain or predict a character's behavior by relying on multiple pieces of text information about that character's history and dispositions, or they might ask readers to describe how the setting of a story contributes to the theme. Integrate and Interpret items might also ask

readers to recognize how specific features of language signal relationships or viewpoints within a text. For example, readers might be asked to make judgments about characters based on the adjectives used to describe them or to rely on signal phrases (e.g., "to the contrary") to understand the connections among ideas.

Analyze and Evaluate. The third Comprehension Target, Analyze and Evaluate, describes the processes associated with examining and assessing one or more texts during and after reading. Readers may analyze by closely examining the choices an author makes about content and form and how those choices affect meaning. Readers may then use those analyses to evaluate a text by judging various aspects of the text as well as its overall effectiveness. In order to engage in Analyze and Evaluate processes, readers must view texts in relation to knowledge from other sources. Sources may include their existing knowledge base (Alexander, 2012; Lee, 2011) or common tools and criteria used in literary analysis, historical reasoning, or scientific argumentation (Lee & Spratley, 2010; Greenleaf et al., 2016; van Drie & van Boxtel, 2008). Readers also draw on their knowledge about and preferences for particular rhetorical strategies, such as the use of language, organization of text, or articulation of claims and evidence.

In items associated with the Analyze and Evaluate target, readers might be asked to evaluate the coherence, credibility, or quality of one or more texts. Readers may be asked to make judgments about the effectiveness of an author's use of figurative language, the degree to which the author provides sufficient evidence to support a claim, or the trustworthiness of the source (e.g., venue and author) (Bråten, Stadtler, & Salmerón, 2018; Meola, 2004; Ostenson, 2014; Wineburg, 1991; Wineberg & McGrew, 2017). For example, readers might use information appearing in one text as the basis for evaluating the ideas or the use of language in a second text.

Use and Apply. The final Comprehension Target, Use and Apply, reflects the culmination of comprehension, in which understandings acquired during reading are used in new situations or applied in the development of novel ideas and products (Goldman, Greenleaf, & Yukhymenko-Lescroart, 2019; Pearson, Palincsar, Biancarosa, and Berman, 2020). This set of targets reflects contemporary understandings that comprehension may involve a series of processes that culminate in readers taking some kind of action in the world outside of text. As they engage in Use and Apply processes, readers must consider how to reframe ideas from their reading and experiences to create a new product for a specific purpose and audience (Marzano, 1988). As readers reflect on how to respond to items that require such processes, they take into account their purposes, norms established by genre and disciplinary conventions, as well as expectations about what is deemed appropriate and compelling to members of the target audience (Gee, 2001; Goldman et al, 2011; Moje, 2005).

Items designed to assess Use and Apply processes will ask readers to use information they acquire through reading to solve a problem or create a new text. For example, after reading a set of commentaries, readers might be asked to produce a blog-type message for a public audience that captures the most relevant information or offers an argument about an issue. Readers might also be asked to use one or more texts as a model for developing a new text or graphic representation. In a literature context, readers might be asked to rewrite an aspect of a story in accordance with a particular, specified goal.

Comprehension Targets and the NAEP Definition of Reading Comprehension. The Comprehension Targets reflect the understanding that the extent to which a reader succeeds at

particular reading tasks is dependent on many factors related to the reader's experiences, knowledge, language development, and motivations. The Comprehension Targets also reflect the centrality of readers' use of reading processes, including a range of different kinds of inferential reasoning, in the meaning they construct. In developing items that target a range of knowledge and skills under conditions that replicate many aspects of authentic reading, the NAEP Reading Assessment provides a more precise and ecologically valid measure of students' reading comprehension.

Contexts and Purposes

As stated earlier in this framework, a central principle of the NAEP Definition of Reading Comprehension is that, as a human meaning-making activity, reading comprehension is a purpose-driven activity, situated within contexts that shape the readers' engagement with text and that influence how readers respond to and learn from the experience of reading. This section describes how two expanded components of the 2026 NAEP Reading Assessment, Disciplinary Contexts and Purposes, contribute to this contextualization.

Disciplinary Contexts. Given recent advances in theory, research, and practice about reading within disciplines, NAEP has elevated the importance of disciplinary reading in literature, science, and social studies to reflect the increased importance of disciplinary reading in schools, state standards, and large-scale reading comprehension assessments. Students will read in each context, and their reading performance on test items will be reported by disciplinary contexts, along with an aggregate score for performance across all three. Reading in such contexts involves reading texts that are drawn from the range that students encounter when reading about literature, science, and social studies. It involves engaging in tasks that yield new understanding, enable problem-solving common to such contexts, and focus on historical and contemporary social issues.

Literature Contexts. Perhaps more than in any other disciplinary domain, reading is the center of literary study and enjoyment. Themes of human experience pervade works of literature—nature and humanity, struggle and survival, love and friendship, loss and betrayal, victory and defeat, mortality and meaningfulness. Reading literary texts, such as poetry, fictional and nonfiction narratives, and criticism, provides opportunities for enjoyment and for reflection and analysis around these themes, including how they shed light on their own experiences and social worlds. Literature also often provides opportunities to connect with cultures and experiences similar to or different from one's own, extending readers' understandings about the world. Literature also invites its readers to examine text as a repository of language, rhetorical moves, and structure; to connect its ideas to those in other texts and those of otherauthors and literary traditions; and to situate problems in contemporary and historical contexts.

Science Contexts. Science contexts are primarily focused on observing and explaining the natural world. Although these scientific activities do not depend exclusively on reading, texts play an important role in learning about and communicating science ideas in school and non-school settings. Learning the concepts and processes of science in school involves the use of varied texts to describe, report, and articulate claims about the natural world (e.g., textbooks) and to record systematic efforts to act upon it (e.g., observation protocols, lab notes, experimental descriptions, journal articles). Outside of schools, individuals often access scientific information (e.g., in newspapers and on internet sites) needed to understand issues and solve problems. Moreover, the application of reading to understanding and acting upon the natural world calls on

Deleted: , and perceptions of self

Deleted: every

Deleted: aspect of

Deleted: As a result of this principle, the 2026 NAEP Reading Assessment contextualizes almost every component of reading comprehension.

an array of reading strategies, as well as understandings about how scientists determine findings and what constitutes credible evidence for those findings.

Social Studies Contexts. Social studies includes history, geography, cultural studies, civics, and government, with less common coverage of disciplines such as sociology and anthropology. These fields offer unique ways of thinking and organizing knowledge and investigating social systems and events, current and past. In schools, social studies texts provide students with an intellectual context for studying how humans have interacted with each other and with the environment over time (College, Career, and Civic Life Framework for Social Studies, 2013). Social studies explores how humans organize societies and governments, how societies make use of available resources, and how cultures develop and change over time. In order to understand social studies texts, readers bring both conceptual tools needed to understand patterns in the social world (e.g., trade-offs, how perspective impacts representation) and understandings about how claims are developed and supported. Reading in social studies also requires the application of a broad range of the reading processes described in the comprehension targets.

Purposes. Purposes are a key component of the 2026 NAEP Reading Assessment. Purposes reflect a commitment on the part of NAEP to ensure that readers know why they are engaging in every part of the assessment, and to reflect the fact that all reading is done in relation to specific purposes. Within the disciplinary contexts described above, the assessment will be oriented toward purposes for reading, and these purposes will be communicated to students throughout the assessment.

Broad Purposes. When students take the 2026 NAEP Reading Assessment, each set of readings and activities they encounter will be situated in one of two broad purposes for reading that reflect standards and curriculum frameworks across the United States—reading to develop understanding and reading to solve a problem.

Reading to Develop Understanding requires students to read texts carefully and respond to comprehension test items generated from the four Comprehension Targets. These items may assess students' understanding of concepts described in a science text or the development of a literary theme, for example. These purposes tend to resemble those associated with items on widely used reading comprehension tests. Readers might read with the purpose of understanding the motives of a particular character in a literary text or read scientific texts to understand the significance of a public health threat.

Reading to Solve a Problem requires that students work across multiple texts and perspectives while solving a problem. These activities entail using information gained during text comprehension in the service of a specific action or to create a product. For example, readers might be asked to use information across four different short texts to develop an argument for or against a city ordinance requiring bicycle lanes on all city streets with a certain traffic load.

Specific Purposes. In addition to these broad purposes, more specific purposes for reading particular texts or engaging in particular tasks will also be communicated to students. For example, within a Literature Context, students may be assigned a role and given a goal, such as working with task characters (avatar collaborators) in a book group to prepare a presentation about which character in a narrative behaved heroically. Or they might be asked to read a brochure for a new bicycle to evaluate how well the claims about the bicycle's qualities are supported with evidence.

Contexts and Purposes and the NAEP Definition of Reading Comprehension. The NAEP Definition of Reading Comprehension describes the role of contexts and purposes in shaping texts and activities related to reading comprehension. This definition relies on research documenting that, when readers taking the assessment know what they are doing, why they are doing it, and what role they are expected to play, the assessment is more likely to serve as a valid proxy for their reading in authentic reading contexts (O'Reilly et al, 2018). Efforts to make contexts and purposes available to students stand in contrast to the practices of many widely used standardized tests of reading comprehension. In some assessments, readers are presented with individual passages and directed to read and answer questions following each passage, with little guidance about the purpose for reading and comprehending the passage. Such tests imply a purpose, namely reading to demonstrate how well one can perform on the test. But they do not explicitly connect with any activity readers might engage with outside of a testing situation. The aim of these components is to reflect the purposes, texts, activities, and resources that influence students' reading in school, home, and community settings.

Texts

Because texts are central to the NAEP Definition of Reading Comprehension, the 2026 NAEP Reading Framework recommends sampling from the large domain of texts that fourth, eighth, and twelfth graders are likely to encounter in school and non-school settings, as is described in more detail in the chapter 3. This portfolio of texts ranges from classic to contemporary text forms that characterize reading within and across varied disciplinary contexts. Texts will be selected with multiple and diverse criteria in mind: cultural diversity, disciplinary representation, and developmental appropriateness with regard to complexity, topic, and modality.

Disciplinary Texts. NAEP will sample texts that are used within the three broad disciplinary contexts described above: literature, science, and social studies. The features of these texts will vary by disciplinary context and include the genres, text types, and discursive, rhetorical, and syntactic structural characteristics specific to texts in those disciplines. Sampling will also consider that such text features are normative rather than absolute, developed to address disciplinary purposes. This means that there is overlap across disciplines regarding the kinds of texts used within disciplines.

Literature Texts. NAEP will draw on literary texts to reflect the range of classic and contemporary genres, text structures, literary language, and cultural traditions that students experience in their classrooms and communities. Literary texts may reflect long-standing cultural traditions, like myths, short stories, novels, drama, and poetry. They can also include current evolving forms, such as fan fiction, author interviews, book reviews, and graphic novels. The challenge of reading literature is also reflected in specific discourse patterns, including word choice, sentence structure, and figurative language. Language used in literature also situates narratives in time and cultural traditions and draws on archetypal characters typical of those traditions. Literature texts may also be ironic, satirical, or narrated from a certain point of view to cue non-literal interpretations (Appleman, 2017; Lee, Goldman, Levine, & Magliano, 2016; Rabinowitz, 1987).

Science Texts. Science texts sampled for NAEP will reflect the formats, language, and structural elements germane to pedagogical, public, and professional science discourse whose purpose is to convey information, findings, and varied applications of scientific ideas. Science

texts include technical information, such as raw data, bench notes, journals, personal communications, handbooks, refereed journal articles, and review articles (Goldman & Bisanz, 2002), as well as more general texts, including press releases, news briefs, websites, and blogs. Such texts draw on varied text structures, such as cause and effect, correlation, problem and solution, sequence, comparison, exemplification, descriptive classification, extended definition, and analogy. Science texts also include many kinds of visuals, including tables, graphs, equations, diagrams, models, and flowcharts, as well as description, exposition, and narrative text (Cromley et al., 2010; Lemke, 1998; van den Broek, 2010). Several challenging language constructions are also common to these texts, including nominalized verbs (e.g., *digest* becomes *digestion*), passive voice (e.g., a liter of hydrochloric acid is added to the solution), and technical and specialized words (e.g., transpiration or metamorphic) (Fang & Schleppegrell, 2010; O'Hallaron, Palincsar & Schleppegrell, 2015).

Social Studies Texts. NAEP will also sample from the varied forms of texts common to the social studies. Selection will represent a wide array of text types, forms of representation, sources of information, and perspectives. These texts document human activity across cultures, societies, and time periods. They include newspaper articles, diaries, letters, speeches, records of sale, advertisements, official government documents, photographs, cartoons, maps, artwork, music, and video and audio recordings. They also include interpretive books and articles about events, time periods, or people, and classroom textbooks. Social studies texts may organize ideas chronologically or thematically to represent time periods, social structures, continuity and change, cause and consequence, and varied social or historical perspectives to consider how the past influences the present (Charap, 2015; Seixas, 2010; Seixas, et al., 2015; Schreiner, 2014). Varied text structures use linguistic frames to mark arguments, persuasion, chronology, cause and effect, perspective, or comparison and contrast. Texts from long ago may even require readers to consider language and the policy contexts within which the texts were generated.

Digital Platform. Like the 2019 NAEP Reading Assessment, the 2026 Assessment will be entirely based in a digital platform. The widespread presence of computers and smart devices in modern society has changed ideas about what counts as text. Students in school are frequently required to read literary, science, and social studies texts that reflect the digital environment, an environment that is different from the world of print on paper. Online newspapers and magazines are replete with graphs that allow readers to simulate different scenarios and see possible outcomes when a causal factor is altered. Digital science texts now in use in schools include simulations that dynamically illustrate what happens to one human body system when variables in the other body systems change.

Digital texts may be static, with no movement of the text on-screen (Barron, 2015) and require readers to make sense of ideas using print and images (e.g., photographs, diagrams, tables) very much like those in a print-on-paper world. Dynamic texts require readers to follow movement across modes (e.g., between print and video or static image) or across nonlinear locations (e.g., clicking a hypertext link that moves you to another section) to construct meaning (Beach & Castek, 2016; Giroux & Moje, 2017; Kinzer & Leander, 2003; Kress, 2013; Manderino, 2012). Reading within and across multiple texts that contain both static and dynamic textual elements makes reading more complex, especially when texts contain conflicting ideas and varying stylistic features that further contribute to complexity. Readers must work actively within and across these text arrangements to construct meaning and create a situation model for a particular reading purpose.

Like the 2019 NAEP Reading Assessment, many state assessments have recently migrated to online digital testing platforms. Widespread use of digital texts was acknowledged by the Common Core State Standards (CCSS) in English Language Arts (NGA-CCSSO, 2010) and by multiple state consortia assessments (including Smarter Balanced and PARCC). Like reading in many of today's classrooms, these assessments include print texts paired with audio clips, podcasts, infographics, and video segments. Even states that moved away from the CCSS and consortium assessments have retained standards and assessments that acknowledge widespread use of digital texts in homes, schools, and communities. Digital platforms offer a range of affordances, including increased attention to principles of Universal Design of Assessment to increase ecological validity and precision in measuring reading comprehension (Coiro, 2020; Fitzgerald, Higgs, & Palincsar, 2020).

Text Complexity. NAEP has long taken a multifaceted approach to assessing the complexity and accessibility of texts to determine which features of text to emphasize in selecting texts. The 2026 NAEP Reading Framework continues this approach, evaluating quantitative and qualitative features of texts, along with reader-text considerations.

Quantitative text complexity measures consider long-standing indicators of complexity, such as the type and number of features that make a text more difficult to read, including such features as familiarity of vocabulary, sentence length and complexity (e.g., Stenner, 1996; Kincaid et al, 1975), and more recent developments, such as the degree of cohesion of ideas across parts of the text, and even the degree to which a given story, for example, exemplifies the classic characteristics of a story (e.g., Graesser, et al., 2014; Sheehan, et al., 2014)

Qualitative tools include careful examination of additional discourse features and conceptual load. Examples might include evaluating the transparency of the relationships between paragraphs or sections (problem-solution, cause-effect), or assessing the quality of a definition and examples provided in a text to help students understand an unfamiliar concept. In reader-text considerations (NGA-CCSSO, 2010), NAEP considers the representativeness of texts for various subgroups by addressing the questions "For whom, in what specific contexts, and with what levels of support are specific texts harder or easier to comprehend?" (Pearson & Hiebert, 2014). With added use of interconnected digital texts, the 2026 NAEP Reading Assessment will also capture navigational complexity (such as the number of links traversed to answer a question) to evaluate the number and nature of moves readers must make within and across digital texts (Coiro, 2020).

Text and the NAEP Definition of Reading Comprehension. Texts are used in the NAEP assessment in ways that tie to all other aspects of the NAEP Definition of Reading Comprehension. The assessment's texts reflect disciplinary contexts, as well as the multiple genres and modalities, used in both school and non-school settings, as well as the many kinds of digital and multimodal texts that make up the textual diets of most students. Broad sampling increases the likelihood that all readers will encounter texts that connect to their experiences and identities, as well as to those texts that are more distant.

Universal Design Elements

The purpose of the 2026 NAEP Reading Assessment is to measure students' reading comprehension across a diverse range of test-takers. To help accomplish this purpose, the 2026 NAEP Reading Assessment employs principles of Universal Design of Assessments (UDA). Universal Design of Assessments calls for the purposeful design of assessments that are

accessible to the greatest number of students possible in order to accurately measure the same construct–in this case, reading comprehension–across the diversity of test takers (Thompson, Johnstone, & Thurlow, 2002; Thompson, Thurlow, & Malouf, 2004). To do this, assessments draw on design features, available to all test takers, called Universal Design Elements (UDEs).

UDEs are design elements of the assessment environment intended to help all test-takers access, organize, analyze, and express ideas when engaging in complex tasks, such as reading comprehension. As such, UDEs aid students' ability to engage with the content that is being tested by reducing the noise (what measurement scholars call *construct-irrelevant variance*) introduced when students lack familiarity with other aspects of assessment. For example, students might not know what the term *synopsis* means when it appears in a test item but could construct one if they knew it was like a summary. Or they might not initially be able to answer questions about the details of an obscure article but would be able to if they knew that the topic was motorcycle design. Or they might not be able to answer a vocabulary question on page 3 of a passage not because they did not know the word, but because scroll bars are a challenge for them.

Importantly, UDEs are designed to improve measurement for students across the performance spectrum rather than for only some students (Johnstone, Altman, & Thurlow, 2006). UDEs minimize but do not eliminate needs for some students' special accommodations, much like access ramps to increase building access may not enable all individuals to enter without added support. Designers validate UDEs before widespread use to ensure that purposes are reliably accomplished, enhancing precise measurement (Johnstone, 2003; Johnstone, Altman, & Thurlow, 2006).

Use of UDEs means that difficult tasks are difficult because they offer rigorous assessment of the construct being measured and not because they introduce unnecessary complexity or other construct-irrelevant sources of variance. For instance, digital test features were employed in the 2019 NAEP, including a look-back button to link test items to points in passages where relevant information was provided to avoid unnecessary searching, scrolling, and page turning; specific directions for approaching the reading of a text; a resetting feature that provided a correct response to a previously answered item so readers could continue without carrying misconceptions from one item to the next; and task partners (e.g., avatar classmates or teachers) to complete tasks in simulation of many classroom assignments. Informed by the use of these features in the 2019 assessment, the 2026 NAEP Reading Assessment uses three expanded categories of UDEs: task-based, motivational, and knowledge-based.

UDEs and the NAEP Definition of Reading Comprehension. Universal Design Elements in the 2026 NAEP Reading Assessment reflect the NAEP Definition of Reading Comprehension in several ways. UDEs enable readers to engage with topics to be read about by providing brief previews and offering instructions on how to complete assessment tasks. They also include lookback buttons and definitions of some words (only those not measured on the assessment), thus reflecting the kinds of navigational aids and tools available in typical reading situations. In addition, UDEs clarify the nature and order of tasks and expected responses. Much more information about UDEs is provided in Chapter Three.

Contextual Variables

In addition to the responses to comprehension items, NAEP also uses questionnaires to gather information about schools and students' interests and experiences. NAEP reports reading

Deleted: Task-based UDEs. Task-based UDEs are designed to clarify requirements and guide readers in their use of available resources.They increase access and sustain readers' attention as they take the assessment. They clarify the expectations for readers and help them examine and use available resources within the assessment blocks (CAST, 2020; Dejong, 2006; Zhang & Quintana, 2012). They maximize the likelihood that readers are able to cognitively engage with complex NAEP-designed reading experiences within the compressed time frame of an assessment. They might include a sequential set of directions to communicate expectations for how and why readers should engage with a collection of texts; they can also help readers plan and monitor their work across multiple texts and tasks (de Jong, 2006). They might also include graphic organizers that allow readers to record and revisit their ideas, reduce time spent on searching and scrolling, and, thus, provide more time for students to read, evaluate, and engage with text content. These UDEs might also include simulated student work examples that offer models of approaches to tasks before students complete similar tasks independently (e.g., Sparks & Deane, 2014). ¶
Motivational UDEs. Motivational UDEs are intentionally embedded into reading activities to encourage and support readers' interest, engagement, and persistence, especially when they encounter challenging tasks. These UDEs are informed by the substantial body of research that describes the beneficial influence of motivation on reading comprehension (Alton & Proctor, 2008; Buehl, 2017; CAST, 2020; Guthrie & Klauda, 2015). They may also maintain readers' interest by communicating explicit connections between the broader purpose for completing a task and the sub-tasks that need to be completed along the way. UDEs in the form of task characters provide written and/or oral directions or serve as experts or peers to provide information or moral support. Task characters may also serve as a simulated target audience with whom readers can communicate new understandings about what they have read and learned (e.g., Use and Apply). Motivational UDEs may also include the kind of resetting feature, described earlier, which has been part of NAEP since 2019. ¶
Knowledge-based UDEs. Knowledge-based UDEs are designed to provide relevant information about topics, concepts, or vocabulary that students may need to make meaning from text as they read. Contemporary models of reading comprehension (Kintsch, 1998; McNamara, 2021; van den Broek & Helder, 2017) describe the significant, positive impact of readers' existing, text-relevant knowledge (especially topic knowledge) on their text comprehension. Wide variations in students' knowledge result in reading comprehension performance scores that reflect differences in background knowledge about specific topics, in addition to differences in comprehension skill. A reader who happens to have knowledge related to the text presented in the ... [35]

Deleted: The provision of knowledge-based UDEs reflects the fact that the 2026 NAEP Reading Framework is directly addressing the decades-old concern about many reading comprehension assessments: that they assume that sampling a wide variety of texts can sufficiently account for inevitable variation in readers' text-related background knowledge. Including these UDEs helps the NAEP assessment to better reflect the conditions of everyday reading situations.

achievement to reflect these data, collectively called contextual variables. These include race/ethnicity, English language proficiency, socio-economic status, and region of the country.. There are many links between these contextual variables and the NAEP Definition of Reading Comprehension. For example, NAEP has issued special reports that summarize performance according to students' experiences (e.g., How often do they read for pleasure, go to the library, and/or read or write on a digital device?).

NAEP collects data via questionnaires that are completed by students and school personnel. The questionnaire items offer many opportunities to gather information about students and their reading. Besides their demographic characteristics and language experiences, questionnaire items can also provide information about students' reading activities in school and community settings, and the encouragement and instructional support they receive from peers, teachers, or community agency leaders. Such information provides information on the backgrounds and supports that students bring to their reading comprehension.

By providing more nuanced reports that display variability within groups, and by measuring disparities in resources and opportunities to learn, the 2026 NAEP Reading Assessment seeks to make variability within groups and variables associated with differences among groups in reading performance more visible. Instead of portraying student groups as unitary and homogeneous, this approach will yield more nuanced reporting of reading disparities.

The digital format, which has been implemented starting in 2017, also allows NAEP to capture students' time on tasks and navigational moves as they complete the assessment. The process data now available because of the data-gathering assets of the digital platform can provide information about student journeys through the texts, directions, UDEs, and items students traverse during the assessment. From these data, NAEP can construct indicators about how students direct their attention (including moment-by-moment shifts in focus) and how long (or how little) they linger on different segments of the texts, the items, the UDEs, or the directions. These indicators can be used to help interpret performance differences in a richer context (Guthrie & Humenick, 2004; Guthrie & Klauda, 2015).

Summarizing the Relationship Between the Definition and Assessment Components

This chapter has described the NAEP Definition of Reading Comprehension and the NAEP Reading Assessment, and the relationship between them. Exhibit 2.1 summarizes these relationships, demonstrating how current understanding of reading comprehension, as embodied in the Definition of Reading Comprehension that opens this chapter, is represented in NAEP through the components of the assessment.

Chapter 3 takes the next step by describing the structure of the assessment and illustrating the use of key design principles and practices that will allow NAEP test developers to create an assessment that includes the components described here.

Deleted: habits and attitudes

Deleted: How much do students like school?

Deleted: to gain insight into contextual variables

Deleted: perceptions of the texts they read, their

Deleted: insights

Deleted: into the knowledge, interest, motivation, engagement, habits, attitudes, language competence, skills, and strategies

Deleted: Reporting results solely by students' demographic characteristics might contribute to a perception that all students within each demographic group are the same. For example, reporting results by students' race/ethnicity might lead the public to infer that the achievement differences between racial groups are attributable only to students themselves rather than to the opportunities to learn which have been presented to them. These ideas are described more fully in Chapter 4.

Deleted: explanatory

Deleted: a more nuanced and complete measure to better understand

Deleted: as the result of a complex of factors

Deleted: (For more information about how contextual variables are reported, see Chapter 4.)

Deleted: Contextual Variables and the NAEP Definition of Reading Comprehension. There are many links between the NAEP Definition of Reading Comprehension and the contextual variables. In general, the questionnaire items allow NAEP to better understand the relationship between performance and different student variables: (a) demographic data (race/ethnicity, socioeconomic status, or community type), (b) perceptions about themselves as readers, or (c) their experiences in school and community contexts. The process data allow NAEP to connect performance to cognitive activities such as attention. Using this information to contextualize results allows for more accurate interpretations of student performance.¶

Exhibit 2.1. Relationships Between the NAEP Definition of Reading Comprehension Definition and the NAEP Reading Assessment

	Features of the NAEP Definition of Reading Comprehension			
Assessment Components	Contexts	Readers	Texts	Activities
Comprehension Items	Reflect a view of the outcomes of reading as influenced by factors within and outside of the assessment.	Address an array of skills and strategies related to comprehension, including literal, inferential, analytical, and critical responses along with items that ask students to apply ideas in the texts.	Query different types of comprehension within and across texts and different aspects of the texts, including local and global features and meanings.	Attend to disciplinary contexts, purposes, and text challenges to determine how items will reflect the four comprehension targets.
Contexts and Purposes	Invoke rich contexts (discipline-related and otherwise) as a way of situating reading in settings that involve reading comprehension.	Communicate purposes for reading, introduce social elements, such as a digital "guide", and enhance engagement by focusing on contemporary issues.	Include varied texts that align with disciplinary contexts and purposes.	Establish authentic contexts, structures, and purposes for reading and formulate tasks that are aligned with those purposes.

Deleted: or peers

	Features of the NAEP Definition of Reading Comprehension			
Assessment Components	Contexts	Readers	Texts	Activities
Texts	Include a variety of texts that represent a range of cultural traditions, disciplinary contexts, and reading purposes.	Select texts that are broadly representative of varied cultural traditions, backgrounds, experiences, and identities.	Include texts from a wide range of genres, modalities, formats, and disciplinary traditions.	Include varied texts that align with the disciplinary contexts, broad purposes, and genres appropriate for the block.
Universal Design Elements	Reflect the kinds of resources that are commonly available during reading in school, workplace, and community contexts.	Provide previews of the topics, information about unknown words that are not the focus of the assessment items, and instructions on how to complete assessment tasks, allowing readers to engage in more challenging reading tasks.	Increase broad access to texts, such as providing definitions of key words not measured on the assessment and offering lookback buttons.	Provide information that clarifies the nature and order of tasks and expected responses.
Contextual Variables *Questionnaire Items*	Gather information about the contexts of readers' lives and experiences in and out of school.	Gather information about demographics, motivation, and in- and out-of-school reading practices.	Gather information about the amount and kinds of texts that readers encounter in and out of school settings.	Gather information about reading activities that readers commonly engage in at school and outside of school.

Deleted: *y*

	Features of the NAEP Definition of Reading Comprehension			
Assessment Components	Contexts	Readers	Texts	Activities
Process variables	Compare pathways when reading in different disciplinary contexts and for different purposes.	Track each participant's navigation through the assessment—reading texts and responding to items.	Compare pathways through the assessment when employing different sorts of texts.	Compare pathways for different sorts of items, both format and Comprehension Targets.

This chapter describes the assessment design components that contribute to best educational measurement practices, as outlined by the National Research Council (2001; AERA/APA/NCME, 2014) and used in previous NAEP Reading assessments (National Assessment Governing Board, 2019). These practices include incrementally augmenting current assessment design with features that are carefully tested and refined over time: a hallmark of NAEP development practices since the inception of the assessment.

The chapter is divided into three sections. The first section provides an overview of considerations related to developing block components of the 2026 NAEP Reading Assessment. This involves situating readers within a disciplinary context, a broad purpose, and a specific purpose and role for each block. The second section discusses the task components and how they can be used to expand the ways in which readers are asked to demonstrate their ability to engage in the comprehension processes outlined in Chapter 2. Task components include texts and comprehension items. The third section details considerations for leveraging digital assessment features, including item response formats, Universal Design Elements (UDEs), and process data. Overall, the design considerations outlined in this chapter are intended to enable the 2026 NAEP Reading Assessment to allow the greatest number of students to participate in ways that result in more valid inferences about their comprehension performance as situated in purposeful, disciplinary contexts.

Deleted: , in line with principles of validity, fairness, and inclusivity (Thompson, Johnstone, & Thurlow, 2002).

Designating Disciplinary Context

A block is the largest organizational unit for the 2026 NAEP Reading Assessment. In a typical NAEP reading session, test-takers engage in two grade appropriate blocks. The design of every block involves situating readers within a disciplinary context, a broad purpose for reading, and a specific purpose and role for the reader working through the block. The distribution of disciplinary contexts by grade level varies according to the approximate amount of time that students in the U.S. are engaged in the respective contexts at grade levels 4, 8 and 12. Exhibit 3.1 shows the design principle and provisional distribution targets for sampling disciplinary contexts at each grade level.

Deleted: **Situating Readers Within Assessment Blocks** ¶
A block is the largest organizational unit for the 2026 NAEP Reading Assessment. In a typical NAEP reading session, test-takers engage in two grade appropriate blocks. The design of every block involves situating readers within a *disciplinary context*, a *broad purpose for reading*, and a *specific purpose* and *role* for the reader working through the block. See Exhibit 2 in Appendix C, which illustrates a range of design features that should be considered when designing assessment components. These features vary along a continuum within a block, from less to more dynamic and cumulative.¶

Deleted: All blocks will sample from a range of grade-appropriate texts within one of three disciplinary contexts, including literature, science, or social studies contexts. The primary context for each block will be identified according to one of these contexts so that NAEP can report reading performance scales for each of these disciplinary contexts, along with an aggregate scale for performance across all three contexts. In some cases, a block may contain texts associated with more than one disciplinary context. In these cases, the block is designed as both a primary reading context that shapes the overall reading purpose and a secondary context identified by one or more interdisciplinary or cross-disciplinary topics or genres.

Exhibit 3.1. Principle and Provisional Distribution Targets for Sampling Disciplinary Contexts by Grade Level

Principle for Sampling Disciplinary Contexts: The percentage of Literature decreases across grades as the percentage of Science and Social Studies increases				
Grade Level		4	8	12
Disciplinary Context	Literature	50	40	33
	Science	25	30	33
	Social Studies	25	30	33

Designating a Broad Reading Purpose

In addition to situating readers in one of the three disciplinary contexts, each assessment block is also designated as having one of two broad purposes: Reading to Develop Understanding or Reading to Solve a Problem. Situating reading in purpose-driven tasks has demonstrated potential for promoting student readers' interest and engagement in existing NAEP reading assessments (Educational Testing Service, 2019).

Reading to Develop Understanding (RDU) blocks are designed to measure what readers do when asked to deeply read and comprehend—literally, inferentially, interpretively, and critically—in or across disciplinary contexts. Reading to Solve a Problem (RSP) blocks are designed primarily to assess what readers do when asked to demonstrate understanding across multiple texts and related perspectives while solving a problem. Reading to Solve a Problem activities entail developing understanding, or comprehending text, but in the service of using this understanding to take a specific action or create a product, such as a written explanation or a classroom presentation.

In both types of blocks, these broad purposes are intended to help readers prepare for reading in order to develop understanding or to solve a problem. The design principle and provisional distribution targets for sampling broad purposes by grade level are depicted in Exhibit 3.2.

Formatted: Indent: First line: 0"

Deleted: ¶
¶
¶
¶
¶

Exhibit 3.2. Principle and Provisional Distribution Targets for Sampling Broad Reading Purposes by Grade Level

Principle for Sampling Broad Purposes. The percentage of Reading to Develop Understanding (RDU) blocks decreases across grades as the percentage of Reading to Solve a Problem (RSP) blocks increases				
Grade Level		4	8	12
Broad Reading Purpose	RDU	60	50	40
	RSP	40	50	60

Identifying Specific Purposes and a Reader Role

Both RDU and RSP blocks also have specific purposes with reader roles that shape how and why readers engage with the tasks, texts, and comprehension items in one of the three disciplinary contexts. These specific purposes differ from the broad block purposes (i.e., RDU or RSP) because the duration of their guidance is limited to the text or texts within a given task in the assessment block. Test developers for the 2026 NAEP Reading Assessment will craft these purpose-driven statements with an eye toward reflecting the real-world contexts and purposes for which readers engage with and make sense of a diverse range of texts.

Reader roles are designed to reflect how readers typically engage with texts and each other in different contexts (e.g., fourth grade classmates and a teacher in a literature circle discussion at school or a group of friends at home reacting to news about a local event in their town). Some blocks may ask readers to take on a simpler, less immersive role that offers fewer

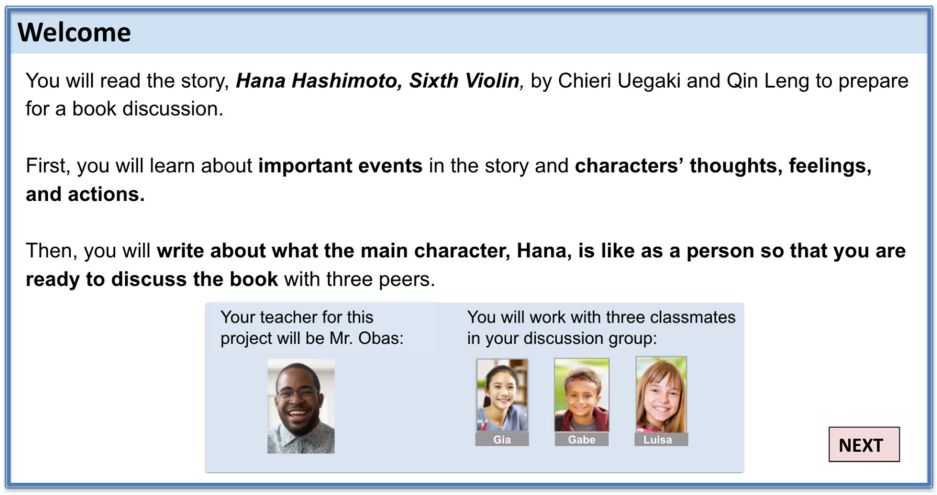
specifications for the kinds of tasks with which readers will engage. Other blocks may assign readers to take on more immersive roles that offer more specifications for how readers should engage with the reading purpose, tasks, and expected outcomes.

Specific purposes and reader roles are explicitly shared with test-takers as part of the directions at one or more locations in the block. Exhibit 3.3 depicts an example of what readers might see when they begin the Grade 4 Reading to Develop Understanding sample block in a literature context. In this block, readers are invited to participate in a book discussion group about the short story *Hana Hashimoto, Sixth Violin* by Chieri Uegaki and Qin Leng with three other fourth grade student task characters (simulated avatar classmates). In addition to reading directions about the discussion goal, students are told they will read the story and respond to items situated in two purpose-driven tasks.

Deleted: (see Appendix C)

The goal of the 2026 NAEP Reading Framework is to immerse readers in discipline-specific blocks for which both reading purpose and reader role are transparent to better simulate the situations in which most readers find themselves in school, workplace, and community situations.

Exhibit 3.3. Task-specific purposes presented at the beginning of a Grade 4 Reading to Develop Understanding block using the text *Hana Hashimoto, Sixth Violin* (a short story) by Chieri Uegaki and Qin Leng



Developing Assessment Tasks: Texts and Items

After readers are situated in the assessment block, they encounter two or more tasks, each with its own specific purpose. A task is a subunit within each block on the 2026 NAEP Reading Assessment. Each NAEP reading block has 2-3 tasks, one or more texts, and related comprehension items. Developers take into consideration time, total passage length, and grade appropriateness when determining the number of texts in each assessment block. Extended pieces of literature or a full argumentative essay might result in only one text with one or two

tasks. Shorter texts such as a haiku poem, photograph, search engine result, or Twitter post might result in more than one text for a particular task.

For example, Exhibit 3.4 from an ePIRLS Grade 4 assessment block illustrates how several texts are embedded into one screen to authentically represent the array of texts young readers encounter when reading on the internet; these texts include a webpage with two tabs and a navigational menu, an embedded hyperlink (which is the source of the answer as displayed in the blue pop-up box when the link is selected), a photo of a rocket, a photo of Mars' surface, a dynamic image of two planets spinning around the sun, and an advertisement with a hyperlink button that leads readers away from the relevant information. The item is intended to assess fourth graders' understanding of how to use embedded hyperlinks to locate and recall important information about the passage.

Deleted: I

Exhibit 3.4. Example of multiple texts readers encounter as part of one task on the ePIRLS (2016) Grade 4 reading assessment

All grade-appropriate blocks will sample from a variety of task-specific purposes and a range of texts, including reading materials that students might use in their everyday lives, in and out of school (see, for example, Creer, 2018; Dobler & Azwel, 2007). The texts can represent one or more genres, modalities, or disciplines. See Exhibit 1 in Appendix A for additional considerations for sampling text formats and modes. See Exhibit 2 in Appendix A for examples of different kinds of text formats and modes.

Selecting Texts

Text Selection Criteria. Passages in the 2026 NAEP Reading Assessment are selected using rigorous criteria that include:

- ***Authenticity***. Do texts represent the types of texts that students encounter in their reading in and out of school?
- ***Diversity***. Do texts reflect an appropriate range of perspectives, geographical regions, gender, and social and cultural traditions characteristic of the diverse U.S. population, and are they written by diverse authors?
- ***Engagement***. Will texts encourage and maintain student interest?
- ***Developmental appropriateness***. Do the texts reflect grade level expectations of the students assessed at grades 4, 8, and 12?
- ***Disciplinary appropriateness***. Do the texts represent the range of genres/text types and text features in the disciplinary contexts of Literature, Science, or Social Studies?
- ***Quality and cohesion***. Are the texts well-written and organized in ways that promote comprehension and learning? Do non-fiction texts, and especially those in a modality other than print, include brief and purposeful topic introductions where appropriate?
- ***Complexity***. Are the language features (vocabulary, syntax, discourse and rhetorical structures) representative of the specific grade and disciplinary context?

Several of these text selection criteria are elaborated below with a number of principles and design considerations.

Authenticity. Most texts included in NAEP Reading will be presented in their entirety, as students would typically encounter them. However, some texts may be excerpted from a novel or a long essay. Excerpted material will be carefully analyzed, and minimally altered if necessary, to ensure that it is coherent in structure. Texts will be selected to evoke the range of reading comprehension processes, or targets. Only in exceptional cases, NCES and its contractors may consider commissioning authors to write a text that satisfies the needs of a particular assessment block. For example, it might become highly challenging to find a text of a particular length that is suitable for a specific grade level for a RSP purpose. In the exceptional cases in which commissioned writing may be required, it should follow the text selection criteria applied to authentic texts. In very rare cases, then, commissioned texts may be used as part of a set of texts. Thus, such commissioned texts will not serve as the main, or anchor, text for a text set, nor will students be asked items focused on evaluating the credibility or accuracy of such texts. See Exhibit 3 of Appendix A for more detail.

Developmental Appropriateness of Texts. Texts included in the assessment will be of different lengths. In grade 4, passage lengths will range from 200-800 words, in grade 8 from 400-1000 words and in grade 12 from 500-1500 words. Differing passage lengths are employed for several reasons, including the total time readers have to complete the block. To gain valid information about students' reading comprehension, stimulus material should be as similar as possible to what students use in their in-school and out-of-school reading. Unlike many common reading tests that use short passages, the 2026 NAEP Reading Assessment will include complete texts of greater length. Such texts require students to use a broader and more complex array of reading strategies, reflecting student reading in authentic in- and out-of-school situations (Goldman, 2018; Paris, Wasik, and Turner 1991).

Deleted: See Exhibit 4 in Appendix A.

Reflecting classroom practice, students in earlier grades generally read shorter texts while older students read longer texts. It is expected that in some cases, two or more texts (with static

and/or dynamic textual features) will be used together to assess students' ability to compare, synthesize, and critique texts in terms of their content, themes, and stylistic features. In these cases, the total number of words will reflect the recommended passage length range for each grade.

Disciplinary Appropriateness of Texts. Selected texts must be representative of the discipline in both content and structure, reflecting the range of genres and discourse features detailed in Chapter 2. Because reporting prompted by the 2026 NAEP Reading Framework will feature scales for the three disciplinary contexts, it is also important to specify both the variability of student reading within contexts and the commonalities across each context. Based on the account provided in Chapter 2 of the range of text types, text structures, and text features, Exhibit 5 in Appendix A shows important text elements that characterize texts in each of the disciplinary contexts, while acknowledging that many text features are common across disciplines. A responsibility of test developers, as they build the portfolio of test blocks and tasks at each grade level, is to try to incorporate the entire array of text types and features in the blocks for each grade level. See *Assessment and Item Specifications for the 2026 NAEP Reading Framework* for more details.

Standards for Cohesion and Complexity of Texts. Efforts should also be made to promote the strategic balance and selection of texts across blocks. This process should be informed by general standards of quality, cohesion, complexity and "considerateness" (including both qualitative and quantitative measures; e.g., conventional readability criteria, reader-text connections, language structures and vocabulary considerations; Anderson & Armbruster, 1984) and reflect contemporary standards applied to digital texts and other contemporary media forms. Because readers use specific knowledge to identify important information in different types of texts, developers attend to variations in organization and cohesion in line with text structures and text features that are found in common across disciplinary contexts. Test developers should strive to select texts with features that cue readers' attention to structure and influence the recall of information (Wixson & Peters, 1987).

The extent to which readers' background knowledge, experiences, and interests connect to a text and its topic will also be considered when evaluating a text's complexity, suggesting that a text is not just complex "in the abstract" but more or less complex for particular groups of readers under specific circumstances (Valencia, et al., 2014). Textual ideas in disciplinary contexts should be represented with appropriate vocabulary and, where needed, texts should have useful supplemental explanatory features such as definitions of technical terms or orthographic features (italics, bold print, headings) and connective signal words (e.g., first, next, because, however). Unfamiliar concepts should be defined with examples provided. Designers should aim for a flexible and diverse representation of language and structures across the blocks.

There is also wide variance in the nature and quality of graphical or multimodal displays of ideas in today's texts. Therefore, in selecting texts, it is important to create a sample that represents the grade-appropriate array of graphical and structural representations (e.g., static, dynamic, multimodal, nonlinear) found in print and digital reading materials. As well, texts often appear, and are used in sets. Thus, it is important to determine grade-appropriate numbers of texts, and the opportunities for readers to engage with ideas within different sections of the same text as well as to process ideas across two or more texts.

Developing Comprehension Items

Deleted: Because videos may be used in NAEP assessments built from the 2026 NAEP Reading Framework, some attention should be given to video length. The length of a video segment will vary in relation to its purpose and to overall block time. Video length may also increase across grade levels. However, because students have greater engagement and perceived retention rates for shorter as compared to longer videos (Slemmons et al., 2018), video length should generally be kept relatively short, especially compared to the length of other written texts within the task. ¶

Deleted: (see Exhibit 6 in Appendix A)

Deleted: A potential difference between traditional and digital texts is the nature of text arrangement and the means with which readers navigate through and across texts (Cho, 2014). In selecting digital texts, it is important to attend to the features that allow for navigating multilayered digital text environments (Afflerbach & Cho, 2017; e.g., search engines, dynamic hypertexts linked within and across documents) to reflect what readers do when they use the Internet. Further, digital texts represent diverse combinations of the information contained in text and the media used to present that information. For example, a digital text may include short (e.g., 30 second), embedded video and links to other sources of information. Thus, it is important to determine that the ideas, perspectives and modes presented in digital media reflect what readers encounter in their academic and everyday lives. ¶

Engaging experts in selecting texts that reflect authentic social and cultural traditions in a range of disciplinary contexts. The text selection process is best conducted by experts with disciplinary, educational, and cultural knowledge about the nature and structure of texts that are representative of particular disciplinary contexts and cultural traditions in specific grade levels. Such experts should represent diverse cultures and languages in order to identify texts that reflect the broad range of student readers' knowledge and experiences.¶

Design Principles. As with the selection of texts, item development is guided by a set of design principles in order to guarantee that readers are asked to respond to important aspects of the text and to use a range of processes that result in successful comprehension. These design principles include:

- ***Importance.*** Items should focus on central textual and intertextual concepts or themes or, on occasion, more specific information related to these themes and concepts. For example, a fact that provides evidence to support a claim or a detail that supports a main idea may be queried.
- ***Balance***. The comprehension targets, as described in Chapter 2, should be proportionally distributed across dimensions of the block (see Exhibit 7 in Appendix A).
  - across grade levels.
  - across the disciplinary contexts of literature, science, and social studies.
  - across broad purposes of blocks.

  While the percentage of comprehension targets may vary across these dimensions, items representing all comprehension targets should be represented at all levels of these dimensions.
- ***Clarity and transparency.*** Items should be accessible and transparent. They should be written in accessible, straightforward language, and accompanied by directions that clearly explain what steps readers should take during the activities (e.g., which texts to read and for what purpose) and how their responses will be evaluated.
- ***Alignment with an array of skills of navigation and inference.*** Across items and in accordance with the focus of the comprehension targets, items should call upon readers to locate information in different multilayered digital text environments (e.g., static and dynamic) and to make different kinds of inferences, from local bridging inferences to more complex inferences across texts and applications of knowledge to a new situation (e.g., Use and Apply). As such, audio and visual features may have items associated with them.
- ***Varied knowledge sources***. Items should invoke a variety of knowledge sources in accordance with the comprehension targets in a given assessment block. Across items, readers should be called upon to employ certain kinds of background knowledge (e.g., knowledge of vocabulary and language structures, knowledge of text structures and features) and to draw information from different sources in the texts (including information at various types of representation [e.g. directly stated in prose, embedded in a visual representation, or implied through symbolism] and across different locations in the text). On the other hand, items should not assess knowledge sources irrelevant to the items and associated comprehension targets in a given block. For example, items should not be answerable by readers only drawing upon text-independent domain and topic knowledge. Knowledge-based UDEs are incorporated into given blocks to maximize students' ability to engage with the content that is being tested. Thus, knowledge-based UDEs are designed to provide orientations to the topical knowledge addressed in the text(s).

Deleted: knowledge, topic knowledge, knowledge of technical vocabulary or idiomatic expressions, or conceptual or domain knowledge in particular subject areas

Deleted: reduce the noise associated with knowledge sources not being assessed in a given block and also

Planning the Distribution and Characteristics of Comprehension Items. The four comprehension targets do not represent a hierarchy of strategies or skills. The difficulty of any particular item, regardless of which comprehension target it is designed to elicit, should be shaped by the content of text(s) (the ideas themselves), the language and structure of the text (the language and relations among ideas), and the cognitive demands of the comprehension target. As a consequence, there can be relatively difficult items representing Locate and Recall comprehension targets and relatively easy items representing either Integrate and Interpret or Analyze and Evaluate targets. The single most important standard that the 2026 NAEP Reading Assessment will meet is asking questions about matters of substance in the texts. Chapter 2 contains examples of what test items might ask readers to do with respect to each of the four comprehension targets.

Considering Navigational Complexity of Texts, Tasks, and Items. Developers should also consider the *navigational complexity of text* as it interacts with the reading task and the specific demands of the comprehension items attached to the text(s) within tasks (see Coiro, 2020). Comprehension items may, for example, vary in difficulty according to the nature of associated comprehension processes (e.g., locating a topically relevant idea is likely easier than inferring the tone of a particular passage or analyzing the impact of an author's word choice on a particular audience). Further, comprehension items may vary in difficulty due to the nature of inferences readers are asked (or required) to make; that is, the *type* of inference (a local, straightforward inference within a paragraph vs. a global inference across ideas in a text) combined with the *number* (one or multiple) and the *distance* of these inferences (within one text, across two texts, or beyond the text) introduce variations in task and item demands that impact the difficulty of a particular comprehension item on the reading assessment. Thus, test developers will follow guidelines from the *Assessment and Item Specifications for the 2026 NAEP Reading Framework* to estimate levels of navigational complexity across an activity block as shaped by the number, levels, and types of inferences as well as the nature of texts, tasks, items, and response types included. In turn, estimated difficulty levels can be used to inform the development of future NAEP reading tasks as NAEP learns more about how reader attributes interact with various task demands to influence comprehension performance.

Language Structures and Vocabulary in the Comprehension Items. Language structures and vocabulary in the 2026 NAEP Reading Framework refers to the application of the reader's understanding of individual words, grammatical structures, and discourse structures characteristic of grade-appropriate texts to text comprehension. Specifically, the 2026 NAEP Reading Assessment will include items designed to evaluate readers' application of their knowledge of useful grade-appropriate words and language structures to their understanding of a text or a set of texts. Because these items target readers' application of the meaning of highly useful language found across grade-appropriate texts to text comprehension, testing items will exclude rare words of limited application across grade-appropriate texts, and idiomatic expressions characteristic of particular cultural and idiosyncratic discourse practices.

A maximum of 15-20 percent of items in any assessment block will assess readers' application of passage-relevant Language Structures and Vocabulary to text comprehension, while concurrently measuring a specific comprehension process. Due to the intricate relation between language understanding and text comprehension, language structures and vocabulary will not be measured independently from comprehension targets. Instead, they will be doubly

Deleted: Exhibit 7 in Appendix A presents guidelines for distributing items mapped to comprehension targets across grade level and blocks. These flexible distributions allow for the possibility of varying the number of items for each target depending on block type. One broad principle is that the percentage of items designed to assess Integrate and Interpret or Analyze and Evaluate ideas increases across grades. In addition, in Reading to Solve a Problem (RSP) blocks, the percentage of items designed to assess Locate and Recall ideas decreases across grades as the percentage of Use and Apply ideas increases. Finally, the distribution targets should never outweigh the other principles in the bulleted list. In other words, for a given text, it is better to fall one item short in the number of items for a target than it is to include one that fails the importance or the clarity standard just for the sake of meeting the distribution goal. ¶

Deleted: (see Exhibit 8 in Appendix A)

Deleted: language known to be part of students' everyday oral proficiency,

Deleted: , discipline-specific concepts

coded for Comprehension Target (e.g., Locate and Recall; or Integrate & Interpret) and Language Structures and Vocabulary.

A note on open-ended responses. Whereas measuring students' understanding of passage-relevant grade-appropriate language is crucial, it is also important not to confuse language dexterity with the demonstration of text understanding in open-ended responses. Thus, consistent with the 2009-2019 NAEP Reading Assessments, the 2026 NAEP Reading Assessment will generate scoring rubrics and training for scorers that are language-conscious so that students are not erroneously penalized for language features irrelevant to the comprehension processes being assessed (for example, a student's written answer that displays accurate comprehension should not be negatively affected by uses of unconventional grammar or misspelled words).

Digital Assessment Features: The Role of Item Response Options, UDEs, and Process Data

An essential goal of the 2026 NAEP Reading Framework is establishing valid assessment tasks that can reliably measure diverse students' real-world reading comprehension. In the 2026 NAEP Reading Assessment, this goal is accomplished in two ways. First, all test components are designed to support ecological validity, which refers to the extent to which assessment elicits students' reading performance as it would be demonstrated in real-world settings. Newer, digital tools in particular allow assessments to situate cognitive acts of reading, to the extent possible, in complex but authentic home, school, and work reading contexts and to do so in ways that are ecologically valid (Mislevy, 2016). Second, by employing newer, digital tools, the 2026 NAEP Reading Assessment supports construct validity by providing more contexualized presentations of test results, thereby increasing awareness of the diversity of test takers (c.f., Mislevy, 2016; Thompson et al., 2002).

Deleted: accurate

Deleted: interpretations

Deleted: the validity of

Deleted: scores across

To undertake these aims, the 2026 NAEP Reading Assessment is grounded in Universal Design of Assessments (UDA). As described in Chapter 2, UDA calls for the purposeful design of assessments that are accessible to the greatest number of students possible in order to accurately measure the same construct across the diversity of test takers (Thompson, Johnstone, & Thurlow, 2002; Thompson, Thurlow, & Malouf, 2004). See Exhibit 3.5 for an overview of UDA principles. The NAEP 2026 Reading Assessment employs UDA (Johnstone et al., 2006; Thompson et al., 2002) to select from a broad range of digital assessment features in order to design an assessment from which stakeholders can make more informed interpretations of assessment scores for all test-takers. Such digital assessment features include the purposeful selection of item response formats, universal design elements, and process data, as described in each of the next three sections. See Exhibit 3.6 for an overview of how these digital features, as well as other aspects of the 2026 NAEP Reading Assessment, align with principles of UDA.

Deleted: valid

Exhibit 3.5. Seven Principles of Universal Design of Assessments (UDA)

Principle Number and Name*	Description of Principle
1. Inclusive Assessment Population	This principle supports equitable participation in, and use of, assessments. Assessments should measure the performance of a wide range of students reflective of the population the assessment aims to represent. The assessment should do so in a way that ensures that students with diverse characteristics have opportunities to "demonstrate competence on the same content" (Johnstone et al., 2002, p. 6). This does not mean that the

	test will be less rigorous or that content should be altered. Rather, this is achieved through accessibility of content using diverse formats (e.g., item formats), technological tools (e.g., Universal Design Elements, or UDEs), and designs that include diverse test-takers.
2. Precisely Defined Constructs	Precisely defined constructs help to ensure that an assessment measures the construct it intends to measure rather than aspects not part of that construct, which creates construct-irrelevant variance. Without a precisely defined construct, it is hard to know whether items and other design features work towards measuring the intended construct or whether they might, in fact, be measuring something else.
3. Accessible, Non-biased Items	The purpose of this principle is to ensure that all test takers can access the content being assessed so that items measure the same construct for all students who take the assessment (i.e., items are "non-biased"). For example, if a passage contains a highly culturally-situated term that might be more familiar to some sub-populations of test takers (e.g., to boys more than to girls), this might result in inaccurate measurement across these subpopulations. Bias is measured statistically by comparing the difficulty of items across subpopulations of students.
4. Amenable to Accommodations	This principle refers to the physical design of the test (e.g., font, colors, graphics) being easily accessible for students' sensory abilities or easily modified (e.g., avoiding vertical text allows for the easier modification of written text into Braille).
5. Simple, Clear, and Intuitive Instructions and Procedures	In accordance with this principle, instructions and procedures of an assessment should be easily understandable regardless of a student's background (e.g., experience, knowledge, language use, concentration level). Instructions that use clear, simple language that is consistent across the assessment serve to maximize the ability of the assessment to measure the intended construct.
6. Maximum Readability and Comprehensibility	This principle refers to the ability of a text to be understood by all test takers so that readability does not interfere with the measurement of other content (e.g., on a math test, a student's ability to read an item stem does not make it harder for them to complete the task). Because readability is systematically varied and assessed in the NAEP reading test, it cannot be maximized as it might be for a math test.
7. Maximum Legibility	This principle refers to test elements (e.g., text, tables, figures, illustrations, and response formats) being easily understood. Developers should consider elements such as contrast, type size, spacing, and typeface when developing a test that is as understandable as possible.

*These UDA principles are drawn from Thompson et al., 2002, where they are referred to as "elements."

Deleted: unfairly advantage these students,

Deleted: ing

Deleted:

Exhibit 3.6 Alignment of the 2026 NAEP Reading Assessment with Principles of Universal Design of Assessments (UDA)

UDA Principle*	Alignment of Aspects of the 2026 NAEP Reading Assessment with UDA Principles
1. Inclusive Assessment Population	***Inclusive Population Assessed in NAEP Reading:*** NAEP Reading aims to measure *reading comprehension* in a way that represents *all* students within the U.S. population at grades 4, 8, and 12 by not excluding any groups from sampling. ***UDEs*** UDEs minimize bias while supporting construct validity by activating students' knowledge, interest, and understanding of tasks across the diverse range of test-takers, helping to ensure that all students can access and understand the items. This supports the ability of the assessment to measure the same construct for all students, aligning with UDA Principles 1, 2 and 3. • Task-based UDEs facilitate students' ability to focus cognitive resources on the assessment tasks and items by providing clear instructions about what to do during the task (but not how to do it). • Motivational UDEs activate interest in the topics of texts and tasks, eliciting motivational processes that typically occur in out-of-test reading situations and thus improving validity of assessment items. • Knowledge-based UDEs preview untested topic knowledge and provide definitions for vocabulary not intended to be assessed. This maximizes the extent to which the assessment can measure the same, intended construct for all, diverse test-takers by minimizing the possibility that one group is advantaged over another and facilitating better measurement for all test-takers.
2. Precisely Defined Constructs	***Definition of Reading Comprehension:*** Chapter 2 of the framework defines the construct of *reading comprehension* and explains how this construct is operationalized using the comprehension targets as situated within the disciplinary contexts and broad purposes. This clearly defined construct helps to ensure that the assessment is measuring what it intends to measure (i.e., construct validity) by outlining exactly what is included and not included, helping to ensure that items can capture this construct and not elements outside of this construct. ***Reader Roles Support Ecological and Construct Validity:*** Reader roles are designed to situate the reader within a disciplinary context and broad purpose, as readers would be during out-of-test reading

	activities. While assessments can never perfectly measure the constructs they intend to measure as those constructs exist in reality, assessments aim to do so to the extent possible (i.e., what is referred to as ecological validity). In so doing, this also supports construct validity, in alignment with the "precisely defined constructs" called for in UDA Principle 2. Situating the reader within a disciplinary context and broad purpose also allows the reader to access the content being measured because it activates the reader's prior understandings relevant to those disciplinary contexts and purposes, allowing for more precise measurement of the construct. ***Specific Purposes:*** Situating readers within specific purposes (e.g., a reader is asked to read a story and participate in a book discussion) activates readers' prior understanding of what it means to read within a given task purpose and in so doing facilitates their ability to engage in the items and tasks. Specific purposes also help make clear to the reader what they are supposed to do with the texts and why. This aligns with "precisely defined constructs" because the specified purposes enable the assessment to do a better job of measuring the student's ability to engage with the construct and not, for example, their ability to figure out what they are supposed to do. ***Item Formats:*** Thoughtful selection of item formats to measure particular comprehension targets within the context of the texts and specific purposes supports students' access to the test construct because they are able to focus limited cognitive resources on tasks aimed to measure the construct. This supports the assessment's ability to measure the construct it intends to measure (Principle 2) by facilitating *all* students' ability to access the construct (Principle 3).
3. Accessible, Non-biased Items	***Regular NAEP Reading Research and Development Process:*** Item bias is tested through NAEP's regular item review and pilot testing procedures to ensure that items are not more or less difficult for students from particular subpopulations. To test item bias, the difficulty of items across different subpopulations of students (e.g., boys and girls) is compared to ensure that items measure the same construct across groups. Biased items are revised until they no longer demonstrate bias. ***Disciplinary Contexts & Purposes:*** Because all students being tested are familiar with the school-based disciplinary contexts of literature, science, and social studies, and with the Reading to Develop Understanding and Reading to Solve a Problem purposes as they are situated within these contexts, sampling texts and tasks from these disciplines and using these purposes helps to minimize bias, since all students can be presumed to be familiar with the kinds of texts used within these three disciplines. ***Range of Texts and Tasks Represented:*** Selection of a diverse range of texts and tasks representing different student identities, interests, knowledge, and other backgrounds helps to

	ensure equity across diverse subpopulations of test-takers. Such broad sampling facilitates equitable test items and scales.
4. Amenable to Accommodations	***UDEs and Item Formats:*** UDEs and thoughtful use of item formats limit the need for special accommodations. For example, task-based UDEs and item formats such as "drag and drop" can limit the need for accommodations such as extended time because they facilitate students' thoughtful use of time and focus on the texts and tasks being measured rather than on unrelated organizational skills.
5. Simple, Clear, and Intuitive Instructions and Procedures	***Instructions:*** Instructions, in simple language, facilitate measurement of the intended construct (in this case, reading comprehension) because they allow readers to focus limited cognitive attention on the items rather than on the instructions. ***Clear Comprehension Items and Tasks:*** Similarly, items written using simple, clear language support the student's ability to engage in the items that are measuring reading comprehension ability aligned to the comprehension targets. Both of these aspects help to ensure that the items are measuring the intended construct (e.g., the student's ability to make meaning from literature) rather than aspects unrelated to the construct (e.g., the student's ability to understand written instructions or to understand the item stem).
6. Maximum Readability and Comprehensibility	***Selection of Grade-Appropriate Texts:*** Texts are selected based on readability and text cohesion elements relevant to the grade levels in which they are tested. This helps to ensure that students taking the test can engage with the texts at these particular levels.
7. Maximum Legibility	***Visual Layout:*** The 2026 NAEP Reading Assessment layout considers elements such as contrast, font type and size, and spacing within the digital environment to facilitate the validity of items because it supports' students' ability to focus limited cognitive resources on the items rather than on visual features. For example, layout should be easily accessible for different students' sensory abilities. Careful consideration of these elements also allows the assessment to be amenable to accommodations (Principle 4) because the layout is easily modified when accommodations do need to be made (e.g., translating the assessment into Braille).

* These UDA principles are drawn from Thompson et al., 2002, where they are referred to as "elements." UDEs are "Universal Design Elements."

Deleted: that is easily understandable regardless of a student's background (e.g., experience, knowledge, language use, interest) …

Deleted: be presumed to be able to read and understand

Item Response Formats

Central to the development of 2026 NAEP Reading Assessment is the careful selection of the ways in which students respond to items. From 1992 through 2016, items on the NAEP Reading Assessment were limited to two formats: multiple choice and constructed response

(write the response with a pen or pencil). In 2017, the term multiple-choice was revised to "selected response" to account for the wider range of item formats available (e.g., "matching") with digitally based assessments. Selected-response items for use on the 2026 NAEP Reading Assessment include a variety of formats. The 2026 NAEP Reading Assessment thus employs Selected Response and Constructed Response options. Additionally, NAEP will be exploring additional kinds of Dynamic Response options. Some examples of item response formats are presented in the next sections.

Deleted: See Appendix D for additional examples.

Selected Response Options. These kinds of responses allow the student to select one or more choices from provided options and include the following types:

- **Single-selection multiple choice** – Students respond by selecting a single choice from a set of given choices.
- **Multiple-selection multiple choice** – Students respond by selecting two or more choices that meet the condition stated in the stem of the item.
- **Matching** – Students respond by inserting (i.e., dragging and dropping) one or more source elements (e.g., a graphic) into target fields (e.g., a table); see Exhibit 3.7.
- **Zones** – Students respond by selecting one or more regions on a graphic stimulus.
- **Grid** – Students evaluate ideas with respect to certain properties. The answer is entered by selecting cells in a table in which rows typically correspond to the statements and columns to the properties checked; see Exhibit 3.8.
- **In-line choice** – Students respond by selecting one option from one or more drop-down menus that may appear in various sections of an item.
- **Select in passage** – Students select one or more ideas in the passage; in some cases, they also drag them into the target fields.

Exhibit 3.7. Example of Matching Response Format from PARCC Grade 8 Literature

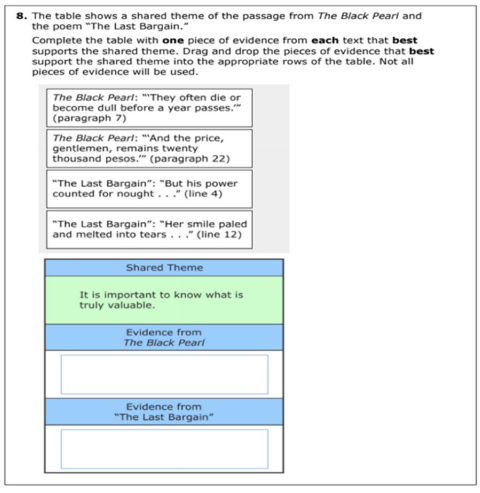
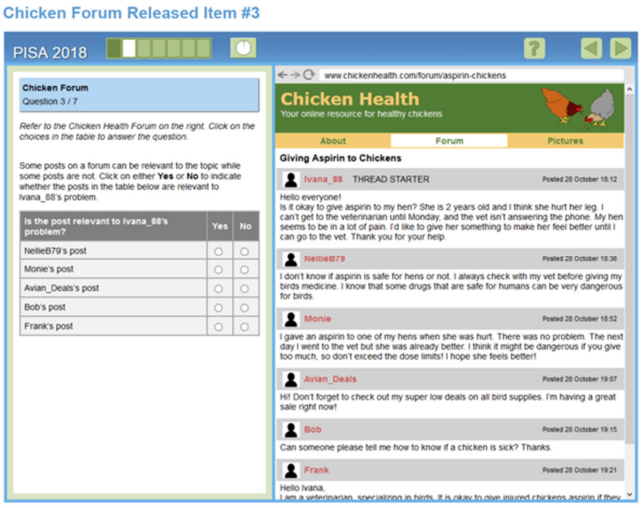


Exhibit 3.8 Example of Grid Response Format from PISA



Constructed Response Options. These kinds of responses allow the student to develop their own response within a given parameter (e.g., a certain number of characters) and include:

- **Short constructed response** – Students respond by entering a short text in a response box that consists of a phrase or a sentence or two.
- **Extended constructed response** – Students respond by entering an extended text in a response box that consists of multiple lines (a paragraph or two).
- **Hybrid constructed response** – Students respond by selecting one or more choices that meet the condition stated in the stem of the item. Then they write a short explanation about their choices.
- **Fill in the blank** – Students respond by entering a short word or phrase in a response box.

  Flexible distributions of item response type across grade level are presented in Exhibit 3.9.

Exhibit 3.9. Flexible Distributions of Item Response Types Across Grade Level

	Selected Response Items	Short Constructed Response Items	Extended Constructed Response Items
Grade 4	40-50%	40-45%	10-15%

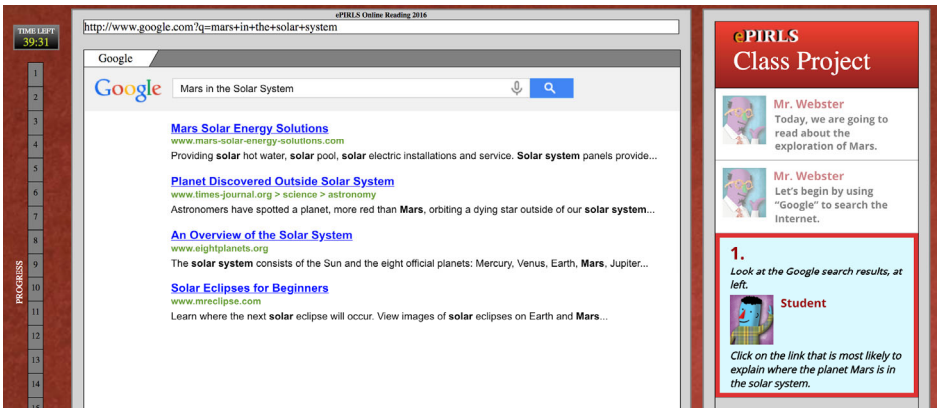
Grade 8	40-50%	40-45%	10-15%
Grade 12	40-50%	40-45%	10-15%

Dynamic Response Options. NAEP is currently exploring the use of dynamic response options to assess comprehension (e.g., graphic organizers and drop-down menus). NAEP should continue this trend in the years ahead by further exploring the use of other interactive or dynamic response formats made possible with emerging digital tools. Many existing state assessments, as well as PARCC and Smarter Balanced, use these kinds of item response formats. Useful frameworks (Scalise & Gifford, 2006) and guidelines (Measured Progress/ETS Collaborative, 2012) introduce a wide variety of innovative item types that should be considered by NAEP in implementing digitally-based facets of the 2026 NAEP Reading Assessment, when it is indicated that such item types bring value to the assessment. For example, dynamic item formats introduce opportunities to assess how readers:

- Search and locate information (e.g., dynamic search engines); (see Exhibit 3.10).
- Select and identify information (e.g., multiple choice items with new media distractors);
- Reorder or rearrange information (e.g., ranking, categorizing, and sequencing items);
- Substitute or correct information (e.g., multiple drop-down menus offering word choices embedded within lines; limited graphical elements that are adjusted or corrected to accurately represent ideas in the passage);
- Categorize or classify information (e.g., tiling, select, and order);
- Construct relationships among information (e.g., dynamic concept maps, multimodal representations); or
- Construct spoken responses (e.g., recorded spoken language in open-ended responses).

When selecting the format of any particular item, developers should be mindful of the cognitive and logistical demands of varied formats and how these may interact with reader familiarity and the time constraints of each activity.

Exhibit 3.10 Example of a Dynamic Search Engine Item from ePIRLS 2016 for Grade 4 Students



Universal Design Elements (UDEs)

Grounded in Universal Design of Assessments (Johnstone et al., 2006; Thompson et al., 2002), the NAEP 2026 Reading Assessment employs design features known as Universal Design Elements (UDEs). UDEs provide orientation, guidance, and motivation to sustain readers' journeys through the block. They are designed to mirror typical (non-testing) reading situations to improve the validity of the assessment. UDEs also offer a way for NAEP to develop fair and inclusive assessment tasks.

Deleted: The *fairness* of an assessment refers to a judgment about the appropriateness of decisions based on test scores (AERA, APA, & NCME, 2014). Research has shown that a student's background, language, and experience is important in how they interpret assessments (Solano-Flores & Nelson-Barber, 2001). Because these influences shape student thinking, they must be taken into account when trying to reduce bias in assessment items and support validity (Lee, 2020; Siegel, Markey, and Swann, 2005).

All readers have access to UDEs. UDEs, or the "built-in features of computer-based assessments," have been increasingly included in NAEP since the introduction of the digital platform in 2017, and are available for *all* students (NCES, 2017). Importantly, UDEs are not the same as legally mandated accommodations. While the use of UDEs might minimize the need for special accommodations, UDEs are not designed to fully address accessibility needs for the full population of students who take the 2026 NAEP Reading Assessment. Other assessment features, called *accommodations*, are legally mandated for *some* but not all students with additional testing needs (see NAEP Accommodations, last updated Oct. 2019). Examples of accommodations available on some assessments include extended time, options for responses in Braille or Sign Language, or having test-items read aloud. Universal Design of Assessments and the inclusion of UDEs are the means to enable *all* readers to validly demonstrate what they know and are able to do.

Types of UDEs. Examples of UDEs already exist in operational NAEP Reading (e.g., highlighters and look-back buttons) to reflect real-world experiences and how readers use technology. Amidst the use of these digital supports by all test-takers, NAEP has effectively maintained the ability to capture trends over time (NCES, 2017). The 2026 NAEP Reading Framework includes three broad categories: task-based UDEs, motivational UDEs, and knowledge-based UDEs. The three categories of UDEs are designed to accomplish three different, yet sometimes overlapping, functions as described next. The next section clarifies the role of each UDE and offers some hypothetical examples of how these might appear in the 2026

Deleted: Increasingly complex reading purposes and more dynamic texts in today's society demand a broad collection of UDEs to enable test-takers to fully engage with the assessment (Mislevy, 2016). Consequently, the

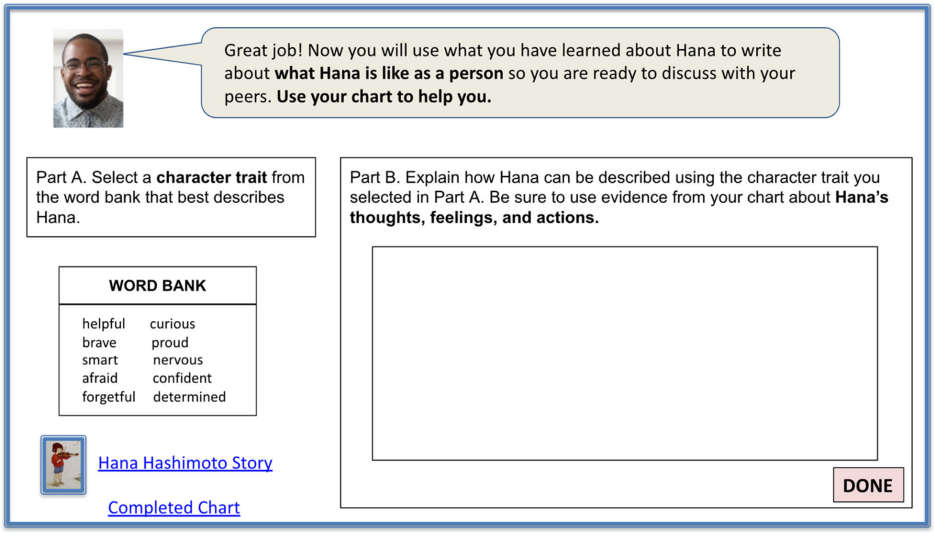
NAEP Reading Assessment. Additional details are provided in the item specifications. Some examples of UDEs are presented in the next sections.

Deleted: See Appendix E for additional examples of UDEs.

Task-based UDEs. In the 2026 NAEP Reading Assessment, task-based UDEs are used to clarify requirements and guide readers in their use of available resources in the testing space. These UDEs are designed to increase access to test content and to sustain readers' attention. A task-based UDE at the beginning of an activity (e.g., a sequential set of directions) might clearly communicate expectations for how and why readers should engage with a collection of texts. Such UDEs might also help readers plan and monitor their work across multiple texts and tasks (de Jong, 2006) by providing guidance on how to move among the texts. As readers move through the block, task-based UDEs might include graphic organizers that allow readers to record and revisit their ideas; these types of UDEs aim to reduce time spent on low-level activities (scrolling to find the location) while providing students more time for higher order activity—reading, evaluating, and engaging with text content (Sparks & Deane, 2014).

Exhibit 3.11 illustrates an example of an Integrate and Interpret item with a task-based UDE that is aligned with UDA principles calling for "assessment instructions and procedures…to be easy to understand, regardless of a student's experience, knowledge, language skills, or current concentration level" (Thompson et al., 2002, p. 13). The item is designed to measure the student's ability to describe, in depth, a character, drawing on specific details in the text. To demonstrate this skill, the student needs to identify a character trait that is relevant, but selecting an accurate trait is insufficient to meet the construct measured. The student needs to be able to connect the selected character trait with a deeper interpretation of the character and the details of the text. In providing the word bank as a task-based UDE, all students have an equivalent opportunity to focus more of their time and attention on the use and apply construct to be measured, rather than on trying to generate a character trait word. This type of task-based UDE is an example of one that aims to assess more challenging comprehension processes while allowing readers to access the item in the relatively short period of time allotted by the assessment. This clarity of expectations also maximizes the likelihood that readers will cognitively engage with complex NAEP-designed reading experiences within the short time frame allotted to each block.

Exhibit 3.11. A Grade 4 Use and Apply item illustrating a task-based UDE in the form of a word bank providing a set of character traits from which readers can select their choice and then use as part of their constructed response



Motivational UDEs. In the 2026 NAEP Reading Assessment, motivational UDEs are designed to facilitate students' interest in assessment content and persistence with challenging tasks (Alton & Proctor, 2008; Buehl, 2017; CAST, 2020; Guthrie & Klauda, 2015).

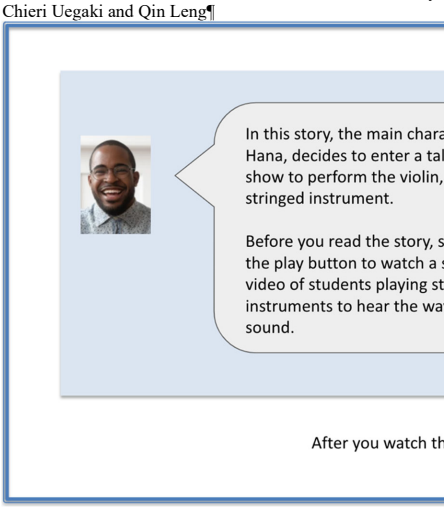
Motivational UDEs may, for example, maintain readers' interest by communicating explicit connections between the broader purpose for completing a block and the sub-tasks that need to be completed along the way. UDEs in the form of task characters may provide written and/or oral directions, or interact directly with readers as experts, teachers, or peers to provide information (see Exhibit 3.13). Task characters may also represent members of an authentic target audience to whom readers can represent and communicate new understandings about what they have read and learned (e.g., Use and Apply). To the extent that assigned purposes (and related texts, tasks and goals) are viewed as meaningful and relevant, readers are more likely to be motivated to engage with or react to the reading activity as a whole (Guthrie & Klauda, 2015; van den Broek, Bon-Gettler, Kendeou, & Carlson, 2011).

Deleted: The use of a word bank as a task-based UDE also aligns with principles calling for "accessible, non-biased items" and the removal of "non-construct oriented...barriers" to the assessment content (Thompson et al., p. 9). In this case, the word bank decreases construct-irrelevance by providing a set of words from which test-takers can *select*, rather than *generate*, a relevant character trait. The provided words allow all readers, and especially English learners, to access the test and validly engage with the item designed to measure their ability to make inferences about character traits and not their ability to generate unfamiliar words in a timed assessment context. Similarly, this task-based UDE aims to reduce testing bias so that all students, regardless of their native language, have an opportunity to make sense of the story and demonstrate how to make inferences about characters and support their answers with evidence from the text. ¶

Formatted: Indent: First line: 0"

Deleted: Motivational UDEs might, for example, provide an engaging pre-reading preview or video that helps to generate a minimal amount of interest in an assessment block. See Exhibit 3.12, where a pre-reading preview and accompanying 15 second video of children playing string instruments serves to pique students' interest in the topic of the reading passage. The passage is about a girl who enters a talent show contest to perform the violin she has just learned how to play. Such UDEs can increase the test's ability to measure the intended construct for all students, regardless of their prior interest and motivation. ¶

Exhibit 3.12. A Motivational UDE in the form of a 15 second video clip of students playing stringed instruments for the Grade 4 short story Hana Hashimoto, Sixth Violin by Chieri Uegaki and Qin Leng¶

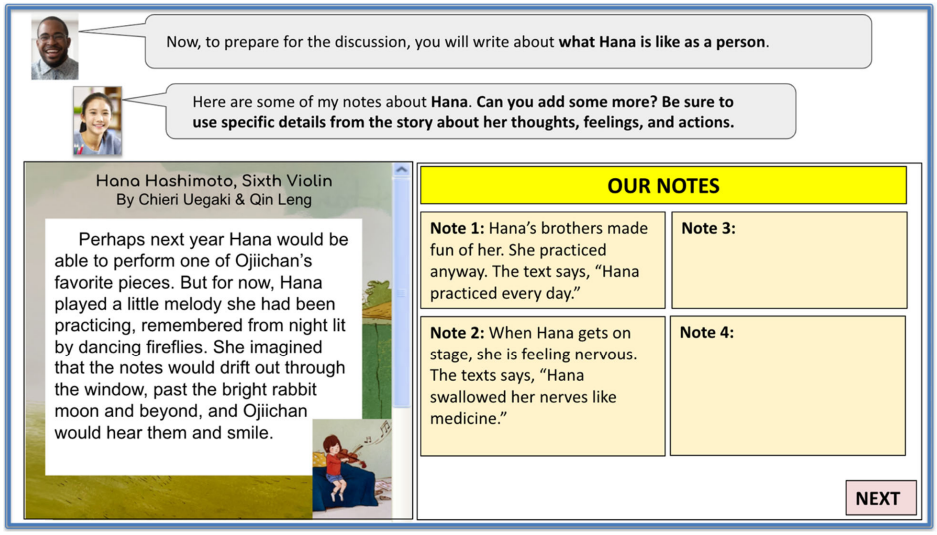


¶

As with task-based UDEs, these kinds of motivational UDEs align with UDA principles calling for "accessible, nc ... [36]

Deleted: also

Exhibit 3.13. Teacher and student task characters remind the reader of the task goal for the second task.



Knowledge-based UDEs. In the 2026 NAEP Reading Assessment, knowledge-based UDEs will provide two types of information: (a) topic previews in the form of short introductions to either the entire block or to a specific task and text, and (b) definitions or examples for unfamiliar vocabulary unless a word is explicitly tested in a comprehension test item. Topic previews may take the form of written texts only, unless video, image, or other kinds of introductions are already part of an authentic source text. Topic previews should be offered as appropriate any time that access to information that is not part of the items being assessed could differentially advantage or disadvantage readers in ways that are outside the relevance of the reading construct being measure. A determination must be made by assessment developers about whether a UDE is construct relevant. Other digital media (e.g., dynamic animations, glossary hyperlinks to related images—with or without language translations—and simulations of interesting or challenging phenomena) can provide visual and multimedia cues to support readers' understanding of words and phrases likely to pose construct irrelevant barriers to comprehension. Please see Exhibit 3.14 for the kinds of knowledge that will and will not be assessed. Finally, as noted in chapter 2, blocks without UDEs, including those without knowledge-based UDEs, are part of the current assessment and will continue to exist in the 2026 NAEP Reading Assessment.

Exhibit 3.14 Reading Knowledge to Be Assessed in the 2026 NAEP Reading Assessment

Knowledge Inherent to Reading Comprehension (to Be Assessed)	**Knowledge Not Intentionally Assessed**
Knowledge of: • Text structures (descriptive, causal, compare and contrast, problem-solution, etc.) • Vocabulary and language structures • Genres and rhetorical structures • Authors' craft	• Text-independent domain knowledge • Topic knowledge • Knowledge of technical vocabulary or idiomatic expressions Conceptual or domain knowledge in particular subject areas
What is Measured on the Assessment Through Comprehension Targets	
Students' Ability to: • Recall specific text information • Use text features to derive meaning • Draw inferences based on information in text • Integrate information within and across texts • Analyze information presented in text • Analyze authors' rhetorical strategies and purposes • Evaluate sources of information in text • Use and apply information from texts	

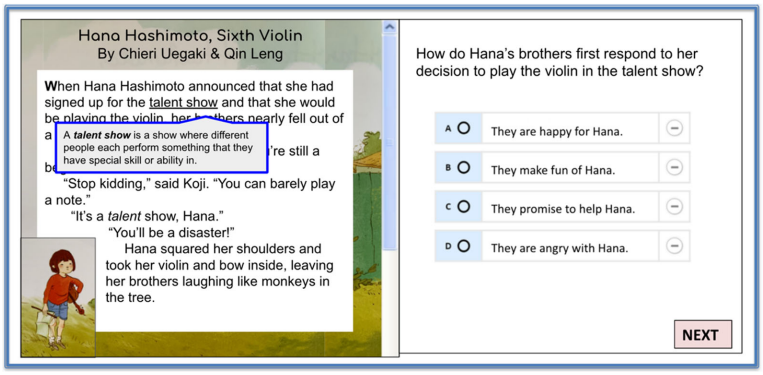
Importantly, knowledge-based UDEs never provide answers to comprehension test items. Instead, they preview untested topic information, activate readers' knowledge, and pique interest in ways that permit readers to engage in the types of literal, interpretive, evaluative, and application processes (i.e., the four comprehension targets described in Chapter 2) required to demonstrate their comprehension of challenging text (Alexander & Jetton, 2000; Buehl, 2017).

Exhibit 3.15 offers one example of a multiple choice Integrate and Interpret item with a Knowledge-Based UDE that aligns with UDA principles calling for "accessible, non-biased items" (Thompson et al., 2002, p. 9). The knowledge-based UDE (a pop-up box defining "talent show") is used appropriately to provide students with background information that does not overlap with the content being assessed. In this case, the multiple-choice item is not intended to measure students' understanding of the phrase "talent show." Rather, the item is intended to measure students' ability to make an inference about how Hana's brothers first respond to her decision to play the violin in the talent show, based on their actions and words (Hana's brothers "nearly fell out of a tree" and they tell her, "you'll be a disaster!"). Since the whole story is

Formatted: Right: 0"

situated in the context of a talent show, the lack of topic knowledge about what a "talent show" is might unfairly disadvantage readers who are not familiar with this term.

Exhibit 3.15. A knowledge-based vocabulary UDE in the form of a pop-up box defining the term "talent show." The pop-up appears when a test-taker clicks on the highlighted term.



Selecting appropriate locations for UDEs. Developers decide on appropriate locations in which to insert UDEs into each block of the assessment. Because some NAEP Reading 2026 tasks involve complexities in response to handling multiple tasks and texts, readers may be asked to check and reflect on their reading progress in the activity and allocate their attention accordingly. Intuitively designed transitions between each task, such as task characters, visual flow charts, or simple written statements may be used to guide readers through the task sequence and structure in any given block.

A major question for block developers is how to decide when to employ and when to forego the deployment of a specific UDE as the potential for added support is weighed against the potential for increased cognitive burden on the reader. Developers will also consider how to populate the grade-appropriate assessment space with UDEs while recognizing that readers have time limits within which to accomplish expected outcomes.

Process Data

Because 2026 NAEP Reading Assessment activities are situated in a fully digital environment, process data involving reader actions (e.g., number of mouse clicks, pathways through a task or hypertext, transcribed voice responses, length of time spent engaged with reading material or responding to an item) can be easily collected in digital log files stored in a database. While these data are not reported for individual students, aggregations of these types of data hold potential power to measure levels of engagement in purpose-driven reading activities (e.g., capturing frequency, density, and intensity of engagement or identifying and comparing novice to expert level of practice). Process data from log files can be aggregated and interpreted to characterize how reader attributes or other explanatory variables influence reading comprehension performance at one or more locations in the NAEP assessment space. Examples of process data developers use to account for reader variations include:

- Timing data (e.g., time on passages and items),

Deleted: . Biases such as this in tests can result in imprecise, inaccurate, and unfair assessments of students' ability to engage in the construct being measured. The NAEP Reading Assessment does not assess what students know about different topics and disciplines; that is the job of disciplinary assessments such as social studies or science. Instead, the NAEP Reading Assessment measures how well students can reason about the information provided in texts as that reasoning is reflected in the comprehension targets used to create comprehension items. Therefore, knowledge-based UDEs like this one orient readers to the topic of the text, without impact on constructs being measured, and reduce testing bias so that all students have an equitable opportunity to make sense of the story and demonstrate how to make inferences about characters.¶
Because the meaning or use of the phrase "talent show" is not directly assessed in this block, this Knowledge-based UDE also aligns with UDA principles calling for "precisely defined constructs" and the removal of "non-construct oriented...barriers" to the assessment content (Thompson et al., p. 9). In this case, the pop-up box defining a talent show is designed to decrease construct-irrelevant variance. That is, the definition allows all readers (and especially those with little knowledge about the kind of show a "talent" show is) to access the text and validly engage with an item designed to measure the reader's ability to make an inference about character actions and words rather than the reader's understanding of what a talent show is.

- Navigation data (e.g., navigating among passages, pages within passages, hyperlinks, using the next button to move through a block); see Exhibit 3.16,
- Data on using other affordances (e.g., the "Look Back Button," glossing), and
- Item response process data (e.g., which answers readers choose, order of selections, answer changes, response mode, use of eliminating options in multiple choice items).

Exhibit 3.16 Example of a Constructed Response Item from ePIRLS 2016 for Grade 4 that Collects Navigational Process Data. The Space Camp image and blast off button serve as a type of distractor item designed to capture process data about readers who click on irrelevant details (i.e., advertisements) on a webpage rather than attending to the comprehension item at hand.

Overall, the strategic use of UDEs and determination of process data collected in each block enables the 2026 NAEP Reading Assessment to fully engage test-takers with complex comprehension tasks while also generating information to better account for the reading performance of fourth, eighth, and twelfth grade students. As knowledge about the use of UDEs becomes more robust and precise, more of these features should be operationalized in the NAEP Reading Assessment in the years ahead.

Conclusion

The opportunities presented by the use of these innovative design features come with a caveat. Pilot offerings of all design features, including the examples above, should be carefully studied, as was noted in the introduction to this chapter. Various reader populations should be sampled carefully in these studies. One reason for this is to ensure that design features yield their intended outcomes for as many students as possible. A second reason is to ensure that new design features do not unintentionally disadvantage some populations of students. In addition to describing how scores will be reported, Chapter 4 illustrates how these new design features allow the 2026 NAEP Reading Assessment to report the reading achievement of the nation's children in new ways that enhance the interpretive use of NAEP results.

Formatted: Space After: 6 pt

Deleted: capacity

Deleted: ¶

Page Break

The purpose of Chapter 4 is to describe how the results of the NAEP Reading Assessment will be communicated to the nation from the year 2026 onward. The chapter addresses the central communication responsibility of NAEP—to report scores in a manner that informs the public about current results and performance trends over time on NAEP Reading Assessment in what has become known as the Nation's Report Card. In addition to describing how scores will be reported, Chapter 4 outlines how the 2026 NAEP Reading Assessment will collect information that can help contextualize and explain the results it reports and serve as a useful resource for informing educational policy related to teaching reading and learning to read.

Reporting Results

Historically, NAEP Reading has reported data for the nation as a whole, for participating states, and for large urban school districts that volunteer to participate in the NAEP Trial Urban District Assessment (TUDA). Results of the NAEP Reading Assessment administrations are reported in terms of average scores for groups of students on the NAEP 0–500 scale and as percentages of students who attain each of the three achievement levels (*NAEP Basic*, *NAEP Proficient*, and *NAEP Advanced*) discussed below. By design, the assessment reports results of overall achievement; it is not a tool for diagnosing the needs of individuals or groups of students. Reported scores are at the aggregate level; by law, scores are not produced for individual schools or students.

In addition to reporting aggregate results for the nation, states, and TUDA school districts, the Nation's Report Card allows for examination of results by school characteristics (urban, suburban, rural; public and nonpublic) and socio-demographic student characteristics (race/ethnicity, gender, English learner status, socioeconomic level (to the extent valid data are available), and disability status, i.e., supported by an Individualized Education Program). The NAEP Data Explorer is a publicly accessible tool that allows users to customize reports and to investigate specific aspects of student reading achievement, such as performance on different comprehension targets or by selected contextual variables. Also, reports of the results of survey questionnaires are produced each year on various topics (e.g., students' Internet access and digital technology at home, instructional emphasis on reading activities, confidence in reading knowledge and skills, teachers' satisfaction and views of school resources).

Legislative Provisions for NAEP Reporting

Under the provisions of the Every Student Succeeds Act (ESSA) legislation, states receiving Title I grants must include assurance in their state plans that they will participate in the reading and mathematics state NAEP at grades 4 and 8. Local districts that receive Title I funds must agree to participate in biennial NAEP reading and mathematics administrations at grades 4 and 8 if they are selected to do so. Their results are included in state and national reporting. Participation in NAEP does not substitute for the mandated state-level assessments in reading and mathematics at grades 3 to 8.

In 2002, NAEP initiated TUDA in five large urban school districts that are members of the Council of the Great City Schools (the Atlanta City, City of Chicago, Houston Independent, Los Angeles Unified, and New York City Public Schools Districts). Ten large districts participated in 2003 and 2005. The number of districts participating in TUDA has grown over

time to a total of 27 beginning in 2017. With student performance results by district, participating TUDA districts can use results for evaluating their achievement trends and for comparative purposes.

Through ESSA and the NAEP TUDA program, the NAEP Reading results report student achievement for the nation, states, and select large urban districts, enabling comparisons between states, large urban districts, and various student demographic groups.

Achievement Levels

Since 1990, the National Assessment Governing Board has used student achievement levels for reporting results on NAEP assessments. Generic policy definitions for achievement at the *NAEP Basic, NAEP Proficient,* and *NAEP Advanced* levels describe in general terms what students at each grade level should know and be able to do on the assessment. Reading achievement levels specific to the NAEP Reading Framework were developed to elaborate on the generic definitions. New reading-specific achievement level descriptors replaced those aligned to the previous framework (NAGB 2009). Exhibit 4.1 presents the generic achievement level descriptors. See Appendix A for the final achievement level descriptions.

Exhibit 4.1. Generic NAEP achievement levels

Achievement Level	Policy Definition
NAEP Advanced	This level signifies superior performance beyond NAEP proficient.
NAEP Proficient	This level represents solid academic performance for each NAEP assessment. Students reaching this level have demonstrated competency over challenging subject matter, including subject-matter knowledge, application of such knowledge to real-world situations, and analytical skills appropriate to the subject matter.
NAEP Basic	This level denotes partial mastery of prerequisite knowledge and skills that are fundamental for performance at the NAEP proficient level.

Reporting Results of the Updated NAEP Reading Assessment

While satisfying legislative requirements and maintaining the scale score and achievement level reporting structures, the 2026 NAEP Reading Framework updates and enhances the assessment and its reporting system to accomplish the following broad goals:

- Emphasize validity throughout the assessment design and the reporting system.
- Revise items included in the reading-specific and the general (i.e., core) part of the questionnaires administered to students, teachers, and administrators whose schools participate in the NAEP Reading Assessment to increase knowledge about opportunities to learn.
- Transform the navigational data (sometimes called process data [Ho, 2017]), referring to how students make their way through the texts and test items) into measures that help explain test performance, as well as student interest and metacognition.
- Increase the capacity of NAEP Reading databases (including enhancements for the NAEP Data Explorer) in ways that encourage educators, policymakers, and researchers to conduct more nuanced analyses of NAEP Reading performance.

Deleted: equity, rigor, precision, and

Deleted: factors that can expand

To provide more nuanced reports and useful data to key stakeholders, the NAEP reporting system will:

1. Disaggregate scores for demographic subgroups in greater detail to provide a more accurate and dynamic description of student performance.
2. Expand the number of categories for reporting the achievement of English learners to better reflect the variability of English language proficiency within this population.
3. Provide information on contextual variables (derived from demographic, questionnaire, and process data) that can contribute to more nuanced interpretations of group results.

Deleted: achieve broader equity goals within an integrated system that

Deleted: s

Deleted: research-based

Deleted: focused on opportunities to learn

Deleted: .

Reporting Categories

The framework reporting system described below provides opportunities to interpret findings from NAEP Reading results by amplifying the demographic and descriptive student categories. To support productive interpretations of results, the reporting of achievement results for the NAEP Reading Assessment will also disaggregate reporting by current and former English learner status.

Deleted: The reporting system expands use of the data derived from the assessment to afford deeper understanding of how socioeconomic status (SES) and race/ethnicity intersect with opportunities to learn in schools and communities (e.g., the availability of libraries or access to challenging curricula). This disaggregation of SES within race/ethnicity allows for examination of diversity within groups.

NAEP Reading Assessment results have provided indispensable information on students' performance with traditional reporting variables parsing results into subgroups to portray how students perform within specific contexts—state, region, access to technology, socioeconomic level, and many more. By expanding reporting categories and adding more contextual variables, NAEP will now be able to point the way to plausible hypotheses for policy makers to consider in crafting reforms. Thus, the 2026 NAEP Reading Framework builds on the strengths of the prior NAEP reporting system by including enhancements to the reporting capacity of NAEP through reporting by disciplinary contexts; disaggregating results within demographic categories; and expanding reporting categories for English learners.

Deleted: and explanatory

Reporting by Disciplinary Contexts

The 2009–2019 framework had two subscales: reading for literary experience and reading for information. The 2026 NAEP Reading Framework uses three subscales to report on reading performance within and across three Disciplinary Contexts: Reading to Engage in Literature, Reading to Engage in Science, and Reading to Engage in Social Studies. In addition to continued reporting of outcomes as a point on a scale from 0-500 and as the percentage of students who score within different achievement level bands (*NAEP Basic*, *NAEP Proficient*, and *NAEP Advanced*), the 2026 NAEP Reading will report additionally on each of the Disciplinary Context scales. This enhancement is informed by increased attention to reading in the content areas in state standards across the nation.

Disaggregating Results Within Demographic Categories

NAEP will continue to report reading scores by selected student subgroups. Student subgroups are defined by the following characteristics: gender; race/ethnicity; family income; disability status; and English language status. In addition, results are reported by school characteristics, such as public/private, urban/rural, and region of the country.

Deleted: , as measured by student eligibility for the National School Lunch Program

Because the 2026 NAEP Reading Framework seeks to capture the dynamic variability within student groups, NAEP disaggregates student group data to show, at a minimum, differences of socioeconomic status within the student subgroup of race/ethnicity. In NAEP

Reading, as in other large-scale assessments, lower levels of achievement historically are correlated with poverty. It is important to note that on international assessments such as PIRLS (Mullis & Martin, 2019) and PISA (OECD, 2019), socioeconomic status (SES) does not predict achievement in reading comprehension as accurately in other countries as it does in the U.S. Enhanced reporting can help policy makers and stakeholders better understand reading performances in context. For example, these data may allow policy makers to consider how access to resources that support rich literacy opportunities may serve as an underlying driver of achievement.

Deleted: Consequently, it seems likely that SES alone does not offer a direct or sufficient explanation for reading performance and that additional contextual variables are crucial to better understand variability in reading

Additional parsing of the results in this way could be important because the results might suggest that what is, on the surface, presumed to be a cohesive and static category may indeed include significant differences in access to resources. Examining SES and race/ethnicity with a more nuanced lens can surface factors that are highly amenable to change, e.g., resource allocation. When the data are disaggregated by states and TUDA districts as described in the 2026 NAEP Reading Framework, they should thus be more helpful to stakeholders for addressing the needs revealed by the assessment.

Expanding Reporting Categories for English Learners

English learners (ELs) are defined by NAEP as students "who are in the process of acquiring English language skills and knowledge" (NAEP Nation's Report Card, 2019). These students have not yet reached state-established standards for grade-level English proficiency and so are at the beginning or intermediate phases of acquiring English. In the prior NAEP reporting system, students were designated either as *not English learners* or *English learners* at the time of the assessment. The results for students who had been classified as ELs but who were no longer classified as such were reported along with students who had never been identified as ELs; hence, there was no way to disaggregate data to observe or track the successes and increases in achievement of former ELs.

The 2026 NAEP Reading Assessment results expand reporting categories in order to present data that is more attuned to the complex composition of today's student populations, and, thus, more informative for states and school communities (Durán, 2006; Hopkins, Thompson, Linquanti, August, & Hakuta, 2013; National Assessment Governing Board, 2014; Kieffer & Thompson, 2018). In keeping with the latest research and current requirements for state-level reporting under ESEA, Section 3121(a), the reporting system for the 2026 NAEP Reading Assessment disaggregates scores by three English proficiency categories for which school systems that participate in NAEP already collect data:

1. *Current English learners* – Students designated as English learners at the time of the assessment;
2. *Former English learners* – Students who have reached grade-level standards of English proficiency within the last two years prior to the assessment and who have formally exited that status;
3. *Non-English learners* – Monolingual students who speak only English; bilingual students who speak English and another language and who were never previously identified as English learners; bilingual students who reached grade-level standards of English proficiency more than two years ago.

Deleted: s

Reporting NAEP results for these three categories will allow more nuanced interpretation of data for students who are designated as current or former ELs and highlight challenges these students may face. Focusing exclusively on the current EL subgroup can obscure the progress that educational systems make in moving students toward English proficiency and higher levels of reading achievement. This expansion of EL reporting categories will shed light on any progress—or lack thereof—that might be detectable in the group of Former ELs. With states increasingly able to collect this information about English learners' histories, and the likelihood that a majority of states will have these data available by 2026, the 2026 NAEP Reading Framework expands reporting categories for English learners in order to more accurately represent the descriptive data states and districts are already using to understand the performance of these students.

Contextual Variables

Students participating in the NAEP assessments respond to survey questionnaires that gather information on variables important to understanding reading achievement nationwide. Teachers and school administrators also complete questionnaires. Questions are intended to be non-intrusive; free from bias; secular, neutral, and non-ideological; and do not elicit personal values or beliefs. To the extent possible and to minimize the burden on those asked to complete the questionnaires, demographic information regarding school and student characteristics is also gathered from non-NAEP sources such as state, district, or school records.

As stated in Governing Board policy, the collection of contextual data on students, teachers, and schools is necessary to fulfill the statutory requirement that NAEP include information whenever feasible that is disaggregated by race or ethnicity, socioeconomic status, gender, disability, and English learner status. Contextual information serves the additional purpose of enriching the reporting of NAEP results by examining factors related to academic achievement in the specific subjects assessed. To satisfy the goal of enriching reports on student achievement in reading, contextual variables are selected to be of topical interest, timely, and directly related to academic achievement. In addition to questionnaires, information on contextual variables is also obtained by analyzing process data derived from computer monitoring of students' navigation within the assessment tasks completed.

The 2026 NAEP Reading Assessment uses an expanded set of research-based contextual variables (Guthrie & Klauda, 2015; Guthrie, Wigfield & Von Secker, 2000) to understand reading achievement. Contextual variables are measurable, and some are also malleable (that is, they can be influenced). These include *reader characteristics* and *environmental characteristics* (students' perceptions about facets of home, community, or school settings, including their perceptions about classrooms and support).

Deleted: (e.g., students' self-perceptions about engagement and motivation, knowledge, self-efficacy, agency, effort, and interest)

Deleted: , sense of belonging,

The current NAEP Reading Framework collects and reports data on contextual variables, factors that shape students' opportunities to learn, including time, content, instructional strategies, and instructional resources. Contextual variables are used to predict or account for variance in the outcome of interest, reading comprehension scores on NAEP. The 2026 NAEP Reading Framework's emphasis on the power of context to shape learning and development leads naturally to the need to identify and expand research-based contextual variables for reading. By measuring students' differential engagement with reading and their access to home and community resources such as libraries, tutoring, and out-of-school programs, the expanded

Deleted: the cultural assets of individuals and

Deleted: taking into account

contextual variable data will support efforts by researchers to interpret students' differential performance on the NAEP Reading Assessment.

The 2026 NAEP Reading Framework envisions an integrated and coherent system of reporting. Research-based contextual variables form an interrelated network intended to capture reader and environmental characteristics. Information on each variable is collected from student, teacher, and administrator questionnaires and process data. Across the different questionnaires, information is collected on school characteristics, socio-demographic student characteristics, and student interests and experiences. Taken together, the network of contextual variables is intended to 1) correlate with performance on the outcome measure of reading comprehension; 2) be malleable (that is, influenced by diffeences in school and community settings); and 3) comply with the provision of the NAEP law that prohibits assessment of personal or family beliefs and attitudes. Specific questionnaire items and process data queries are selected or created to address the variables in light of each one's potential contribution to the whole.

Reader Characteristics

Research demonstrates that when students do not see an assessment as meaningful or relevant, it may not adequately capture what they know and are able to do (Valencia, Wixson, & Pearson, 2014). With respect to reader characteristics, the 2026 NAEP Reading Framework seeks to describe the role of students' perception of the interest, difficulty, and familiarity of texts, tasks, and contexts on their performances (Pintrich and Schrauben 1992; Eccles, O'Neil et al. 2005; Valencia, Wixson et al. 2014). Reader characteristic data to be collected from questionnaires and process data include the following:

Cognition and Metacognition

1. **Cognitive strategies** in reading comprehension refer to skills used to understand a text, such as drawing inferences to connect sentences together and checking to be certain that text information is fully understood (OECD, 2011).
2. **Metacognitive strategies** in reading comprehension refer to, for example, a student's use of a mental guidance system to perform such operations as deciding which sections of text are most relevant to an assigned reading goal, how to link two sections, and/or when to reread to seek more information or clarify understanding (Cho & Afflerbach, 2017).
3. **Topical knowledge** refers to students' use of their pre-existing knowledge of the reading topic to enable them to understand text information and construct new knowledge (O'Reilly &Wang, 2019).

Engagement and Motivation

1. **Volume of reading** refers to the amount of reading a student does for personal interest, pleasure or learning (Schaffner, Schiefele, Ulferts, 2013).
2. **Reading for enjoyment** refers to the goals, uses, purposes, reasons and benefits students have for reading in school and out of school (Pitzer, & Skinner, 2017).
3. **Motivations for reading** refer to students' attention, effort, interest, and value for reading a particular text with a unique set of tasks and questions related to it (NAEP Reading Special Study, 2019).

Environmental Characteristics

Deleted: are intended to

Deleted: help contextualize and explain

Deleted: The 2026 NAEP Reading Framework expands the scope of contextual variable data collected in conjunction with the NAEP Reading Assessment to reflect expanded knowledge in the field regarding cultural validity in assessment (Solano-Flores, 2010). Cultural validity refers to "the effectiveness with which the assessment addresses the sociocultural influences that shape student thinking and the ways in which students make sense of [test] items and respond to them" (Solano-Flores, 2010; Solano-Flores & Nelson-Barber, 2001, p. 555). Attention to cultural validity in assessments can guide the development of instruments to capture the proposed contextual variables by anticipating how students with different background experiences will interpret what is being asked of them. This approach to assessment acknowledges that reading as a social and cultural practice influences how readers approach, engage with, and make meaning from texts (Pacheco, 2015, 2018). Readers' values, beliefs, experiences, and ways of communicating and thinking are all shaped by their everyday experiences (Lee, 2007, 2016a). Readers' histories of engagement with texts also affect how often they read, the types of texts they read, and their purposes for reading (Cazden, 2002; Heath, 1983, 2012; Lee 1993, 2005; 2019). ¶

Deleted: With the aim of supporting policymakers and stakeholders to understand student performance and craft effective policy and education practice, t

Deleted: predict

Deleted: it can be

Deleted: and 3) avoid unnecessary or inappropriate intrusions into the lives of students' families

Deleted: The assessment construct (reading comprehension) calls for better understanding the role of student self-efficacy in carrying out particular tasks (Bandura 1993; Pajares 1996) and the relevance of such tasks for students' motivation and engagement (Guthrie and Wigfield, 2000).

Deleted: , sense of self-efficacy

Environmental characteristics are equally important in accounting for student performance. For example, students vary in their participation in cultural communities that may value reading in varied ways and integrate reading into their lives for different purposes (Skerrett, in press). Students' histories of engagement and participation constitute resources readers accumulate across their lifetimes and bring to bear on reading tasks, including those on NAEP assessments. Furthermore, what it means to read has evolved over time as cultural communities and societies have employed texts for different purposes and goals. Understanding students' differential access to community resources that support literacy development (i.e., libraries, tutoring, out-of-school programs) is important, since as these environmental contexts shift, so do the roles of reading and texts in students' lives. The degree to which schools and communities offer access to out-of-school resources influences, to some degree, students' opportunities to learn, including their own self-initiated learning, which may vary considerably. These characteristics are surveyed with regard to students' perceptions of them. Environmental characteristic data to be collected from questionnaires and process data include the following:

School and Community Resources

1. **School social support** refers to the extent to which students report that their teachers and peers contribute positively to classroom reading (through listening, speaking and interacting well with others) (Vaux, Phillips, Holly, Thompson, Williams, & Steward, 1986).
2. **Belonging in school** refers to the extent to which students report being accepted members of the school community (Faircloth, & Hamm, 2005).
3. **Participation in out-of-school reading/literacy activities** refers to the degree to which students report that they have access to resources (i.e., books, computers, media centers, camps, and community organizations) that utilize literacy for enjoyment, communication, learning, and pursuing a variety of activities (Bowen, Bowen & Ware, 2002).

Teacher, Instructional, and Classroom Supports

1. **Teacher support for reading engagement** refers to the extent to which students report that their teacher(s) provide materials and tasks that encourage the development of their reading competence and engagement (Afflerbach, Hurt, & Cho, 2020).
2. **Teacher support for motivation** refers to the degree to which students that their teacher(s) support their interests, self-efficacy, and reading goals (Wigfield & Wentzel, 2007).
3. **Teacher support for students' background experiences** refers to the degree that students report that their teacher recognizes and uses students' cultural, language, and social knowledge during reading instruction (Shin, Daly & Vera, 2007).
4. **Program and curricular support for reading development** refers to the extent to which teachers and administrators report that the school's reading program and curriculum enables them to support students' development of effective reading practices.

The NAEP 2026 Reading Framework expands collecting and reporting of contextual variables via use of refined survey item design, thereby allowing policy makers and stakeholders to gain more actionable insights regarding the variables' influences on students' efforts and their performances. For example, students' reported sense of reading engagement and motivation

Deleted: **Perceptions of**
Deleted: perceive
Deleted: believe they
Deleted: perceive
Deleted: themselves to be
Deleted: **Perceptions of**
Deleted: perceive
Deleted: as
Deleted: ing
Deleted: perceive
Deleted: to
Deleted: ' perceptions
Deleted: perceive

could be positively related to higher levels of NAEP Reading performance (Guthrie, Wigfield & You, 2012). Students' positive perceptions of their teachers' support and classroom climate could also be associated with higher NAEP Reading performance (Pitzer & Skinner, 2017). If relations such as these emerge from NAEP, they could have meaningful implications for the need to attend to perceptions, identity, and affect to support reading comprehension and achievement, recognizing that the causal nature of these variables cannot be demonstrated with NAEP cross-sectional data.

Data Sources

Beyond expanding the coverage of contextual variables, the 2026 NAEP Reading Framework also updates the method for collecting such information. In addition to items in the *questionnaires* that are routinely completed by students, teachers, and administrators from participating schools or drawn from available state, district, or school records, information about some variables will be obtained from the *process data* (computer-generated records of navigational data collected automatically as students engage with the assessment) (Ho, 2017; Bergner & Davier, 2018). Exhibit 4.2 provides a list of variables, along with their source in the revised contextual variable plan.

Exhibit 4.2. Contextual Variables

Variables	Source		
	Student Questionnaire	Teacher/ Administrator Questionnaires	Process Data
Reader Characteristics			
Cognition and Metacognition			
Cognitive strategies	√	√	√
Metacognitive strategies	√		√
Topical knowledge	√	√	
Engagement and Motivation			
Volume of reading	√	√	√
Reading for enjoyment	√	√	
Motivations for reading	√	√	
Environmental Characteristics			
Perceptions of School and Community Resources			
School social support	√	√	
Belonging in school	√	√	
Participation in out-of-school reading/literacy activities	√		
Perceptions of Teacher, Instructional, and Classroom Supports			
Teacher support for reading engagement	√	√	
Teacher support for motivation	√	√	

Deleted: . Consideration of such factors is consistent with research on the importance of social and emotional well-being to learning (Durlak et al., 2015; Elias, 2019; Guthrie & Klauda, 2016; Guthrie, Wigfield, & You, 2012; Mahoney et al., 2019; Taylor et al., 2017), the incorporation of social-emotional learning into the design of classroom and school climate (Farmer et al., 2019; Farrington et al., 2012), and approaches that build on and engage students' out-of-school identities and interests to make learning meaningful and relevant (Katz et al., 2019; Shin et al., 2007).

Deleted: These variables can also add deeper explanations for surface level findings. For example, girls are often higher achievers than boys, but this information is of limited utility for pedagogical or curricular improvement. Girls often exhibit higher motivation than boys, and they spend more time reading than do boys. When boys and girls are compared, controlling for reading time, the gender performance gap disappears (Torppa, Eklund, Sulkunen, Niemi & Ahonen, 2018). Since both reading time and motivation are malleable factors that can be impacted by interventions, the more nuanced explanation of the gender difference could inform educators about the need to reorganize instruction and improve support for reading opportunities for boys in schools. Availability of such contextual variables disaggregated within race/ethnicity and SES also provide opportunities to understand malleable factors that can be impacted by the organization of instruction. ¶

Teacher support for students' background experiences	√	√
Program and curricular support for reading development	√	√

Enhancing NAEP's Reporting Capacity

The importance and visibility of NAEP results are unquestioned within the educational policy arena, both at the national and state level. When the NAEP Report Card for Reading is issued every two years, policy makers and the public pay attention, particularly to trend data. Yet, NAEP results have also been subject to misinterpretation (Linn and Dunbar 1992; Jaeger 2003; National Research Council 2017). Because results are reported in broad categories (Race by Grade or Language Status by School Setting – Urban/Rural), they can be inappropriately interpreted. In addition, in the past, achievement results have seldom been reported in the contextmalleable factors, either for reader characteristics (e.g., student motivation) or environmental characteristics (e.g., opportunity to learn factors). Implementing the changes summarized below can mitigate potential misinterpretations and increase the usefulness of NAEP data. Reframing and expanding the reporting system is as important as the assessment construct itself in enhancing the appropriateness of inferences based on NAEP results.

1. **Revise Questionnaires**. NAEP seeks to revise and refresh questions to better reflect current research. A thorough review of current surveys—both the reading-specific and core questionnaires for the three categories of participants (students, teachers, and administrators)—will determine questions that need to be revised, replaced, or discarded. While continuing its history of ensuring the appropriateness and sensitivity of all NAEP questionnaire items, this review also enables development of questions that reflect improvements in survey item design and that will allow for better data (i.e., the data reflect the constructs outlined for questionnaires in Exhibit 4.2).
2. **Disaggregate Scores to Achieve More Nuanced and Explanatory Reporting.** Just as international, state, and formative/benchmark assessments have increased disaggregation of data in reporting, it is essential to add nuance to the reporting of performance for the major demographic categories (e.g., SES within race/ethnicity) to keep NAEP reporting structures current and useful.
3. **Expand Reporting Categories for English Learners**. Expanding the number of categories for reporting the achievement of ELs enables NAEP to track the progress of different subgroups, importantly for the added category of former ELs. By reporting the performance of non-ELs and former ELs separately, it will be possible to determine whether the two groups perform at similar levels on the NAEP Reading Assessment.
4. **Mine Process Data for Evidence of Cognitive and Metacognitive Processing**. Initial forays evaluating the utility of the process (logfile) data for NAEP (Bergner & von Davier, 2018) and other digitally delivered assessments and instructional programs (Ho, 2017) suggest that there is substantial potential for using these navigational data as indirect indices of cognitive and metacognitive processes. These indices can be used, perhaps in triangulation with measures of the same variables from reading questionnaire responses, to understand comprehension performance more deeply. Simple bar graphs

Deleted: Explanatory

Deleted: This chapter provides evidence for the potential of NAEP's reporting system to both report on and offer insights into relations between reading outcomes, students' cognitive processes and perceptions about factors that contribute to reading comprehension.

Deleted: as a function of

Deleted: , yet it is the understanding and attention to malleable factors that are most likely to lead to improved policies and practices that can shift student outcomes.

Moved (insertion) [1]

Deleted: 's explanatory power

Deleted: and its key role in promoting equity in the nation's education.

Deleted: <#>**Reframe the Reporting System Within the Larger Assessment Construct.** As discussed in preceding chapters, the 2026 NAEP Reading Assessment is guided by a commitment to equity, rigor, precision, and validity while grounded in scholarship about the nature of all learning and human development. The assessment reflects the field's evolving understanding of reading comprehension, cognitive processes, and the changing nature of reading demands in today's society. Importantly, it optimizes readers' opportunities to demonstrate reading comprehension that reflect the changing demands of our increasingly complex world (Mislevy, 2016; National Research Council, 2018). Reframing and expanding the reporting system is as important as the assessment construct itself in enhancing NAEP's explanatory power and its key role in promoting equity in the nation's education.¶

Moved up [1]: <#>Reframing and expanding the reporting system is as important as the assessment construct itself in enhancing NAEP's explanatory power and its key role in promoting equity in the nation's education.¶

Deleted: <#>To increase the capacity to examine the impact of contextual variables related to readers and their environments,

can be displayed in the Report Card, and data can be related to reading performance in the NAEP Data Explorer.

5. **Enhance the Visibility and Utility of the NAEP Reporting Portfolio.** An effort to expand, energize, and advertise the untapped resources of the NAEP reporting portfolio would allow for more nuanced data analyses. The NAEP Data Explorer, for example, permits users to go online and generate more sophisticated analyses than typically appear in the Report Card, which, by its nature, can only provide foundational reporting. In the NAEP Data Explorer for the 2019 Reading Assessment, a user can query the database to obtain a report which, for fourth graders in the nation, breaks down the performance of low- versus high-SES students on the cognitive targets of Locate and Recall, Integrate and Interpret, and Critique and Evaluate when reading literary and informational text. For sound psychometric reasons, NAEP results are not reported separately for the comprehension targets; regardless, NAEP data can be used to obtain more in-depth reports beyond the standard ones offered by the Nation's Report Card.

Conclusion

Reading comprehension performances vary depending on the combination of individual and contextual factors at the time of the assessment. Thus, NAEP Reading scores provide only a snapshot of the nation's students' reading comprehension performance as displayed in a particular testing situation at a certain moment in time. Recognizing these inherent limitations, the assessments derived from the 2026 NAEP Reading Framework nonetheless offer increased opportunities to understand the validity, efficacy, and utility of students' assets and needs as readers.

This update of the NAEP Reading Framework provides opportunities to examine malleable contextual variables that can help explain comprehension scores. The identification of malleable factors by the 2026 NAEP Reading Assessment reporting system also provides information that educators and policy makers can use to guide the improvement students' reading comprehension instruction and performance. Moreover, the disaggregation of reporting that examines heterogeneity within groups (e.g., race/ethnicity, SES, gender, English learners) will also be important. Efforts to disaggregate scores beyond what has been done in past iterations of the NAEP Reading Assessment provide opportunities for further explanatory power and greater utility for practice and research and help the field and the nation to avoid some common misinterpretations of data (e.g., overgeneralizing about groups).

The enhanced reporting system for NAEP will provide a wealth of new data sources for policymakers at state and district levels. Having access to reporting by states and networks of districts, such as TUDA, can inform state- and district-level initiatives about factors that not only predict performance but that are also malleable. Such state- and district-level reporting allows policymakers to re-examine policies intended to support students and teachers. Finally, the updated reporting system offers opportunities for researchers who will have access to a wider range of data for exploring foundational questions around the dynamic nature of reading comprehension.

Deleted: , statistically reliable

Deleted: Moreover, NAEP has a long tradition of funding small grants for secondary analyses that permit scholars to answer, in a statistically robust design, the sorts of questions that users can query with the Data Explorer tool. Increasing the funding for these initiatives would dramatically increase the portfolio of the more nuanced explanatory analysis suggested by this framework. It would be useful to replicate the 1998 study conducted by the National Validity Studies Panel (Jaeger, 1998) regarding how NAEP results are used by policy makers and educational leaders, with a focus on whether the inferences that users draw from the NAEP Report represent valid interpretations of the evidence. ¶ Implementing these steps, including a systematic study of the NAEP reporting portfolio, could serve to create an integrated system designed to better explain student performance. Such a process would use reporting variables, contextual variables, and the all-important outcome variable of comprehension, to create and evaluate the efficacy and utility of just such a system, including consideration of its costs, benefits, and feasibility. ¶

Deleted: The NAEP Reading Assessment attempts to address the role of background knowledge, readers' perceptions about the relevance and social utility of comprehension tasks, use of cognitive and metacognitive strategies, and socioemotional factors.

Deleted: Ultimately, the focus on equity, rigor, precision, validity, and the definition of reading comprehension informing the NAEP 2026 Reading Framework can shape future investments in expanding student access to robust opportunities for reading and literacy engagement in and beyond schools.¶

Accessibility: Designed or made available so all test-takers can participate or be engaged with the texts and/or assessment.

Accommodations: Modifications to the administration of an assessment that allow students with special needs or English Learners to meaningfully participate in the assessment without conveying any test advantages.

Achievement Level Descriptors: Descriptions of student performance at official NAEP achievement levels (NAEP Basic, NAEP Proficient, and NAEP Advanced), detailing what students should know and be able to do in terms of reading comprehension on the NAEP Reading Assessment

Activity (reading): Everything that readers do when they comprehend, apply and communicate their understanding of texts.

Agency: Individuals' power or control over their performance or efforts.

Assessment blocks: Largest organizational unit of the NAEP Reading Assessment, which includes a disciplinary context, broad reading purpose, 2 or more tasks, 1 or more texts, and 9-12 comprehension items.

Authentic text: Communication or composition produced by an author for publication purposes.

Avatar: Assessment task character acting as a simulated task partner.

Background knowledge: Previously acquired information and understanding about a concept, event, procedure, process, or topic. See prior knowledge.

Cognitive model (of reading comprehension): Theoretical construct that identifies mental operations to show the relationship between knowledge and reading comprehension.

Component: The parts of the reading comprehension assessment, specifically comprehension items, disciplinary contexts, broad purposes, texts, universal design elements, and contextual variables.

Comprehension item: Question or task that test-takers answer or complete to demonstrate how well they understand and can use what they read.

Constructed response: An open-ended response (short or long) to a comprehension item; includes a scoring guide to evaluate students' answers.

Construction-integration model: Theoretical account that depicts the multiple models of meaning that readers create and employ to comprehend: surface level (accurate decoding or

literal meaning); text-based (key ideas and inferences within the text); situation model (the links that readers make between their knowledge and text ideas).

Context: The physical, temporal, historical, cultural, or linguistic setting for an event, performance, statement, or idea; latter fully understood and assessed in terms of context.

Contextual variables: Factors in the home, school, community, or workplace setting that shape students' opportunities to learn, including time, content, instructional strategies, and instructional resources.

Cultural assets: The strengths students bring with them to the classroom or to the assessment, including academic and personal background knowledge, life experiences, skills and knowledge used to navigate everyday social contexts, and world views.

Cultural validity: Effectiveness with which an assessment addresses the sociocultural influences that shape student thinking and how students make sense of assessment items and respond to them.

Decoding: Applying letter sound knowledge to a letter or string of letters to translate it into a sound representation.

Design principle: Guideline for how the assessment is structured or created (e.g., guidelines for the distribution of disciplinary contexts or purposes for 4th, 8th, and 12th grades).

Developmental appropriateness: Items, tasks, or texts that are suitable for readers at certain ages, grade levels or maturity stages in terms of content, how they are written, and cognitive or academic demands.

Digital assessment feature: A characteristic of an electronic, online, or computerized evaluation.

Digital platform: Electronic location or environment on the internet or computer where a technologically enabled assessment is operated.

Digital text: Electronic print, communication (e.g., audio, visual, images) or composition on a computer.

Digitally-based assessment: Electronic, computer-based, or online evaluation of individuals' performance.

Disaggregation: Separated into parts or elements. In the 2026 Framework, considering the effects of one variable, such as income, within another, such as race/ethnicity.

Discipline/ Disciplinary Context: Specialized academic domain (e.g., Literature, science, social studies) with specific purposes, tasks, ways of thinking, vocabulary, rhetoric, and discourse conventions.

Discrete tasks: Stand-alone text passages and related questions.

Distribution: How an item is divided, spread or organized.

Domain knowledge: Information or understanding about a particular academic field (e.g., geography) or discipline or concept (e.g, rock formation).

Dynamic text: Non-static digital format. Involves movement or navigation across modes (e.g., print, images, or video) or nonlinear locations (e.g., a hypertext link).

Ecological validity: The extent to which an assessment elicits students' reading performance as demonstrated in real-world settings, such as school, home, community or workplace.

English Learner: Second-language learner of English who speaks minority language at home, but enrolled in a bilingual education or English-as-a-second-language (ESL) program at school to develop grade-level English proficiency.

English-language proficiency: An English Learner's assessed level of speaking, writing, listening, and reading in English. Includes the use of English in academic and social settings.

Equity: The state of being fair, just, and free from bias or favoritism.

Expository text (exposition): Nonfiction composition or classification of discourse. Presents information or ideas, instructs.

Figurative language: Employed by authors of literature to create images or associations that extend beyond literal meaning of words (e.g., metaphors, hyperbole, personification, and simile).

Fluency: Quick and accurate oral reading with expression or prosody that reflects the meaning of the text.

Former English Learners: Second-language learners of English exited from bilingual education or ESL programs within the last two years and participants in all-English classrooms.

Foundational reading skills: The basic competences needed for English reading comprehension, such as word recognition (decoding and vocabulary knowledge), sight word reading, and fluency.

Global inference: Reader's assumption or conclusion based on ideas or evidence drawn from prior knowledge and across the text.

Historical reasoning: Critical thinking about the past that involves evaluating the credibility of primary sources. May be assessed by the Analyze and Evaluate Comprehension Target when students read texts in the disciplinary context of social studies.

Hypertext: Interconnected documents or sources of information that readers can immediately access on the internet through diverse actions (clicking on a word, a link, etc.)

Inferential reasoning: Act or process of deriving logical conclusions from premises known or assumed to be true; the conclusions drawn from this process. In 2026 NAEP reading assessment, involved in all four Comprehension Targets.

Foreshadowing: Use of hints or clues in a narrative to suggest future action.

Knowledge-based UDE: A type of Universal Design Element (UDE) that includes topic previews/introductions and vocabulary pop-up definitions.

Linguistic knowledge: Native-speakers' unconscious understanding of the language(s) (vocabulary, syntax, etc.) spoken in their homes and communities. What is taught to students about English in school.

Malleable factors: Conditions, items or issues that can be changed or modified in students' schools or communities.

Metacognition: Awareness and analysis of one's own learning, reading, or thinking processes.

Modality: Different ways that information is presented (e.g., auditory, visual, tactile, kinesthetic).

Motivational UDE: A type of Universal Design Element (UDE) that encourages and supports readers' interest, engagement and persistence, especially when encountering challenging tasks.

Multimodal text: Meaning conveyed through still and moving images, animations, color, words, music, and sound.

Navigational complexity: The difficulty of progressing through assessment components and modalities to demonstrate comprehension based on what test takers encounter and have to do. Includes the number and types of texts to read, inferences to make, tasks to complete, items to answer, responses to provide, and modes (print, visual, images, audio, etc.).

Operationalization: To put into action or to realize.

Opportunities to learn (OTL): Inputs and processes that enable student achievement of intended outcomes.

PISA: The Programme for International Student Assessment, an international assessment that measures 15-year-old students' reading, mathematics, and science literacy every three years.

Prior knowledge: Previously acquired information and understanding about a concept, event, procedure, process, or topic. See background knowledge.

Process data: Information collected as students navigate the digital assessment, including the time taken to read texts and respond to questions, how often they return to the text to answer questions, and their use of optional digital tools.

Scenario-based tasks: Simulated settings in which students read passages while following steps to accomplish a particular purpose, especially to solve a problem.

Selected response: Answers in which a student selects one or more options from a given, limited set of answer choices.

Situation model: Part of the Construction-Integration model of reading comprehension (Kintsch, 1988). The level where readers make links between text ideas and their own knowledge.

Social Emotional Learning (SEL): How humans "develop healthy identities, manage emotions and achieve personal and collective goals, feel and show empathy for others, establish and maintain supportive relationships, and make responsible and caring decisions" (CASEL, https://casel.org/what-is-sel/).

Sociocultural context: The environments and experiences that shape individuals' thinking, learning, and development, including reading comprehension. Diverse communities' values, beliefs, experiences, communication patterns, and styles of teaching and learning.

Static text: Non-moving print, graphics, or images.

Student identity: A student's evolving view of self in a given social context influenced by his or her experiences, personal history, and other events.

Syntax: The organization of words or phrases into sentences in a text, composition, or speech.

Task-based UDE: A type of Universal Design Element that clarifies requirements and guides readers in their use of available resources; increases readers' access and sustains their attention as they take an assessment.

Text complexity: The conceptual, structural and linguistic features that create comprehension challenges for readers. Includes density and nuance of ideas and language structures, word frequency, passage length, syntactic complexity, and stylistic features. Typically monitored by research-based quantitative measures of readability and qualitative analyses of semantic, syntactic, and discourse elements.

Text genre: Category used to classify literary and other works by form, technique, or content.

Text structure: Organization of ideas in a composition. In narrative compositions, according to a sequential, event-driven story grammar; in expository compositions, according to rhetorical structures (e.g., description, comparison-contrast, sequence, problem-solution, or conflict-resolution).

Deleted: ¶

Deleted: **Reader self-efficacy:** An individual's belief in his or her capacity to read effectively to accomplish reading tasks.¶

Text-based inference: Act or process of deriving logical conclusions or assumptions based on information stated in the composition.

Topic knowledge: Understanding or information about the specific subject of a text or text segment, such as dinosaurs or river formation. Tends to be more specific than domain knowledge or world knowledge or prior/background knowledge.

Trait: A distinguishing feature or quality.

Universal Design Element (UDE): A feature of the assessment environment provided to help all test takers access, organize, analyze, and express ideas when engaged in complex tasks.

Universal Design for Assessment: Principles for creating and administering evaluations or tests so accessible, include as many types of students as possible, and result in valid inferences or scores in terms of grade-level performance.

Validity: How accurately a method measures what it is intended to measure.

Variance: A statistical measurement of the spread between numbers in a data set.

Vocabulary pop-up: A knowledge-based UDE in NAEP that a test taker can access to obtain the meaning of a word important for understanding the overall text but not assessed in the comprehension items.

World knowledge: Global information about other cultures, countries, and people. See background and prior knowledge.

REFERENCES

(NCES, 2018).

Abedi, J. (2004). The No Child Left Behind Act and English language learners: Assessment and accountability issues. *Educational Researcher, 33*(1), 4-14, doi.org/10.3102/0013189X033001004

AERA/APA/NCME. (2014). Standards for educational and psychological testing. Washington, DC: Author.

Afflerbach, P., Biancarosa, G., Hurt, M., & Pearson, P. D. (2020). Teaching reading for understanding: Synthesis and reflections on the curriculum and instruction portfolio. In P. D. Pearson, A. S. Palincsar, G. Biancarosa, & A. Berman (Eds.), (2020). *Reaping the rewards of the Reading for Understanding Initiative* (pp. 215-250). Washington, DC: National Academy of Education.

Afflerbach, P., Hurt, M., & Cho, B. (2020). Reading comprehension strategy instruction. In D. L. Dinsmore, L. A. Fryer, & M. M. Parkinson (Eds.), *Handbook of strategies and strategic processing* (pp. 99-118). New York: Routledge.

Afflerbach, P., Pearson, P. D., & Paris, S. G. (2007). Clarifying differences between reading skills and reading strategies. *The Reading Teacher, 61*(5), 364-373, doi.org/10.1598/RT.61.5.1

Aguilar, G., Uccelli, P., & Galloway, E. P. (2020). Toward biliteracy: Unpacking the contributions of mid-adolescent dual language learners' Spanish and English academic language skills to English reading comprehension. *TESOL Quarterly* (early view), DOI: 10.1002/tesq.570

Ajzen, I. (2002). Perceived behavioral control, self-efficacy, locus of control, and the theory of planned behavior. *Journal of Applied Social Psychology, 32*(4), 665-683, doi.org/10.1111/j.1559-1816.2002.tb00236

Alexander, P. A., & Grossnickle, E. M. (2016). Positioning interest and curiosity within a model of academic development. In K. Wentzel & D. Miele (Eds.), *Handbook of motivation at school* (2nd ed., pp. 188–208). New York: Routledge.

Alexander, P. A., Kulikowich, J. M., & Schulze, S. K. (1994). How subject-matter knowledge affects recall and interest. *American Educational Research Journal, 31*(2), 313-337, doi.org/10.3102/00028312031002313

Alvermann, D. E. & Wilson, A. A. (2011). Comprehension strategy instruction for multimodal texts in science. *Theory Into Practice, 50*(2), 116-124.

Ambruster, B. B. (1984). Learning from content area textbooks: The problem of "inconsiderate text." In G. Duffy, L. Roehler, & J. Mason (Eds.), Comprehension instruction (pp. 202-217). New York, NY: Longman.

Anderson, R. C. (2019). Role of the reader's schema in comprehension, learning and memory. In D.E. Alvermann, N.J. Unrau, & R.B. Ruddell (Eds.), *Theoretical models and processes of literacy* (pp. 136-145). New York: Routledge.

Anderson, R. C., & Pearson, P. D. (1984). A schema-theoretic view of basic processes in reading comprehension. In P. D. Pearson, R. Barr, M. L. Kamil, P. Mosenthal (Eds.), *Handbook of Reading Research* (Vol. I) (pp. 255-291). New York: Routledge.

August, D., & Shanahan, T. (Eds.) (2006). *Developing literacy in second-language learners: Report of the National Literacy Panel on Language-Minority Children and Youth* (pp. 583-596). Mahwah, NJ: Lawrence Erlbaum Associates.

Deleted: *APPENDIX A: ADDITIONAL ASSESSMENT DESIGN FEATURES*¶

¶

Exhibit 1. Principle and Provisional Distribution Targets for Sampling Assessment Design Elements: Text Formats and Modes¶

For All Grade Levels ... [37]

Bakken, J. P., & Whedon, C. K. (2002). Teaching text structure to improve reading comprehension. *Intervention in School and Clinic, 37*(4), 229-233.
Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Upper Saddle River, NJ: Prentiss Hall.
Bandura, A. (1993). "Perceived self-efficacy in cognitive development and functioning." Educational Psychologist 28(2): 117-148.
Bandura, A. (2006). Guide for constructing self-efficacy scales. In F. Pajares & T. Urdan (Eds.), *Self-efficacy beliefs of adolescents* (pp. 307-337). Greenwich, CT: Information Age Publishing.
Barzilai, S., & Zohar, A. (2012). Epistemic thinking in action: Evaluating and integrating online sources. *Cognition and Instruction, 30*(1), 39-85, DOI: 10.1080/07370008.2011.635495
Barzilai, S., & Zohar, A. (2016). Epistemic (meta)cognition: Ways of thinking About knowledge and knowing. In J. A. Greene, W. A. Sandoval & I. Bråten (Eds.), *Handbook of epistemic cognition* (pp. 409-424). New York: Routledge.
Beach, R. (2000). Critical issues: Reading and responding to literature at the level of activity. *Journal of Literacy Research, 32*(2), 237-251, doi.org/10862960009548075
Beach, R., & Castek, J. (2015). Use of apps and devices for fostering mobile learning of literacy practices. In B. Guzzetti & M. Lesley (Eds.), *Handbook of research on the societal impact of digital media* (pp. 343–370). Hershey, PA: IGI Global.
Bergner, Y. & von Davier, A. A. (2018). Process data in NAEP: Past, present, and future. Journal of Educational and Behavioral Statistics, 20(10), 1-27. DOI: 10.3102/1076998618784700
Bowen, N., Bowen, G. & Ware, W. (2002). Neighborhood Social Disorganization, Families, and the Educational Behavior of Adolescents. *Journal of Adolescent Research*, 17 (5), 468-490.
Bråten, I., Braasch, J.L.G., Salmerón, L. (2020). Reading multiple and non-traditional texts: New opportunities and new challenges. In E. B. Moje, P. Afflerbach, P. Enciso, & N. K. Lesaux (Eds.), *Handbook of reading research* (Vol. V) (pp. 79-98). New York: Routledge.
Bråten, I., Britt, M. A., Strømsø, H. I., & Rouet, J. (2011). The role of epistemic beliefs in the comprehension of multiple expository texts: Toward an integrated model. *Educational Psychologist, 46*(1), 48-70, doi.org/10.1080/00461520.2011.53847
Bronfenbrenner, U., & Morris, P. A. (2006). The bioecological model of human development. In R. M. Lerner & W. Damon (Eds.), *Handbook of child psychology: Theoretical models of human development* (pp. 793-828). John Wiley & Sons.
Breakstone, J., McGrew, S., Smith, M., Ortega, T., & Wineburg, S. (2018). Why we need a new approach to teaching digital literacy. *Phi Delta Kappan, 99*(6), 27-32.
Bryant, A. C., Triplett, N. P., Watson, M. J., & Lewis, C. W. (2017). The browning of American public schools: Evidence of increasing racial diversity and the implications for policy, practice, and student outcomes. *The Urban Review*, *49*(2), 263-278.
Bunch, G. C., Walqui, A. & Pearson, P. D. (2014). Complex Text and New Common Standards in the United States: Pedagogical Implications for English Learners. *TESOL Quarterly, 48*(3), 533-559.
Business Roundtable. (2017). *Work in progress: How CEOs are helping close America's skills gaps.* Retrieved May 2018, from https://www.businessroundtable.org/skills

Calfee, R. C., & Miller, R. G. (2005). Comprehending through composing: Reflections on reading assessment strategies. In S. Paris & S. Stahl (Eds.), *Children's reading comprehension and assessment* (pp. 215-233). Mahwah, NJ: Lawrence Erlbaum Associates.

Cantor, P., Osher, D., Berg, J., Steyer, L., & Rose, T. (2019). Malleability, plasticity, and individuality: How children learn and develop in context. *Applied Developmental Science, 23*(4), 307-337, DOI: 10.1080/10888691.2017.1398649

Carroll, J. M., & Fox, A. (2017). Reading self-efficacy predicts word reading but not comprehension in both girls and boys. *Frontiers in Psychology, 7*(25), 1-9, DOI: 10.3389/fpsyg.2016.02056

CAST. (2020). About Universal Design for Learning. *CAST.* http://www.cast.org/our-work/about-udl.html#.XrWphi-z3kI

Cazden , C.B. (2002). Classroom discourse: The language of teaching and learning (2nd ed.). Portsmouth, NH : Heinemann .

Cazden, C. B. (2001). *Classroom discourse: The language of teaching and learning*. Portsmouth, NH: Heinemann.

Cervetti, G. N. (2020). The nature and development of reading for understanding. In P. D. Pearson, A. S. Palincsar, G. Biancarosa, & A. Berman (Eds.), (2020). *Reaping the rewards of the Reading for Understanding Initiative* (pp. 39-64). Washington, DC: National Academy of Education.

Cervetti, G.N., & Wright, T.S. (2020). The role of knowledge in understanding and learning from text. In E. B. Moje, P. Afflerbach, P. Enciso, & N. K. Lesaux (Eds.), *Handbook of reading research* (Vol. V) (pp. 237-260). New York: Routledge.

Chen, C-M, & Chen, F-Y. (2014). Enhancing digital reading performance with a collaborative reading annotation system. *Computers & Education, 77,* 67-81.

Cho, B. (2014). Competent adolescent readers' use of internet reading strategies: A think-aloud study. *Cognition and Instruction, 32*(3), 253-289, DOI: 10.1080/07370008.2014.918133

Cho, B.-Y., & Afflerbach, P. (2017). An evolving perspective of constructively responsive reading comprehension strategies in multilayered digital text environments. In S. E. Israel (Ed.), *Handbook of research on reading comprehension* (2nd ed., pp. 109–134). New York, NY: Guilford.

Coiro, J. (2011). Predicting reading comprehension on the Internet: Contributions of offline reading skills, online reading skills, and prior knowledge. *Journal of Literacy Research*, 43(4), 352–392. doi :10.1177/1086296X11421979

Coiro, J. (2020). Toward a multi-faceted heuristic of digital reading to inform assessment, research, practice, and policy. *Reading Research Quarterly* (early view), doi.10.1002/rrq.302

Coiro, J., & Dobler, E. (2007). Exploring the online reading comprehension strategies used by sixth-grade skilled readers to search for and locate information on the Internet. *Reading Research Quarterly*, 42(2), 214–257. doi:10.1598/RRQ.42.2.2

Coiro, J., Knobel, M., Lankshear, C., & Leu, D. J. (2014). Handbook of research on new literacies. Routledge.

Coiro, J., Sparks, J., Kiili, C., Castek, J., Lee, C., & Holland, B. (2019). Capturing dimensions of collaborative online inquiry and social deliberation with multiple-source inquiry tasks in face-to-face and remote contexts. *Literacy Research: Theory, Method, and Practice.* Online first version available at https://doi.org/10.1177/2381336919870285

Cole, M. (1996). *Cultural psychology: A once and future discipline.* Cambridge, MA: Harvard University Press.
Cote, N., Goldman, S.R. & Saul, E.U. (1998). Students making sense of informational text: Relations between processing and representation. *Discourse Processes, 25*, 1-53, DOI: 10.1080/01638539809545019
Creer. A. (2018). Introducing everyday 'digital literary practices' into the classroom: An analysis of multi-layered media, modes and their affordances. *Journal of New Approaches in Educational Research, 7*(2), 131-139.
Dalton, B., & Proctor, C. P. (2008). The changing landscape of text and comprehension in the age of new literacies. In J. Coiro, M. Knobel, C. Lankshear, & D. Leu (Eds.), *Handbook of research on new literacies* (pp. 297-324). Mahwah, NJ: Lawrence Erlbaum.
Daniel, S. M., & Pacheco, M. B. (2016). Translanguaging practices and perspectives of four multilingual teens. *Journal of Adolescent & Adult Literacy, 59*(6), 653-663, doi.org/10.1002/jaal.500
Dinsmore, D., L., & Parkinson, M. M. (2013). What are confidence judgments made of? Students' explanations for their confidence ratings and what that means for calibration. *Learning and Instruction, 24*, 4-14, doi.org/10.1016/jlearninstruc.2012.06.001
Dobler, E., & Azwell, T. (2007). Real world reading: Making sense of the texts that matter in our everyday lives. Kansas Career & Technical Education Resource Center. https://www.ksde.org/Portals/0/CSAS/CSAS%20Home/CTE%20Home/Instructor_Resources/FinalRealWorldReading.pdf
Duke, N. K., & Pearson, P. D. (2002). Effective practices for developing reading comprehension. In A. E. Farstrup & S. J. Samuels (Eds.), *What research has to say about reading comprehension* (3rd ed.) (pp. 205-242). Newark, DE: International Reading Association.
Durán, 2006
Durlak, J., Domitrovich, C., Weissberg, R., & Gullotta, T. (Eds.). (2015). *Handbook of social and emotional learning: Research and practice*. Guilford Press.
Dweck, C., & Molden, D. (2005). Self-theories: Their impact on competence motivation and acquisition. In A. Elliot, and C. Dweck (Eds.), *Handbook of competence and motivation* (pp. 122-140). New York: Guilford Press.
Eccles, J. O'Neill, et al. (2005). Ability self-perceptions and subjective task values in adolescents and children.
Eccles, J., et al. (2005). Ability self-perceptions and subjective task values in adolescents and children. Moore, K.
Educational Testing Service (2019). NAEP Reading Special Study: Scenario-Based Task and Discrete Task Report. Deliverable in Response to ID Task 9.2.1. Note: This report is not available to the general public because it contains secure information.
Educational Testing Service. (January 22, 2019). NAEP reading special study: Scenario-based task (SBT) and discrete (DI) task report. 2018 special study grades 4, 8, and 12.
Efklides, A. (2006). Metacognition and affect: What can metacognitive experiences tell us about the learning process. *Educational Research Review, 1*(1), 3-14, DOI: 10.1016/jedurev.2005.11.001
Elias, M. (2019). What if the doors of every schoolhouse opened to social-emotional learning tomorrow: Reflections on how to feasibly scale up high-quality SEL. *Educational Psychologist, 54*(3), 233–245.

Enciso, P., Volz, A., Price-Dennis, D., & Durriyah, T. (2010). Story club and configurations of literary insights among immigrant and non-immigrant youth. In R. Jiménez, D. Rowe, V. Risko, & M. Hundley (Eds.), *59th Yearbook of the National Reading Conference* (pp. 354–366). Oak Creek, WI: National Reading Conference.

Faircloth, B., & Hamm, J., (2005). Sense of Belonging Among High School Students Representing 4 Ethnic Groups. *Journal of Youth and Adolescence*, 34, 4, 293–309

Fang, Z., Shleppegrell, M., & Cox, B. (2006). Understanding the language demands of schooling: Nouns in academic registers. Journal of Literacy Research, 38(3), 247-273.

Farmer, T., Hamm, J., Dawes, M., Barko-Alva, K., & Cross, J. (2019). Promoting inclusive communities in diverse classrooms: Teacher attunement and social dynamics management. *Educational Psychologist, 54*(4), 286–305.

Farmer, T. W., Lines, M. M., & Hamm, J. V. (2011). Revealing the invisible hand: The role of teachers in children's peer experiences. *Journal of Applied Developmental Psychology*, 32, 247–256. doi:10.1016/j.appdev.2011.04.006

Farrington, C. A., Roderick, M., Allensworth, E., Nagaoka, J., Keyes, T. S., Johnson, D. W., & Beechum, N. O. (2012). *Teaching adolescents to become learners: The role of noncognitive factors in shaping school performance.* Chicago, IL: Consortium on Chicago School Research.

Fisher, D., Frey, N., & Lapp, D. (2012). *Text complexity: Raising rigor in reading*. Newark, DE: International Reading Association.

Fitzgerald, M., Higgs, J., & Palincsar, A. (2020). *New media, new literacies: Implications for reading for understanding.* Washington, DC: National Academy of Education.

Fletcher, J. (2015). Performing digital literature. *Caracteres: Estudios Culturales y Críticos de la Esfera Digital, 4*(2), 18-42.

Francis, D. J., Lesaux, N. K., & August, D. (2006). Language of instruction. In D. August & T. Shanahan (Eds.), *Developing literacy in second-language learners: Report of the National Literacy Panel on Language-Minority Children and Youth* (p. 365–413). New York: Lawrence Erlbaum Associates.

Frankel, K. K., Becker, B. L. C., Rowe, M. W., & Pearson, P. D. (2016). From "What is reading?" to what is literacy? *The Journal of Education, 196*(3), 7-17, doi.org/10.1177/002205741619600303

García, G. E, & Godina, H. (2017). A window into bilingual reading: The bilingual reading practices of fourth-grade, Mexican-American children who are emergent bilinguals. *Journal of Literacy Research, 49*(2), 273–301, doi.org/10.1177/1086296X17703727

García, G.E. (1991). Factors influencing the English reading test performance of Spanish-speaking Hispanic children. Reading Research Quarterly, 26(4), 371-392

García, G.E., Saco, L.J., & Guerrero-Arias, B.E. (2020). Cognate instruction and bilingual students' improved literacy performance. *The Reading Teacher, 73*(5), 617-625.

García, O. (2009). *Bilingual education in the 221st century: A global perspective*. Malden, MA: Basil/Blackwell.

Gee, J. P. (2001). Reading as situated language: A sociocognitive perspective. *Journal of Adolescent and Adult Literacy, 44*(8), 714-725.

Goldman S. (2012). Adolescent literacy: learning and understanding content. *The Future of Children / Center For the Future of Children, the David and Lucile Packard Foundation*. 22: 89-116. PMID 23057133

Goldman, S. R. (2018). Discourse of learning and the learning of discourse. Discourse Processes,55(5–6), 434–453.
Goldman, S. R., & Lee, C. D. (2014). Text complexity: State of the art and the conundrums it raises. *The Elementary School Journal, 115*(2), 290-300, doi/10.1086/678298
Goldman, S. R., et al. (2016). "Disciplinary Literacies and Learning to Read for Understanding: A Conceptual Framework for Disciplinary Literacy." Educational Psychologist: 1-28.
Goldman, S., Britt, M. A., Brown, W., Cribb, G., George, M., Greenleaf, C., Lee, C.D., Shanahan, C., & Project READI. (2016). Disciplinary literacies and learning to read for understanding: A conceptual framework for disciplinary literacy. *Educational Psychologist, 51*(2), 219-246. http://dx.doi.org/10.1080/00461520.2016.1168741.
González, N., Moll, L. C., & Amanti, C. (Eds.). (2005). Funds of Knowledge: Theorizing practices in households, communities, and classrooms. Mahwah, NJ: Erlbaum.
Greenleaf, C. Schoenbach, R., Cziko, C., and Mueller, F. (*2001*) Apprenticing Adolescent Readers to Academic Literacy. *Harvard Educational Review*: April 2001, Vol. 71, No. 1, pp. 79-130.
Gribben, M.; Patelis, T. & Schultz, S. (2020) *Review of the Impact of Student Choice in Assessment,* Report 2020:037. Alexandria, VA: HumRRO.
Guan, S. S. A., Nash, A., & Orellana, M. F. (2016). Cultural and social processes of language brokering among Arab, Asian, and Latin immigrants. *Journal of Multilingual and Multicultural Development*, *37*(2), 150-166.
Guthrie, J. T. and A. Wigfield (2000). Engagement and motivation in reading. Handbook of reading research. M. Kamil, P. Mosenthal, P. D. Pearson and R. Barr. Mahwah, NJ, Lawrence Erlbaum. III: 403-422.
Guthrie, J. T., & Klauda, S. (2014). Effects of classroom practices on reading comprehension, engagement, and motivations for adolescents. *Reading Research Quarterly, 49*(4), 387-416, DOI: 10.1002/rrq.81
Guthrie, J. T., Klauda, S. L. (2016). Engagement and motivational processes in reading. In Afflerbach, P. (Ed.), *Handbook of individual differences in reading: Reader, text, and context* (pp. 41–53). Routledge.
Guthrie, J. T., Klauda, S. L., & Ho, A. N. (2013). Modeling the relationships among reading instruction, motivation, engagement, and achievement for adolescents. *Reading Research Quarterly*, *48*(1), 9-26
Guthrie, J. T., Wigfield, A., & You, W. (2012). Instructional contexts for engagement and achievement in reading. In S. Christensen, A. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 601-635). New York: Springer Science
Guthrie, J. T., Wigfield, A., Metsala, J. L., & Cox, K. E. (1999). Motivational and cognitive predictors of text comprehension and reading amount. *Scientific Studies of Reading, 3*, 231–256, doi..org/10.1207/s1532799xssr0303_3
Guthrie, J. T., Wigfield, A., Von Secker, C. (2000) Effects of integrated instruction on motivation and strategy use in reading. Journal of Educational Psychology, 92(2), 331-341.
Guthrie, J.T. & Klauda, S. (2015). Engagement and motivational processes in reading. In P. Afflerbach, Handbook of Individual Differences in Reading. (pp. 41-54). Routledge Publishers, New York.

Gutiérrez, K. D., Morales, P. Z., & Martinez, D. C. (2009). Re-mediating literacy: Culture, difference, and learning for students from nondominant communities. *Review of Research in Education, 33,* 212-245, DOI: 10.3102/0091732X08328267

Hain, B.A. & Piper, C. (2016). PARCC as a case study in understanding the design of large-scale assessment in the era of Common Core State Standards. In H. Jiao & R.W. Lissetz (Eds). The next generation of testing: Common Core Standards, Smarter-Balanced, PARCC, and the nationwide testing movement (pp. 29-48). Information Age Publishing.

Hall, L. A. (2016). The role of identity in reading comprehension development. *Reading & Writing Quarterly*, *32*(1), 56-80.

Harris, Y. R., & Schroeder, V. M. (2013). Language deficits or differences: What we know about African American vernacular English in the 21st century. *International Education Studies, 6*(4), 194-204. DOI: 10.5539/ies.v6n4p194

Hasbrouck, J. & Tindal, G. A., (2006). Oral reading fluency norms:A valuable assessment tool for reading teachers. The Reading Teacher, 59(7), 636-644.

Heath, S. B. (1983). *Ways with words: Language, life, and work in communities and classrooms*. Cambridge: Cambridge University Press.

Heath, S. B. (2012). "So what's it about? Your book, I mean?" *Journal of Adolescent & Adult Literacy, 56*(4), 266-270, doi-org.libezproxy2.syr.edu/10.1002/JAAL.00137

Heath, S.B. (2012). Words at work and play: Three decades in family and community life. Cambridge, UK: Cambridge University Press.

Ho, A. (2017). Advancing Educational Research and Student Privacy in the "Big Data" Era. Washington, DC: National Academy of Education.

Hofer, P. J. (1997). Multicultural assessment of IQ. *Archives of Clinical Neuropsychology*, *12*(4), 335.

Hopkins, Thompson, Linquanti, August, & Hakuta, 2013

Hruby, G. G., & Goswami, U. (2019). Educational neuroscience for reading researchers. In D.E. Alvermann, N.J. Unrau, & R.B. Ruddell (Eds.), *Theoretical models and processes of literacy* (7th ed.) (pp. 252-278). New York: Routledge.

Hruby, G. G., Goswami, U., Frederiksen, C. H., & Perfetti, C. A. (2011). Neuroscience and reading: A review for reading education researchers. *Reading Research Quarterly, 46*(2), 156–172, dx.doi.org/10.1598/RRQ.46.2.4

Ilter, I. (2019). Efficacy of note-taking skills instruction supported by self-monitoring on the reading comprehension. *Education and Science, 44*(198), 229-253, DOI: 10.15390/EB.2019.7862

Immordino-Yang, M. H., & Gotlieb, R. (2017). Embodied brains, social minds, cultural meaning: Integrating neuroscientific and educational research on social-affective development. *American Educational Research Journal*, *54*(1_suppl), 344S-367S, doi.org/10.3102/0002831216669780

Jaeger 2003

Jaeger, 1998

Jiménez, R. T., García, G. E., & Pearson, P. D. (1995). Three children, two languages, and strategic reading: Case studies in bilingual/monolingual reading. *American Educational Research Journal, 32*(1), 67–97, doi.org/10.2307/1163214

Jiménez, R. T., García, G. E., & Pearson, P. D. (1996). The reading strategies of bilingual Latina/o students who are successful English readers: Opportunities and obstacles. *Reading Research Quarterly, 31*(1), 90-112, doi.org/10.1598/RRQ.31.1.

Johnson, S. B., Riis, J. L., & Noble, K. G. (2016). State of the art review: Poverty and the developing brain. *Pediatrics, 137*(4), e20153075, DOI: 10.1542/peds.2015-3075
Katz, M.L., Brynelson, N., and Edlund, J.R. (2019). "Enacting Rhetorical Literacies: The Expository Reading and Writing Curriculum in Theory and Practice." seventh edition, *Theoretical Models and Processes of Literacy*, D. Alvermann, N. Unrau, and M. Sailors (Eds.). New York: Routledge (pp. 533-562).
Kendeou, P., & O' Brien, E. J. (2016). Prior Knowledge: Acquisition and revision. In P. Afflerbach (Ed.), *Handbook of individual differences in reading: Text and context* (pp. 151-163). New York: Routledge.
Kendeou, P., Van den Broek, P., Helder, A., & Karlsson, J. (2014). A cognitive view of reading comprehension: Implications for reading difficulties. *Learning Disabilities Research & Practice*, *29*(1), 10-16.
Kieffer, M. J., & Thompson, K. D. (2018). Hidden Progress of Multilingual Students on NAEP. *Educational Researcher, 47,* pp. 391–398. doi: 10.3102/0013189X18777740
Killi, C., Laurinen, L., & Marttunen, M. (2008). Students evaluating internet sources: From versatile evaluators to uncritical readers. *Journal of Computing Research, 39*(1), 75-95, doi.org/10.2190/EC.39.1.e
Kim, Y.-S. G., Petscher, Y., Wanzek, J., & Al Otaiba, S. (2018). Relations between reading and writing: A longitudinal examination from Grades 3 to 6. *Reading and Writing: An Interdisciplinary Journal, 31*(7), 1591–1618, doi.org/10.1007/s11145-018-9855-4
Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge, Cambridge University Press.
Kintsch, W. (2019). Revisiting the construction-integration model of text comprehension and its implications for instruction. In D.E. Alvermann, N.J. Unrau, & R.B. Ruddell (Eds.), *Theoretical models and processes of literacy* (7th ed.) (pp. 178-203). New York: Routledge.
Kress, G. (2013). Recognizing learning. In I. deSaint-Georges & J. J. Weber (Eds.), Multilingualism and multimodality (pp. 119–140). Rotterdam, ND: Sense Publishers.
Kress, G., & van Leeuwen, T. (2001). Multimodal discourse: The modes and media of contemporary communication. London, UK: Arnold.
Kress, G., Jewitt, C., & Tsatsarelis, C. (2000). Knowledge, identity, pedagogy: pedagogic discourse and the representational environments of education in late modernity. *Linguistics and Education, 11*(1), 7-30, doi.org/10.1016/S0898-5898(99)00015-7
Kuhn, D., & Weinstock, M. (2002). What Is epistemological thinking and why does it matter? In B. Hofer, & P. Pintrich (Eds.), *Personal epistemology: The psychology of beliefs about knowledge and knowing* (pp. 121-144). New York: Routledge.
Lafontaine, D., Baye, A., Vieluf, S., & Monseur, C. (2015). Equity in opportunity-to-learn and achievement in reading: A secondary analysis of PISA 2009 data. *Studies in Educational Evaluation, 47*, 1-11.
LaRusso, M., Kim, H. Y., Selman, R., Uccelli, P., Dawson, T., Jones, S., et al. (2016). Contributions of academic language, perspective taking, and complex reasoning to deep reading comprehension. *Journal of Research on Educational Effectiveness, 9*(2), 201-222.
Latini, N., Bråten, I., Anmarkrud, Ø., & Salmerón, L. (2019). Investigating effects of reading medium and reading purpose on behavioral engagement and textual integration in a

multiple text context. *Contemporary Educational Psychology*, *59*, 101797, doi.org/10.1016/j.cedpsych.2019.101797.
Lee, 2019
Lee, C. D. (1993). *Signifying as a scaffold for literary interpretation: The pedagogical implications of an African American discourse genre*. Urbana, IL, National Council of Teachers of English.
Lee, C. D. (1997). Bridging home and school literacies: A model of culturally responsive teaching. In J. Flood, S. B. Heath, & D. Lapp (Eds.), *A handbook for literacy educators: Research on teaching the communicative and visual arts* (pp. 330–341). New York, NY: McMillan.
Lee, C. D. (2005). Culture and language: Bi-dialectical issues in literacy. In P. L. Anders & J. Flood (Eds.), *The literacy development of students in urban schools*. Newark, DE: International Reading Association.
Lee, C. D. (2006). Every good-bye ain't gone: Analyzing the cultural underpinnings of classroom talk. I*nternational Journal of Qualitative Studies in Education, 19*(3), 305–327.
Lee, C. D. (2007). *Culture, literacy and learning: Taking bloom in the midst of the whirlwind.* NY, Teachers College Press.
Lee, C. D. (2014). A voyeuristic view of possibilities and threats: Neurosciences and education. *Human Development*, *57*(1), 1-4, doi.org/10.1159/000360166.
Lee, C. D. (2016). Examining conceptions of how people learn over the decades Through AERA presidential addresses: Diversity and equity as persistent conundrums. *Educational Researcher, 45*(2), 73-82, doi.org/10.3102/0013189X16639045
Lee, C. D. (2016). Influences of the experience of race as a lens for understanding variation in displays of competence in reading comprehension. In P. Afflerbach, (Ed.), *Handbook of individual differences in reading: Reader, text, and context* . NY, Routledge: 286-304.
Lee, C. D. (2020). Social and cultural diversity as lens for understanding student learning and the development of reading. In E. B. Moje, P. Afflerbach, N. K. Lesaux, & P. Enciso (Eds.), *Handbook of reading research, Volume V* (37-56). New York: Routledge.
Lee, C. D. (in press). Social and Cultural Diversity as Lens for Understanding Student Learning and the Development of Reading Comprehension. In Handbook of Reading Research,Volume V. Moje, E.B., Afflerbach, P., Encisco, P. & Lesau, N. NY, Routledge.
Lee, C. D. and A. Spratley (2009). *Reading in the disciplines and the challenges of adolescent literacy.* NY, Carnegie Foundation of New York.
Leu, D. J., Kinzer, C. K., Coiro, J., Castek, J., & Henry, L. A. (2017). New literacies: A dual-level theory of the changing nature of literacy, instruction, and assessment. *Journal of Education, 197*(2), 1-18.
Linn and Dunbar 1992
Lippman, L. NY, Springer: 237-270.
List A., & Alexander, P. A. (2018). Cold and warm perspectives on the Cognitive Affective Engagement Model of Multiple Source Use. In J. L. Braasch, I. Bråten, & M. T., McCrudden (Eds.), *Handbook of multiple source use* (pp. 46-66). New York: Routledge.
Mahoney, J.L., Durlak, J.A., & Weissberg, R.P. (2018). An update on social and emotional learning outcome research. *Phi Delta Kappan, 100*(4), 18–23.
Manderino, M. (2012). Disciplinary literacy in new literacies environments: Expanding the intersections of literate practice for adolescents. In P. Dunston, L. Gambrell, K. Headley,

S. Fullerton, & P. Stecker (Eds.), *Sixty-first yearbook of the literacy research association* (pp. 69–83). Oak Creek, WI: Literacy Research Association.
Markman, E. M. (1979). Realizing that you don't understand: Elementary school children's awareness of inconsistencies. *Child Development, 50*(3), 643-655, DOI: 10.2307/1128929
Marzano, R. J., Brandt, R. S., Hughes, C. S., Jones, B. F., Presseisen, B. Z., Rankin, S. C., & Suhor, C., (XXX). Dimensions of thinking: A framework for curriculum and instruction. Alexandria, VA: Association for Supervision and Curriculum Development.
Massey, D. D. (2009). Self-regulated comprehension. In S. E. Israel & G. D. Duffy (Eds.), *Handbook of research on reading comprehension* (pp. 389-399). New York: Routledge.
McNamara, D. S. and W. Kintsch (1996). "Learning from text: Effects of prior knowledge and text coherence." Discourse Processes 22: 247-287.
Measured Progress/ETS Collaborative. (2012). *Smarter Balanced Assessment Consortium: Technology-enhanced item guidelines.* https://www.measuredprogress.org/wp-content/uploads/2015/08/SBAC-Technology-Enhanced-Items-Guidelines.pdf
Meyer, B. J. F. (1975). Identification of the structure of prose and its implications for the study of reading and memory. *Journal of Reading Behavior, 7,* 7-47.
Michalsky, T., Mevarech, Z. R., Haibi, L. (2009). Elementary school children reading scientific texts: Effects of metacognitive instruction. *Journal of Educational Research, 102*(5), 363-376, DOI: 10.3200/JOER.102.5.363.376
Mislevy, R. J. (2016). How development in psychology and technology challenge validity argumentation. *Journal of Educational Measurement, 53(3),* 265-292.
Moje, E. B. (2000). "To be part of the story": The literacy practices of gangsta adolescents. *Teachers College Record, 102*(3), 651–690.
Moje, E. B. (2009). A call for new research on new and multi-literacies. *Research in the Teaching of English, 43*(4), 348-362.
Moje, E. B. (2015). Doing and teaching disciplinary literacy with adolescent learners: A social and cultural enterprise. *Harvard Educational Review, 85*(2), 254–278. doi:10.17763/0017-8055.85.2.254
Moll, L. C. (2014). L. S. *Vygotsky and education.* New York, NY: Routledge.
Moll, L. C., & Gonzalez, N. (1994). Lessons from research with language-minority children. *Journal Of Reading Behavior, 26*(4), 439–456.
Moll, L. C., Amanti, C., Neff, D., & Gonzalez, N. (1992). Funds of knowledge for teaching: Using a qualitative approach to connect homes and classrooms. *Theory into Practice*, *31*(2), 132-141.
Mullis, I. V. S., & Martin, M. O. (2019). *Chapter 1 PIRLS 2021 Reading assessment framework.* Boston: TIMSS & PIRLS International Study Center.
Mullis, I. V. S., & Martin, M. O. (Eds.). (2019). *PIRLS 2021 assessment frameworks.* Boston, MA: International Association for the Evaluation of Education Achievement, & TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College. http://pirls2021.org/wp-content/uploads/sites/2/2019/04/P21_Frameworks.pdf
Mullis, I.V.S., Martin, M.O., Foy, P., & Drucker, K.T. (2012). Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College
NAEP Nation's Report Card, 2019
NAEP Reading Special Study, 2019

Nagy, W., & Townsend, D. (2012). Words as tools: Learning academic vocabulary as language acquisition. *Reading Research Quarterly, 47* (1), 91–108. doi: 10.1002/RRQ.011
Nasir, N. S., & Hand, V. M. (2006). Exploring sociocultural perspectives on race, culture, and learning. *Review of Educational Research, 76*(4), 449-475, doi.org/10.3102/0034655430/6004449
National Academies of Sciences, Engineering, and Medicine. (2018). *How People Learn II: Learners, Contexts, and Cultures*. Washington, DC: The National Academies Press. https://doi.org/10.17226/24783.
National Academy of Sciences. (2018). *The science of science communication III: Inspiring novel collaborations and building capacity: Proceedings of a colloquium.* Washington, DC: The National Academies Press, doi.org/10.17226/24958
National Assessment Governing Board, 2014
National Assessment Governing Board. (2017). *Reading framework for the 2017 National Assessment of Educational Progress.* Washington, DC: National Assessment Governing Board and U.S. Department of Education. https://www.nagb.gov/content/nagb/assets/documents/publications/frameworks/reading/2017-reading-framework.pdf
National Assessment Governing Board. (2019). Reading Framework of the 2019 National Assessment of Educational Progress. Washington, DC: Author
National Center for Education and the Economy [NCEE]. (2013). What does it really mean to be college and workforce ready? The English literacy required of first year community college students. Washington, DC: NCEE. Retrieved June 20, 2018 from http://ncee.org/wp-content/uploads/2013/05/NCEE _EnglishReport_May2013.pdf.
National Center for Education Statistics. (2018). Going digital: NAEP assessments for the future. *National Assessment of Educational Progress.* Washington, D.C.: U.S. Department of Education, Institute of Education Sciences. https://nces.ed.gov/nationsreportcard/subject/dba/pdfs/2018_dba_brochure.pdf
National Center for Education Statistics. (2020). Digitally based assessments. *National Assessment of Educational Progress.* Washington, D.C.: U.S. Department of Education, Institute of Education Sciences. https://nces.ed.gov/nationsreportcard/dba/
National Research Council 2017?
National Research Council, 2018?
National Research Council. (2009). *Learning science in informal environments: People, places, and pursuits.* Washington, DC: The National Academies Press.
National Research Council. (2010). *Language Diversity, School Learning, and Closing Achievement Gaps: A Workshop Summary.* Washington, DC: The National Academies Press.
New London Group. (1996). A pedagogy of multiliteracies: Designing social futures. *Harvard Educational Review, 66*, 60–92.
Newell, G. E., Beach, R., Smith, J., VanDerHeide, J., Kuhn, D., & Andriessen, J. (2011). Teaching and learning argumentative reading and writing: A review of research. *Reading Research Quarterly, 46*(3), 273–304.
O'Reilly T., Wang, J., Sabatini, J. (2019). How Much Knowledge Is Too Little? When a Lack of Knowledge Becomes a Barrier to Comprehension Psychological. *Science*, Vol. 30(9) 1344–1351

O'Reilly, T., & McNamara, D. S. (2007). The impact of science knowledge, reading skill, and reading strategy knowledge on more traditional "high-stakes" measures of high school students' science achievement. American Educational Research Journal, 44, 161–196.
O'Reilly, T., Wang, Z., & Sabatini, J. (2019). How much knowledge is too little? When a lack of knowledge becomes a barrier to comprehension. Psychological Science, 30(9), 1344–1351.
OECD. (2019). *PISA 2018 assessment and analytical framework.* Paris: OECD Publishing. https://doi.org/10.1787/b25efab8-en
Organisation for Economic Co-operation and Development. (2019). *Education at a glance 2019: OECD indicators. Paris, FR: OECD.*
Osher, D., Cantor, P., Berg, J., Steyer, L. & Rose, T. (2020). Drivers of human development: How relationships and context shape learning and development. *Applied Developmental Science, 24*(1), 6-36, DOI: 10.1080/10888691.2017.1398650
Pacheco, M. (2015). Bilingualism-as-participation: Examining adolescents' bi(multi)lingual literacies across out-of-school and online contexts. In D. Molle, E. Sato, T. Boals, & C. Hedgespeth (Eds.), *Multilingual learners and academic literacies: Sociocultural contexts of literacy development in adolescents* (p. 135- 165). New York: Routledge.
Pacheco, M. (2018). Learning and becoming writers: Meaning, identity, and epistemology in a newsroom community of practice. *Mind, Culture, and Activity*, 25(2), 105-124.
Pacheco, M. B., & Miller, M. E. (2016). Making meaning through translanguaging in the literacy classroom. *The Reading Teacher, 69*(5), 533-537.
Pacheco, M. B., & Smith, B. E. (2015). Across languages, modes, and identities: Bilingual adolescents' multimodal codemeshing in the literacy classroom. *Bilingual Research Journal, 38*(3), 292-312.
Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research, 66*(4), 543–578, doi.org/10.3102/0034654306600543
Paris, D. (2009). "They're in my culture, they speak the same way": African American Language in multiethnic high schools. *Harvard Educational Review, 79*(3), 428–448.
Paris, S. G., Wasik, B. A., & Turner, J. C. (1991). The Development of Strategies of Readers. In R. Barr, M. Kamil, P. Mosenthal, & P. D. Pearson (Eds.), *Handbook of Reading Research* (Vol. 2, pp. 609-640). Mahwah, NJ: Lawrence Erlbaum Associates.
Pearson, P. D. and E. H. Hiebert (2014). "The State of the Field: Qualitative Analyses of Text Complexity." Elementary School Journal.
Pearson, P. D., & Camparell, K. (1981). Comprehension of text structures. In J. Guthrie (Ed.), *Comprehension and teaching* (pp. 27-54). Newark DE: International Reading Association.
Pearson, P. D., & Cervetti, G. N. (2015). Fifty years of reading comprehension theory and practice. In P. D. Pearson, E. H. Hiebert,& N. K. Duke (Eds.), *Research-based practices for teaching Common Core literacy* (pp. 1-24). New York: Teachers College Press.
Pearson, P. D., & Hiebert, E. (2014). The state of the field. *The Elementary School Journal, 115*(2), 161-183.
Pearson, P. D., Hansen, J., & Gordon (1979). The effect of background knowledge on young children's comprehension of explicit and implicit information. Journal of Reading Behavior, 11, 201-210.

Pearson, P. D., Palincsar, A. S., Biancarosa, G., & Berman, A. (2020). *Reaping the rewards of the Reading for Understanding Initiative.* Washington, DC: National Academy of Education.

Pearson, P. D., Valencia, S., & Wixson, K. (2014). Complicating the world of reading assessment: Better assessments for better teaching. *Theory into Practice, 53*(3), 236-246. DOI: 10.1080/00405841.2014.916958

Pérez, C. C. (2017). *The first rule of punk.* New York: Penguin Random House.

Peura, P., Viholainen, H. J. K., Aro, T. I., Raikkonen, E. M., Usher, E. L., Sorvo, R. M. A., Klassen, R. M. & Aro**,** M. T; (2019). *Specificity of reading self-efficacy among primary school children. Journal of Experimental Education, 87,* 496-516, doi.org/10.1080/00220973.2018.1527279

Phillips Galloway, E., Uccelli, P., Aguilar, G. & Barr, C. (2020) Exploring the cross-linguistic contribution of Spanish and English academic language skills to English text comprehension for middle-grade dual language learners. AERA Open, 1, 1-20. doi: 10.1177/2332858419892575

Pintrich, P. R. and B. Schrauben (1992). Students' motivational beliefs and their cognitive engagement in classroom academic tasks. Student perceptions in the classroom: Causes and consequences D. Schunk and J. Meece. Hillsdale, NJ, Erlbaum: 149-183.

Pintrich, P. R. and B. Schrauben (1992). Students' motivational beliefs and their cognitive engagement in classroom academic tasks. Student perceptions in the classroom: Causes and consequences D. Schunk and J. Meece. Hillsdale, NJ, Erlbaum: 149-183.

Pitzer, J & Skinner, E. (2017). Predictors of changes in students' motivational resilience over the school year: The roles of teacher support, self-appraisals, and emotional reactivity. *International Journal of Behavioral Development*, 41(1) 15–29.

Pressley, M., & Afflerbach, P. (1995). *Verbal protocols of reading: The nature of constructively responsive reading.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Proctor, C. P., Silverman, R.D., Harring, J. R., Jones, R.L., & Hartranft, A. M. (2020) Teaching Bilingual Learners: Effects of a Language-Based Reading Intervention on Academic Language and Reading Comprehension in Grades 4 and 5 *Reading Research Quarterly 55* (1), 95-122 doi: 10.1002/rrq.258

Purcell-Gates, V., Duke, N.K., & Stouffer, J. (2016). Teaching literacy: Reading. In D.H. Gitomer & C.A. Bell (Eds.), *The AERA handbook of research on teaching* (5th ed.) (pp. 1217– 1267). Washington, DC: American Educational Research Association.

RAND Reading Study Group. (2002). *Reading for understanding: Toward an R&D program in reading comprehension.* Santa Monica, CA: RAND. https://www.rand.org/pubs/monograph_reports/MR1465.html

Rose, DH, & Meyer, A (2002) *Teaching every student in the digital age: Universal design for learning.* Alexandria, VA: Association for Supervision and Curriculum Development.

Roth, W. M., & Lee, Y.-J. (2007). "Vygotsky's neglected legacy": Cultural-historical ctivity theory. *Review of Educational Research*, *77*(2), 186–232, doi.org/10.3102/0034654306298273

Sabatini, J. P., O'Reilly, T., Halderman, L. K., & Bruce, K. (2014). Integrating scenario-based and component reading skill measures to understand the reading behavior of struggling readers. *Learning Disabilities Research & Practice, 29*(1), 36–43.

Sabatini, J., et al. (2018). Scenario-based assessment of multiple source use. Handbook of Multiple Source Use, Routledge: 447-465.

Sabatini, J., O'Reilly, T., Weeks, J., & Wang, Z. (2020). Engineering a twenty-first century reading comprehension assessment system utilizing scenario-based assessment techniques. *International Journal of Testing*, *20*(1), 1-23. https://doi.org/10.1080/15305058.2018.1551224

Sanchez, I. G., & Orellana, M. F. (2006). The construction of moral and social identity in immigrant children's narratives-in-translation. *Linguistics And Education: An International Research Journal, 17*(3), 209–239.

Schaffner, E., Schiefele, U., & Ulferts, H. (2013). Reading Amount as a Mediator of the Effects of Intrinsic and Extrinsic Reading Motivation on Reading Comprehension. *Reading Research Quarterly*, 48, 369-385.

Schallert, D. L. (2002). Schema theory. In B. J. Guzzetti (Ed.), *Literacy in America: An encyclopedia of history, theory, and practice* (pp. 556-558). Santa Barbara, CA: ABC-Clio.

Schleppegrell, M.J. (2004). *The language of schooling: A functional linguistics perspective.* Mahwah, NJ: Erlbaum.

Schultz, K. (2002). Looking across space and time: Reconceptualizing literacy learning in and out of school. Research in the Teaching of English, 36(3), 356–390.

Scribner, S. and Cole, M (1981). *The Psychology of Literacy*. Cambridge, Mass.: Harvard University Press.

Scribner, S., & Cole, M. (1978). Literacy without schooling: Testing for intellectual effects. *Harvard Educational Review, 48*(4), 448–461.

Serafini, F. (2004). Images of reading and the reader. *The Reading Teacher, 57*(7), 610-617.

Shin, R. Q., Daly, B. P., & Vera, E. M. (2007). The relationships of peer norms, ethnic identity, and peer support to school engagement in urban youth. *Professional School Counseling*, 10, 379– 388.

Singer, L. M., & Alexander, P. A. (2017). Reading across mediums: Effects of reading digital and print texts on comprehension and calibration. *The Journal of Experimental Education, 85*(1), 155-172.

Singer, L. M., & Alexander, P. A. (2017). Reading on paper and digitally: What the past decades of empirical research reveal. *Review of Educational Research, 87*(6), 1007-1041, doi.org/10.3102/0034654317722961

Skerrett, A. (2011). English teachers' racial literacy knowledge and practice. *Race, Ethnicity, and Education*, 14(3), 313-330.

Skerrett, A. (2012). Social and Cultural Differences in Reading Development 36 Skerrett, A. "We hatched in this class": Repositioning of identity in and beyond a reading classroom. *The High School Journal*, 95(3), 62-75.

Skerrett, A. (2015). *Teaching transnational youth: Literacy and education in a changing world*. New York, NY: Teachers College Press.

Skerrett, A. (2020). Social and cultural differences in reading development: Instructional approaches, learning gains, and challenges. In E. B. Moje, P. Afflerbach, P. Enciso, and N. Lesaux. (Eds.). *Handbook of reading research* (Vol. V) (pp. 328-344). New York, NY: Routledge.

Skerrett, A. (in press) *Handbook of Reading Research*. Social and Cultural Differences in Reading Development: Instructional Processes, Learning Gains, and Challenges

Slemmons, K., Anyanwu, K., Hames, J., Grabski, D., Mlsna, J., Simkins, E., & Cook, P. (2018). The impact of video length on learning in a middle-level flipped science setting:

implications for diversity inclusion. Journal of Science Education and Technology, 27(5), 469-479.

Smagorinsky, P. (2001). If meaning is constructed, what is it made from?: Toward a cultural theory of reading. *Review of Educational Research, 71*(2), 133–169.

Snow, C. E., & Uccelli, P .(2009). The challenge of academic language. In Olson, D. R., & N. Torrance (Eds.), *The Cambridge handbook of literacy* (pp.112-133). Cambridge: Cambridge University Press.

Solano-Flores, G. (2011). Assessing the cultural validity of assessment practices: An introduction. In M. del Rosario Basterra, E. Trumbull, and G. Solano-Flores (Eds.), *Cultural validity in assessment* (pp. 19-37). New York: Routledge.

Solano-Flores, G., & Nelson-Barber, S. (2001). On the cultural validity of science assessments. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, *38*(5), 553-573.

Sparks, J. R., & Deane, P. (2014). *Cognitively-based assessment of research and inquiry skills: Defining a key practice in the English language arts*. Manuscript in preparation.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21*(4), 360–407, doi.org/10.1598/RRQ.21.4.1

Stanovich, K. E. (2013). Why humans are (sometimes) less rational than other animals: Cognitive complexity and the axioms of rational choice. *Thinking and Reasoning, 19*(1), 1-26, DOI: 10.1080/13546783.2012.713178

Stenner, A.J. (1996). Measuring reading comprehension with the Lexile framework. Durham, NC: Metametrics Inc.

Street, B. V. (1984). *Literacy in theory and practice.* NY: Cambridge University Press.

Tatum, A. W. (1999). Reading and the African American male: Identity, equity, and power. *Journal of Adolescent & Adult Literacy, 43*(1), 62–64.

Taylor, R. D., Oberle, E., Durlak, J. A. & Weissberg, R. (2017). Promoting positive youth development through school-based social and emotional learning interventions: A meta-analysis of follow-up effects. *Child Development*, *88*(4), 1156–1171.

The Council of Chief State School Officers and The National Governors Association Center for Best Practices (2010). *Common core state standards for English language arts and literacy in history/social studies & science*. http://www.corestandards.org/Standards/K12/

Tomasello, M. (2016). The ontogeny of cultural learning. *Current Opinion in Psychology*, *8*, 1–4, doi:10.1016/j.copsyc.2015.09.008

Torppa, Eklund, Sulkunen, Niemi & Ahonen, 2018

Uccelli, P., & Phillips Galloway, E. (2017). Academic language across content areas: Lessons from an innovative assessment and from students' reflections about language. *Journal of Adolescent & Adult Literacy, 60*(4), 395-404.doi:10.1002/jaal.553

Uccelli, P., Barr, C. D., Dobbs, C. L., Phillips Galloway, E., Meneses, A., & Sánchez, E. (2015). Core academic language skills (CALS): An expanded operational construct and a novel instrument to chart school-relevant language proficiency in pre-adolescent and adolescent learners. *Applied Psycholinguistics, 36,* 1077-1109. doi:10.1017/S014271641400006X

Uccelli, P., Galloway, E. P., Barr, C. D., Meneses, A., & Dobbs, C. L. (2015). Beyond vocabulary: Exploring cross-disciplinary academic-language proficiency and its association with reading comprehension. *Reading Research Quarterly, 50*(3), 337-356, doi.org/ 10.1002/rrq.104

Valencia, S. W., Wixson, K. K., & Pearson, P. D. (2014). Putting text complexity in context: Refocusing on comprehension of complex text. *Elementary School Journal*, *115*(2), 270-289.

van den Broek, P., Bohn-Gettler, C. M., Kendeou, P., Carlson, S., & White, M. J. (2011). When a reader meets a text: The role of standards of coherence in reading comprehension. In M. T. McCrudden, J. P. Magliano, & G. Schraw (Eds.), *Text relevance and learning from text* (p. 123–139). IAP Information Age Publishing.

van Dijk, T. A. & Kintsch, W. (1983). *Strategies of discourse comprehension.* New York, Academic Press.

Vaughn, S., Solis, M., Miciak, J., Taylor, W. P., & Fletcher, J. M. (2016). Effects from a randomized control trial comparing researcher and school-implemented treatments with fourth graders with significant reading difficulties. *Journal of Research on Educational Effectiveness, 9*(1), 23-44, doi.org/10.1080/19345747.2015.1126386

Vaux, A., Phillips, J., Holly, L., Thompson, B., Williams, D., & Steward, D. (1986). The social support appraisals scale: Studies of reliability and validity. *American Journal of Community Psychology,* 14, 195–219. doi:10.1007/BF00911821.

Veenman, M. V. J., Van Hout-Wolters, B. H. A. M., & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. Metacognition and Learning, 1, 3-4, doi.org/10.1007/s11409-006-6893-0

Vygostky, L. (1978). *Mind in society: The development of higher psychological processes.* Boston: Harvard University Press.

Wang, Z., Sabatini, J., O'Reilly, T., & Weeks, J. (2019). Decoding and reading comprehension: A test of the decoding threshold hypothesis. *Journal of Educational Psychology, 1*1(3), 387–401.

Wever & Kelemen, 1997 Weaver, C. A., & Kelemen, W. L. (1997). Judgments of Learning at Delays: Shifts in Response Patterns or Increased Metamemory Accuracy? *Psychological Science*, *8*(4), 318–321, doi.org/10.1111/j.1467-9280.1997.tb00445.x

Wexler, N. (April 13, 2018). Why American students haven't gotten better at reading in 20 years. *The Atlantic.* Downloaded from https://www.theatlantic.com/education/archive/2018/04/-american-students-reading/557915/

Wigfield, A., & Wentzel, K. R. (2007). Introduction to motivation at school: Interventions that work. *Educational Psychologist, 42*(4), 191–196.

Wigfield, A., Gladstone, J., & Turci, L. (2016). Beyond cognition: Reading motivation and reading comprehension. *Child Development Perspectives 10*(3), 190-195*,* DOI: 10.1111/cdep.12186

Wigfield, A., Guthrie, J. T., Perencevich, K. C., Barber, A. M. T., Klauda, S., McRae, A., & Barbosa, P. (2008). Role of reading engagement in mediating effects of reading comprehension instruction on reading outcomes. *Psychology in Schools, 45*(5), 432-445, DOI: 10.1002/pits.20307

Wixson, K. K., & Peters, C. W. (1987). Comprehension assessment: Implementing an interactive view of reading. *Educational Psychologist, 22*, 333-356.

Wu, JY (2013) Gender differences in online reading engagement, metacognitive strategies, navigation skills and reading *Journal of Computer Assisted Learning*

Yeari, M., & van den Broek, P. (2011). A cognitive account of discourse understanding and discourse interpretation: The Landscape Model of reading. *Discourse Studies, 13*(5), 635-643, 10.1177/1461445611412748

Zelazo, P. D. (2013). Developmental psychology: A new synthesis. In P. D. Zelazo (Ed.), *The Oxford handbook of developmental psychology* (Vol. 1: Body and Mind). https://doi.org/10.1093/oxfordhb/9780199958450.013.0001

Page iii: [1] Deleted	**russ**	**5/1/2021 9:02:00 AM**
Page iii: [2] Deleted	**russ**	**5/1/2021 9:02:00 AM**
Page iii: [3] Deleted	**russ**	**5/1/2021 9:02:00 AM**
Page iii: [4] Deleted	**russ**	**5/1/2021 9:02:00 AM**
Page iii: [5] Deleted	**russ**	**5/1/2021 9:02:00 AM**
Page iii: [6] Deleted	**russ**	**5/1/2021 9:02:00 AM**
Page iii: [7] Deleted	**russ**	**5/1/2021 9:02:00 AM**
Page iii: [8] Deleted	**russ**	**5/1/2021 9:02:00 AM**
Page iii: [9] Deleted	**russ**	**5/1/2021 9:02:00 AM**
Page iii: [10] Deleted	**russ**	**5/1/2021 9:02:00 AM**
Page iii: [11] Deleted	**russ**	**5/1/2021 9:02:00 AM**
Page iii: [12] Deleted	**russ**	**5/1/2021 9:02:00 AM**
Page iii: [13] Deleted	**russ**	**5/1/2021 9:02:00 AM**
Page iii: [14] Deleted	**russ**	**5/1/2021 9:02:00 AM**
Page iii: [15] Deleted	**russ**	**5/1/2021 9:02:00 AM**
Page iii: [16] Deleted	**russ**	**5/1/2021 9:02:00 AM**
Page iii: [17] Deleted	**russ**	**5/1/2021 9:02:00 AM**
Page iii: [18] Deleted	**russ**	**5/1/2021 9:02:00 AM**
Page iii: [19] Deleted	**russ**	**5/1/2021 9:02:00 AM**
Page iii: [20] Deleted	**russ**	**5/1/2021 9:02:00 AM**

Page iii: [21] Deleted	**russ**	**5/1/2021 9:02:00 AM**
Page iii: [22] Deleted	**russ**	**5/1/2021 9:02:00 AM**
Page iii: [23] Deleted	**russ**	**5/1/2021 9:02:00 AM**
Page iii: [24] Deleted	**russ**	**5/1/2021 9:02:00 AM**
Page iii: [25] Deleted	**russ**	**5/1/2021 9:02:00 AM**
Page iii: [26] Deleted	**russ**	**5/1/2021 9:02:00 AM**
Page iii: [27] Deleted	**russ**	**5/1/2021 9:02:00 AM**
Page iii: [28] Deleted	**russ**	**5/1/2021 9:02:00 AM**
Page iii: [29] Deleted	**russ**	**5/1/2021 9:02:00 AM**
Page 1: [30] Deleted	**russ**	**4/30/2021 5:06:00 PM**
Page 3: [31] Deleted	**russ**	**4/30/2021 5:17:00 PM**
Page 8: [32] Deleted	**russ**	**4/30/2021 5:31:00 PM**
Page 9: [33] Deleted	**russ**	**4/30/2021 5:34:00 PM**
Page 10: [34] Deleted	**russ**	**4/30/2021 5:38:00 PM**
Page 18: [35] Deleted	**russ**	**4/30/2021 5:48:00 PM**
Page 41: [36] Deleted	**russ**	**5/2/2021 6:25:00 PM**
Page 62: [37] Deleted	**russ**	**5/1/2021 7:40:00 AM**

Orientation to edits in the draft reading framework

The edits in the associated document do not require any changes in the plans for the 2026 reading assessment. Rather they address sections of the narrative that have proved divisive and they shorten the document considerably to make it more accessible to the general reader. A high priority for the edits was to avoid putting words in the mouth of the development panel. The goal was editorial: to clarify and to adjust content that critics of the framework have found objectionable.

Substantive edits, whether by addition, modification, or deletion of original text, address one or more of following goals:

- To reframe the socio-cultural perspective that is elevated to conceptual preeminence by the development panel so that it describes an important set of variables that deserve more attention in NAEP rather than an all-encompassing point of view

  The edits treat the socio-cultural "theory" as referring to sources of important environmental variables (family, home, and school) that will be better measured and reported under the new framework. This reframing allows appropriate attention to variables that are in the socio-cultural sphere without privileging them with respect to the many other categories of variables that influence reading comprehension, e.g., curriculum and instruction, teacher quality, social media, individual differences in neuro-cognitive processing, background knowledge, and so forth. It avoids forcing NAGB and users of NAEP to accept a particular point of view of what is most important in learning to read. And it allows NAGB to steer clear of the politically charged and divisive issue of whether our nation's educational policies should support a salad bowl of socio-culturally distinct outcomes or a common core of shared knowledge and understandings.

- To acknowledge that a reader's background knowledge is a critical component of reading comprehension while emphasizing that the assessment should avoid items on which student answers are significantly affected by background knowledge that most students being tested would not have had an opportunity to acquire

  We have spent hours as a full board on debates about how to handle background knowledge. The framework authors want to conceptualize background knowledge as separate from reading comprehension. But as the David Pearson, the chair of the development panel, said in the NAGB webinar of April 29th, "background knowledge is to some degree always a factor in reading comprehension". The goal, then, is to handle background knowledge in ways that strengthen the validity of the assessment, rather than trying to define it out of existence as a factor in reading comprehension.

  One way to do this is exactly as background knowledge has been handled in previous NAEP assessments and is handled in nearly every state assessment -- assure that text-independent knowledge relevant to a particular test item is shared by most test takers. For example, item developers should not be hesitant to include references to melting icebergs in a climate change passage in the science portion of the reading assessment of eighth graders -- text-independent knowledge of what an iceberg is can be shown to be very common knowledge in much younger children. But comprehension of the passage should not depend on the text-independent knowledge

of types of icebergs because few eighth graders will have had the opportunity to learn how a Glacier berg is different from other types.

- To remove references to questionnaire items and context variables that require that students report on their personal beliefs and attitudes. The NAEP law (Sec 303(c)(5)) specifically requires that the tests "not evaluate or assess personal ... beliefs and attitudes"

  The clearest violation of the NAEP prohibition on assessing personal beliefs in the draft framework is the intent to add a measure of student self-efficacy to the context variables. Self-efficacy is tested by asking individuals about their belief in their capabilities to organize and execute courses of action. For example, students taking NAEP might be asked how strongly they agree with the following statement: "I believe I can succeed at almost any endeavor to which I set my mind." Such assessment items are without doubt assessments of personal beliefs. As such they are proscribed by the NAEP law.

  Where the violation is obvious the relevant text has been deleted. Where it is borderline, e.g., questions to students about the support they perceive they receive from their teachers, the text has been rephased to describe the testing activity as inquiring about what student's observe rather than about their beliefs and perceptions.

- To eliminate assertions and suggestions that the addition of new context variables will better explain group differences in NAEP scores

  Correlation is not causation. Assertions by the authors of the draft framework that additional context variables will provide better explanations of NAEP results are false. For example, assume that there is a healthy correlation between students reading scores and new NAEP questionnaire items that ask test takers how much support they get from their teachers. Assume further that students from lower income homes report less of such support than students from higher income families. The draft framework suggests that consumers of NAEP data would then have an explanation, or "deeper understanding", of why students from low-income families don't read as well as those from higher income families – they don't get as much support from teachers.

  But run-of-the-mill students in any college research methods course could easily dismantle this conclusion and its underlying assumptions about the requirements for making causal claims. Just one of many issues is the chicken and the egg problem: we can't know from NAEP data whether teachers give more attention to good readers because they are good readers or whether students are good readers because they get more attention from teachers.

  Conclusions about what causes differences in children's education outcomes require experiments whenever possible. Weaker methods may produce useful information when experiments are not possible. But correlations among variables obtained at a single point in time from a snapshot assessment do not pass any responsible threshold for support of conclusions about what causes differences in NAEP scores. The Education Sciences Reform Act, of which the NAEP law is a part, requires that dissemination of scientifically valid statistics by the Institute of Education Sciences present "findings and makes claims that are appropriate to and supported by the methods that have been employed." The treatment of context variables as explanatory in the draft framework is a flagrant violation of this.

List of Whitehurst Proposed Changes and Project Officer Notes

May 5, 2021

This document summarizes edits recommended by Board member Russ Whitehurst, as transmitted to Board Chair Haley Barbour on May 3, 2021. The purpose of this document is to clarify how the Project Officer would need to direct the Development Panel (on behalf of the Board) to execute these edits, if directed to do so by the Governing Board. A copy of the edits recommended by Whitehurst are attached here.*

*A copy that includes the project officer notes listed below is available upon request.

No.	Recommended Edit	Project Officer Note
1.	Delete section describing the current assessment (Chapter 1)	The text (suggested for deletion) describes the current assessment as implemented by NCES and aligned to the current framework. It demonstrates continuity between the current assessment and the proposed updates, as requested in the 2019 Board-adopted charge to the Panel and as requested again in July-December 2020 feedback from the Governing Board and the Summer 2020 public comment period.
2.	Revise text describing the Board-adopted charge (Chapter 1)	The revised text is inconsistent with the 2019 Board-adopted charge. It also uses the term "proficiently" in a way that does not align with the policy definition for the NAEP Proficient achievement level.
3.	Delete reference to the Visioning Panel's guidelines for the framework update and the research base prompting these guidelines (Chapter 1)	• The text (suggested for deletion) lists the guidelines from the 33-person Visioning Panel to the 17-person Development Panel. These guidelines are typically provided in NAEP framework documents, e.g., see most recently adopted NAEP frameworks for Mathematics, Technology and Engineering Literacy, and Science. • The text (suggested for deletion) cites professional standards for testing and a consensus report from the National Academies. • References to equity and fairness are deleted.
4.	Delete summary of research base supporting the current NAEP Reading Framework and the proposed NAEP Reading Framework update (Chapter 1)	The text (suggested for deletion) demonstrates continuity between the current assessment and the proposed updates, as requested in the 2019 Board-adopted charge to the Panel and as requested again in July-December 2020 feedback from the Governing Board and the Summer 2020 public comment period.

No.	Recommended Edit	Project Officer Note
5.	Delete reference to how cognitive processes of reading relate to social and cultural influences (Chapter 1)	The text (suggested for deletion) indicates why the NAEP definition of reading comprehension was revised.
6.	Rephrase the NAEP Definition of Reading Comprehension to replace social and cultural influences with synonymous concrete terminology (Chapter 1)	The revised text is an editorial change.
7.	Delete reference to how cognitive processes of reading relate to social and cultural influences (Chapter 1)	The text (suggested for deletion) demonstrates continuity between the current assessment and the proposed updates, as requested in the 2019 Board-adopted charge to the Panel and as requested again in July-December 2020 feedback from the Governing Board and the Summer 2020 public comment period.
8.	Rephrase references to sociocultural aspects of reading with synonymous terminology (Chapter 1)	The revised text is an editorial change.
9.	Delete recommendations for contextual variables (Chapter 1)	The contextual variables (suggested for deletion) have already started to be reported on via NAEP questionnaires. See NAEP Website.
10	Delete references to increased measurement precision (Chapter 1)	The revised text is less specific than the previous text.
11	Delete reference to equity, non-bias, validity, and Universal Design of Assessments (Chapter 1)	The text (suggested for deletion) explains the basis for updates proposed in the April 21, 2021, version of the draft Framework (referred to as "proposed updates" hereafter).
12	Delete reference to Universal Design Elements (Chapter 1)	The text (suggested for deletion) lists a major component of the framework update alongside other components of the framework update.
13	Delete reference to equity, rigor, precision, and validity (Chapter 1)	The text (suggested for deletion) explains the basis for proposed updates.
14	Delete reference to socioeconomic status within race/ethnicity as a feature of NAEP reporting (Chapter 1)	This is a core reporting recommendation from the Panel to improve NAEP Reporting.
15	Delete recommendations for contextual variables (Chapter 1)	The contextual variables (suggested for deletion) have already started to be reported on via NAEP questionnaires. See NAEP Website.
16	Delete reference to how cognitive processes of reading relate to social and cultural influences (Chapter 1)	The text (suggested for deletion) indicates why the NAEP definition of reading comprehension was revised.
17	Added text to constrain use of pop-up definitions (Chapter 1)	Specifying pop-up notes for rare words only would be a new requirement. Pop-up notes are not proposed for words that are part of the comprehension target being tested.

No.	Recommended Edit	Project Officer Note
18	Added text to elaborate on when/how socioeconomic status information can be reported (Chapter 1)	Based on attendance at Development Panel meetings, this was implied in the Panel's initial recommendation. The added text is more specific/explicit.
19	Rephrase the NAEP Definition of Reading Comprehension to replace social and cultural influences with synonymous concrete terminology (Chapter 2)	The revised text is an editorial change.
20	Delete definitions of key terms in the NAEP Definition of Reading Comprehension (Chapter 2)	The revised text is less specific than the previous text.
21	Moved text describing the importance of reader's knowledge (Chapter 2)	This text was moved from an earlier section of the chapter.
22	Delete reference to how cognitive processes of reading relate to social and cultural influences (Chapter 2)	The text (suggested for deletion) indicates why the NAEP definition of reading comprehension was revised.
23	Delete reference to how cognitive processes of reading relate to social and cultural influences (Chapter 2)	The text (suggested for deletion) indicates why the NAEP definition of reading comprehension was revised.
24	Delete references to research and assessments that relate to sociocognitive processes (Chapter 2)	The text (suggested for deletion) indicates why the NAEP definition of reading comprehension was revised.
25	Rephrase references to precision of inferences from NAEP (Chapter 2)	The revised text is less specific than the previous text.
26	Delete reference to Universal Design Elements (Chapter 2)	The text (suggested for deletion) lists a major component of the framework update alongside other components of the framework update.
27	Delete one factor related to reader experiences (Chapter 2)	This text (suggested for deletion) represents the consensus of the Visioning and Development Panels.
28	Delete reference to contextualizing the assessment (Chapter 2)	The text (suggested for deletion) explains rationales for the proposed assessment updates.
29	Delete reference to Universal Design Elements (Chapter 2)	The text (suggested for deletion) lists a major component of the framework update alongside other components of the framework update.
30	Rephrase references to previous special reports issued by NAEP (Chapter 2)	The revised text is less specific than the previous text.
31	Delete recommendations for contextual variables and related reporting (Chapter 2)	The contextual variables (suggested for deletion) have already started to be reported on via NAEP questionnaires. See NAEP Website.
32	Rephrase benefits of reporting recommendations (Chapter 2)	The revised text is an editorial change.
33	Delete recommendations for contextual variables and process data (Chapter 2)	The contextual variables (suggested for deletion) have already started to be reported on via NAEP questionnaires. See NAEP Website. The same goes

No.	Recommended Edit	Project Officer Note
		for process data, which NAEP has also started to include in recent report cards.
34	Delete the recommendations that peers might serve in an assessment context (Chapter 2)	This text (suggested for deletion) represents the consensus of the Visioning and Development Panels.
35	Delete references to validity, fairness, and inclusivity (Chapter 3)	The text (suggested for deletion) explains the basis for proposed updates.
36	Moved text describing how disciplinary contexts and purposes will be integrated into blocks (Chapter 3)	This text was moved from an earlier section of the chapter. It is also shortened.
37	Delete note that video may be included as an assessment component (Chapter 3)	This text removes video from any part of the assessment. (Video is already removed from passage introductions in the April 2021 draft. Video does appear in other parts of the current NAEP Reading Assessment. So, this removal would be inconsistent with the current assessment.)
38	Delete text selection guidance and example that includes video (Chapter 3)	This text removes references to video in any part of the assessment and deletes guidance for text selection, including the types of experts that should be used to select texts.
39	Delete prohibition that items relate to technical vocabulary, idiomatic expressions, and subject area knowledge (Chapter 3)	The revised text is less specific than the previous text.
40	References to appendices are removed (Chapter 3)	The text (suggested for deletion) describes content in one of the appendices, which are recommended for deletion in the proposed edits.
41	Delete prohibition that items relate to students' everyday oral proficiency and subject area (discipline-specific) knowledge (Chapter 3)	The revised text is less specific than the previous text.
42	Delete references to more accurate interpretations and validity across diverse test takers (Chapter 3)	The revised text instead promises more contextualized presentations of NAEP results.
43	Delete references to valid interpretations of test scores (Chapter 3)	The revised text instead sets a goal of more informed interpretations.
44	Delete reference to unfairly advantaging students in the assessment itself (Chapter 3)	The text (suggested for deletion) alters principles articulated from Universal Design of Assessment.
45	Added text that readability cannot be maximized (Chapter 3)	The text (suggested for addition) alters principles articulated from Universal Design of Assessment.
46	Deleted reference to language that is easily understandable regardless of student's background (Chapter 3)	The text (suggested for deletion) alters principles articulated from Universal Design of Assessment.
47	Rephrase guidance for selection of grade-appropriate text (Chapter 3)	The revised text is an editorial change.

No.	Recommended Edit	Project Officer Note
48	Deleted references to fairness and bias in testing as well as related references, including national testing standards (Chapter 3)	The revised text removes references to fair and unbiased assessments and associated professional standards.
49	Deleted reference to rationale for having Universal Design Elements in a reading assessment (Chapter 3)	The revised text removes the rationale and research basis for having Universal Design Elements in a reading assessment, but does not eliminate or change the Universal Design Elements themselves.
50	Deleted reference to example of a word bank as a task-based Universal Design Element (Chapter 3)	The revised text removes a particular example of a task-based Universal Design Element.
51	Deleted reference to multiple examples of motivational Universal Design Elements (Chapter 3)	The revised text removes particular examples of motivational Universal Design Elements.
52	Deleted references to bias, fairness, and equity in connection with Universal Design Elements (Chapter 3)	The revised text removes references to fair and unbiased assessments and equitable opportunities for students to engage with the assessment.
53	Deleted reference to example of a pop-up definition as a knowledge-based Universal Design Element (Chapter 3)	The revised text removes a particular example of a knowledge-based Universal Design Element.
54	Rephrase “capacity of NAEP results” to instead say “use of NAEP results” (Chapter 3)	The revised text is an editorial change.
55	Added text to elaborate on when/how socioeconomic status information can be reported (Chapter 4)	Based on attendance at Development Panel meetings, this was implied in the Panel’s initial recommendation. The added text is more specific/explicit.
56	Delete references to equity, rigor, and precision, with validity reference remaining (Chapter 4)	The text (suggested for deletion) explains the basis for proposed updates.
57	Delete indication that NAEP reporting can increase knowledge about factors that can expand opportunities to learn (Chapter 4)	The revised text articulates a different vision for the use and potential impact of NAEP results.
58	Delete references to equity as a reporting goal (Chapter 4)	The text (suggested for deletion) explains the basis for proposed updates.
59	Delete reference to socioeconomic status within race/ethnicity as a feature of NAEP reporting (Chapter 4)	This is a core reporting recommendation from the Panel to improve NAEP Reporting.
60	Delete a reference to enhancing the explanatory capacity of NAEP (Chapter 4)	The revised text articulates a different vision for the use and potential impact of NAEP results.
61	Delete reference to the National School Lunch Program as the current NAEP measure of family income (Chapter 4)	The revised text is less specific than the previous text.

No.	Recommended Edit	Project Officer Note
62	Delete and rephrase comments about the measurement of socioeconomic status (Chapter 4)	The revised text is an editorial change, and removes rationale for contextual variables recommendations.
63	Delete recommendations for contextual variables (Chapter 4)	The contextual variables (suggested for deletion) have already started to be reported on via NAEP questionnaires. See NAEP Website.
64	Delete reference to cultural assets of individuals (Chapter 4)	The text (suggested for deletion) explains the basis for proposed updates.
65	Rephrasing terminology about goals of NAEP contextual variables (Chapter 4)	The revised text articulates a different vision for the use and potential impact of NAEP results.
66	Deleting research references supporting contextual variables recommendations (Chapter 4)	The text (suggested for deletion) explains the basis for proposed updates.
67	Deleting indication that NAEP reporting can assist policymakers and other stakeholders in crafting policy and practice (Chapter 4)	The revised text articulates a different vision for the use and potential impact of NAEP results.
68	Rephrases NAEP prohibition on intrusiveness of questionnaires to instead cite law (Chapter 4)	The revised text is an editorial change.
69	Delete references to self-efficacy (Chapter 4)	The contextual variables (suggested for deletion) have already started to be reported on via NAEP questionnaires. See NAEP Website.
70	Rephrasing contextual variables from self-perception to self-reporting (Chapter 4)	The revised text is an editorial change.
71	Rephrasing terminology about goals of NAEP contextual variables (Chapter 4)	The revised text articulates a different vision for the use and potential impact of NAEP results.
72	Deleting research references supporting contextual variables recommendations (Chapter 4)	The text (suggested for deletion) explains the basis for proposed updates.
73	Delete a reference to enhancing the explanatory capacity of NAEP (Chapter 4)	The revised text articulates a different vision for the use and potential impact of NAEP results.
74	Rephrasing terminology about goals of NAEP contextual variables (Chapter 4)	The revised text articulates a different vision for the use and potential impact of NAEP results.
75	Delete reference to improving statistical reliability of NAEP data (Chapter 4)	The revised text is less specific than the previous text.
76	Delete references to the existence of periodic secondary analyses (Chapter 4)	The revised text is less specific than the previous text and articulates a different vision for the use and potential impact of NAEP results.
77	Rephrasing terminology about goals of NAEP contextual variables (Chapter 4)	The revised text articulates a different vision for the use and potential impact of NAEP results.
78	Delete rationales for recommended updates (Chapter 4)	The text (suggested for deletion) explains the basis for proposed updates.

No.	Recommended Edit	Project Officer Note
79	Delete references to equity, rigor, precision, and validity as focus areas for NAEP (Chapter 4)	The text (suggested for deletion) explains the basis for the framework. The revised text articulates a different vision for the use and potential impact of NAEP results.
80	Delete references to self-efficacy (Chapter 4)	The contextual variables (suggested for deletion) have already started to be reported on via NAEP questionnaires. See NAEP Website.
81	All Appendices are deleted	The revised text deletes all appendices, including sections (mandated by Board policy) describing the achievement levels and sample-items.

NAEP Mathematics and Reading Framework Processes

Periodically, the Assessment Development Committee (ADC) takes stock of lessons learned from implementing the Governing Board Framework Development Policy. In prior discussion, the ADC affirmed that one role of the Committee is to assure that the framework update process is carefully followed to produce a high quality framework for each NAEP assessment. To execute this responsibility, the ADC monitors framework processes via routine project updates and provides direction to the framework panels, as needed. This guidance is intended to assure compliance with the NAEP law, Governing Board policies, Department of Education and government-wide regulations, and requirements of the contracts used to implement the framework project.

As framework panels engage deeply in the issues specific to the subject area, the Board must exercise policy oversight by considering a wider context. This includes consideration of the role and purpose of NAEP in informing the public about student achievement, the legislative parameters for NAEP, constraints of a large-scale assessment, technical assessment standards, and issues of burden and cost-effectiveness in designing the assessment. This wider context also includes the Board's priorities, as articulated in the Governing Board's Strategic Vision and through plenary deliberations.

The following list of critical questions has supported the ADC as it monitored recent framework update processes, assuring compliance with the Governing Board's Framework Development Policy. Accordingly, key outcomes from the Board's policy are also listed.

Process

The process must be comprehensive, inclusive, and deliberative. Based on the Governing Board Framework Development Policy, process questions for the Committee's monitoring efforts for each framework include:

- Does the Development Panel have a proportionally higher representation of content experts and educators (compared with the Visioning Panel)?
- Does the Development Panel's content expertise collectively address all grade levels designated for the assessment?
- Did the framework update project begin with an extensive review of the current framework?
- Does the process engage a broad spectrum of stakeholders in developing recommendations for the knowledge and skills NAEP should assess?
- Is the process informed by a broad, balanced, and inclusive set of factors, delicately balancing current curricula and instruction, research, and the nation's future needs?
- Is the process being conducted in an environment that is open, balanced, and even-handed?
- Is the Development Panel considering all viewpoints raised and debating all pertinent issues?

Outcomes

In accordance with the Board's policy, the final framework must:

- Be inclusive of content valued by the public
- Reflect high aspirations
- Focus on important, measurable indicators
- Avoid endorsing or advocating a particular instructional approach
- Be clear and accessible to educators and the general public
- Define the construct(s) to be assessed and reported upon
- Articulate item formats, sample items, and sub-content weightings to demonstrate the construct is to be measured
- Describe how much of the content domain relates to the NAEP Basic, NAEP Proficient, and NAEP Advanced levels for each grade to be tested
- Align to widely accepted professional testing standards
- Support fair and accurate measurement of student academic achievement
- Support NAEP assessment items that will be secular, neutral, and non-ideological and free from racial, cultural, gender, or regional bias

Session Objectives

At the May 7 ADC meeting, Governing Board contractor WestEd will provide an overview of how framework processes were implemented for the NAEP Mathematics update (completed in 2019) and the NAEP Reading update (ongoing). The goal of this session is to encourage ADC discussion regarding: ***What are potential process improvements that should be considered for future framework projects?***

As context for this discussion, the attached paper provides a historical overview of how NAEP framework development has evolved over the years.

This discussion will set the foundation for: (a) an upcoming related joint session with the Governing Board Committee on Standards, Design and Methodology (COSDAM); and (b) work plans related to drafting a procedures manual to accompany the Board's Framework Development Policy.

HISTORY, POLICY, AND DECISION POINTS

Developing NAEP Assessment Frameworks

April 2021
Cornelia S. Orr, Ph.D.

History, Policy, and Decision Points for Developing NAEP Frameworks

Table of Contents

I. Introduction and Historical Overview

The National Assessment Governing Board (Governing Board) is an independent, bipartisan organization that sets policy for the National Assessment of Educational Progress (NAEP), commonly known as The Nation's Report Card. Since its creation by Congress in 1988, the Governing Board has overseen and set policy for NAEP by identifying subjects to be tested, determining and approving the assessment content, setting achievement levels for each assessment (i.e., NAEP Basic, NAEP Proficient, and NAEP Advanced), improving the reporting of results, and planning and executing initial releases of NAEP Report Cards.

The 26 members of the Governing Board includes governors, state legislators, state and local school officials, educators, researchers, business representatives, and members of the general public, who are appointed by the U.S. Secretary of Education. As part of the Governing Board's policy setting role, it adopts policy statements and resolutions for NAEP which provide guidance about the implementation of NAEP to persons and organizations working with and on behalf of the Governing Board. The Governing Board's policies align with the purpose of NAEP to provide fair and accurate measurement of student academic achievement. Members of the Governing Board and the National Center for Education Statistics (NCES), working in tandem, conduct activities to implement NAEP and communicate NAEP results to diverse audiences.

This paper provides a summary of the history of the Governing Board framework development processes and the evolution of the policy that now governs how the Governing Board determines the content for NAEP. It explains how changes have occurred over time and the implications for current and future framework development. This paper also describes key decision points in this process, for example, when the Board involves external partners and stakeholders in updating or revising frameworks, and describes the Board's role in approving frameworks.

What Is a NAEP Assessment Framework?

In the 2009 publication *A History of NAEP Assessment Frameworks*, Carol Jago provides this definition.

> NAEP frameworks describe the assessment objectives and design for national tests in reading, mathematics, writing, science, history, civics, economics, foreign languages, geography, and the arts. Governing Board policy dictates that these assessments must be valid, reliable, and based on widely accepted professional standards. (Jago, 2009, p. 1.)

NAEP assessment frameworks "are conceptual, overview documents that lay out the basic structure and content of a domain of knowledge and thereby serve as a blueprint for assessment development." (Haertel, et al., 2012, p. 14) Framework documents typically define the content area in two dimensions: (1) the content and skills to be tested, and (2) the cognitive processes and complexity assessed within the content area. Further, the framework specifies

the types of test questions to be used and the balance of content (weighting) to be assessed. More specific details about developing items to measure the content and cognitive processes at differing levels of cognitive complexity are contained in a companion "specifications" document for each framework. NAEP assessment frameworks provide both the "what" and the "how" for NAEP and have been used by the Governing Board since its inception in 1988.

NAEP before the Governing Board

Since the initial administration of the NAEP in 1969, much has changed in the education landscape and the assessment itself. In the early years, the assessment was developed to provide content-specific information useful to educators. The NAEP reports were designed to provide data on the success levels on a task (percent correct) and not an overall score. Summary scores were avoided because there were concerns about federal government intrusion into state and local school district decisions about education. (Lehmann, 2004; Selden, 2004) Similar concern exists today and probably always will.

In 1969, the responsibility for implementing the national assessment was given to the Education Commission of the States (ECS)—an organization of state leaders that could be "trusted" not to infringe on the rights of its members. While this arrangement continued successfully for several years, a 1976 government report issued by the Comptroller General contained a plea to "make NAEP more useful." (U.S. General Accounting Office, 1976) New federal legislation in 1978 brought changes to the oversight and organization of NAEP and established an Assessment Policy Committee of 17 members (the precursor to the Governing Board). In 1982, a major study critical of NAEP was published which said NAEP was underdeveloped and underutilized, and of "apparently negligible influence." (Wirtz & Lapointe, 1982)

In 1986, then Secretary of Education William J. Bennett formed a distinguished group of state leaders, called the Alexander-James study group. The group questioned the narrow range of subjects that NAEP was covering—due mainly to inadequate funding. Their report was reviewed by the National Academy of Education, and their review was incorporated in the report prior to publication. (Alexander & James, 1987) The debate which followed resulted in revised legislation and more changes for NAEP. The 1988 reauthorization of NAEP not only created the National Assessment Governing Board, it gave the Board specific responsibilities in regard to NAEP. One of these responsibilities was determining what would be assessed and how.

Anticipating the 1988 legislation that would permit voluntary state participation in NAEP, the National Assessment Planning Project (NAEP, 1988, pp. 5-6) was established to make recommendations for the 1990 mathematics assessment. The project utilized a process for developing objectives similar to that described in the legislation which authorized NAEP through June 30, 1988. However, it was expanded to ensure careful attention to formal mathematics objectives of states and some local school districts, and to elicit the opinions of practitioners at the state and local level about the content that should be assessed. This

involvement was seen as a key component to encourage the participation of states, particularly given that NAEP would produce state report cards. The effort to identify and review the objectives provided the assurance states wanted about the content being assessed. (Selden 2004, pp. 195-199)

1987-1990 Overlap: NAEP and the Governing Board[1]

The first assessments administered after the 1988 establishment of the Governing Board were in reading and mathematics in 1990. Those assessments utilized the NAEP reading and mathematics objectives being developed in anticipation of the 1988 law. These objectives were developed and reviewed as part of the NAEP National Assessment Planning Project. The 1990 NAEP Mathematics Framework and Reading Framework were published in November 1988 and April 1989, respectively, by ETS on behalf of NAEP. (NAEP, 1988; NAEP, 1989)

The development of the frameworks utilized a consensus development process. The 1988 Mathematics Framework described these elements. (NAEP, 1988, pp. 6-9).

- A seventeen-member Steering Committee included policy makers nominated by national organizations. One member was also on the Mathematics Objectives Committee.
- An eleven-member Mathematics Objectives Committee comprised of a teacher, a school administrator, mathematics education specialists from various states, mathematicians, parents, and citizens recommended objectives for the assessment.
- The draft objectives were distributed to the mathematics supervisor in each of the 50 states and also to 25 mathematics educators and scholars for their review.
- Incorporation of comments and revisions were made by the Mathematics Objectives Committee with the final recommendations approved by the Steering Committee.
- After the objectives were submitted to NCES, they were provided to the Assessment Policy Committee which approved the Project recommendations.[2]

Because NAEP would now produce state report cards, both the reading and mathematics process to develop objectives paid careful attention to the formal objectives of states and to the opinions of practitioners at the state and local level. In particular, efforts were made to integrate new theory and research on the learning and teaching of these subjects and to reflect the innovative approaches of assessments being developed. (NAEP, 1989, p. 7)

The Governing Board Framework Development Policy Overview

Beginning with assessment frameworks adopted for the 1992 assessment, Governing Board staff managed the process of soliciting and engaging contractors, and overseeing the work of

[1] A more detailed presentation of the historical activities related to the history of NAEP and the Governing Board is found in Appendix A.

[2] The Assessment Policy Committee provided policy oversight for NAEP and was established in the 1978 NAEP reauthorization. Also see discussion on page 2 and Appendix A.

committees charged with identifying the content for the assessments. A Governing Board staff member attending the second meeting of the Governing Board observed, "One of the most important issues considered at the January 1989 meeting was developing a 'consensus process' for determining the content of the 1992 reading assessment." (Bourque, 2004, p 205) The development of the framework was to be carried out via a contract with the Council of Chief State School Officers (CCSSO). The CCSSO staff recommended the principles summarized below which were contained in the January 1989 Governing Board meeting materials.

1. The process should be participatory, visionary, iterative, structured, explicit, stable, and supported by adequate resources.
2. The management of consensus committees should be in a value-free way, to encourage opinions and avoid curtailing or intimidating the participants.
3. The process should be mutually educational for those involved.
4. Values and constraints for the process should be stated up front.
5. Changes in the structure or rules of the consensus process during the process must be avoided.
6. Solicitation of comments representing the field is needed only in response to the draft recommendations.
7. Board members must decide carefully with which people they will work.
8. Work on subject-matter objectives, procedural, and analytic plans should be a staff function of the governance process, and review by the field should be part of the process.
9. The consensus process should be self-evaluating.
10. The planning process should have a built-in buffer to ensure that the recommendations are thoughtful and appropriate.

Bourque, the Governing Board Assistant Director for Psychometrics from 1989 to 2001 and an observer of the consensus processes for reading, writing, U.S. history, world geography, science and civics indicated these 10 principles were "in large measure what govern the work of the groups" who make the framework recommendations. (Bourque 2004, p. 206) The CCSSO report at the January 1989 meeting also included the recommendation that the Governing Board develop an explicit policy to direct those developing objectives for NAEP. When one considers the Governing Board workload to adopt frameworks between 1989 and 2002[3], it is not surprising that the explicit policy did not emerge until 2002. It is reassuring that similar practices as those ultimately included in the 2002 Framework Development Policy were in place before they were codified.

In 2018, the Governing Board revised the Framework Development Policy, primarily to add a provision for updating frameworks when a complete framework revision was not needed. The policy had originally been conceived for the development of new frameworks. This revision

[3] The Governing Board adopted the following frameworks between 1989 and 2002: Reading (1990), Writing (1990), Science (1991), U.S. History (1992), Geography (1992), Arts (1994), Civics (1996), Writing (1996), Mathematics (2001), Foreign Language (2000), Economics (2002).

also included streamlining some wording and moving procedural details to the contracting documents called statements of work. Details about these revisions will be discussed in a later section.

II. Legal Requirements for Assessment Frameworks

Are "frameworks" required in the law?

Technically, no. The current and previous versions of the Congressional authorization do not use the term framework. 'Assessment framework' is a construct used to distinguish what will be tested from what is taught (curriculum standards or instructional objectives). Some assessment programs use the term test blueprint or test specifications. While the construct of an assessment framework is not unique to the Governing Board, it is the term that was chosen. The NAEP assessment frameworks do not cover every aspect a content area, especially what students should be taught and how; they simply describe which aspects of the content area will be tested on NAEP and the how that content will be assessed.

By implication, yes. The NAEP legislation in effect just prior to the establishment of the Governing Board in 1988 included the requirement that the content to be assessed be defined. Specifically, the law required that "each learning area assessment shall have goal statements devised through a national consensus approach, providing for active participation of teachers, curriculum specialists, subject matter specialists, local school administrators, parents and members of the general public." (NAEP, 1988, p. 6) This process was used to develop the content-by-process matrix used for the assessments prior to the 1988 legislation, which are now largely referred to as the Long-Term Trend assessment (Mullins, 2017). The language related to assessment content in the current congressional authorization (P.L. 107-297, 2002) does not use the term "framework," but it has similar meaning.

What are the Legal Responsibilities of the Governing Board?

The responsibilities for the Governing Board as defined in the authorizing legislation (P.L. 107-297) are about more than developing assessment frameworks for NAEP. In Table 1 below, all of the requirements of the law are listed for clarity with the **distinctly framework-related ones shown in bold**. It should be noted that P.L. 107-279 is also about more than the Governing Board. It provides authorization for both the Governing Board (Section 302) and NAEP (Section 303). One requirement in Table 1 (No. 8) is from Section 303 and is included because it has implications for the policies and work for which the Governing Board is responsible. Also, references to Section 303 are found throughout Section 302 in acknowledgement of the necessity to coordinate all aspects of NAEP. While the requirements for the Governing Board in Table 1 are organized into an easier to read list than is typical presentations of laws, the correct legal citations are provided in brackets after each item.

Table 1

Legal Responsibilities of the Governing Board from P.L. 107-279

(Emphasis added for distinctly framework-related responsibilities)

1. There is established the National Assessment Governing Board which shall …" [Section 302(e)(1)]

i. **formulate policy guidelines** for the National Assessment (carried out under section 303). [Section 302(e)(1)(A)]

ii. **select the subject areas to be assessed** (consistent with section 303(b)); [Section 302(e)(1)(B)]

iii. **develop appropriate student achievement levels** as provided in section 303(e); [Section 302(e)(1)(C)]

iv. **develop assessment objectives** consistent with the requirements of this section and test specifications that produce an assessment that is valid and reliable, and are based on relevant widely accepted professional standards; [Section 302(e)(1)(C)]

v. **develop a process for review of the assessment which includes the active participation of teachers, curriculum specialists, local school administrators, parents, and concerned members of the public**; [Section 302(e)(1)(D)]

vi. **design the methodology of the assessment to ensure that assessment items are valid and reliable**, in consultation with appropriate technical experts in measurement and assessment, content and subject matter, sampling, and other technical experts who engage in large scale surveys; [Section 302(e)(1)(E)]

vii. consistent with section 303, **measure student academic achievement in grades 4, 8, and 12 in the authorized academic subjects**; [Section 302(e)(1)(F)]

viii. develop guidelines for reporting and disseminating results; [Section 302(e)(1)(G)]

ix. develop standards and procedures for regional and national comparisons;

x. take appropriate actions needed to improve the form, content, use, and reporting of results of any assessment authorized by section 303 consistent with the provisions of this section and section 303; [Section 302(e)(1)(I)] and

xi. plan and execute the initial public release of National Assessment of Educational Progress reports. [Section 302(e)(1)(J)]

2. The National Assessment of Educational Progress data shall not be released prior to the release of the reports described in subparagraph (J). [Section 302(e)(1)]

3. The Assessment Board may delegate any of the Assessment Board's procedural and administrative functions to its staff. [Section 302(e)(2)]

4. **The Assessment Board shall have final authority on the appropriateness of all assessment items**. [Section 302(e)(3)]

5. **The Assessment Board shall take steps to ensure that all items selected for use in the National Assessment are free from racial, cultural, gender, or regional bias and are secular, neutral, and non-ideological**. [Section 302(e)(4)]

6. In carrying out the duties required by paragraph (1), **the Assessment Board may seek technical advice, as appropriate, from the Commissioner for Education Statistics and other experts**. [Section 302(e)(5)]

7. Not later than 90 days after an evaluation of the student achievement levels under section 303(e), the Assessment Board shall make a report to the Secretary, the Committee on Education and the Workforce of the House of Representatives, and the Committee on Health, Education, Labor, and Pensions of the Senate describing the steps the Assessment Board is taking to respond to each of the recommendations contained in such evaluation. [Section 302(e)(6)]

8. Such agreement (with the Secretary to participate in state assessments) shall contain **information sufficient to give States full information about the process for decision-making (which shall include the consensus process used), on objectives to be tested,** and the standards for random sampling, test administration, test security, data collection, validation, and reporting. [Section 303(b)(3)(B)(II)]

Have the legal requirements for frameworks changed over time?

The duties of the National Assessment Governing Board were initially authorized in the legislation establishing the Board in 1988 and have remained quite stable throughout periodic reauthorizations, the latest of which is P.L.107-279 (2002). This law provides authorization for both the Governing Board (Section 302) and NAEP (Section 303).

In each iteration of the law the subsections have been rearranged slightly and language was added, deleted or clarified. The requirements, however, have remained essentially the same. Two unique elements were added in 2002. The first was Section 302(e)(1)(D), [No. 1.v. in Table 1], which calls for an inclusive review process for the assessment that is now addressed both by a Governing Board policy (NAGB, 2002i)[4] and by the framework review/revision process involving panels of experts and the solicitation of public comments before each framework is adopted. The other addition was Section 302(e)(1)(F), [No. 1.vii. in Table 1], which provides a linkage to Section 303 – the NAEP section. Appendix B presents all of the legal requirements in a side-by-side arrangement. Each requirement is presented with the legal numbering used in each reauthorization and identifies changes that occurred in each revision.

III. Board Policy Work Impacting Assessment Frameworks

This section of the report takes a broad look at the policy work of the Governing Board and how these efforts have influenced the development of NAEP Assessment Frameworks and the Framework Development Policy.

Before the Governing Board Framework Policy

As noted previously, the 1990 NAEP Mathematics and Reading Frameworks were the first frameworks issued after the Board's establishment. These objectives initially were developed and published (1988 and 1989 respectively) under the NAEP National Assessment Planning Project. The project, just like NAEP in prior years, used the accepted professional practices for test development. However, this project was more political than previous NAEP assessments had been. That is, the opinions and endorsements of local and state education leaders became more important than ever before. As objectives-based assessments had grown in the states throughout the 1970's and 1980's, these leaders wanted to be sure that the NAEP assessments covered the content they considered important and that it was tested in ways they thought appropriate. Of course, NAEP had always considered the advice of the subject area experts, but the advent of state report cards heightened NAEP's importance to states and resulted in more scrutiny for the assessments. These leaders wanted to ensure that what was tested would be reflective of the essential content being taught in their schools.

[4] The Governing Board policy statement, *Review of the National Assessment of Educational Progress*, adopted August 3, 2002, included six guiding principles that describe expectations for the rigorous review of the National Assessment of Educational Progress and actions of the Governing Board.

Historical Processes Impacting Governing Board Policies

The Governing Board became an operational entity in October 1988 with six members from the existing Assessment Policy Committee and other members appointed to staggered terms by Secretary of Education William J. Bennett in September 1998. (Vinovskis 1998, p. 20) The first Board meeting occurred on November 18–19, 1988, just seven weeks after the law went into effect. Some of the first activities included hiring staff, establishing a way of work (adopting by-laws), and planning for the 1990 Reading and Mathematics Assessments. Two working groups (organizational and policy) were formed at the very first meeting of the National Assessment Governing Board, and work was begun to develop by-laws which were adopted a year later.

The early years of the Governing Board were spent addressing the responsibilities contained within the authorizing legislation, including plans for reporting, setting achievement levels, and preparing frameworks. Assessment frameworks were adopted in 1990, 1991, 1992, 1994, 1996, 2000, and 2001. The *Redesigning the National Assessment of Educational Progress Policy Statement* (NAGB, 1996) was adopted at a time when Congress had codified National Education Goals, and it was the expectation that the NAEP would be a primary means for monitoring progress in student achievement. The new National Education Goals called for more subjects to be assessed than in the past and, not surprisingly, assessment frameworks were addressed throughout the policy. Although the legislation has now been replaced by the *No Child Left Behind Act of 2002* (P.L. 107-097), some of the principles in that policy remain (e.g., inclusive process and stable frameworks).

The greatest impact on Governing Board policy development was the No Child Left Behind Act of 2002 (P.L. 107-097). That year was very busy and many policies were codified, including the *Framework Development* and *Item Development and Review* policies.[5] In his letter to Board members about the August 1-3, 2002 meeting, then Executive Director, Roy Truby, summarized these actions in the selected quotes which follow.

> Actually, the Governing Board's work on No Child Left Behind began more than a year ago at the Board's special meeting in Houston on June 28, 2001. It was then, … adopting the design changes that make it possible for 2003 to be the base year for the mandatory state NAEP. … At the March and May meetings, the Board adopted a new schedule of assessments, eight new policies, several changes in its by-laws, and one white paper to implement the law. At this meeting, three

[5] Governing Board policies codified after the passage of the *No Child Left Behind Act of 2002 included*: *NAEP and the No Child Left Behind Act* (NAGB 2001b), *Framework Development* (NAGB 2002a), *Item Development and Review* (NAGB 2002b), *Long-term Trend* (NAGB 2002c), *Plan for Study of NAEP Sampling* (NAGB 2002d), *Policies and Procedures for Complaints Related to the National Assessment of Educational Progress* (NAGB 2002e), *Prohibition on Using NAEP to Influence State and Local Standards, Tests, and Curricula* (NAGB 2002f), *Public Access to Test Questions, Item Release, and Confidentiality of Data for NAEP* (NAGB 2002g), *Resolution on Participation of the Commonwealth of Puerto Rico in NAEP* (NAGB 2002h), and *Review of the National Assessment of Educational Progress* (NAGB 2002i).

> more policies and a study plan have been prepared for Board action. (NAGB, 2002l)

A more complete history of the early days of the Governing Board can be found in the resource *Overseeing the Nation's Report Card* (Vinovskis, 1998).

Ongoing Governing Board Policy Work

Governing Board policies have operationalized the requirements in the law. They have, for example, determined how the work of setting achievement levels would be completed. Governing Board policy work is an ongoing activity and will require the attention of Board members and staff again and again.

Governing Board polices have been responsive to the law, but specific policies have not been required by the law. The need for a policy is solely determined by the Governing Board. As mentioned earlier, the *Redesigning the National Assessment of Educational Progress* policy included guidance related to framework development which is still being used today. The excerpts below are examples of Governing Board decisions to codify in policy topics that are not explicitly required in the law.

> Test frameworks and test specifications developed for NAEP generally shall remain stable for at least 10 years.
>
> In rare circumstances, such as where significant changes in curricula have occurred, the Governing Board may consider making changes to test frameworks and specifications before 10 years have elapsed.
>
> NAEP shall be designed so that others may access and use NAEP test frameworks, specifications, scoring guides, results, questions, achievement levels, and background data. (NAGB, 1996, pp. 14-16)

The Governing Board does continue to update policies. Recent examples, in addition to *Framework Development Policy*, are the *Reporting, Release, and Dissemination of NAEP Results Policy Statement* (NAGB, 2017a) and the policy on *Developing Student Achievement Levels for the National Assessment of Educational Progress* (NAGB, 2018c).[6]

Some policies originally established in 2002, such as the Framework Development Policy, have been updated but others have remained intact and are still relevant today. A primary example is the policy on the *Prohibition on Using NAEP to Influence State and Local Standards, Tests, and*

[6] Ongoing work on updating the *Item Development and Review Policy* (NAGB, 2002b) and the *NAEP Testing and Reporting on Students with Disabilities and English Language Learners Policy* (NAGB, 2010, 2014) has been severely impacted by the restrictions the COVID-19 Pandemic has imposed on the Governing Board and others across the country who would have participated.

Curricula (NAGB, 2002f). The law gave this admonition, but the Governing Board decided to codify its position in a policy.

Influence of Professional Standards

Implementing NAEP and Governing Board policy is not done in a vacuum. External influences such as changes in the content standards of professional organizations or the instructional practices for a content area are a consideration when developing or revising frameworks. For example, changes were made in the 1996 Mathematics Framework "which would better align the NAEP program in mathematics with the National Council of Teachers of Mathematics Standards (NCTM, 1989) and the Professional Standards for Teaching Mathematics (NCTM, 1991)." (NAGB, 1992, p. 2) Another example was the nationwide emphasis on the preparedness of high school graduates for the workplace and college. A review of the mathematics and reading assessment frameworks was conducted and changes were made. (Achieve, 2005; Achieve, 2006)

There are also professional standards in the field of tests and measurements, known as psychometrics. As the Governing Board has developed policies, the staff and contractors have worked to adhere as closely as possible to these standards and also to the statistical standards of the National Center for Education Statistics. Both editions of the Framework Development Policy make reference to the following standards. The 2018 edition of the policy states it this way. (NAGB, 2018b)

> This Policy complies with the National Assessment of Educational Progress Authorization Act of 2002 (P.L. 107-279) and the documents listed below which express widely accepted technical and professional standards for test development. These standards reflect the agreement of recognized experts in the field, as well as the policy positions of major professional and technical associations concerned with educational testing.
>
> *The Standards for Educational and Psychological Testing*. (2014). Washington, DC: American Educational Research Association, American Psychological Association, and National Council on Measurement in Education.
>
> *Code of Fair Testing Practices in Education*. (2004). Washington, DC: Joint Committee on Testing Practices.
>
> *Center for Education Statistics (NCES) Statistical Standards*. (2012).

These standards emphasize features of tests including, for example, the content to be assessed and the statistical information that should be provided about test items and tests as a whole. If these standards are updated, the Board must work to address any new components that are applicable to NAEP and update the Governing Board policies, practices, and procedures, as may be needed. Contractors are expected to implement framework development projects in a manner that honors and is congruent with these standards. The requirements document for

the most recent frameworks procurement describes the procedures expected of contractors so that an assessment consistent with the standards will be implemented. (NAGB 2018a)

One challenge should be noted. The documents cited above focus primarily on the assessment and reporting of individual student scores. NAEP does test individual students but does not report individual scores. Thus, the professionals working in these areas must interpret how these standards are intended to apply to the unique situation of NAEP. While these standards are updated from time to time, it is infrequent. The most recent editions emphasize collecting many types of validity evidence in order that the validity claims of an assessment can be supported. Validity has always been important to NAEP and the Governing Board, and to the organizations which have evaluated NAEP. (National Research Council, 1999; Buckendahl, et.al., 2009; National Academies of Sciences, Engineering, and Medicine, 2017) Therefore, collecting validity evidence for NAEP and implementing other applicable portions of the standards will continue to be an important consideration for the Governing Board. In this regard, the Board examines the overlap between the NAEP framework and the standards used by other organizations and states. Recently, comprehensive reviews of state standards were conducted for mathematics and science. (AIR, 2018a, 2018b, 2018c, 2018d; HumRRO 2021)

IV. Board Policy for Framework Development

This section of the report focuses on the Governing Board Framework Development Policy, its origins, components, and changes over time. In addition, a list of Board decision points for framework development are presented.

2002 Framework Development Policy

The first Framework Development Policy was adopted on May 18, 2002 (NAGB, 2002a). As described earlier, the framework development activities conducted from 1988 to 2002 utilized processes similar to those codified in 2002. In particular, an iterative process was followed that used committees of content specialists from the field, a consensus process, opinions solicited from stakeholders, and the involvement of the Governing Board. The intent of the Assessment Development Committee (ADC) to incorporate similar guidance into the policy is manifest in their March 1, 2002, meeting minutes. (NAGB, 2002i)

> … the Executive Committee delegated this issue to the ADC since it involved the area of framework development and item review. ADC members discussed the current Board practice of "casting a wide net" to have broad representation on the framework development panels. The new policy language should make this explicit, perhaps by setting targets for representation of various NAEP constituencies. Strategies for involvement and feedback from the general public should also be stipulated. A draft policy will be prepared for discussion at the May Board meeting. (NAGB, 2002j)

At the May 2002 meeting, the Governing Board reviewed the policy ADC recommended for adoption. The ADC minutes of that meeting contain the following statements.

> This policy was reviewed and discussed in detail at the ADC's April 29 meeting in Detroit, Michigan. Committee members had no further changes to the draft policy. Action Item: The Assessment Development Committee recommends Board approval of the Policy on Framework Development. (NAGB, 2002l)

After receiving the ADC report and recommendation, the first Framework Development Policy was adopted. (NAGB, 2002a) The purpose of establishing this policy was to incorporate the requirements of the authorizing legislation and professional best practices into an official policy that provided explicit guidance for Governing Board staff and contractors to follow in framework development projects. The original 2002 policy was organized around seven principles with additional guidance about how to implement each of the principles. Simply stated, the policy provided for the following.

Principle 1 – the definition of a framework and what is to be included

Principle 2 – the process and participants for developing the frameworks

Principle 3 – the inclusion in the review process of current theory and practice standards within the discipline as defined by a variety of organizations

Principle 4 – the role of the Governing Board in approving the framework and the role of its designees including committees, staff, and contractors that might be hired by the Governing Board, and the required documents to be presented to the Board for approval

Principle 5 – the inclusion of preliminary achievement level descriptions and intended uses of them

Principle 6 – specific instructions, to be used by others, for the design of the test and constructing items

Principle 7 – the expectation that frameworks would remain stable for at least 10 years

2018 Framework Development Policy

In 2018, the Governing Board made a revision to the 16-year-old Framework Development Policy. (NAGB, 2018b) In addition to some minor reorganization and rewording, primary distinctions between the 2002 and 2018 editions included four changes that will be discussed in this section: (1) updating frameworks, (2) reviewing frameworks, (3) participants/stakeholders, and (4) framework panels/committees. Additionally, the current policy maintains a focus on the overarching principles to be followed, with the details and procedures moved to procedural documents and requirements for contractors. (NAGB, 2018a)

This section first describes the general contents of the 2018 policy and subsequently provides more detail about the four changes mentioned above. The two versions have similar content, although they are arranged somewhat differently. Appendix C contains a more detailed comparison of the policy principles for both versions in a side-by-side display. Although Appendix C does not capture all of the edits which occurred to remove redundancy and procedures, it does provide some examples of the specific wording changes.

The 2018 policy was organized around six principles, each containing additional guidance about how to implement the principle. Simply stated, the policy provides for the following.

Principle 1 – Elements of Frameworks: the scope of the domain to be measured, delineating the knowledge and skills to be tested at each grade, the format of the NAEP assessment, and the achievement levels. (Note: Combines 2002 Principles 1 and 5.)

Principle 2 – Development and Update Process: develop and update frameworks through a comprehensive, inclusive, and deliberative process that involves active participation of stakeholders. (Note: Updating frameworks was added to this section.)

Principle 3 – Framework Review: determine whether an update is needed to continue valid and reliable measurement of the content and cognitive processes reflected in evolving expectations of students and anticipates a framework review at least once every 10 years. (Note: This section was added to describe the process for determining if a framework update is needed and to address timing included in 2002 Principle 7.)

Principle 4 – Resources for the Process: take into account state and local curricula and assessments, widely accepted professional standards, exemplary research, international standards and assessments, and other pertinent factors and information.

Principle 5 – Elements of Specifications: shall be developed for use by NCES as the blueprint for constructing the NAEP assessment and items.

Principle 6 – Role of the Governing Board: shall monitor all framework development and updates. The result of this process shall be recommendations for Governing Board action in the form of three key documents: the framework; assessment and item specifications; and contextual variables that relate to the subject being assessed.

Updating Frameworks. The original Framework Development Policy in 2002 was stated in terms of developing new frameworks because this had been the primary focus of the work at the time the policy was adopted. Only Principle 7 referred to revising frameworks, but provided little guidance about the process. Therefore, the 2018 revision of the original policy was undertaken to include provisions for updating frameworks when a complete revision might

not be necessary. References to updating frameworks were added throughout the policy and guidance about the update process was included in Principle 2.d.

> The scope and size of a framework development project shall determine the size of framework panels and the number of panel meetings needed. A framework update project may require smaller panels and fewer meetings if a smaller scope is anticipated for recommended revisions. Each project shall begin with a review of major issues in the content area. For a framework update, the project shall also begin with an extensive review of the current framework, and the Visioning Panel shall discuss the potential risk of changing frameworks to trends and assessment of educational progress. (NAGB, 2018b, p. 6)

An important consideration for making decisions to update a framework is the potential impact on NAEP reporting. This concern was addressed under Principle 6.d. "In initiating a framework update, the Governing Board shall balance needs for stable reporting of student achievement trends. Regarding when and how an adopted framework update will be implemented, the Board may consider the NAEP Assessment Schedule, cost and technical issues, and research and innovations to support possibilities for continuous trend reporting." (NAGB, 2018b, p. 9)

Reviewing Frameworks. In the 2018 Framework Development Policy, a process was included for reviewing frameworks to determine if/when an update was needed. Principle 7 of the 2002 policy emphasized the importance of holding a framework stable for 10 years. The 2018 new Principle 3 calls for reviewing frameworks at least once every 10 years. Further, this new principle describes the review as considering the current relevance of the assessments and frameworks, input from experts, and the risk of changing the reporting of trends. The policy makes clear the decision to update involves the full Board's recommendation and describes the process for conducting an approved update.

Principle 3 also explains that ADC, within the 10-year period, may observe major changes in the states' or nation's education system related to NAEP frameworks and when/if these changing conditions warrant recommending an update to the full Board. The Board's decision may involve convening a Visioning Panel to examine the issues including commissioning special research and analysis to inform the updates under consideration. Based on these findings, a determination will be made about next steps and the processes to be implemented as described in the policy.

Participants/Stakeholders in Framework Panels. The 2018 policy identifies the various stakeholders in a comprehensive list (page 2) that applies to all aspects of the framework development or update processes. In the 2002 policy, stakeholders were identified under various principles and consistent terms were not always used. The 2018 policy, also provides more specificity about the participants in the framework development panels. While both policies call for the use of content experts, curriculum specialists, state and local educators, and policy makers, the 2018 policy is more specific about involving members with classroom teaching experience. The 2018 policy specifies that at least 20% of the members have

classroom teaching experience, perhaps in recognition that it may be difficult for current classroom teachers to make the time commitments required for these projects, even though funds for substitute teachers are included. For example, a recent framework project required approximately 15 days of meetings. The bottom line as described in the contract requirements document is that anyone chosen to serve on these panels "must be well qualified by content knowledge and familiarity with the knowledge, skills, and abilities in the respective subject, while addressing all grade levels designated for the assessment." (NAGB, 2018a, p. 16)

Additionally, the 2018 policy identified an upper limit for the number of participants in panels. Although the 2018 policy does not provide a rationale for these limits, perhaps this change was to facilitate the consensus process, as well as shorten timelines and reduce expenses. The number of panel members working on past projects has sometimes been much larger than 30. For example, the project for the 2009 NAEP Science Framework development used a total of 57 panelists, with no duplication across committees. A challenge with using only 30 panel members will be to attain the desired diversity for the framework panels as described on page 5 of the policy (NAGB, 2018b). Balancing these competing priorities will be an ongoing consideration. Fortunately, the 2018 policy recognizes that it may be necessary to add additional members. This option will be most needed for projects that are large in scope, that is, all three grade levels and multiple areas of expertise required.

It should be noted that the participants in framework development panels are identified by the contractor hired to conduct the assessment development activities. This is not a nominations process. Governing Board staff (sometimes Governing Board members) review the proposals and monitor the implementation of contract activities. For example, if the diversity or classroom experience goals indicated in the policy are not present in the names submitted as panelists, staff would ask the contractor to augment the panel to account for identified deficiencies.

Table 2, which is found at the end of the next section, includes a summary of the stakeholders discussed in this section and their expected panel assignments.

Framework Committee/Panel Functions. The 2002 and the 2018 policies are both nominally and substantively different: nominally in terms of the panel names and substantively in their composition. Both policies utilize two framework development groups and they have separate functions – the first function is to develop the high-level guidance for the work and the second function is to develop drafts of the documents that are consistent with the guidance. The more substantive difference is their composition and division of labor. The 2002 policy provides for separate groups of individuals and the 2018 policy provides for overlapping participants in the visioning and development activities. Although the policy does not specify the rationale for the overlap, it is likely the development panel will more fully understand the vision and guidelines for completing the work without having to be informed about it separately.

A third group of panelists is the technical advisors, primarily testing specialists. The 2018 policy describes their involvement as a resource to the framework development work rather than as a

committee. This approach permits different experts to be involved on different topics when their expertise is needed. For example, expertise about assessing certain types of content or expertise about the impact of changes on maintaining trends. The framework panels would be able to get expert advice as needed during their deliberations rather than waiting for a meeting of the technical advisors to be scheduled. The work of the technical advisors is expected to be conducted by representatives who participate in framework development meetings and as a group in separate meetings for more in-depth technical discussions.

Table 2 below provides a comparison of the functional working groups and the participants in each which were discussed in the previous sections.

Table 2
Framework Development Groups Comparison

2002 Policy (NAGB 2002a)	2018 Policy (NAGB 2018b)
Policy Oversight/Steering Committee • Represents key policy groups, etc. • At least 30% users and consumers • Formulates guidelines for the process consistent with law and NAGB charge • Monitors progress of project • Reviews final product before Governing Board	**Framework Visioning Panel** • Represents all stakeholders, including policy makers and users/consumers • At least 20% have classroom teaching experience • Formulates initial guidance for framework development • Includes up to 30 members (including up to 15 on Development Panel) • Additional members as needed
Planning Committee • Content experts & educators, etc. • Consider NAGB Charge and project guidelines • Develop deliverables • No overlap with Steering • Classroom teachers "well represented"	**Framework Development Panel** • Subset of Visioning Panel • Proportionally higher content experts & educators than the Visioning Panel • Detailed deliberations to resolve issues & recommend framework • Up to 15 members • Additional members as needed
Committee of Technical Experts (TAC) • Primarily testing experts • Involved where appropriate • Respond to technical issues raised by the committees • Review documents, esp. specifications • Provide guidance to project staff	**Technical Experts (TAC)** • Primarily testing experts • A resource to framework panels • Respond to technical issues raised during deliberations and meet separately, as needed • Review documents, esp. specifications

Natural Tension Points

The Framework Development Policy recognizes several natural tensions that exist in the education community at large. Education disciplines and the professionals who work within them are not unidimensional. Professionals naturally have different viewpoints about what is most important, what is most important to assess, and how that content should be assessed and reported. The policy provides the following guidance about the consensus process for developing or updating an assessment framework as broadly inclusive as possible.

> In balancing the relative importance of various sources of information, framework panels shall consider direction from the Governing Board, the role and purpose of NAEP in informing the public about student achievement, the legislative parameters for NAEP, constraints of a large-scale assessment, technical assessment standards, issues of burden and cost-effectiveness in designing the assessment, and other factors unique to the content area. (NAGB, 2018b, p. 8)

Additionally, there are frequently concerns about the scope of the content or range of content difficulty included in a framework. The Framework Development Policy recognizes this as natural tension point and provides the following guidance about addressing this concern and resolving it through the panel consensus process.

> The NAEP framework development and update processes shall be informed by a broad, balanced, and inclusive set of factors. The framework shall reflect current curricula and instruction, research regarding cognitive development and instruction, and the nation's future needs and desirable levels of achievement. This delicate balance between "what is" and "what should be" is at the core of the NAEP framework development process. (NAGB, 2018b, p. 7)

These are not all of the possible tension points that can arise in a broad-based committee process where varying opinions naturally exist. However, they do illustrate the Board's acknowledgment of them and guidance about resolving issues when they arise.

Resolving Points of Disagreement

Clearly, the Board acknowledges that different people and groups have different opinions about even the simplest constructs. In every framework adoption process, there is always some disagreement about the decisions represented in framework documents. The Framework Development Policy anticipates that there will be differences of opinion and provides guidance in this regard.

> Panels shall consider all viewpoints and debate all pertinent issues in formulating the content and design of a NAEP assessment, including findings from research.

> Reference materials shall represent multiple views. For each project, protocols shall be established to support panel deliberations and to develop a unified proposal for the content and design of the assessment. Written summaries of all hearings, forums, surveys, and panel meetings shall be made available in a timely manner to inform deliberations. (NAGB, 2018b, p. 6)

This is not a new challenge. Resolving these differences is what was envisioned by use of the term "consensus process" in the authorizing legislation. As mentioned in an earlier section of this report, the very first Reading Framework contains this statement.

> While objectives resulting from such a consensus process reflect neither a narrowly-defined theoretical framework nor every view of every participant, they do represent the thinking of a broad cross section of individuals who are expert in the areas of literacy research and reading instruction and who are deeply committed to the improvement of reading in our schools. (NAGB, 1990, p. 8)

Another example is the statement made by Charles Smith, then Executive Director, at the August 2004 Board meeting about the adoption of the 2009 Reading Framework which was two years in the making.

> Thousands of hours of effort have been devoted to the initiative, and the result awaiting your decision is, I understand, the most scrutinized framework ever to come before this Board. (NAGB, 2004e)

As the Governing Board has become more experienced in the process of identifying the content to be assessed, the framework documents themselves have become more thorough and more thoroughly and openly discussed. The Governing Board has expanded the involvement of experts in the field, utilized the research base within each discipline, and provided more opportunities for public comment. These activities are discussed in the next section of this report.

V. Framework Development and Implementation Activities

The legislation and Framework Development Policy have not changed substantially since enacted, but the activities to implement a new framework or update an existing one are much more extensive today than they were in the early 1990's. Some of the important changes are highlighted in this section.

Developing and Updating Assessment Frameworks

The development of a framework for a new assessment or updating one is guided by the schedule of NAEP assessments adopted by the Governing Board. (NAGB, 2018b) The

assessment schedule is a forward-looking document and identifies when changes in a framework might be expected. When development of a new framework or a framework update is initiated, several concerns must be balanced. For example, the need for stable reporting of student achievement trends, cost, specific changes in the discipline, relevant research, and innovations or new initiatives in impacting the field. These concerns are mostly objective considerations, but there are also more subjective elements. For example, when the subject area includes competing ideologies for which there is no obvious consensus, it can lengthen the timeframe for completing the framework. Making a decision to develop or update a framework is a complex process and involves many decision points as discussed in the following section.

Framework Decision Points

The framework policy broadly describes the process for developing a new framework and updating an existing one. It does not prescribe an order of events, although one may be logically inferred from the policy. Throughout the process of framework development, there are a number of important interactions between the Governing Board and its committees, subject area experts, stakeholders, the general public, and the panels convened to make recommendations to the Board.

The Governing Board by-laws assign responsibility for implementing the processes involved in framework development to the Assessment Development Committee (ADC). Their duties in this area include: developing and implementing a broadly inclusive process, developing content objectives, ensuring the active participation various stakeholders, developing assessment specifications, and providing for the review of test frameworks and specifications by other groups. (NAGB, 2010b, page 7) Additionally, the by-laws assign to ADC the responsibility of reviewing subject-specific background questions and all cognitive test items.

Consistent with the by-laws, Principle 6 of the 2018 policy describes the role of the Governing Board and ADC for framework development. (NAGB, 2018b, page 9) ADC's role is to monitor all the activities leading up to a framework development or update project and the ongoing project work. The Board's role is to approve and adopt the charge to the Visioning Panel and final framework documents prior to their handoff to NCES for developing the test questions. Although the Assessment Development Committee has the primary role for oversight of framework development/updating processes, other committees of the Board and NCES are involved as needed. Typically, COSDAM is involved in technical issues (scoring, scaling, trend reporting, etc.), R&D is involved in discussions about reporting and contextual data collection, and NCES is involved in issues related to item development, test construction, test scoring, data analysis, and reporting.

The discussion below provides a brief summary of important decision points and offers fundamental questions to be answered during the process of developing or updating a framework. It does not include every possible question or interaction between the Board, its committees, and other organizations. Appendix D supplements the information provided

below with a little more detail about the range of actions and the involvement of the Board, the Assessment Development Committee, contractors, and external reviewers.

1. **Should a framework revision or update be considered?** At least once every 10 years the Assessment Development Committee determines the timing for review of frameworks based on two key variables – the NAEP Assessment Schedule and lead time needed to implement a new/revised framework, including developing and field-testing new items for the assessment. The committee considers the relevance of assessments and their underlying frameworks, and any changes occurring in the field in making this decision. In their deliberations, the Assessment Development Committee may solicit input from experts, hear testimony or review white papers, discuss and determine what action should be recommended to the full Governing Board. Recently, comprehensive reviews of state standards were conducted for mathematics and science to document the overlap between the NAEP frameworks and the array of State standards before deciding to pursue a framework update. (AIR, 2018a, 2018b, 2018c. 2018d; HumRRO 2021)
2. **Is a new framework or update needed?** The Board receives a report from the Assessment Development Committee about their discussion and recommendations about the framework. Depending on the issues and interest, the Board may also hear presentations from various experts. If the Board agrees with the Assessment Development Committee recommendation, they will review, revise (if needed), and adopt the charge to the Visioning Panel. Many other actions will follow including contracts, working panels, and revised framework documents. See Appendix D for additional detail on these activities.
3. **Is the draft framework ready to be evaluated by external reviewers?** As the work to develop the framework proceeds, Governing Board staff carefully monitor the entire process. They have weekly conference calls with the project team and attend all the meetings of the Visioning and Development Panels. Others also attend the panel meetings, including the project technical advisors and representatives from NCES. This involvement throughout the project identifies and resolves potential issues. The Assessment Development Committee receives regular reports from the Framework Development Project staff and Governing Board staff, who in turn provide updates to and seek input from other Committees of the Governing Board on issues related to their areas of expertise and responsibility. Governing Board staff, in consultation with the Assessment Development Committee, determine when the contractor can begin the process of conducting external reviews. Agreements with the contractor describe how feedback will be solicited, reviewed, and incorporated.
4. **What feedback should be incorporated in the Framework?** The Framework Development Panel must consider all viewpoints, debate all pertinent issues about the content, including findings from research, and make revisions to the framework accordingly. This will likely be an iterative process, that is, reviewing and revising framework documents may occur more than once. After feedback is incorporated, the final draft is shared with staff and the Assessment Development Committee who review and recommend revisions or approval by the full Board.

5. **Should the framework be adopted and implemented?** In making a final decision, the Board should consider the process used to develop the framework, the role and purpose of NAEP to inform the public about student achievement, the legislative parameters for NAEP, constraints of a large-scale assessment, technical assessment issues (for example, the continuation of trend lines), issues of burden and cost-effectiveness in designing and implementing the assessment, and other factors unique to the specific content area. After the framework is approved, the next logical steps will be the development of item specifications and contextual variables for the assessment. Although it is likely the Panels have been considering these elements throughout their deliberations, they will formalize a document containing the prescribed information and submit it to the Board for review and approval through the Assessment Development Committee. Once approved, NCES and their contractors will begin item development and other planning for the assessment.

Appendix D supplements the information provided above with a little more detail about the range of actions and the involvement of the Board, ADC, contractors, and external reviewers. It highlights the major questions/decisions and other subordinate ones needed for framework development, approval, and adoption by the Board. Many smaller decisions and steps are behind these major decision points, but cannot be captured in this simplistic presentation. While the decision points are presented in an orderly manner, they may not always be implemented in the chronology implied by this list.

Need for Subject Area Updates

The 2018 Framework Development Policy added an entire section on how framework reviews would be conducted. For example, "the ADC shall solicit input from experts to determine if changes are warranted, making clear the potential risk of changing frameworks to trends and assessment of educational progress." (NAGB, 2018b, page 6) In making a decision about updating a framework, the Board needs to have explicated how extensive the revisions to a framework are likely to be, e.g., if substantive change would be required in the content being reported. For example, a major change would be changing the content areas and subscores reported. A more minor update could keep the test design and reporting intact, but recommend changes in how the content is assessed or which elements of the content are no longer relevant. Obtaining clarity about the need for an update in a subject area could involve the solicitation of white papers from subject matter experts about how the subject area should be assessed and important elements that should be considered. Another alternative could involve a panel discussion at an Assessment Development Committee or a full Board meeting. In either case, it will be the Board's responsibility to determine if a revision or update is needed.

Framework Panelists

The Board has always valued the opinions of and made every attempt to include classroom teachers, curriculum specialists, school administrators, policy specialists, subject-matter experts, and representatives of the general public in framework development projects.

However, balancing the membership of panels is not easy. The current Framework Development Policy provides the following guidance.

> In accordance with the NAEP statute, framework development and update processes shall be fair and open through active participation of stakeholders representing all major constituents in the various NAEP audiences, as listed in the introduction above.
> Framework panels shall reflect diversity in terms of gender, race/ethnicity, region of the country, and viewpoints regarding the content of the assessment under development. (NAGB, 2002a, pg. 5)

The role of the Governing Board, in particular the staff, and the Assessment Development Committee, is to review the panelists recommended by the contractor and ensure they meet the rigorous requirements of the contract. "All panelists must be well qualified by content knowledge and familiarity with the knowledge, skills, and abilities in the respective subject, while addressing all grade levels designated for the assessment." (NAGB, 2018a, p. 16) If there are concerns about panelists individually or collectively, it is incumbent upon the Governing Board to communicate these concerns and ensure they are addressed promptly.

The Framework Development Policy adopted in 2018 made some changes to the composition of the panels. Please refer to that earlier section for those details.

Public Comment Opportunities

It has always been the practice of the Board to seek public comment on the framework to be adopted. Sometimes, this included only advertising a comment opportunity in the Federal Register which may have limited the number of comments received. Since the early 2000's, the Board has expended much more effort in seeking feedback. Examples include public forums, meetings with state leaders in the content area and assessment directors, and working collaboratively with policy advisory groups and professional associations. The current policy guidelines emphasize the importance of a broad reach in obtaining public comment.

> Public comment shall be sought from various segments of the population to reflect many different views, as well as those employed in the specific content area under consideration. (NAGB, 2002a, pg. 5)

People who comment on a framework usually represent a constituency and have a particular viewpoint to be expressed. Their opinions may be minute or major and may be raised quietly or loudly. No matter, their opinions are important and hearing them is important. This does not mean the Governing Board is compelled to implement all recommendations made during the public comment period.

Constraints – Cost, Contracting, and Timelines

In addition to the decision about developing or updating a framework, the Governing Board must also contend with matters of budget, contracting, and timelines. These concerns are interrelated and difficult to parse.

Cost Factors. The Governing Board budget is constrained by the appropriation of funds from Congress. The cost of a framework development project depends on a number of factors including the complexity of the requirements, the competitiveness of the marketplace, the timeframe for completing the project, the extensiveness of revisions requested, and the unexpected. As might seem obvious, the more complex the project and the longer it takes to complete, the more expensive it will be. Some of these factors are predictable, but others, like the COVID-19 pandemic, are more are difficult to anticipate. In general, the Governing Board budget is sufficient to cover the cost of developing new or updating existing frameworks when done one at a time. Circumstances requiring multiple contracts in the same year may entail extensive advance planning to accommodate.

Framework Contracts. Contracts with organizations experienced in developing educational assessments have been used by the Board since it was established in 1988. The very first frameworks were supported by contracts with the Council of Chief State School Officers (CCSSO) that established the National Assessment Planning Project. Over the history of framework development, contracts have been awarded to the American Institutes for Research; American College Testing; the College Board; the Council of Chief State School Officers; the National Center for Research on Evaluation, Standards, and Student Testing (CRESST) at UCLA; and WestEd, and others. (Jago, 2009)

In recent years, the number of contractors bidding on NAEP Assessment Framework Development contracts has dwindled. The failure to have multiple bidders is a disadvantage because choice in vendors is desirable, as is competitive bidding. The root cause of the reduction in bidders is unknown, but reasons can be assumed to include the uniqueness of the project, lack of prior experience, changing or realigned corporate capabilities, availability, conflict of interest, potential for controversy, lack of interest, or other factors.

Contracting Procedural Requirements. The sophistication of the framework development procedures and contracting requirements has grown over time. The Framework Development Policy implies a number of processes that should be completed by those developing frameworks, but the contract requirements are much more detailed. For example, the policy is contained in nine pages, but the current Governing Board procedural requirements for contractors is 35 pages long. These requirements were recently Attachment A to the Governing Board procurement *Update of National Assessment of Educational Progress (NAEP) Frameworks for Mathematics, Reading, and Other Subjects*. (NAGB, 2018a)

The length is necessary because of the number of detailed requirements contained therein. The current work calls for regular monitoring of the project by Governing Board staff, and

regular reporting to the Assessment Development Committee throughout the scope of the contract. Attention is also given to the identification of panel members and the processes being implemented. A process report is required which summarizes all procedures implemented and issues encountered. This detailed information is used to support the validity of the recommended framework, specifications, and contextual variables. The Table of Contents from the most recent Statement of Work is found in Appendix E and shows the extensiveness of the requirements covered.

Timelines. This discussion about timelines will be considered from two perspectives: the time required to develop and adopt a new framework, and the lead time to implement changes to the assessment. These are related in that the latter cannot be accomplished without the former.

The lead time for changes to the assessment will be considered first because it has a fixed end point because of the NAEP assessment schedule. According to information NCES has communicated at Board meetings, the timelines for creating new assessment items and including them in a NAEP assessment can take from five to six years, whether the assessment framework is new or is being updated and applies equally to developing a new digital-based assessment or digital items for an existing assessment. This timeline is long because items must be developed and reviewed, tried out with small groups of students, analyzed, added to existing assessments, and then administered in an actual NAEP assessment. Because NAEP is not administered every year this timeline is longer than is typical for most assessment programs.

In understanding this timeline, it might be helpful to think about developing assessment items in three phases.

- The first phase is to develop questions for cognitive skills to be assessed, including reviews by experts in the field and conducting cognitive labs to ensure the questions are assessing the cognitive skills intended by the framework. Sometimes, several rounds of review and revision are needed to develop questions that meet the NAEP framework and review criteria. These questions also must be formatted for the platform on which they will be presented and reviewed in that same manner.
- The second phase involves collecting data from students which is called pilot testing. This is usually done during a regular NAEP testing window. Questions for this phase must be formatted and presented as they ultimately will appear on NAEP. Sufficient quality control steps must be performed to ensure data capture and scoring are accurate. Additionally, data must be collected from a significant number of students so that results can be correctly interpreted and used to develop future forms of NAEP. Another round of reviews occurs after these data are collected which includes examining item and test statistics, including item bias. If questions are rejected at this point, they may be revised and recycled through the first two phases.
- The third phase involves administering forms (blocks) in the actual NAEP assessment, administering them to students, scoring questions, and summarizing the data to be reported.

The schedule may also depend on when the Board authorized the work to begin as well as the level of innovation represented in the items identified in the framework. After the Governing Board approves the assessment framework, item specifications, and contextual variables, work can begin. After item writing is completed and items are reviewed by standing committees of content experts and the Governing Board, the approved items can be field tested (item tryouts) with the target group of students. Field testing will be done during the regular NAEP assessment window with a special sample of students. Those items which survive statistical standards and another round of reviews are assembled into forms and reviewed by NCES and the Governing Board. Because the field testing is done in one calendar year and the actual test administration is done in another, the minimum amount of time needed is two years. However, if new item types or constructs are contained in the framework, or if an innovative delivery of item content must be explored, more time will be required to try out items and analyze them before they are deemed valid for their intended purpose. It is not the purpose of this paper to discuss cognitive labs or other methodologies useful in determining item validity. It is enough to say this takes much longer.

The most obvious statement to be made about developing frameworks is that developing a new framework should take longer than updating an existing framework; however, that statement is very misleading. The more agreement there is in a subject area is probably a better factor for predicting how much time will be involved in developing a new framework or updating an existing one. As the Framework Development Policy prescribes, the Governing Board is seeking a consensus project; therefore, the longer it takes to reach consensus the longer the framework project will take. In thinking about the timeline for a framework project, one cannot think only about the framework panels who make content recommendations to the Board. One also must consider the time required to hire contractors on the front end of the work, as well as the public comment period and Governing Board deliberations/actions on the back end. In the best-case scenario where there is a great deal of consensus about the content to be assessed and when the public commentary is also agreeable, a period of one to two years can be expected for developing a charge, issuing a procurement, hiring a contractor, convening panels, etc. In the worst-case scenario where there is contentious debate, much more time is required. Finally, if the Board cannot support the recommended framework and reach a compromise that the Visioning and Development panels can support, then the entire process must begin again.

VI. Issues for the Future

In recent years the Governing Board has been having strategic discussions and reflecting on the data NAEP has been reporting over the last 40+ years. These discussions were designed to focus the Board's work on the strategic priority of providing NAEP information in the most innovative and effective ways. The Governing Board Strategic Vision for 2020 was adopted in November 2016 and the Strategic Vision for 2025 was adopted in September 2020 (NAGB, 2020b). Both of these efforts have included a vision for assessment frameworks. In both vision

statements, the reference to frameworks is found in the goal area "to innovate." Both versions are shown below with emphasis added.

2020 Strategic Vision The National Assessment Governing Board will revise the design, form, and content of The Nation's Report Card using advances in technology to keep NAEP at the forefront of measuring and reporting student achievement. The Governing Board will develop new approaches to update NAEP subject area frameworks to support the Board's responsibility to measure evolving expectations for students, while maintaining rigorous methods that support reporting student achievement trends.
2025 Strategic Vision The National Assessment Governing Board will ensure The Nation's Report Card remains at the forefront of assessment design and technology by refining design, content, and reporting, increasing relevancy for NAEP users and inspiring action to improve achievement for all. The Governing Board will optimize the utility, relevance, and timing of NAEP subject-area frameworks and assessment updates to measure expectations valued by the public.

As the Board continues implementing their Strategic Vision for 2025, they will establish priorities for the ongoing assessment framework activities. Consequently, discussing the issues about future framework development seems appropriate in this paper.

Framework Responsiveness

For the development of the Board's 2020 Strategic Vision described above, work groups were formed to consider avenues for advancing NAEP. These working groups and committees explored new approaches that could be utilized. One of the discussions focused on how the NAEP frameworks could become more responsive to small changes in the discipline area. The aim was to make adjustments in a manner that could reduce the timeframe typically required to change a NAEP framework and assessment.

At their joint "strategic vision" planning meeting in November 2016, the Assessment Development Committee (ADC) and the Committee on Standards, Design, and Methodology (COSDAM) discussed the concept of making the frameworks more responsive to the current curriculum standards being implemented on a broad scale (e.g., the Common Core State Standards). Other topics discussed included maintaining trends, valid alignment with student learning activities (e.g., writing using word processing), lead time for changes, the extent of NAEP's alignment (or lack thereof) with state and other content standards, changes in the field that might not be detected by the static nature of NAEP, communicating incremental changes to the public, not creating moving targets for school systems, and the concept of dynamic frameworks. (NAGB, 2016) (NAGB, 2017, p. 36)

At the joint meeting of these two committees in March 2017, there was a more in-depth discussion of the "dynamic framework" concept. The Governing Board committees agreed that the term "dynamic frameworks" was not the best way to characterize this effort because it implied that the frameworks would constantly be in flux, and such fluidity or the perception of it could have unintended consequences as well as miscommunicate the nature of the updates which might occur. There also was agreement that more discussion and study about this topic was important with the goal of learning how frameworks could become more responsive without affecting NAEP's trend reporting. (Haertel, et.al., 2012, pp. 3, 16-17) (NAGB, 2017, pp. 28-29)

The concept of "dynamic frameworks" as presented in the *Future of NAEP Panel White Paper*, is intriguing. The paper suggests these considerations.

> Dynamic frameworks would balance dual priorities of trend integrity and trend relevance. … it would be important to establish and to enforce clear policies concerning the reporting of significant changes in assessment frameworks, so as to alert stakeholders when constructs change and to reinforce the crucially important message that not all tests with the same broad content label are measuring the same thing. (Haertel, et.al., 2012, p. 17)

This discussion is ongoing.

Standing Subject-Matter Committees

Another idea for identifying changes needed in a framework is to make use of NAEP standing subject-matter committees. NCES contractors establish standing committees of content experts, state and local education agency representatives, teachers, parents, and representatives of professional associations to review the items developed for NAEP. Each standing committee considers: the appropriateness of the items for the particular grade; the representative nature of the item set; the match of the items with the framework and test specifications; and the quality of items and scoring rubrics. (NCES, 2020b)

The *Future of NAEP Panel White Paper* makes the case for using such committees as follows.

> Under our proposal, standing committees would review field test data, for example, and be aware when "after-the-fact" distortions of the intended domain occur because more ambitious item types fail to meet statistical criteria. Standing committees could also update assessment frameworks incrementally, at the same time assuring that the constructs underlying NAEP reporting scales did not drift to the point where new trend lines were indicated. In particular, assessment frameworks would be updated to accommodate changing learning environments. Inquiries with dynamic knowledge representations and simulations in science would be one example. (Haertel, et.al., 2012, pp. 17, 44)

The NAEP contractors already use standing subject-matter committees, particularly for item reviews. However, they are not charged with the explicit functions described by Haertel, et al. It is customary for Governing Board staff to attend the debriefing sessions of these committees, so some consideration could be given to seeking input as suggested.

Digital-Based Assessment Frameworks and Policy

NAEP transitioned to digital based assessments in 2017. Updating frameworks in this context should provide clarity about whether the construct of the assessment is changed by the digital-based format. Additionally, it is important to clarify how the content is to be assessed differently using digital techniques. Although, the new platform may not substantially alter the construct being assessed, the design implications of the digital-based formats should be elaborated so that the revised framework is consistent with this new delivery system.

The Assessment Framework Development Policy does not address delivery systems or related procedural details, rather these details are addressed in procedural requirements included in framework procurements. (NAGB, 2018a, p. 19) One of the rationales for seeking framework updates going forward includes incorporating new items that will more fully capitalize on current advances in digital-based assessment. The ADC and Governing Board staff need to determine if the policy should contain guidance specifying the extent to which frameworks should include content addressing platform-specific elements. (NAGB, 2018b)

VII. Reflections and Recommendations

Reflections on Framework Development Changes

Over time, the procedures for implementing frameworks have evolved in several important ways. Beginning with the frameworks developed since the early 2000s, the frameworks and process reports have demonstrated the broad representation in this work, have included more thorough documentation of the activities conducted, and have validated the increased public comment. While the authorizing legislation and the Governing Board Framework Development Policy are important, their influence on the frameworks has not really changed. In my opinion, the law and the policy have not been the primary drivers of these changes. The greatest influencer in these changes has been the increased utilization of test information for accountability decisions and the increased expectations for test publishers, including NAEP, because of this increased use.

Broad Representation. The framework committees have always included representation of subject-area experts (academicians and curriculum specialists), educators (teachers, local and state administrators), policy makers, parents, and the general public. Additionally, they were diverse in terms of gender, ethnicity/race, region, and representation of public-private school students, high-poverty students, and low-performing school students. When the participation

of all students in NAEP and accommodations were added to the assessment, persons who specialize in assessing students with disabilities and English learners also were included. Documentation of participants in framework committees and in the public comment opportunities provides evidence of this broad representation.

More Thorough Documentation. The framework documents produced today provide much more detail than the first framework documents, especially in terms of item examples and information about achievement levels. An example is found in the 1996 and the 2019 Mathematics Assessment Frameworks for NAEP. The 1996 Mathematics Framework includes three example items, one for each type of item to be included in the assessment: multiple-choice, open-ended, and extended open-ended. In contrast, the 2019 Mathematics Framework includes 14 unique items, five to describe the types of items included in the assessment (multiple-choice, short constructed-response, and extended constructed-response), and nine to provide examples of pure mathematics items (four items), calculator involved items (three items), and items using manipulatives (two items). In addition, the 2019 Mathematics Framework included a separate discussion of accessibility to item content for students with disabilities and English learners, after the examples of items. More detailed information about item design and accommodations is found in the *Assessment and Item Specifications for the NAEP Mathematics Assessment*. (NAGB, 1992; NAGB, 2006a; NAGB, 2007)

Another example of more thorough documentation in framework documents is the description about NAEP achievement levels. The 1996 framework describes the achievement levels in a single paragraph.

> The new NAEP Mathematics Framework was considered in light of the three NAEP achievement levels basic, proficient, and advanced. These levels are intended to provide descriptions of what students should know and be able to do in mathematics. Established for the 1992 mathematics scale through a broadly inclusive process and adopted by the Governing Board, the three levels per grade are a major means of reporting NAEP data. The new mathematics assessment was constructed with these levels in mind to ensure congruence between the levels and the test content. (NAGB, 1992, p. 3)

However, the 2019 Mathematics Framework, provides much more information, including achievement level descriptions. An introduction to achievement levels and the policy definitions are provided in the overview section (page 2) and an entire appendix is devoted to the achievement level descriptions (pages 71-76). Descriptions are provided for each grade level and for each of the three levels (basic, proficient, and advanced) within each grade level. Also provided are the scale score points associated with each achievement level. A great deal of detail is provided in these descriptions; in fact, the grade twelve descriptions require three pages. (NAGB, 2006a)

Greater Visibility and Debate. The advent of reporting scores on NAEP which were associated with individual locales has been a huge driver for the visibility of and debate about what is

assessed. When the Governing Board was authorized in legislation, preparations had been made to provide an opportunity for states to participate voluntarily in NAEP and receive scores for their own state. One of the major concerns about the Trial State NAEP project was the content, or framework, for the assessment. In fact, a mathematics content committee was formed and they developed an objectives-based approach similar to what states would have used. Although NAEP had always been developed under the scrutiny of subject matter experts, this became the most visible and extensive review process for the assessment content up to that time.

The greatest visibility and debate about NAEP came as a result of the No Child Left Behind Act (NCLB) in 2001. Some states had been participating in NAEP voluntarily for several years, however NCLB required all states to participate. Further the NCLB requirements revealed that NAEP would be used to evaluate the progress being reported by states on their own state tests and based on their own proficiency definitions. The publication of state-by-state NAEP results, especially in terms of the percent proficient, became controversial and the topic of much debate. In 2003, NCES began comparing each state's standard for proficient performance in reading and mathematics at grades 4 and 8 by placing the state standards onto a common scale of the National Assessment of Educational Progress (NAEP). The periodic report, *Mapping State Proficiency Standards* onto the NAEP scales also created much discussion and debate in the educational assessment community. (NCES, 2009; Ho and Haertel, 2007a; Ho and Haertel, 2007b)

There were claims that the NAEP content was different from state content and that the levels of proficiency for NAEP were higher than typical grade level expectations for students. There was partial truth in these claims, but the claims did not acknowledge the intentional design differences between NAEP and state assessments, including the intended meaning of the achievement levels, especially proficient. From the beginning NAEP frameworks had avoided matching its framework to a single set of content objectives and had strived to be broadly representative of the content domain. The NAEP frameworks were never intended to be a curriculum framework, like the standards states use, and never claimed to be. In addition, in setting the NAEP achievement levels, the Governing Board did not want them necessarily to reflect only the current level of student achievement. The desire was to define the content students should know across a range of achievement. Therefore, educators were asked to identify content expectations for basic, proficient, and advanced levels of achievement. The debates about the use of the word “proficient” and the alignment of it with state definitions of proficiency, and the alignment of NAEP frameworks with state standards will continue as long as comparisons of results are made across different locales, different assessments, and using different performance level definitions.

Another concern about the content defined in the NAEP assessment frameworks was how to consider the impact of the Common Core State Standards and their subsequent adoption/implementation in numerous states. The National Governors Association supported this initiative and the U.S. Education Department provided grants (via several consortia projects) to support states in revising their standards and assessments to align with the

"common core." During this period, there also were calls for the NAEP frameworks to be aligned with the common core and alignment studies were conducted by groups external to the Governing Board. (Daro, et.al., 2015) Recently, comprehensive reviews of state standards were conducted by the Governing Board for mathematics and science. (AIR, 2018a, 2018b, 2018c. 2018d; HumRRO 2021) Prior to wide-spread adoption of the "common core," there was much less convergence across state standards and expectations for students. This variability had historically impacted the feasibility and understandability of studies about the relationship of NAEP to state standards.

External Input/Public Comment. Input into the first NAEP content frameworks was obtained both from the committee members who recommended the content to the Board and from individuals and national organizations external to this work. Staff solicited comments on frameworks as well as posted notices of the Board's intended actions in the Federal Register, a legal requirement still in effect. Today, proactive outreach activities for the purpose of obtaining feedback on the draft frameworks are required in the procurements issued by the Governing Board (NAGB, 2018a, p. 18). Contractors conduct these activities and document them in process reports prepared for the Governing Board. (WestEd, 2006, 2010, 2021)

The 2018 Framework Development Policy recognizes that external input is important. In fact, the policy calls for the identification of substantive issues at the beginning of the process to review the framework so these can be addressed during the project to develop or update the framework. "… the ADC shall solicit input from experts to determine if changes are warranted, making clear the potential risk of changing frameworks to trends and assessment of educational progress." (NAGB, 2018b, p. 6) Additionally, framework development project staff conduct extensive external reviews of the draft framework before a final draft is presented to the Board for adoption.

The excerpts below from the most recent process report for the NAEP Mathematics Framework illustrate the extensiveness of the outreach efforts conducted before the Board is presented a final draft for adoption. (WestEd, 2021, pp. E-3-4)

> "Outreach to organizations and individuals … was conducted with assistance from a number of collaborating organizations including the Council of Chief State School Officers (CCSSO), Conference Board for the Mathematical Sciences (CBMS) and its member organizations, National Council of Teachers of Mathematics (NCTM), TODOS: Mathematics for ALL (TODOS), Benjamin Banneker Association, National Council of Supervisors of Mathematics (NCSM), Association of Mathematics Teacher Educators (AMTE), Mathematical Association of America (MAA), and Mathematical Sciences Research Institute (MSRI).
>
> "Organizations (e.g., NCTM, AMTE, TODOS, MAA) disseminated information about the project website (naepframeworkupdate.org) and through flyers, email newsletters, social media, website announcement, hosted webinars, and

podcasts. In conjunction with partnership organizations, WestEd facilitated six live webinars, five in-person presentations, and one podcast recording.

"Across in-person and live venues, more than 1,000 people participated in outreach activities from the target stakeholder groups: Teachers, Curriculum Specialists, Content Experts, Assessment Specialists, State Administrators, Local School Administrators, Instructional Leaders, Policymakers, Business Representatives, Parents, Students, Users of Assessment Data, Researchers and Technical Experts, and other interested Members of the Public.

"Across digital communications, … email and social media dissemination of information reached more than 25,000 people across the target audiences … ."

Important Policy Updates

When the *Framework Development Policy* was revised in 2018, adding a process for updating frameworks was conceptually important. Time will tell if it is of any practical significance. The Governing Board is such a deliberative body, it is not assumed that the time for completing an update will be substantially shorter than for creating a new framework. Additionally, it is unknown how receptive the users of NAEP will be to "minor" revisions to the framework. Of course, this is both a perception and a communication challenge, and only the communication concern can be addressed by Board actions.

Removing procedures from policy is a good practice, because policy documents should provide guidance about processes and describe desirable outcomes (e.g., a valid and reliable assessment). Changes in methodology and processes should be informed as much as possible by current research and accepted best practice. If these were to become embedded in a policy, frequent revisions might be necessary and become very burdensome. A policy should focus on the big picture. The 2018 changes to the policy successfully addressed this concern.

The updates to the *Framework Development Policy* made in 2018 included: incorporating the Development Panel as part of the Visioning Panel, specifying the expected size of the panels, and utilizing technical experts in a different manner. Each of these changes are important and should facilitate the process of framework development going forward. Incorporating the Development Panel into the Visioning Panel will facilitate the ongoing work of the panelists who will be revising the framework itself. Since these panelists will have heard and participated in the discussion of issues and rationales, they should be well prepared to implement the vision for the new framework. Limiting the size of the panels will facilitate the communication of panel members with one another and be more conducive to the consensus building process. Finally, having the technical advisors available or participating in the Visioning Panel and Development Panel meetings will expedite the resolution of any technical concerns. All of these changes seem fitting and logical.

The revised 2018 *Framework Development Policy* has carefully addressed the use of classroom teaching expertise in the work of revising/updating NAEP frameworks. Almost everyone agrees that the involvement of classroom teachers is critical. That said, doing the work of revising a framework is time-consuming. Although framework projects include funds for substitute teachers' pay, it is likely that few active teachers or their administrators will be open to extended out-of-classroom time (approximately 15 days for a recent framework development process). The revised policy has addressed this tension by placing the importance on having classroom teaching experience on the Visioning Panel which requires less out-of-classroom time than the Development Panel. All members of both panels must be well qualified by content expertise and familiarity with the knowledge, skills, and abilities in the respective subject. Classroom teaching experience ensures that familiarity with the assessed grade levels will be included.

Recommendations

After reviewing mountains of minutes and many reading and mathematics framework iterations, as well as some historical documentation and reports, there are a few changes which seem worth considering.

Digital-based Assessments. Some questions in this area come to mind. Do the frameworks and specifications adopted by the Board adequately address both paper-based and digital-based assessments, especially in regard to the sample items included? Is an assessment in the digital space something about which the Governing Board needs a separate policy? A staff and committee discussion of these topics would be worthwhile.

Item Review Feedback. The Governing Board and NCES staff should discuss and develop a feedback loop process utilizing the item review standing committees. In particular, this feedback loop should focus on identifying elements in the framework that could be revised because the assessment of them lacks fidelity to the desired outcome as intended in the framework.

Continued Discussion Needed. Although the construct of "dynamic frameworks" is alluring, it has not been defined operationally in a sufficient enough manner to evaluate its practicality for the Governing Board. At this point, a recommendation for future consideration is all that can be offered. Further study and implementation details are definitely necessary to make such a proposal viable. Perhaps the standing committee feedback loop is a first step for identifying small changes that are needed in a framework to clarify how the content will be assessed.

Suggestions

The following list of suggestions are related to Framework publications. They are not presented in any order of importance and are offered for consideration of the Board and staff.

- The professional assessment standards cited in the Framework Development Policy also should be cited in framework documents because readers of these should not be left to wonder if they were utilized and implemented where applicable.
- The framework documents typically include a section of major changes. It would be helpful if these were expanded to include the rationale for the changes that were made.
- While it is important to issue framework documents corresponding to each administration of NAEP, more clarity is needed about when the Board actually adopted the framework represented in the publication. Having this embedded in the report is fine, but not sufficient for easy historical clarity. The title of the document should be augmented to contain the adoption date.
- Given the 2018 Framework Development Policy about updating frameworks, the framework document should clarify if the framework represents a major revision that may impact trend or if only minor updates were made, i.e., to incorporate digital-based items. While this is may be an empirical issue, the framework document should indicate whether special analyses will be conducted to make this determination.
- The framework documents need to include a little more about the "big picture" process followed in producing the framework, including references and links to expert testimony and public hearings which led to adoption by the Governing Board. This need not detract from the presentation of the content, but could be included as an appendix along with the names of panel members.

References

The references reviewed for this report are organized into five categories. The general category, appearing first, includes all resources that did not fall under the other titles. The other categories are: Legislation, Assessment Frameworks and Reports, National Assessment Governing Board Policies, and Governing Board Meeting Materials, Minutes and Transcripts.

General

Alexander, L., and James, H.T., (1987). *The Nation's Report Card: Improving the Assessment of Student Achievement*. (Report of the Study Group with a review of the report by a Committee of the National Academy of Education.) Washington, D.C.: National Academy of Education.

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (2014). *Standards for Educational and Psychological Testing.* Washington, D.C.

Bourque, M. L. (2004). Chapter 8: A History of the National Assessment Governing Board. In Jones, L. V., and Olkin, I. (Eds.). (2004). *The nation's report card: evolution and perspectives*. Bloomington, IN: Phi Delta Kappa Educational Foundation.

Buckendahl, C.W., Davis, S.I., Plake, B.S., Hambleton, R.K. Zenisky, A.L., and Wells, C.S. (2009). *Evaluation of the National Assessment of Education Progress: Study Reports*. Washington, DC. Department of Education.

Daro, P., Hughes, G. B., & Stancavage, F. (2015). Study of the Alignment of the 2015 NAEP Mathematics Items at Grades 4 and 8 to the Common Core State Standards (CCSS) for Mathematics. Retrieved from https://www.air.org/project/naep-validity-studies-nvs-panel

Haertel, E., Beauregard, R., Confrey, J., Gomez, L., Gong, B., Ho, A., Horwitz, P., Junker, B., Pea, R., Rothman, R., Shepard, L. (2012). *NAEP: Looking Ahead, Leading Assessment into the Future*. A panel report on ideas and recommendations for the future of NAEP. Retrieved from https://nces.ed.gov/nationsreportcard/about/future_of_naep.asp.

Ho, A. D., and Haertel, E. H., (2007a). (Over)-Interpreting mappings of state performance standards onto the NAEP scale. Paper commissioned by the Council of Chief State School Officers. Retrieved from: https://scholar.harvard.edu/andrewho/publications/over-interpreting-mappings-state-performance-standards-naep-scale

Ho, A. D., and Haertel, E. H. (2007b). Apples to apples? The underlying assumptions of state–NAEP comparisons. Paper commissioned by the Council of Chief State School Officers. Retrieved from https://scholar.harvard.edu/andrewho/files/ho_haertel_apples_to_apples.pdf

Joint Committee on Testing Practices. (2004). *Code of Fair Testing Practices in Education*. Washington, D.C.

Jones, L. V., and Olkin, I. (Eds.). (2004). *The nation's report card: evolution and perspectives*. Bloomington, IN: Phi Delta Kappa Educational Foundation in cooperation with the American Educational Research Association.

Lehmann, I. J. (2004). Chapter 1: The Genesis of NAEP. In Jones, L. V., and Olkin, I. (Eds.). (2004). *The nation's report card: evolution and perspectives*. Bloomington, IN: Phi Delta Kappa Educational Foundation.

Mullins, I.V.S. (2017) *Content of the Long-Term Trend Assessments Compared to Main NAEP*. A reaction paper prepared for the National Assessment Governing Board. Retrieved from https://www.nagb.gov/news-and-events/news-releases/2017/2017-long-term-trend-symposium.html.

National Academies of Sciences, Engineering, and Medicine. (2017). *Evaluation of the Achievement Levels for Mathematics and Reading on the National Assessment of Educational Progress*. Washington, DC: The National Academies Press. doi: 10.17226/23409.

National Center for Education Statistics-NCES. (2012). *Statistical Standards*. Retrieved from https://nces.ed.gov/statprog/2012/.

National Center for Education Statistics-NCES. (2020a). *Beginning of NAEP.* Retrieved from https://nces.ed.gov/nationsreportcard/about/newnaephistory.aspx#beginning.

National Center for Education Statistics-NCES. (2020b). *NAEP Item Development Process.* Retrieved from https://nces.ed.gov/nationsreportcard/contracts/item_dev.asp.

National Center for Education Statistics-NCES, Institute of Education Sciences. (2009). *Mapping State Proficiency Standards*. http://nces.ed.gov/nationsreportcard/studies/statemapping/.

National Council of Teachers of Mathematics. (1989). *Curriculum and Evaluation Standards for School Mathematics*. Reston, VA: NCTM.

National Council of Teachers of Mathematics. (1991). *Professional Standards for Teaching Mathematics*, Reston, VA: NCTM.

National Research Council. (1999). *Grading the Nation's Report Card: Evaluating NAEP and Transforming the Assessment of Educational Progress*. Committee on the Evaluation of National and State Assessments of Educational Progress. Pellegrino, J.W., Jones, L.R., Mitchell, K.J. (Eds.). Board of Testing and Assessment, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academy Press.

Selden, R. (2004). Chapter 7: Making NAEP State-by-State. In Jones, L. V., and Olkin, I. (Eds.). (2004). *The nation's report card: evolution and perspectives*. Bloomington, IN: Phi Delta Kappa Educational Foundation.

U.S. General Accounting Office (1976). *The National Assessment of Educational Progress: Its Results Need to be Made More Useful*. HRD–76–113; Washington, DC: U.S. Government Printing Office.

Vinovskis, M.A. (1998). *Overseeing the nation's report card: the creation and evolution of the National Assessment Governing Board (NAGB)*. Paper prepared for the National Assessment Governing Board. Washington, DC. http://www.nagb.org/publications/95222.pdf.

Wirtz, W., and Lapointe, A. (1982). *Measuring the Quality of Education: A Report on Assessing Educational Progress*. Washington, D.C. (Willard Wirtz and Archie E. Lapointe)

Legislation

Public Law 100-297. (1988). *Hawkins-Stafford Elementary and Secondary School Improvement Amendments of 1988*, § 100-297 U.S.C., Title III, Part C-National Assessment of Educational Progress Improvement Act § 3403 (pp. 215-220). Retrieved from https://www.govinfo.gov/content/pkg/STATUTE-102/pdf/STATUTE-102-Pg130.pdf.

Public Law 103-382. (1994) *Improving America's Schools Act of 1994*, § 103-382 U.S.C. § 412 (pp. 522-525). Retrieved from https://www.congress.gov/public-laws/103rd-congress.

Public Law 107-110. (1994). *No Child Left Behind Act of 2002,* § 107-110 U.S.C. § 412 (pp. 1904-1907). Retrieved from https://www.govinfo.gov/app/details/PLAW-107publ110.

Public Law 107-279. (2002). *Education Sciences Reform Act of 2002*, § 107–297 U.S.C. § 302 (pp. 44-45). Retrieved from https://www.govinfo.gov/link/plaw/107/public/279.

Assessment Frameworks and Reports

American Institutes for Research-AIR (2018a). Review of State Curricular Standards in Mathematics: Executive Summary. Unpublished manuscript. Retrieved from the National Assessment Governing Board.

American Institutes for Research-AIR (2018b). Review of State Curricular Standards in Mathematics: Methodological Narrative. Unpublished manuscript. Retrieved from the National Assessment Governing Board.

American Institutes for Research-AIR (2018c). Review of State Curricular Standards in Mathematics: Study Report. Unpublished manuscript. Retrieved from the National Assessment Governing Board.

American Institutes for Research-AIR (2018d). Review of State Curricular Standards in Mathematics: Supplemental Exhibits. Unpublished manuscript. Retrieved from the National Assessment Governing Board.

Achieve, Inc. (2005). *Recommendations to the National Assessment Governing Board on Aligning 12th Grade NAEP with College and Workplace Expectations: Reading*. Retrieved from the National Assessment Governing Board.

Achieve, Inc. (2006). *Recommendations to the National Assessment Governing Board on Aligning 12th Grade NAEP with College, Workplace and Military Expectations: Mathematics*. Retrieved from the National Assessment Governing Board.

Educational Testing Service. (2010a). *Final Report on the Study to Draft Achievement-Level Descriptions for Reporting Results of the 2009 National Assessment of Educational Progress in Reading for Grades 4, 8 and 12*. Prepared under contract to and in conjunction with the National Assessment Governing Board. Retrieved from https://www.nagb.gov/focus-areas/reports.html?q0=ets.

Educational Testing Service. (2010b). *Final Report on the Study to Draft Achievement-Level Descriptions for Reporting Results of the 2009 National Assessment of Educational Progress in Mathematics for Grade*

12. Prepared under contract to and in conjunction with the National Assessment Governing Board. Retrieved from https://www.nagb.gov/focus-areas/reports.html?q0=ets.

Human Resources Research Organization-HumRRO (2021). Comparative Analysis of the NAEP Science Framework and State Science Standards: Technical Report. Unpublished manuscript. Retrieved from the National Assessment Governing Board.

Jago, Carol (2009). *A History of NAEP Assessment Frameworks.,* A paper commissioned for the 20th Anniversary of the National Assessment Governing Board 1988-2008. Retrieved from https://www.nagb.gov/news-and-events/calendar/special-events/20th-anniversary.html.

National Assessment Governing Board-NAGB. (1990). *Reading Framework for the 1992 National Assessment of Educational Progress*. Retrieved from the National Assessment Governing Board. (Note: this framework was also applicable to the assessments conducted in 1994, 1998, 2000, 2002 and 2003, although separate editions were published for each assessment year.)

National Assessment Governing Board-NAGB. (1992). *Mathematics Framework for the 1996 National Assessment of Educational Progress*. Retrieved from the Education Resources Information Clearinghouse – Document ED 404 184. (Note: This framework was also applicable to the assessments conducted in 2000 for all grades and in 2003 for grades 4 and 8, although separate editions were published for each assessment year.)

National Assessment Governing Board-NAGB. (2001a). *Mathematics Framework for the 2005 National Assessment of Educational Progress*. Retrieved from the National Assessment Governing Board. (Note: This framework was also applicable to the assessments conducted in 2007, although separate editions were published for each assessment year.)

National Assessment Governing Board-NAGB. (2004a). *Reading Framework for the 2005 National Assessment of Educational Progress*. Retrieved from the National Assessment Governing Board. (Note: This framework was also applicable to the assessment conducted in 2007, although a separate edition was published for that year.)

National Assessment Governing Board-NAGB. (2006a). *Mathematics Framework for the 2009 National Assessment of Educational Progress*. Retrieved from the National Assessment Governing Board. (Note: This framework was also applicable to the Grade 12 assessments conducted 2013 and 2016, and to the Grades 4 and 8 assessments conducted in 2011, 2013, 2015, 2017, 2019, although separate editions were published for each assessment year.)

National Assessment Governing Board-NAGB. (2006b). *Reading Framework for the 2009 National Assessment of Educational Progress*. Retrieved from https://www.nagb.gov/naep-frameworks/reading.html. (Note: This framework was also applicable to the assessments conducted in 2011, 2013, 2015, 2017 and 2019, although separate editions were published for each assessment year.)

National Assessment Governing Board-NAGB. (2007). *Assessment and Item Specifications for the NAEP 2009 Mathematics Assessment*. Retrieved from https://www.nagb.gov/naep-frameworks/mathematics.html.

National Assessment Governing Board-NAGB. (2018a). Attachment A: Statement of Work. *Update of NAEP Frameworks for Mathematics, Reading and Other Subjects (RFP# 91995918R0002)*. Retrieved from https://www.nagb.gov/governing-board/contract-opportunities.html.

National Assessment Governing Board-NAGB. (2020a). *Mathematics Framework for the 2025 National Assessment of Educational Progress*. Unpublished Manuscript. Retrieved from the National Assessment Governing Board.

National Assessment Governing Board-NAGB (2020b). *Strategic Vision 2025*. Retrieved from https://www.nagb.gov/governing-board/strategic-vision.html.

National Assessment Governing Board-NAGB. (2021). *Draft: Process Report – 2025 NAEP Mathematics Framework Update*. Unpublished Manuscript. Retrieved from the National Assessment Governing Board.

National Assessment of Educational Progress-NAEP (1988). *Mathematics Objectives: 1990 Assessment*. Princeton, N.J.: Educational Testing Service. Retrieved from the National Assessment Governing Board. (Note: this framework

National Assessment of Educational Progress-NAEP (1989). *Reading Objectives: 1990 Assessment*. Princeton, N.J.: Educational Testing Service. Retrieved from the National Assessment Governing Board.

WestEd (2006). *Final Report: NAEP Science Assessment 2009*. Retrieved from the National Assessment Governing Board.

WestEd (2010). *Final Report: Technology and Engineering Literacy Assessment for the 2014 National Assessment of Educational Progress.* Retrieved from the National Assessment Governing Board.

WestEd (2021). *Process Report: 2025 NAEP Mathematics Framework Update*. Retrieved from the National Assessment Governing Board.

National Assessment Governing Board Policies

National Assessment Governing Board-NAGB. (1996). *Redesigning the National Assessment of Educational Progress*. Retrieved from https://www.nagb.gov/about-naep/board-policies-for-naep.html.

National Assessment Governing Board-NAGB. (2001b). *NAEP and the No Child Left Behind Act*. Letter to Secretary of Education from the Board Chair, on behalf of the National Assessment Governing Board. Retrieved from https://www.nagb.gov/about-naep/board-policies-for-naep.html.

National Assessment Governing Board-NAGB. (2002a). *Framework Development – Policy Statement*. Retrieved from the National Assessment Governing Board.

National Assessment Governing Board-NAGB. (2002b). *Item Development and Review – Policy Statement*. Retrieved from https://www.nagb.gov/about-naep/board-policies-for-naep.html.

National Assessment Governing Board-NAGB. (2002c). *Long-term Trend – Policy Statement*. Retrieved from https://www.nagb.gov/about-naep/board-policies-for-naep.html.

National Assessment Governing Board-NAGB. (2002d). *Plan for Study of NAEP Sampling*. Retrieved from https://www.nagb.gov/about-naep/board-policies-for-naep.html.

National Assessment Governing Board-NAGB. (2002e). *Policies and Procedures for Complaints Related to the National Assessment of Educational Progress*. Retrieved from https://www.nagb.gov/about-naep/board-policies-for-naep.html.

National Assessment Governing Board-NAGB. (2002f). *Prohibition on Using NAEP to Influence State and Local Standards, Tests, and Curricula – Policy Statement*. Retrieved from https://www.nagb.gov/about-naep/board-policies-for-naep.html.

National Assessment Governing Board-NAGB. (2002g). *Public Access to Test Questions, Item Release, and Confidentiality of Data for NAEP – Policy Statement*. Retrieved from https://www.nagb.gov/about-naep/board-policies-for-naep.html.

National Assessment Governing Board-NAGB. (2002h). *Resolution on Participation of the Commonwealth of Puerto Rico in NAEP – Policy Statement*. Retrieved from https://www.nagb.gov/about-naep/board-policies-for-naep.html.

National Assessment Governing Board-NAGB. (2002i). *Review of the National Assessment of Educational Progress – Policy Statement*. Retrieved from https://www.nagb.gov/about-naep/board-policies-for-naep.html.

National Assessment Governing Board-NAGB. (2003a). *Informing Parents About Participation in NAEP – Policy Statement*. Retrieved from https://www.nagb.gov/about-naep/board-policies-for-naep.html.

National Assessment Governing Board-NAGB. (2004b). Resolution on the NAEP 2009 Reading Framework. Retrieved from https://www.nagb.gov/about-naep/board-policies-for-naep/resolution-09.html.

National Assessment Governing Board-NAGB. (2010a, 2014). *NAEP Testing and Reporting on Students with Disabilities and English Language Learners – Policy Statement.* Retrieved from https://www.nagb.gov/about-naep/board-policies-for-naep.html.

National Assessment Governing Board-NAGB. (2010b). By-Laws of the National Assessment Governing Board. Retrieved from https://www.nagb.gov/governing-board/overview.html.

National Assessment Governing Board-NAGB. (2015). R*esolution on Maintaining NAEP Trends with the Transition to Digital-Based Assessments (DBA)*. Retrieved from https://www.nagb.gov/about-naep/board-policies-for-naep.html.

National Assessment Governing Board-NAGB. (2018b). *Framework Development – Policy Statement*. Retrieved from https://www.nagb.gov/about-naep/board-policies-for-naep.html.

National Assessment Governing Board-NAGB. (2018c). Developing Student Achievement Levels for the National Assessment of Educational Progress – Policy Statement. Retrieved from https://www.nagb.gov/about-naep/board-policies-for-naep.html.

Governing Board Meeting Materials, Minutes and Transcripts

National Assessment Governing Board-NAGB. (2002j). *Materials for the May 2002 Board Meeting.* Retrieved from the NAGB archive upon request.

National Assessment Governing Board-NAGB. (2002k) *Meeting Transcript from the May 17-18, 2002 Board Meeting Open Session*. Retrieved from the National Assessment Governing Board.

National Assessment Governing Board-NAGB. (2002l). *Materials for the August 2002 Board Meeting.* Retrieved from the National Assessment Governing Board.

National Assessment Governing Board-NAGB. (2002m). *Meeting Minutes from the August 2-3, 2002, Board Meeting*. Retrieved from the National Assessment Governing Board.

National Assessment Governing Board-NAGB. (2002n). *Meeting transcript from the August 2-3, 2002, Board Meeting Open Session*. Retrieved from the National Assessment Governing Board.

National Assessment Governing Board-NAGB. (2002o). *Meeting Minutes from the November 15-16, 2002, Board Meeting*. Retrieved from the National Assessment Governing Board.

National Assessment Governing Board-NAGB. (2003b). *Materials for the March 2003 Board Meeting: NAEP 2007 Reading Framework Project Update - Reading Issues Papers Synthesis*. Retrieved from the National Assessment Governing Board.

National Assessment Governing Board-NAGB. (2003c). *Meeting Minutes from the March 7-8, 2003, Board Meeting*. Retrieved from the National Assessment Governing Board.

National Assessment Governing Board-NAGB. (2003d). *Meeting Minutes from the May 16-17, 2003, Board Meeting*. Retrieved from the National Assessment Governing Board.

National Assessment Governing Board-NAGB. (2003e). *Meeting Minutes from the August 1-2, 2003, Board Meeting*. Retrieved from the National Assessment Governing Board.

National Assessment Governing Board-NAGB. (2003f). *Meeting Minutes from the November 14-15, 2003, Board Meeting*. Retrieved from the National Assessment Governing Board.

National Assessment Governing Board-NAGB. (2004c). *Meeting Minutes from the March 5-6, 2004, Board Meeting*. Retrieved from the National Assessment Governing Board.

National Assessment Governing Board-NAGB. (2004d). Meeting transcript from the April 19, 2004, Public Forum on the Draft 2009 Reading Framework. Retrieved from the National Assessment Governing Board.

National Assessment Governing Board-NAGB. (2004e). *Meeting transcript from the August 6, 2004, Board Meeting Open Session*. Retrieved from the National Assessment Governing Board.

National Assessment Governing Board-NAGB. (2004f). *Meeting Minutes from the November 19-20, 2004, Board Meeting*. Retrieved from the National Assessment Governing Board.

National Assessment Governing Board-NAGB. (2005a). *Meeting Minutes from the March 4-5, 2005, Board Meeting*. Retrieved from the National Assessment Governing Board.

National Assessment Governing Board-NAGB. (2005b). *Meeting Minutes from the May 20-21, 2005, Board Meeting*. Retrieved from the National Assessment Governing Board.

National Assessment Governing Board-NAGB. (2005c). *Meeting Minutes from the August 5-6, 2005, Board Meeting*. Retrieved from the National Assessment Governing Board.

National Assessment Governing Board-NAGB. (2005d). *Meeting Minutes from the November 18-19, 2005, Board Meeting*. Retrieved from the National Assessment Governing Board.

National Assessment Governing Board-NAGB. (2006c). *Meeting Minutes from the March 3-4, 2006, Board Meeting*. Retrieved from the National Assessment Governing Board.

National Assessment Governing Board-NAGB. (2006d). *Meeting Minutes from the May 19-20, 2006, Board Meeting*. Retrieved from the National Assessment Governing Board.

National Assessment Governing Board-NAGB. (2006e). *Meeting Minutes from the August 4-5, 2006, Board Meeting*. Retrieved from the National Assessment Governing Board.

National Assessment Governing Board-NAGB. (2016). Meeting minutes from the November 2, 2016 Board Meeting. Retrieved from https://www.nagb.gov/governing-board/quarterly-board-meetings/archive.html.

National Assessment Governing Board-NAGB. (2017). Meeting minutes from the March 2, 2017 Board Meeting. Retrieved from https://www.nagb.gov/governing-board/quarterly-board-meetings/archive.html.

National Assessment Governing Board-NAGB. (2018c). *Materials for the March 2018 Board Meeting: 14-framework-development-policy-revision March 2018*. Retrieved from https://www.nagb.gov/governing-board/quarterly-board-meetings/2018/2018-march.html.

Appendix A
Historical Context[7] for Framework Development of the National Assessment of Educational Progress

Dates	Historical Activities	Assessment Development
1960-70's ECS era	• The 1960s were a formative time for the development of NAEP. (NCES website: https://nces.ed.gov/nationsreportcard/about/newnaephistory.aspx#beginning) • 1964-68 – The Education Commission of the States (ECS), managed and conducted the first national assessments. They established an Exploratory Committee for the Assessment Progress in Education (ECAPE) and established a National Assessment Planning Project. • 1969 – First national assessment data collection, now known as the National Assessment of Educational Progress (NAEP), was the 1969 trial assessment of the citizenship, science, and writing performance of 17-year-old in-school students in the spring of that year. In the fall, 9- and 13-year-old students as well as out-of-school 17-year-olds were assessed. • The frameworks for the early NAEP utilized a content-by-process matrix to develop items for the assessment, most of which were released with the reporting.	The assessment was based on a content-by-process matrix set of objectives developed by representatives for the Education Commission of the States (ECS).
1976-1988 Early national assessment and NAEP era[8]	• The Comptroller General (GAO) Report, Make NAEP More Useful, was released in 1976. • The original national assessment legislation in 1978 brought changes to the oversight and organization of the assessment (now NAEP) and specified an Assessment Policy Committee of 17 members (the precursor to the National Assessment Governing Board). • A major study critical of NAEP (Wirtz & Lapointe, 1982) said NAEP was underdeveloped and underutilized, and of apparently negligible influence. • In 1983, a non-profit organization (Educational Testing Service, ETS) was selected as the NAEP Contractor and a redesigned assessment (more sophisticated sampling, scaling & analyses) was developed. • The 1986 reauthorization of the Elementary and Secondary Education Act (ESEA) included provisions for voluntary state assessments and referred to the national assessment as the National Assessment of Educational Progress, the name that continues today. It also, continued the requirement for an Assessment Policy Committee of 19 members, adding two additional members representing elementary and secondary school principals.	Because of the desire by some state members of ECS, two policy pushes changed NAEP. (1) Voluntary participation and reporting on states (2) A move to an objectives-based approach instead of the content-by-process matrix approach previously used for the assessments.

[7] A thorough examination of the establishment and early years of the National Assessment Governing Board can be found in the report, *Overseeing the Nation's Report Card: The Creation and Evolution of The National Assessment Governing Board (NAGB)*. Vinovskis, M.A. (1998). http://www.nagb.org/publications/95222.pdf.

[8] A thorough examination of the evolution of the National Assessment of Educational Progress is found in the book, The Nation's Report Card: Evolution and Perspectives (Jones & Olkin, 2004).

Appendix A
Historical Context[7] for Framework Development of the National Assessment of Educational Progress

Dates	Historical Activities	Assessment Development
	• In 1986, the Trial State Assessments were begun in cooperation with the ECS and the Southern Region Education Board (SREB). The planning for this effort was advised by a mathematics content committee which wanted to develop an objectives-based approach that could lead instruction instead of the content-by-process matrix approach previously used for the assessments.	
1988 – Present NAEP-NAGB era	• The 1988 reauthorization of ESEA & NAEP (Hawkins-Stafford Act, 1988) included provisions the establishment of a separate policy board of 24 members, the National Assessment Governing Board. The Governing Board was to be of similar composition to the Assessment Advisory Committee (specifying the additional inclusion of two curriculum specialists, a non-public educator, two governors, and an ex officio member). It also included a requirement to set feasible achievement goals – achievement levels, as they have come to be called. • The 1994 reauthorization of ESEA, Improving America's Schools Act, updated the membership of the Board to 26 by adding one more test and measurement expert and delineating the general public representatives as including two parent representatives (one additional). • The 2001 reauthorization of ESEA required state participation in NAEP Reading and Mathematics if the state received Title I funds, and called for biennial testing of Reading and Mathematics, as well as the school accountability provision known as adequate yearly progress. The content and all aspects of NAEP were now being scrutinized much more strenuously. • A 2003 authorization of the NAEP legislation provided for the voluntary inclusion of urban district level reports, included additional funding for their participation which increased from six in 2003 to 27 presently. • The 2015 reauthorization of ESEA, the Every Student Succeeds Act (ESSA), maintained the requirements for NAEP, including required state participation and biennial testing in Reading and Mathematics.	The National Assessment Governing Board was established. The 1988 legislation included provisions for trial assessments in mathematics at 8th grade (1990) and 4th and 8th grade (1992) and in reading at 4th grade (1992). The first assessment frameworks were developed for these grades/subject areas. The policy and practices for developing the NAEP Assessment Frameworks was now the responsibility of the Governing Board.

Appendix B
Governing Board Duties in Legislation Over Time

The National Assessment Governing Board was authorized by Federal legislation in 1988 and has been reauthorized twice. The duties of the National Assessment Governing Board were initially authorized in the legislation establishing the Board in 1988 and have remained quite stable throughout the periodic reauthorizations, the latest of which is P.L.107-279 (2002).[9] This law provides authorization for both the Governing Board (Section 302) and NAEP (Section 303). Appendix B presents only the Governing Board section, but does contain references to the NAEP section.

In each iteration of the law the subsections have been rearranged slightly and language was added, deleted or clarified. The requirements, however, have remained essentially the same. Two unique elements were added in 2002. The first, 302(e)(1)(D), called for an inclusive review process for the assessment that is now addressed both by a Governing Board policy (NAGB, 2002i)[10] and by the extensive external reviews conducted before each framework is adopted. The other addition, 302(e)(1)(F), provided a linkage to the NAEP section. Appendix B presents all of the legal requirements in a side-by-side arrangement. Each requirement is presented with the legal numbering used in each reauthorization and identifies changes that occurred in each revision.

Appendix B **Governing Board Duties in Legislation Over Time** (New wording is underlined. Notes in red are not included in the legislation.)		
1988 P.L. 100-297 **Sec. 3403. (6)(A)**	**1994 P.L. 103-382** **SEC. 412 (e)(1)**	**2002 P.L. 107-279** **SEC. 302. (e)(1)[11]**
6(A) In carrying out its functions under this subsection, the Board shall be responsible for-	(1) In General. -- In carrying out its functions under this section the Board shall	(1) IN GENERAL- In carrying out its functions under this section the Assessment Board shall—
(i) selecting subject areas to be assessed (consistent with paragraph (2)(A));	(A) select subject areas to be assessed (consistent with section 411(b)(1));	(A) select the subject areas to be assessed (consistent with section 303(b));

[9] The 1988 authorization, Public Law 100-297, was part of the *Hawkins-Stafford Elementary and Secondary School Improvement Amendments of 1988.* The 1994 reauthorization, Public Law 103-382, was part of the *Improving America's Schools Act of 1994*.

[10] The Governing Board policy statement, *Review of the National Assessment of Educational Progress*, adopted August 3, 2002, included six guiding principles that describe expectations for the rigorous review of the National Assessment of Educational Progress and actions of the Governing Board.

[11] Public Law 107-279, the Education Sciences Reform Act of 2002, provided amendments to the original No Child Left Behind Act of 2002, Public Law 107-110.

Appendix B Governing Board Duties in Legislation Over Time (New wording is underlined. Notes in red are not included in the legislation.)		
1988 P.L. 100-297 Sec. 3403. (6)(A)	**1994 P.L. 103-382 SEC. 412 (e)(1)**	**2002 P.L. 107-279 SEC. 302. (e)(1)[11]**
(ii) identifying appropriate achievement goals for each age and grade in each subject area to be tested under the National Assessment;	(B) develop appropriate student performance levels as provided in section 411(e);	(B) develop appropriate student achievement levels as provided in section 303(e);
(iii) developing assessment objectives; (iv) developing test specifications;	(C) develop assessment objectives and test specifications through a national consensus approach which includes the active participation of teachers, curriculum specialists, local school administrators, parents, and concerned members of the public; Note: Consensus process was incorporated here from 1998 section (E).	(C) develop assessment objectives consistent with the requirements of this section and test specifications that produce an assessment that is valid and reliable, and are based on relevant widely accepted professional standards; Note: Reference to a consensus approach was moved from the NAGB, Section 302, to the NAEP Section 303(b)(3)(B)(II) but still applies to the content of NAEP for which the Board is responsible.
		(D) develop a process for review of the assessment which includes the active participation of teachers, curriculum specialists, local school administrators, parents, and concerned members of the public;
(v) designing the methodology of the assessment;	(D) design the methodology of the assessment, in consultation with appropriate technical experts, including the Advisory Council established under section 407;	(E) design the methodology of the assessment to ensure that assessment items are valid and reliable, in consultation with appropriate technical experts in measurement and assessment, content and subject matter, sampling, and other technical experts who engage in large scale surveys;

Appendix B Governing Board Duties in Legislation Over Time (New wording is underlined. Notes in red are not included in the legislation.)		
1988 P.L. 100-297 Sec. 3403. (6)(A)	**1994 P.L. 103-382 SEC. 412 (e)(1)**	**2002 P.L. 107-279 SEC. 302. (e)(1)[11]**
		(F) consistent with section 303, measure student academic achievement in grades 4, 8, and 12 in the authorized academic subjects;
(vi) developing guidelines and standards for analysis plans and for reporting and disseminating results;	(E) develop guidelines and standards for analysis plans for reporting and disseminating results;	(G) develop guidelines for reporting and disseminating results; Note: 'Standards for analysis plans" was removed from this section.
(vii) developing standards and Procedures for interstate, regional and national comparisons; and	(F) develop standards and procedures for interstate, regional, and national comparisons; and	(H) develop standards and procedures for regional and national comparisons; Note: 'interstate' was removed from this section.
(viii) taking appropriate actions needed to improve the form and use of the National Assessment.	(G) take appropriate actions needed to improve the form and use of the National Assessment.	(I) take appropriate actions needed to improve the form, content, use, and reporting of results of any assessment authorized by section 303 consistent with the provisions of this section and section 303; and
		(J) plan and execute the initial public release of National Assessment of Educational Progress reports. The National Assessment of Educational Progress data shall not be released prior to the release of the reports described in subparagraph (J).

Appendix B Governing Board Duties in Legislation Over Time (New wording is underlined. Notes in red are not included in the legislation.)		
1988 P.L. 100-297 Sec. 3403. (6)(A)	**1994 P.L. 103-382 SEC. 412 (e)(1)**	**2002 P.L. 107-279 SEC. 302. (e)(1)[11]**
(B) The Board may delegate any functions described in subparagraph (A) to its staff.	(2) Delegation. -- The Board may delegate any of the Board's procedural and administrative functions to its staff.	(2) DELEGATION- The Assessment Board may delegate any of the Assessment Board's procedural and administrative functions to its staff.
(C) The Board shall have final authority on the appropriateness of cognitive items.	(3) Cognitive Items. -- The Board shall have final authority on the appropriateness of cognitive items.	(3) ALL COGNITIVE AND NONCOGNITIVE ASSESSMENT ITEMS- The Assessment Board shall have final authority on the appropriateness of all assessment items.
(D) The Board shall take steps to ensure that all items selected for use in the National Assessment are free from racial, cultural, gender, or regional bias.	(4) Prohibition Against Bias. -- The Board shall take steps to ensure that all items selected for use in the National Assessment are free from racial, cultural, gender, or regional bias.	(4) PROHIBITION AGAINST BIAS- The Assessment Board shall take steps to ensure that all items selected for use in the National Assessment are free from racial, cultural, gender, or regional bias and are secular, neutral, and non-ideological.
(E) Each learning area assessment shall have goal statements devised through a national consensus approach, providing for active participation of teachers, curriculum specialists, local school administrators, parents and concerned members of the general public.	(5) Technical. -- In carrying out the duties required by paragraph (1), the Board may seek technical advice, as appropriate from the Commissioner and the Advisory Council on Education Statistics and other experts. Note: the stakeholder list and consensus approach were moved to Section 412 (e)(1)(C).	(5) TECHNICAL- In carrying out the duties required by paragraph (1), the Assessment Board may seek technical advice, as appropriate, from the Commissioner for Education Statistics and other experts.

Appendix B Governing Board Duties in Legislation Over Time (New wording is underlined. Notes in red are not included in the legislation.)		
1988 P.L. 100-297 Sec. 3403. (6)(A)	**1994 P.L. 103-382 SEC. 412 (e)(1)**	**2002 P.L. 107-279 SEC. 302. (e)(1)[11]**
	(6) Report. -- Not later than 90 days after an evaluation of the student performance levels under section 411(e), the Board shall make a report to the Secretary, the Committee on Education and Labor of the House of Representatives, and the Committee on Labor and Human Resources of the Senate describing the steps the Board is taking to respond to each of the recommendations contained in such evaluations.	(6) REPORT- Not later than 90 days after an evaluation of the student achievement levels under section 303(e), the Assessment Board shall make a report to the Secretary, the Committee on Education and the Workforce of the House of Representatives, and the Committee on Health, Education, Labor, and Pensions of the Senate describing the steps the Assessment Board is taking to respond to each of the recommendations contained in such evaluation. Note: This change provides an update to the House and Senate Committee names at the time.

Appendix C
Framework Development Policy Revision 2002 to 2018

The NAGB Framework Development Policy was developed initially in 2002 and revised 16 years later in 2018. The original policy was based on the accepted best practice NAGB had been following since 1988. Although many changes occurred in assessment methodologies and education policy, the 2002 policy served the Board will, even with some redundancies and procedural details not usually found in policies. Revisions to the Framework Development Policy in 2018 addressed these issues.

In addition to some minor reorganization and rewording, primary distinctions between the 2002 and 2018 editions included four changes that are discussed in more detail within this report: (1) updating frameworks, (2) reviewing frameworks, (3) participants/stakeholders, and (4) framework panels/committees. Additionally, the current policy maintains a focus on the overarching principles to be followed, with the details and procedures moved to procedural documents and requirements for contractors.

Basically, the two versions address the same content, although they are arranged somewhat differently and with fewer procedural elements in 2018. The summary below compares the principles in each version, in a side-by-side manner, and summarizes the changes that were implemented in 2018 (shown in red). Italicized words show *2002 language that was changed* and underlining shows new wording in 2018. Of course, this summary does not capture all changes as the text under each principle also was revised in a similar manner to remove redundancy and procedures, and for more clarity and efficiency in wording. A few are noted in the table. The only substantive change is the addition of a framework update process which is not intended to be as extensive as the development of a new framework.

Policy Elements	5/18/02 Framework Development Policy	03/18/18 Framework Development Policy
Preface: Purpose	It is the policy of the National Assessment Governing Board to conduct a comprehensive, inclusive, and deliberative process to determine the content and format of all subject area assessments under the National Assessment of Educational Progress (NAEP).	No change
Preface: Desired Outcome	Objectives developed and adopted by the Governing Board as a result of this process shall be used to produce NAEP assessments that are valid and reliable, and that are based on widely accepted professional standards. *The process shall include the active participation of educators, parents, and members of*	The primary result of this process shall be an assessment framework (hereafter, "framework") with objectives to guide development of NAEP assessments for students in grades 4, 8, and 12 that are valid, reliable, and reflective of widely accepted professional standards.

Policy Elements	5/18/02 Framework Development Policy	03/18/18 Framework Development Policy
	the general public. The primary result of this process shall be an assessment framework to guide NAEP development at grades 4, 8, and 12	Rewording & reorganization of italicized details
Preface: Process	*The process shall include the active participation of educators, parents, and members of the general public.*	This process detail is contained in the introduction and in Principle 2
Preface: Board Delegation to ADC	The Governing Board, through its Assessment Development Committee, shall *carefully* monitor the framework development process to ensure *that all Governing Board policies are followed; that the process is comprehensive, inclusive, and deliberative; and* that the final Governing Board-adopted framework, specifications, and background variables documents *are congruent with the Guiding Principles, Policies, and Procedures that follow.*	The Governing Board, through its Assessment Development Committee, shall monitor the framework development and update processes to ensure that the final Governing Board-adopted framework, specifications, contextual variables documents, and their development processes comply with all principles and guidelines of the Governing Board Framework Development Policy. Rewording, reorganization of italicized details
Intro: Legal Authorization	P.L. 107-279 Section 302(e)(1) and Restatement of law requirements	No change in citation, but requirements not explicitly listed
Intro: Involvement of Stakeholders	Stakeholders were given in the restatement of the law	Expanded description of compliance with the law and identification of specific stakeholders
Intro: Professional Standards	Adherence to standards acknowledged with current publications cited.	No change except for the editions cited
The Principles	Seven (7) principles included with *policies and procedures* for implementing each. Order is shown in relation to the 2018 policy.	Six (6) principles included with guidelines for implementation. Essentially the same principles and guidelines as in 2002 (with some combining and rewording), titles were added to each principle.
	1. The Governing Board is responsible for developing an *assessment* framework for each NAEP subject area. The framework shall define the scope of the domain to be measured by delineating the knowledge and skills to be tested at each grade, the format of the NAEP assessment, and preliminary achievement level descriptions. *5. Through the framework development process, preliminary*	1. Elements of Frameworks: The Governing Board is responsible for developing a framework for each NAEP assessment. The framework shall define the scope of the domain to be measured by delineating the knowledge and skills to be tested at each grade, the format of the NAEP assessment, and the achievement levels. Define what will be tested and how, as well as

Policy Elements	5/18/02 Framework Development Policy	03/18/18 Framework Development Policy
	achievement level descriptions shall be created for each grade being tested. These preliminary descriptions shall be an important consideration in the item development process and will be used to begin the achievement level setting process.	how much students should know at each achievement level. 2002 Principle 5 incorporated with this principle
	2. The Governing Board shall develop an assessment framework through a comprehensive, inclusive, and deliberative process that involves the active participation of *teachers, curriculum specialists, local school administrators, parents, and members of the public*. (Note: This 2002 principle contained guidelines for panel members which did not explicitly require classroom experience for the subject area. *"At least 30 percent of this committee shall be composed of users and consumers in the subject area under consideration.")*	2. Development and Update Process: The Governing Board shall develop and update frameworks through a comprehensive, inclusive, and deliberative process that involves active participation of stakeholders. Addition of 'update'; redundancy in wording reduced; and move of stakeholders list to the introduction This principle more clearly identified the various panels, their purposes, shared membership expectation, classroom teaching experience (20%) in the subject area, and expected discussions about the impact on trend reporting when content changes.
	7. *NAEP assessment frameworks and test specifications generally shall remain stable for at least 10 years*.	3. Framework Review: Reviews of existing frameworks shall determine whether an update is needed to continue valid and reliable measurement of the content and cognitive processes reflected in evolving expectations of students. The addition of this principle provides an emphasis on the work of reviewing/updating frameworks and contains guidelines about reviewing/updating frameworks at least once every 10 years.
	3. The framework development process shall take into account state and local curricula and assessments, widely accepted professional standards, exemplary research, international standards and	4. Resources for the Process: Framework development and update processes shall take into account state and local curricula and assessments, widely accepted professional standards, exemplary research, international standards

Policy Elements	5/18/02 Framework Development Policy	03/18/18 Framework Development Policy
	assessments, and other pertinent factors and information.	and assessments, and other pertinent factors and information. Addition of 'update' This principle contains expanded guidance on ways to identify curricular content issues in the field.
	6. The specifications document shall be developed *during the framework process* for use by NCES *and the test development contractor* as the blueprint for constructing the NAEP assessment and items *in a given subject area.*	5. Elements of Specifications: The specifications document shall be developed for use by NCES as the blueprint for constructing the NAEP assessment and items. Reduce unnecessary words
	4. The Governing Board, through its Assessment Development Committee, shall *closely* monitor all *steps in the* framework development *process*. The result of this process shall be recommendations for Governing Board action in the form of three key documents: the assessment framework; assessment and item specifications; and *background* variables that relate to the subject being assessed.	6. Role of the Governing Board The Governing Board, through its Assessment Development Committee, shall monitor all framework development and updates. The result of this process shall be recommendations for Governing Board action in the form of three key documents: the framework; assessment and item specifications; and contextual variables that relate to the subject being assessed. Addition of 'update' & change of term from background to contextual variables. This principle contains guidelines about balancing the maintenance of trends with including new content.

Appendix D
Decision Points and Roles for Framework Development

Appendix D highlights the major questions/decisions and other subordinate ones needed for framework development, approval, and adoption by the Board. Also included are the likely roles and involvement of contractors and external reviewers, that is, stakeholders and the general public. Many smaller decisions and steps are behind these major decision points, but cannot be captured in this simplistic presentation. While the decision points are presented in an orderly manner, they may not always be implemented in the chronology implied by this list.

Appendix D Decision Points and Roles for Framework Development				
Activity	**Full Board**	**Assessment Development Committee***	**Contractor Activities**	**External Reviews**
① Should a framework revision or updating be considered?		- Identify need for review - Recommend going forward with review		
Experts make presentations to the Assessment Development Committee.		- Convene experts - Review relevant research		
Formulate a recommendation about update/replacement of framework and draft charge		- Formulate recommendation - Draft charge		
② Is a new framework or update needed?	**Review-Approve charge**			Via public information and open meetings
Conduct procurement and select contractor to manage workload		- Issue procurement - Review proposals - Initiate Contract - Monitor*	- Begin contract and implement as required	Via public postings and notices
<u>Visioning Panel Deliberations</u> (includes Development Panel members) <u>Purpose</u>: to provide the initial high-level guidance about the state of the discipline and recommendations (guidelines or goals) for developing the framework		- Review/approve panels - Provide charge & direction - Review guidelines and goals - Regularly monitors progress*	- Identify panel chair & participants - Facilitate Process - Regularly reports progress	

Appendix D Decision Points and Roles for Framework Development				
Activity	**Full Board**	**Assessment Development Committee***	**Contractor Activities**	**External Reviews**
Development Panel Deliberations (overlap with Visioning Panel) Purpose: to draft the three project documents, engage in the detailed deliberations about how issues outlined by the Visioning Panel should be reflected in the framework		- Regularly monitors progress*	- Identify panel chair & participants - Facilitate Process - - Regularly reports progress	
Technical Experts Involved Purpose: to uphold the highest technical standards and as a resource to the framework panels to respond to technical issues raised during panel deliberations.		- Participate as needed* - Regularly monitors progress	- Identify participants - Facilitate Process - Produce Reports	
③ Is the draft framework ready to be evaluated by external reviewers? Public comment will be sought from various segments of the population to reflect many different views, and targeted feedback will be solicited from those employed in the content area under consideration, especially educators and policy makers.		- Regularly monitors progress* - Recommend going forward with external review and public comment	- Provide drafts & make revisions - Produce Reports	Via public information and open meetings
Framework – Define what, how and how much of the content domain is to be included on the NAEP assessment, and desirable levels of achievement		- Monitor* - Approve	- Facilitate Process - Produce Reports	
④-What feedback should be incorporated in Framework? The Framework Development Project must consider the policy impact and provide advice about changes needed based on the feedback, weighing all of the issues.		- Recommend activities -- Participate in activities - Review feedback - Recommend next steps	- Identify participants - Facilitate Process -- Incorporate feedback - Produce Reports	Provide verbal and written comments about the framework & other issues

Appendix D Decision Points and Roles for Framework Development				
Activity	**Full Board**	**Assessment Development Committee***	**Contractor Activities**	**External Reviews**
⑤ Should the framework be adopted and implemented? After considering the revisions made to the framework, the Board formally adopts the framework and approves the next steps.	**- Review** **- Approve or modify**	- Recommend adoption - Identify next steps (item specification and contextual variables)		
5.2 (Later) Item specifications – the blueprint for constructing the NAEP assessment in sufficient detail for developing high-quality questions based on the framework	**- Review** **- Approve or modify**	- Monitor* - Approve	- Facilitate Process - Produce Reports	
5.2 (Later) Contextual variables **–** recommendations on related contextual variables to be collected from students, teachers, and school administrators	**- Review** **- Approve or modify**	- Monitor* - Approve	- Facilitate Process - Produce Reports	
Implement Assessment in collaboration with NCES.		- Monitor* - Approve items	NCES contractors	

* Although the Assessment Development Committee has the primary role for oversight of framework development/updating processes, other committees of the Board and NCES are involved as needed. Typically, the Committee on Standards, Design, and Methodology (COSDAM) is involved in technical issues (scoring, scaling, trend reporting, etc.), the Reporting and Dissemination Committee (R & D) is involved in discussions about reporting and contextual data collection, and the National Center for Education Statistics (NCES) is involved in issues related to item development, test construction, test scoring, data analysis, and reporting.

TABLE OF CONTENTS

Framework Development Processes

Under the leadership of the Assessment Development Committee (ADC), the Board updated its Framework Development policy in March 2018. One of the primary revisions reflected in the current policy was to account for the process of updating existing frameworks; the previous policy emphasized the development of new frameworks and contained little explicit guidance on monitoring and revising frameworks without starting from scratch.

The current policy has now been in place for three years and has guided the updates of the NAEP Mathematics Framework (adopted by the Board in November 2019) and the NAEP Reading Framework (currently under Board consideration). Leadership of ADC and COSDAM have identified a need to evaluate the extent to which the current policy and procedures are meeting the intended goals and determine whether any aspects need to be revisited.

To support a joint ADC-COSDAM session on this topic, Board staff commissioned two papers:

- As a consultant, former Governing Board Executive Director Cornelia Orr synthesized historical information on NAEP framework development, including:
    - Initial NAEP legislation and how it has evolved in its requirements for framework processes and outcomes
    - Board policy and how it has evolved in its requirements for framework processes and outcomes
    - Policy contexts and professional standards that have shaped framework processes
    - Procedures the Board has used to adhere to law/policies/professional standards
    - Description of how framework procedures have evolved over time
    - Reflections on why framework procedures have evolved the way they have, in light of policy contexts, professional standards, laws, etc.

- As part of the Board's contract for Technical Support in Psychometrics, Assessment Development, and Preparedness for Postsecondary Endeavors, the Center for Assessment (under subcontract to the Human Resources Research Organization) prepared information on how NAEP framework development relates to procedures for developing other assessments, including:
    - Summarizing elements of framework processes for state, national, and international assessments
    - Comparing these framework processes, articulating similarities and differences
    - Listing and describing best practices in framework processes, in general
    - Evaluating which best practices are appropriate for NAEP's legislative mandates, e.g., curricular-neutrality, pedagogical-neutrality, etc.
    - Describing how current NAEP framework processes reflect or do not reflect these NAEP-appropriate best practices

The papers have been completed and will be the focus of a joint ADC-COSDAM meeting that will occur in June. A copy of the first paper is in Attachment B. A copy of the second paper is attached hereto. The ADC will have the opportunity to discuss its initial feedback on both papers at the May 7 ADC meeting.

2021 No. 022

Assessment Framework Development Processes

Final Report

Prepared for:	National Assessment Governing Board 800 North Capitol Street N.W., Suite 825 Washington DC 20002	**Prepared under:**	ED-NAG-17-C-0002
Authors:	Will Lorié Brian Gong Center for Assessment	**Date:**	March 31, 2021

Headquarters: 66 Canal Center Plaza, Suite 700, Alexandria, VA 22314 | Phone: 703.549.3611 | humrro.org

Assessment Framework Development Processes

Table of Contents

List of Tables

Assessment Framework Development Processes

Executive Summary

By describing what is to be assessed and how to assess it, assessment frameworks play a pivotal role in testing programs. In February 2021, the National Assessment Governing Board (Governing Board), which oversees the National Assessment of Educational Progress (NAEP), invited a technical memo to discuss the processes that large-scale assessment sponsors initiate, conduct, or commission to develop, review, or update assessment frameworks. The Governing Board was particularly interested in how the framework processes of other large-scale assessment programs and framework process standards/best practices might inform the framework processes for the NAEP.

In this technical memo, we present an organizer that enumerates the elements of assessment processes. These elements and their components classify all the decisions relevant to shaping framework processes. We developed the organizer while reviewing framework process-relevant documents for NAEP and other testing programs, such as assessment frameworks themselves, technical reports, and process reports.

Although there are no recognized standards for framework processes, we also reviewed standards or other widely consulted sources that might address aspects of framework processes, such as the *Standards for educational and psychological testing* (AERA/APA/NCME, 2014). Apart from documenting what is available regarding framework process best practices, this review informed the organizer.

Our review has two significant implications for NAEP and similar large-scale testing programs. The elements of framework processes imply a set of options that will substantially shape framework processes for a program, the resulting framework, and ultimately the resulting assessment. Assessment sponsors can make choices concerning these options, delegate those choices, or a combination.

We conclude that a sound principle of best practice in this area is for test sponsors to be aware of the framework process elements/components and their associated options. Moreover, test sponsors should be deliberate in their specification of requirements. They should provide a rationale for their choices.

A second implication is that much of the quality of the framework product depends upon the *process* used to develop the framework. Because there are few established criteria to evaluate the quality of assessment frameworks, it becomes more essential that the processes be specified well and carried out well. Programs should document, evaluate, and try to improve their framework development processes.

For NAEP and all the programs reviewed, this takes on greater importance when multiple assessment frameworks are developed and there is a desire to have similar features, specificity, and/or process quality across frameworks. Consistency in product and/or process will be a matter of deliberate design and careful implementation.

We end with seven recommendations regarding further work in this area. They include investigations of:

1. The structure of domain descriptions across different assessment frameworks.
2. The different kinds of sources informing assessment frameworks.
3. The structure of assessment objectives across different assessment frameworks.
4. Different approaches to ensuring curriculum neutrality in assessment framework development.
5. The scope of the assessment design component across different assessment frameworks.
6. Best practices for implementation fidelity evaluation and documentation for group-based processes.
7. Best practices in effective committee work, especially processes for generating, discussing, and resolving issues.

Background and Approach

Assessment Frameworks

Every modern assessment program has some definition of the intended construct to be measured, including a definition of the domain. That is typically referred to as the content framework. In addition, there will be a specification of what and how to assess to produce sufficient evidence to support the intended assessment interpretations and uses. That is typically referred to as test specifications or the test blueprint. In the NAEP program, an "assessment framework" is produced that combines definition of the content and the essential assessment specifications. The assessment framework is produced under the direction of the Governing Board, typically by committees of persons with desired expertise. The assessment frameworks specify the basic architecture of the assessment to be developed.

Statement of Work

The Center and the Governing Board developed the following statement of work at the outset of the program. It is presented here without edits.

The National Assessment Governing Board (Governing Board) invited a paper to discuss how framework/standards development processes are conducted to specify the content to be covered in an assessment (hereafter, noted as "framework processes"). In consultation with HumRRO and the National Center for the Improvement of Educational Assessment (Center), Governing Board agreed that the paper should:

1. Summarize elements of framework processes for state, national, and international assessments.
2. Compare these framework processes, articulating similarities and differences.
3. List and describe common practices for developing frameworks.

4. Evaluate which practices are appropriate for NAEP's legislative mandates, e.g., curricular-neutrality, pedagogical-neutrality, etc.

5. Describe how current NAEP framework processes reflect or do not reflect these NAEP-appropriate practices.

6. Recommend possible additional work to inform Board considerations.

Approach

To accomplish the six goals of this paper as delineated in the statement of work, we began by reviewing initial documentation provided by Governing Board. Next, we read assessment frameworks and related documentation for selected assessment programs. A set of guiding questions (presented below) informed our reading.

We selected assessment programs based on their potential relevance to the NAEP context, which assesses achievement of students' domain-specific knowledge and skills across populations governed by different educational standards or curricula.

Next, we discussed dimensions that can describe different framework process choices and their interrelationships across assessment programs. Then, we created an organizer for these choices. In the process, we proposed working definitions of key terms.

We posit that assessment program sponsors should make conscious choices concerning these features. NAEP's mandates and traditions have implications for these choices, especially when compared to other programs' framework processes. Our recommendations build upon these implications.

Scope of the Review of Framework Processes

Our review of framework processes is limited to large-scale content area-based or skills-based assessments in K-12, with mandates issued by national, (U.S.) state, or international agencies. We focused on relatively recent assessment programs (or the most recent framework processes of those programs) with publicly available documentation. We shared a list of programs to review with the Governing Board early in the project through an annotated outline. Our list is presented here as originally communicated to the Governing Board:

- NAEP
- A national assessment operating in a setting where there is a national curriculum, such as the U.K.
- A national assessment operating in a multi-curricular setting like the U.S. (if there is one)
- SAT
- ACT
- An assessment for states responding to a multi-state or national-level consensus, e.g., Common Core State Standards (CCSS)-based or Next Generation Science Standards (NGSS)-based content standards for assessment

- A non-consortium state assessment example where the state developed content standards and explicitly did not substantially adopt a widely used set of content standards
- A potential state example operating under very different constraints
- Two leading international assessment programs operating under very different conceptual relationships to curriculum
  - Programme for International Student Assessment (PISA)
  - Trends in Mathematics and Science Study (TIMSS)

We subsequently identified a non-U.S.-based national program operating in a multi-curricular setting like the U.S., with the relevant documentation publicly available. This program is the Pan-Canadian Assessment Program (PCAP). Because of the similarity between the PCAP and NAEP contexts, we conducted a relatively more thorough review of PCAP and included that review as a case study in an appendix.

We did not locate a NAEP-like program in the U.K. We subsequently reconsidered the relevance of national assessment programs in countries where there is a national curriculum.[1] Our final list excluded state testing programs that develop their own standards outside the context of a consortium. In general, state testing programs do not report much about the processes they use to derive their assessment frameworks. A useful proxy may be how state curriculum or academic content standards are developed and adopted. A review of these, however, was beyond the scope of this technical memo.

Guiding Questions for Review of Framework Processes

The following questions guided our review of framework processes for NAEP and other programs.

1. What documentation is publicly available concerning framework processes for large-scale assessments, and how thoroughly does it describe those processes?

2. What are the different legislative or other mandates for framework processes, and what do these directly or indirectly imply about those processes?

3. What are the processes for selecting steering group members and authors of assessment frameworks?

4. What are the processes for securing internal agreement during authorship, and how is dissent managed?

5. What are the parameters governing review by stakeholders or other constituencies, and how are differences of opinion managed in the review process?

6. What standards or other external guidance, if any, are referenced or consulted to guide framework processes?

[1] This is why, for example, we did not investigate Australia's National Assessment Program – Literacy and Numeracy (NAPLAN). Australia has a national curriculum and so NAPLAN would not have to contend with curricular neutrality in the same way as NAEP.

7. What are common features of framework processes across all programs, and what appears to be unique to programs or programs with specific characteristics?

8. Which features of framework processes seem most appropriate to those assessment programs with a legislative mandate similar to NAEP?

9. To what extent have NAEP framework processes reflected those features?

Definitions

The language associated with framework development processes are not often very precise, therefore we articulate some working definitions below: An *assessment framework* is a document or set of documents containing (at minimum) an assessment-oriented description of the domain assessed. A domain description is ass*essment-oriented* if it can guide assessment developers to produce assessment blueprints, item and test specifications, and similar intermediate products of assessment development. An assessment framework may also contain descriptions of construct claims (such as achievement level descriptions), specific assessment design elements (such as blueprints or acceptable item formats), and process documentation (a report of how the framework was developed). Frameworks typically also include special requirements, constraints, or criteria. (See also Martineau, Dadey, & Marion, 2018, p. 4).

A framework process is a process that results in either an approved assessment framework, an update or revision to a framework, or a decision to revise, replace, or leave a framework in place. Thus, for example, a framework process might be instantiated to determine to what extent a framework is still relevant.

An *element of a framework process* is a significant dimension of a framework process. We derived a list of elements after reviewing several assessment frameworks and related documents. We identified six elements: Initiating conditions, work product, work process, owner, timeframe, and approval.

A *specification of requirements* is a document (or a part of one) that states at least one constraint or requirement of at least one element of a framework process. By contrast, elements of framework processes may be *reported* with or without reference to any requirements. A hypothetical example of a requirements specification, which might be found in a statement of work, “The framework must include four achievement levels with descriptions of what students know and can do at the upper three levels.”

Mandate is an overarching term that covers laws, memorandums of understanding, charters, and other agreements. Even though we classify mandates as “documents,” a mandate may be verbal – for example, a charge delivered by an authority to a group in person counts as a mandate. A mandate does not have to be “documented.” A hypothetical example of an undocumented mandate is a program sponsor telling a working group to prioritize content standards above studies of how content is actually taught, assuming this instruction does not make it into any document.

Methodology

Overview of Methodology

Our goal was to develop an organizer to describe framework processes. We proceeded by reviewing the initial (NAEP) documentation provided by the Governing Board. We discussed internally salient dimensions or aspects of these processes, compared to what we knew of framework processes from other assessment programs. We drew up a list of programs to review and then scanned available documentation for references to framework processes. We continued to refine our articulations of the general "elements" of framework processes, developing some definitions to guide our approach. We did an in-depth review of one additional assessment program, after which we finalized our organizer. Finally, we collated and summarized what we could find concerning professional standards for framework processes.

Initial Documentation

We received documentation relevant to NAEP framework processes at the outset of this project. These documents include the NAEP law, NAEP's framework development policy statement, select NAEP frameworks, design documents, schedules, and studies relevant to framework processes. These documents are listed in References and Appendix A and are denoted by a single asterisk.

Rationale for Selection of Assessment Programs to Review

We looked at assessments operating at national, state, and international levels. Our goal was to select assessment programs with contexts like NAEP. Specifically, we sought out achievement assessment programs where test-takers learn through different curricula and possibly under educational authorities with varying content standards.

There are two major programs with these characteristics at the international level – the Programme for International Student Assessment (PISA) and Trends in Mathematics and Science Study (TIMSS). At the national level outside of the U.S., we discovered one other national assessment program operating in contexts like NAEP. This is the Pan-Canadian Assessment Program (PCAP). At the national level within the U.S., the ACT and SAT are the prime candidates. Finally, at the state level, there are at least as many testing programs as states. We chose to focus on processes for developing consortium-based frameworks because states otherwise rely on their own academic content standards, which inform both assessment and instruction. That context differs from NAEP, which cannot make explicit connections to instruction.

Additional Documentation Reviewed

We reviewed additional documentation from other assessment programs. There are two kinds of documents: (1) documents that may *specify requirements* for elements of framework processes, *report* them, or both; and (2) documents that purport to address standards and best practices for the elements of framework processes.

The difference between *specifying requirements for* a framework process and reporting an element of a framework process is that the former states, for example, how the framework should be structured or how the product should unfold.

The difference between a document specifying requirements and a document purporting to address standards is that the first is typically written by a test sponsor and outlines what they want the product to contain and how the process should unfold. The second type of document would include principles or guidance that should apply to *every* framework process, regardless of sponsor.

Table 1. Documents Addressing Framework Processes

Documents *specifying requirements for* or *reporting elements of* framework processes	Documents addressing or potentially addressing standards or best practices
• Mandates (Laws, memorandums of understanding, charters, and other agreements – see definitions) • Statements of work • Work plans • Assessment frameworks • Reports • Communiques • Other (websites, presentations, briefs, etc.)	• Standards • Guidelines • Assessment frameworks • Reports • Communiques • Other (websites, presentations, briefs, etc.)

We present a complete list of specific documents reviewed for this technical memo in References and Appendix A. The double-asterisked references are relevant to our review of the Pan-Canadian Assessment Program (PCAP), the closest comparison to a NAEP-like program that we could find.

Organizer: Elements of Framework Processes

We developed the following organizer during our review of framework processes for NAEP and other assessment programs. We employ the highlighted terms in the manner defined in the section on working definitions. Developing, reviewing, or updating an *assessment framework* (the "work") implies the following *elements of framework processes*. A potential source of confusion is that *work process* is an element of framework processes. "Framework processes" is an over-arching term for the many aspects of developing an assessment framework.

Note that both "work product" and "work process" are considered elements of framework processes. The first addresses the critical questions about what gets included in a framework document. One way framework documents differ is how far they go in addressing test design, for example. Broadly speaking, deciding what is in the framework document and how it should be organized is a framework process. In contrast, the second element – "work process" – is about the steps to follow to produce the framework document. These two elements are independent: It is possible for test sponsors to specify requirements for components either, neither, both.

Table 2. Framework Processes Key Components and Questions Addressed by Element

Element	Key Components*	Questions addressed**
A. Initiating conditions	None	Under what conditions will this work be initiated?
B. Work product	None	What are to be the components of the final work product?
B. Work product	Domain description	What is to be the format of an assessment-oriented description of this domain?
B. Work product	Descriptions of achievement levels	What claims about student knowledge or ability are intended?
B. Work product	Assessment design	What aspects of assessment design are to be included in the work product?
B. Work product	Documentation of process	How much of the process for producing the work product is to be included in the work product itself?
B. Work product	Basis for decision to revise/retain	In the case of a review, what is to be the basis for revising or retaining an existing framework?
B. Work product	Special requirements, constraints, and criteria	What additional requirements or constraints must be reflected in the final work product?
C. Work process	None	What is the process to be followed in producing the work product?
C. Work process	Commissioning procedures	How will a contractor be selected to produce the work?
C. Work process	Selection of authors, consultants, and working groups	How will authors, consultants, etc. be selected by the contractor?
C. Work process	Timelines and milestones	What is the timeline for the work and milestones (if any milestones)?
C. Work process	Sources informing framework; their role in the work	What other sources should inform the framework, and in what way?
C. Work process	Reconciliation	What will be the process for addressing competing views on the domain or competing requirements, such as fidelity to the domain and practical assessment constraints?
C. Work process	Internal drafting and review	What will be the process for drafting the work product? Who is to be responsible? How is internal review to be managed?
C. Work process	Role of external consultants and owners in shaping the work	How will external expertise be solicited, and from whom? How will sponsors/owners provide input, if at all, prior to work product finalization? How will feedback from these parties be incorporated?
C. Work process	External review, response, and finalization	How will external (including constituency) review be conducted? How will input from the parties be responded to? What is the process for incorporating that input into the final work product?
C. Work process	Documentation requirements	What is to be documented about the work process components?

Table 3. Framework Processes Key Components and Questions Addressed by Element (Continued)

Element	Key Components*	Questions addressed**
D. Owner	None	Who is the client or sponsor of the work product?
E. Timeframe	None	What is the timeframe for producing the work product?
F. Approval	None	What is to be the process for approving the work product?
F. Approval	Approving party	Who will be approving the work product?
F. Approval	Decision process	By what process will the work product be approved (or not)?
F. Approval	Criteria for judging the work product and process	What will be the criteria for judging the quality of the work product and process?
F. Approval	Contingencies	What procedures will be followed if the work is not approved?

Note: **Please note that a component is a subdivision of an element. *The questions are written in a format anticipating *requirement specifications* for that element or component. They could also be written to anticipate *reporting* of that element or component.

Key Aspects of Framework Processes Relevant to NAEP

Several key aspects of framework processes are particularly relevant to a large-scale assessment such as NAEP.

Table 4. Key Aspects of Framework Processes Relevant to NAEP

Key aspect of framework process	Relevant framework process elements	Documents typically specifying (S) or reporting (R) this aspect
The authority or legislative mandate for developing an assessment framework	Mandates can address all framework process elements	Mandates (S)
Framework derivation*– i.e., a description of how, given authority, legislative mandate, sources, or working groups, a person or group should derive (or derived) the assessment frameworks.	C** – The process to follow/all components	Mandates (S) Statements of work (S) Frameworks (R)
Intended relationship to academic standards or curricula of the assessed population	C – The process to follow/Sources informing the framework, and their role in the work	Mandates (S) Statements of work (S) Frameworks (R)
Intended role of standards/curricula of the assessed population	C – The process to follow/Sources	Mandates (S) Statements of work (S) Frameworks (R)
Role of education research in the content area	C – The process to follow/Sources	Statements of work (S) Frameworks (R)
Role of other frameworks	C – The process to follow/Sources	Statements of work (S) Frameworks (R)
Articulating the dividing line between the aspects of test design to be covered in the framework, from those that will be in other documents, such as test or item specifications	B – Work product/Assessment design	Statements of work (S)
Sources for the assessment design	C – The process to follow/Sources	Statement of work (S) Frameworks (R)
Authorship of framework documents	*Who authors?* is addressed in C – The process to follow/Selection of authors *How?* is addressed under the same element/Reconciliation; Internal drafting and review; External review, response, and finalization	Statements of work (S) Frameworks (R)

Notes: **Derivation of a framework means developing a new framework or reviewing an existing framework and, if applicable, revising/updating that framework. *Letters refer to labels for elements in the organizer. The format in this column is “label -element / component.”

Descriptions of Assessment Programs Reviewed

The descriptions below focus on the programs' relation to the assessed population's curricula or content standards and the extent of available documentation relevant to framework processes. We describe who is involved in drafting frameworks to the extent that such information is publicly available.

National Assessments

National Assessment of Education Progress (NAEP)

Of the programs reviewed, the National Assessment of Educational Progress (NAEP) has the most extensive documentation of framework processes.

Initiating Conditions

Conditions for initiating a particular NAEP program's framework process are not specified in the National Assessment of Educational Progress Authorization Act of 2002 ("NAEP law"). Principle 3 of the NAEP Framework Development Policy Statement ("NAEP framework policy", Governing Board, 2018), however, notes that:

> "At least once every 10 years, the Governing Board, through its Assessment Development Committee (ADC), shall review the relevance of assessments and their underlying frameworks. […] Within the 10 year period for an ADC review, major changes in the states' or nation's educational system may occur that relate to one or more NAEP frameworks. In this instance, the ADC will determine whether and how changing conditions warrant an update […]" (p. 6)

As part of our review, the Governing Board responded to the question "What triggers a framework review?" with "[F]ramework reviews often occur when there are major developments in the field, developments that need to be incorporated into the assessment. Major consensus reports from groups such as the National Academies may prompt Board discussion, etc." [personal communication (email) February 16, 2021].

While this places a timeframe within which a review must occur, it underspecifies the conditions for timing such a review.

Work Product

The NAEP framework policy specifies several components of the framework process element *work product*. If framework processes are treated broadly to include the development of test specifications, then Principle 5 (Element of Specifications) specifies aspects of the "Assessment design" component of the *work product*. Principle 1 (Elements of Frameworks) explains that the frameworks should contain a description of the domain.

However, the NAEP framework policy does not specify how descriptions should be formatted or structured to fit within specific measurement paradigms – for example, it might be an implicit requirement that items must be nested within the smallest units of the framework and that tests should conform to unidimensional IRT with 3-5 major groupings of items.[2] NAEP framework

[2] This is only an example, not a recommendation from the authors.

policy Principle 5, Guideline (c), implies that the framework should have "content" and "process" dimensions.

Some components of the *work product* are further specified in NAEP framework revision statements of work, such as that attached to RFP# 91995918R0002 (Governing Board, 2018).

Work Process

As with the *work product*, the NAEP framework policy addresses several components of the framework process element *work process*. Principles 2 (Development and Update Process), 3 (Framework Review), and 4 (Resources for the Process) all address *work process* components. Two Guidelines, (b) and (d), under Principle 6 (Role of the Governing Board), also address the *work process*.

In general, the NAEP framework policy guidelines provide parameters for the components of processes but do not specify them. For example, Principle 2 highlights the need to represent a variety of viewpoints regarding the content of the assessment. However, the NAEP framework policy does not prescribe a panel-selection process to ensure this outcome. This leaves open the question of how the panel selection process should actively include those who hold minority or less popular views on the content assessed. The same applies to the framework review guidelines under Principle 3. The choice of experts from whom the Assessment Development Committee (ADC) is to solicit input can make a difference in determining whether changes are warranted, as there are often significant differences of opinion among experts. These considerations pertain to the *work process* component "Selection of authors, consultants, and working groups."

Guideline (f) of Principle 2 indicates that "protocols shall be established to support panel deliberations and to develop a unified proposal for the content and design of the assessment." (p. 6) A critical component left unaddressed at the NAEP-wide level is the process by which differences will be resolved to move forward in case consensus is not reached, called "Reconciliation" in the organizer.

A recent NAEP design document lays out a three-step approach to reconciliation, which might serve as a starting point for a cross-program reconciliation protocol:

> The first strategy will involve a process for reconciling differences in points of view relevant to the assessment framework. An overview of panel norms will be presented at the Visioning Panel meeting, with emphasis placed on building consensus. The second strategy will include a process to follow when agreement cannot be reached. For example, when the Development Panel cannot agree, it will define and document the contentious issues and differences that cannot be reconciled. If differences are technical and related to measurement, the issues will be brought to the TAC [Technical Advisory Committee]. Other issues will be sent to the project expert advisory group, who will consider the arguments and provide advice on reconciliation. If, after consulting with the TAC and/or advisory group, differences persist, the Development Panel will generate alternative options with the pros and cons articulated and priorities suggested, which can be reviewed during the public comment phase of the project. (WestEd, 2019, pp. 14-15)

(Note that reconciliation protocols should anticipate potentially unreconcilable differences of opinion at every stage where multiple individuals, including experts and the public, provide input or feedback.)

For NAEP, the *work product* includes descriptions of achievement levels (ALDs). Principle 1 of the NAEP framework policy indicates that framework development entails answering "how much" of content domain students should know and be able to do at the three NAEP levels. Still, aside from needing to be based on the Governing Board's very general policy definitions, there is little guidance on how to derive these descriptions. The Governing Board's Policy on achievement levels (Governing Board, 2018) explains that achievement levels consist of three parts: ALDs, cut scores, and exemplar items or tasks. That policy indicates early in the document that the development of ALDs "shall be completed initially through the process that develops the assessment frameworks." (p. 5). The remainder of the Policy on Achievement Levels appears to focus on standard setting, a process into which ALDs serve as *input*. The NAEP framework policy does not specify a process for developing ALDs.

The NAEP framework policy partially addresses the *work process* component "Sources informing the framework, and their role in the work" under Principle 4 (Resources for the Process). Several resources are mentioned, including:

> An initial compilation of resources" that "summarize[s] relevant research, advantages and disadvantages and latest developments, and trends in state standards and assessments in the content area. […And] curriculum guides and assessments developed by states and local districts, widely accepted professional standards, scientific research, other types of research studies in the literature, key reports having significant national and international interest, international standards and assessments, other assessment instruments in the content area, and prior NAEP frameworks. (p. 7)

The universe of documents represented in this list is monumental for any given content area. No aspect of the process for selecting what to include in this library is specified. The NAEP framework policy provides some guidance on factors to "balance" in prioritizing source documents but is otherwise silent on the way that this library should shape panel deliberations and, ultimately, the framework being developed or reviewed.

The "Commissioning procedures" component of the *work process* element is not specified in any NAEP source reviewed.

As with *work product*, requirements for several aspects of the *work process* are specified in statements of work. Also, process reports of NAEP framework development or update [e.g., WestEd, 2006; WestEd, 2010; WestEd (draft), 2021] provide detailed schedules and accounts of meetings but only general statements about discussion topics, how consensus was reached, or how differences of opinion were addressed.

Owner, Timeframe, and Approval

The *owner* or client of NAEP assessment frameworks is the Governing Board. The *timeframe* for producing frameworks does not appear to be specified in general. Contract lengths or schedules in specific statements of work *report* desired timeframes.

The NAEP framework policy addresses the "Approving party" component of the *approval* element of framework processes. It does not specify an approval process or criteria for judging the quality of the *work process* or *product*. The policy does not specify the procedures to follow in case a framework project is not approved.

Pan-Canadian Assessment Program (PCAP)

The Pan-Canadian Assessment Program (PCAP) resembles NAEP in context: It is a national survey in a country without a single set of national-level academic standards or national curricula. The PCAP is given every three years in reading, mathematics, and science. PCAP was the first program that we reviewed, and this review greatly informed the development of our organizer for framework processes. Our review of this program is in Appendix B.

The SAT and the ACT

Two long-standing and well-recognized testing programs in the U.S. are the SAT and the ACT. Many colleges and universities require or accept these tests for admission. Recently, several states have adopted one or another of these tests to meet the ESEA requirement for testing in high school. The SAT is revised or redesigned every few years.

Due to these testing programs' national user base, the test takers they serve have been learning under different standards and curricula. Neither of these programs claims to be neutral with respect to curriculum, although the ACT more explicitly claims to incorporate information about the different curricula of the population of test-takers: Every three to five years, ACT conducts a national curriculum survey that asks K-12 and postsecondary educators to rate the importance of several discrete skills in their teaching or as a prerequisite to their course. ACT conducted the last such survey in 2020 (ACT, 2020 a).

Neither the SAT nor ACT programs provide detailed documentation of their assessment framework processes. ACT offers some highlights of the process in its most recent technical manual, particularly the sources or factors informing the ACT frameworks. These include subject-matter experts, academic research, ACT data, the ACT national curriculum survey, and a survey of other content standards – such as the Next Generation Science Standards (NGSS). (ACT, 2020 b, p. 1.6) However, most framework components listed in the organizer of this technical memo are not reported by ACT.

College Board documentation on framework processes for the redesigned SAT reveals a more hierarchical organization of committees and working groups involved in these processes. Their membership is not specified except in general terms (for example, "The Higher Education Advisory Working Group is composed of 30 representative higher education leaders from institutions across the nation." (College Board, 2015, p. 15). Available documentation on the input provided by these groups highlights *role* and not *process*. For example, "The group provides direct, in-depth feedback on such matters as implementation and reporting, scores and validation, and communications." (p. 15) Like the ACT, the SAT does not report on most framework process elements and their components.

Frameworks for State Assessments

Common Core State Standards (CCSS)

The Common Core State Standards (CCSS, NGA/CCSSO, 2010) are a seminal set of content standards in K-12 English language arts and mathematics, intentionally anchored in "college/career readiness," developed under the sponsorship of the National Governors Association (NGA) and the Council of Chief State School Officers (CCSSO). Published in 2010, the CCSS were adopted by over 40 states, districts, and territories by 2013.

The CCSS are unusual in that their sponsorship by the NGA and CCSSO was as close to a set of "national, not federal" content standards created in modern times. The development process

involved four sets of contributors: a core team of lead authors that designed the architecture and key aspects of what became the CCSS, a "work team" heavily involved in writing the CCSS— first college/career readiness standards, and then K-12 standards— and several review groups, including an official "feedback group." There was also a "validation group" that considered the evidential and argumentative basis for the CCSS. And finally, multiple drafts of the CCSS were released for comment—both targeted (e.g., state departments of education, professional organizations) and public—and those comments were considered in creating the final versions of the CCSS. The lead authors and work groups for the CCSS were primarily university academics or people from business organizations; there was no specific call for active teachers or school administrators to be on the committees. None were, although some committee members had been elementary/secondary teachers previously, and several had worked with other sets of content standards. The "lead writers" consisted of three persons each for ELA and mathematics; the "work group" consisted of 24 total persons. The validation committee consisted of 29 members, primarily university- or institute-based academics, although there were also five teachers and principals, as well as a few employees of testing companies.

The CCSS were conceived as content standards for instruction, not assessment specifications. The intent of the CCSS—for example, for assessment—was commented on by individual lead authors and by an organization established by a few of the CCSS lead authors—Student Achievement Partners. However, these were not treated as authoritatively reflecting the consensus of the CCSS authors and development process. States and others developing assessments were able to treat the CCSS as academic content standards and develop different assessment constructs, blueprints, and other specifications. For example, two federally funded consortia, each joined by many states, developed quite different assessment specifications using quite different development processes, resulting in the two different operational assessments by the Partnership for Assessment of Readiness for College and Careers (PARCC) and the Smarter Balanced Assessment Consortium.

There is little documentation available regarding the processes of how the committee number, structure, or membership were determined; or the processes by which the CCSS were conceptualized or developed in terms of how committee work was allocated, how leadership took place, or differences reconciled. Also, although a public comment process was engaged in by the developers of the CCSS, we could not find documentation of the process by which comments were solicited or responded to. Some of this may be attributed to the fact that NGA, CCSSO, and the work groups wanted to control the development without undue outside influence until formal feedback was instituted. Some may also be attributed to the subsequent controversial nature of the CCSS; for example, neither NGA, CCSSO, nor the website they established for the Common Core have listings of the various committee members, let alone primary documentation of the CCSS developmental process on their websites.

Next Generation Science Standards (NGSS)

The Next Generation Science Standards (NGSS) are a widely popular set of K-12 science content/assessment standards. Over 30 states had adopted some version of the NGSS by 2021. The NGSS have two foundational documents: A framework document and a standards document, authored and published independently.

The *Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas* (National Research Council, 2012) was authored by a group sponsored by the National

Research Council, National Academy of Sciences.[3] The committee responsible for the *Framework* consisted of 18 persons, including "practicing scientists, including two Nobel laureates, cognitive scientists, science education researchers, and science education standards and policy experts." (Achieve, n.d. a) There was no charge for specific groups to be represented on the writing committee; no elementary/secondary educators were included.

The Framework document included applications of the *Framework* to specific science domains. "In addition, the NRC used four design teams to develop the *Framework*. These four design teams, in physical science, life science, earth/space science, and engineering, developed the *Framework* sections for their respective disciplinary area." (Achieve, n.d. a) The development process included gathering public comments. "After releasing a public draft in July of 2010, the NRC reviewed comments and considered all feedback prior to releasing the final *Framework*." (Achieve, n.d. a)

The *Next Generation Science Standards* document provides specific content standards reflective of the *Framework* in grades K-5, middle school, and high school. Thirteen appendices provide additional information regarding rationale, additional information, and discussion of relevant issues in extending the Framework into Standards. The Standards were produced by a group of 26 Lead State Partners, managed by Achieve (Achieve, n.d. b). One of the key Achieve staff persons and another member of the NGSS writing team had been members of the *Framework* committee. The writing committee for the Standards included many state department of education employees, but there was not a charge for specific representation from specific groups. Educator input was specifically and actively sought during the feedback and comment processes.

The NGSS have a conceptual *Framework* document developed separately from the *Standards* document. One disadvantage is that the two committees were not together to work out issues. A prime example is that the *Framework* delineates a domain much larger than is possible to assess practically, or even perhaps to learn. The developers of the *Standards* had to make choices about what to include and what to leave out, without the authoritative agreement of the *Framework* authors. And although the authors of the *Standards* aimed them at assessment specifications, they worked at the level of individual standards rather than defining what would be adequate for a construct or domain. The result has been that states that have adopted the NGSS have adopted different things: notably, some have adopted the *Framework*, while other have adopted the *Standards*; some consider the performance expectations in the *Standards* to be the standards, while others consider the performance expectations merely examples. States and their partners have struggled to use the documentation to create practical assessment blueprints, and there has been considerable variation across states.

The NGSS publicly available documentation does not include information regarding the processes of how the committee number, structure, or membership was determined; or the processes by which the *Framework* or *Standards* were conceptualized or developed in terms of how committee work was allocated, how leadership took place, or differences reconciled. Also, although a public comment process was engaged in by the developers of both the *Framework* and the *Standards*, documentation did not include detailed description of the process by which comments were solicited or responded to.

[3] A starting point for documentation about the Framework development is https://www.nationalacademies.org/our-work/conceptual-framework-for-new-science-education-standards#sectionCommittee

International Assessments

The assessment frameworks of the two leading international assessment programs have very different conceptual relationships to curricula.

Programme for International Student Assessment (PISA)

The Programme for International Student Assessment (PISA) is a sample-based assessment headed by the Organization for Economic Cooperation and Development (OECD) and administered to 15-year-olds in participating countries and economies (79 in 2018) once every three years. The first PISA assessment was in 2000. Domains assessed include reading, mathematics, science, and financial literacy. PISA assesses an innovative domain in each cycle. In 2018, that was global competence (OECD, 2019). PISA does not purport to align to any curricular or content standards. Instead, it aims to assess "the extent to which 15-year-old students near the end of their compulsory education have acquired the knowledge and skills that are essential for full participation in modern societies." (OECD, 2019, p. 11). The PISA Governing Board (OECD, n.d.) has members from each participating country. Framework and related documents are available through the PISA website.

The most recently published framework (for 2018, when reading was the "major domain" assessed) lists the chair and members (total of 6) of the reading framework working group. The same information is provided for the global competence working group (total of 5). All members are affiliated with universities or similar organizations. The global competence framework was developed by a member of the OECD Secretariat working with a university collaborator (OECD, 2019, pp. 18-19). Publicly available documents do not indicate which, if any, elements or components of framework processes were shaped by requirements specifications. The *work process* components are not reported.

Trends in International Mathematics and Science Study (TIMSS)

The Trends in International Mathematics and Science Study (TIMSS) has been assessing mathematics and science in fourth and eighth grade every four years since 1995. In 2019 – the most recent year of administration – 64 countries and 8 "benchmarking participants" (generally, cities) participated in TIMSS (Mullis et al., 2020). TIMSS assesses mathematics and science in grades 4 and 8.

The TIMSS assessment frameworks highlight the importance of curriculum as the basis for the domain description. The most recent assessment frameworks indicate they are updates of earlier frameworks. Framework documents list names of members of the framework revision committees. These also serve as members of item review committees. In the most recent revision of the TIMSS framework (2019), there were 7 members per content area; most are university staff and are described as "internationally recognized mathematics and science experts." (Mullis & Martin, 2017, p. 96). However, the frameworks also present an extensive list of TIMSS national research coordinators (at least one per participating country) who "participated in a series of reviews of the updated frameworks." (p. 98) As with PISA, available documents (assessment frameworks, technical reports, etc.) do not indicate which, if any, elements or components of framework processes were shaped by requirements specifications. The *work process* components are not reported.

Professional Standards and Framework Processes

Processes for framework development are not covered extensively in widely available professional standards that deal with test development or validation. The *Standards for educational and psychological testing* (*Standards*, AERA/APA/NCME, 2014) address select aspects of framework processes in Chapter 4, Test design and development, Test specifications (pp. 75-81). In the *Standards*, test development begins with developing test specifications. In many ways, this places the framework processes beyond the scope of the *Standards* because the essential component of assessment frameworks (the domain description) precedes test specifications. Note, however, that most assessment frameworks contain at least some assessment design aspects. The *Standards* apply to *these* parts of assessment frameworks and thus framework processes more generally:

> The term *test specifications* is sometimes limited to description of the content and format of the tests. In the *Standards*, test specifications are defined more broadly to also include documentation of the purpose and intended uses of the test, as well as detailed decisions about content, format, test length, psychometric characteristics of the items and test, delivery mode, administration, scoring, and score reporting. (p. 76)

The Standards have little to say about appropriate processes for deriving domain descriptions (also called *content specifications* and *content frameworks* in the *Standards*) for achievement tests such as NAEP: “The delineation of the content specifications can be guided by theory or by an analysis of the content domain (e.g., an analysis of job requirements in the case of many credentialing and employment tests).” (p. 76)

The *ETS Standards for quality and fairness* (ETS, 2015) closely follow the *Standards* and do not explicitly address framework processes. One ETS standard speaks to settings where information about the construct is not readily available, indicating that “obtaining the information may be part of the test developers’ (typically, a contractor) task.” The standard continues, “If the information has to be obtained, work collaboratively with clients, subject-matter experts, and others as appropriate.” (p. 29) But the ETS standards go no further in discussing appropriate framework processes.

Guidance published by the Department of Education for the assessment peer review process addresses some requirements for state (Every Student Succeeds Act, ESSA) assessment framework processes. State assessment programs must show that they have “challenging academic content standards in reading/language arts, mathematics, and science” that are “aligned with entrance requirements for credit-bearing coursework in the system of public higher education in the State and relevant State career and technical education standards.” (U.S. Department of Education, 2018, pp. 30). Among the examples evidence that states can provide to meet this requirement, the guidelines cite:

> A detailed description of the strategies the State used to ensure that its academic content standards adequately specify what students should know and be able to do;
>
> Documentation of the process used by the State to benchmark its academic content standards to nationally or internationally recognized academic content standards; Reports of external independent reviews of the State’s academic content standards by content experts, summaries of reviews by educators in the State, or other documentation to confirm that the State’s academic content standards adequately specify what students should know and be able to do;

> Endorsements or certifications by the State's network of institutions of higher education (IHEs), professional associations and/or the business community that the State's academic content standards represent the knowledge and skills in the content area(s) under review necessary for students to succeed in college and the workforce. (pp. 30-31)

These examples suggest some principles or standards for framework processes in the context of ESSA, especially around vetting or approval. However, this is a special context in which there is an independent criterion (college and career readiness) built into the mandate for ESSA.In either case, there is a principle implied by the peer review guidance: When there is an external referent in the mandate, then framework development should incorporate some process to ensure that the content to be assessed is related to that criterion.

The previously referenced NAEP framework policy (Governing Board, 2018) comes closer to supplying professional standards for framework processes than any other source. Principles 1 (Elements of Frameworks) and 5 (Elements of Specifications) address some of the components of the framework process element *work product*. Similarly, some components of *work process* are addressed in Principles 2 (Development and Update Process), 3 (Framework Review), and 4 (Resources for the Process). Principle 6 (Role of the Governing Board) covers components of *work process*, *owner*, and *approval*.

Key Findings

Five elements of framework processes answer foundational questions about framework development. These elements are: The conditions for initiating a framework (or review), what is to be included in a framework, what are the steps or rules to be followed in putting a framework together, who owns the framework process, what is the timeline for the process, and what is the process for approval.

There is considerable variation among assessment programs in the framework process elements that programs report. Some programs specify general requirements for some elements (or components thereof). No program we know of specifies requirements for all components.

Although most programs have a structure for framework development, such as a sequence of panels or working groups, no assessment program we reviewed specifies systematic processes for (a) selecting panel members or authors, (b) selecting source documents, (c) addressing competing views about what should be in the framework, (d) integrating source documents, expert judgment, and public review to derive a framework, and (e) approving the final product, together with a contingency plan in case the work is not approved.

Implications of NAEP Legislative Mandate for NAEP Framework Processes

Here we address implications of three aspects of NAEP law and tradition: Curricular neutrality, representation of diverse views, and the role of professional standards.

Curricular Neutrality

By tradition and by law, NAEP has been guided by a criterion of curricular neutrality.

The concept is applied to framework processes in NAEP's framework development policy statement, which includes as a guideline that:

> The framework shall focus on important, measurable indicators of student achievement to inform the nation about what students know and are able to do without endorsing or advocating a particular instructional approach. (Governing Board, 2018, p. 4)

However, the standards, curriculum, and teaching practices in the U.S. are relevant to the NAEP framework, even if NAEP adopts a neutral stance. (See, for example, the list of resources that the NAEP framework policy Principle 4 asks panelists to consider.)

The principle of curricular neutrality has implications for the NAEP framework development process. Whatever those may be, they are not explicit in the NAEP documentation we reviewed. Among our recommendations for future work, we offer some considerations towards more precise definition of curricular neutrality to inform framework processes on a NAEP-wide level.

Diversity of Views

The NAEP framework policy indicates that framework panels "shall reflect diversity in terms of gender, race/ethnicity, region of the country, and viewpoints regarding the content of the assessment under development." (Governing Board, 2018, p. 5)

Ensuring representation of diverse viewpoints regarding assessment content implies that the process for selecting framework panel members should be informed of both existing viewpoints and candidate panelists' views. It may be that in practice, this is or has been part of the panelist selection process.

"[D]iversity in terms of […] viewpoints regarding the content of the assessment" would likely include experts who have strong opinions not only about the nature of the construct but also about the appropriateness, for their content domain, of measures largely composed of multiple-choice test items.

The representation of diverse viewpoints on panels is likely to result in perspectives that cannot always be reconciled into one framework. How should impasses be handled? Rules of order might be specified ahead of time.

Role of Professional Standards

NAEP law references "professional standards" or "professional assessment standards" several times. Three instances have implications for framework processes. In the first, "professional standards" are referenced as the basis for the development of "assessment objectives," "test specifications," or both:

> IN GENERAL – In carrying out its functions under this section the Assessment Board shall—[…] develop assessment objectives consistent with the requirements of this section and test specifications that produce an assessment that is valid and reliable, and are based on relevant widely accepted *professional standards* [Section 302, (e)(1)(C), emphasis ours]

The second and third instances concern the determination of achievement levels:

> IN GENERAL- Such levels shall-- be determined by—(I) identifying the knowledge that can be measured and verified objectively using widely accepted *professional assessment standards*; and (II) developing achievement levels that are consistent with

> relevant widely accepted *professional assessment standards* and based on the appropriate level of subject matter knowledge for grade levels to be assessed, or the age of the students, as the case may be. [Section 303, (e)(2)(A)(i)(I-II), emphasis ours]

The importance of professional standards is evident in the NAEP law. However, a central question is to what extent do they apply to framework processes as understood in this technical memo? If they apply at all, then the lack of a robust set of professional standards for framework processes poses a real challenge for assessing the extent to which any NAEP program involving framework processes was properly designed and implemented.

How this Review Might Inform NAEP Framework Processes

This review might inform NAEP framework processes primarily through the organizer we developed. We believe that all elements and components should certainly be documented for any framework project. More importantly, the NAEP program may benefit from more deliberate consideration of the extent to which it wishes to specify requirements for those components, and whether (or when) it will delegate such requirements specification to others, such as contractors.

Delegation of requirements specification may lead to different requirements for different testing programs. This may be appropriate for some elements/components – for example, insisting on content-by-process organization of all domain descriptions could run counter to current or future conceptualizations of domains. But there doesn't seem to be an obvious rationale for diverse requirements specifications for some other components, such as all *work process* components.

Towards Best Practices for Framework Processes

The absence of professional standards for most components of framework processes leaves much room for proposing principles, guidelines, and standards.

We propose that sponsors make deliberate choices regarding which components to specify requirements for and to document the rationale for those choices.

When sponsors consider delegating requirements specification for a component to other groups or contractors, it may be useful to prepare for the different ways in which the component may unfold, possibly resulting in very different work products.

A good analogy for what a systematic framework development process might look like is standard-setting. There are many standard-setting methods, and no consensus about which is best in every case. However, the more mature methods prescribe a step-by-step process, contingency planning, specific documentation requirements, and success criteria. Disagreements are addressed through rounds of conversation and voting procedures.

As with standard-setting, it may be possible to outline a standard set of procedures for some special cases of framework development.

Standard-setting needs an external criterion, or has to very heavily rely on process and internal coherence. A reliance on what has sometimes been called "procedural validity"—that is, the quality and evaluation of quality are dependent upon having a good process—needs to show reasonable process for producing work products and evaluation showing implementation fidelity.

For example, suppose that (by sponsor-level specification or by contractor-level specification) it is decided that the process for generating NAEP assessment objectives will involve sub-setting from a broader set of content standards. One can imagine a few ways to approach this general task, involving discussions and voting. Those approaches can be cast as systematic framework development methods.

When the sources are many and varied and the actual task of creating a framework less certain, sponsors can still indicate how each type of source should inform framework development. Sponsors might also specify what the resulting assessment objectives should look like individually – in terms of syntax, length, the extent of performance description (see "content/performance continuum" in the section on recommendations for additional work), and similar properties – as well as collectively.

Recommendations for Additional Work to Inform Governing Board Considerations

This section proposes additional studies, reviews, or conceptual work to help inform how the Governing Board addresses framework processes. We elaborate on some of the proposals.

Proposal 1. Every assessment program has a definition or description of the domain to be assessed; this is part of every assessment framework. (See framework process element *work product*, component "Domain description.") There is considerable variation in how frameworks arrive at these descriptions, however. The Governing Board might explore the structure of domain descriptions in different assessment frameworks to decide which is most appropriate NAEP-wide.

Proposal 2. Review the different kinds of sources informing assessment frameworks to develop a systematic way to incorporate those sources into the framework development process.

Commentary. One class of sources includes content standards that may differ in terms of their educational orientation.

All assessment frameworks report domain descriptions that are assessment-oriented. This means that they were developed for the purpose of creating an instrument to determine what students know and can do. By contrast, domain descriptions can be oriented toward instruction – that is, primarily for the purpose of getting students to know and be able to do the knowledge/skills that are indicated. Some content standards, such as the high-level academic content standards that states adopt, purport to inform both uses. The sources from which an assessment framework might draw may be instruction-oriented, assessment-oriented, over-arching, or some combination of these.

Academic content standards adopted by states are good examples of over-arching domain descriptions: States typically adopt content standards to specify what, at a minimum, students should learn and be able to do. These content standards are intended to provide guidance for educators as they select or develop curricula and as they design their associated instruction. Instructional and over-arching domain descriptions generally encompass more than those for large-scale assessments.

Domain descriptions for instruction include more than those for assessment in that the former often specify:

- More complex content than can feasibly be assessed in large-scale assessments – such as the full writing process, including research projects; and
- Skills that do not fit well within the tradition of assessment of work products produced by individuals working alone, such as mental math, problems solved in groups, cross-curricular learning targets, non-standardized learning targets such as individual projects, and learning arising from extended experiences such as reading specific novels in a literature class.

The content standards that go into a domain description for assessment will typically be a subset of over-arching standards or those with a (primarily) instructional orientation.

Whenever the process for generating a domain description in an assessment framework involves sub-setting from a broader set of content standards for learning, the sponsors for an assessment program might specify how that is done (element *work process*, component *sources*). At minimum, they should require that the process by which it is done be documented (element *work process*, component *documentation requirements*). For transparency purposes, the sponsor may require that this documentation be included in the framework itself (element *work product*, component *documentation of process*).

Proposal 3. Consider the *content/performance continuum* of assessment objectives, to specify which is most appropriate for NAEP.

Commentary. In most assessment programs, the foundational unit of content specifications (typically found in assessment design documents) is called a “content standard.” However, there is considerable variation in what is included in a content standard across assessment programs. Content standards always contain the content of the construct (if the construct is a skill, the description of that skill to be assessed would be the “content” of the content standard). Important variations occur around what else is included in the content standard—particularly, how much of a performance description is included in the content standard.

Content standards used by assessment programs can be classified on a continuum reflecting increasingly elaborate performance descriptions. Assessment sponsors can choose to specify in advance where on this continuum to target the resulting content standards, and direct assessment framework authors to write frameworks in such a way that assessment content standards derived from those frameworks will be at their chosen level:

1. Content only. The content standard describes what students should know or understand or be able to do but does not include how a student is supposed to demonstrate that knowledge, understanding, or skill.

2. Content with minimal performance descriptions. The content standard includes description of the content and indicates what the student is supposed to be able to do with that knowledge, understanding, or skill. Minimal detail is provided in this performance description. Very many U.S. state content standards use this structure.

3. Content with detailed performance descriptions. The content standard includes description of the content and indicates in some detail what the student is supposed to be able to do with it or how the student is supposed to demonstrate the desired level of expertise. The Next Generation Science Standard’s (NGSS) Performance Expectations (P.E.s) are a widely known example of this approach.

4. Content with multiple detailed performance descriptions at different levels. The content standard includes content and descriptions of multiple levels of expertise and/or how the student demonstrates those levels of expertise. Examples of content standards using this approach include those developed in the "learning progressions" approach. Dynamic Learning Maps (DLM) precursors and NWEA for Nebraska range ALDs employ this approach.

This aspect of the structure of content standards has far-reaching implications for assessment specifications, designs, and activities. NAEP can choose to specify what to include about it, both in terms of content and process, in its framework process guidance across programs. This would lead to assessment content standards written at parallel levels of specificity across content areas.

Proposal 4. Explore the ways in which assessment programs attempt to remain "neutral" with respect to curriculum, to state how NAEP will provide guidance (requirements specification) so its resulting assessment frameworks are all "curriculum neutral" in the same ways.

Commentary. Most large-scale U.S. state assessments aim to be more general than a specific curriculum. States resolve this issue through the mechanism of common content standards. Other contexts, such as some national and all international assessment programs, however, operate across jurisdictions with different curricular/content standards. These programs also aim to be more general than a specific set of curricular/content standards, and thus must adopt some conceptual relationship to the curricula/content standards of the assessed population.

How they go about that varies. Some programs, such as PCAP, provide a general criterion (what is common across the curricula for the different jurisdictions in the population tested). However, PCAP does not go further in specifying how that commonality is to be judged or determined. NAEP does not provide a specific criterion, nor a specific process for considering the curricula (or academic content standards) of the assessed population.

Some approaches to help ensure an assessment is not tied too closely with a particular curriculum or state content standards:

- Determine what is common across the curricula/content standards of the assessed population. An assessment may focus on those things which all curricula agree on; that might be found through a systematic survey of relevant curricula. This is done explicitly for at least one non-U.S. assessment program. (We note that NAEP also has conducted such studies but, to our knowledge, not expressly to test what is common.) Note that the methodology for determining what is common, and assessing whether the process results in something meaningful, is a separate and non-trivial matter that could be addressed ahead of time.

- Refer to education research in the content domain and deliberately ignore curricula/content standards. An assessment may build its content specifications from research only, if available, without referencing curricula. If the research literature is extensive and detailed enough, it may provide sufficient basis to generate content standards, especially if there is broad consensus about the research base. Note: This seems like the least practical to us and the most difficult to specify requirements for. We include it here anyway for completeness.

- Refer to other authoritative content frameworks, without referencing curricula. If there is a widely accepted content framework outside the assessment program, that content framework may be adopted for the assessment program, especially if that content framework does not reference specific curricula. This is what was done by states adopting the Common Core State Standards, the Next Generation Science Standards, and other content standards generated by national or professional consensus such as the NCTM content standards and the previous National Science Standards. There is at least one challenge for NAEP here: An assessment framework derived from an authoritative content framework is difficult to distinguish from an assessment framework for the curriculum implied by that authoritative content framework (and thus potentially not "curriculum neutral").

- Refer to international assessment frameworks for assessments in which many countries participate. Some challenges: (1) How would NAEP not simply be a different instantiation of that international program? And is it a problem if it were? (2) This option may or may not be consistent with different readings of the NAEP law. (3) There are likely strong political views, pro and con, about the relevance of education in other countries to an assessment of educational progress for U.S. students. What is the scope of NAEP's curricular relevance/neutrality? Is it curricula in the U.S. or curricula throughout the world?

Proposal 5. Study what goes into the assessment design component of frameworks for different assessment programs and consider whether developing test specifications should also be part of the framework development task involving the same group or groups.

Commentary. There typically are two levels of specifications for assessments. One level is more foundational. The other is more detailed. The more foundational may be thought of as defining the core validity claims for the assessment, while the other level specifies how those claims are to be supported in terms of assessment evidence. In many large-scale assessment programs, such as state assessment programs, there is an explicit division in who is responsible for developing which level of specifications. The state is explicitly responsible for developing the first level of specification without input from possible vendors, because the first level of specifications often constitutes the core of a request for proposals. Bidders then propose the second set of specifications—or how to develop them—as the vendor's responsibility. Of course, the vendor's proposals must be approved by the program sponsor; often there is iterative consultation between the program sponsor and vendor to arrive at this second level of specification. Explicit in this organization is the assumption that there are multiple possible ways the second level can be specified, once work at the foundational level is complete. Some of those ways may not reflect the intentions of those who developed the foundational level frameworks.

Proposal 6: Investigate best practices for including implementation fidelity evaluation and documentation.

Commentary. Since NAEP's development of assessment frameworks are so dependent on processes being specified and followed well, the development process might benefit from incorporating means to formatively check on the quality of the process while the framework is being developed, as well as a summative evaluation. For example, if the purpose of recruiting a diverse committee is to ensure diverse perspectives contribute to the framework development, then a formative evaluation would check whether committee members feel comfortable during the process. This could be accomplished through a survey with items such as, "I feel my voice is

being heard," "I am clear about the objectives of our committee work," "The work is well-organized," "I think committee assignments are fair," etc. An external evaluator could support the formative evaluations. Similarly, a summative evaluation should include evaluation of the process. This should incorporate documentation of "procedural validity" that would support the quality of the assessment framework. The summative evaluation of the process should also draw lessons learned to help inform future NAEP assessment frameworks.

Proposal 7: Draw on the best available knowledge to inform effective committee work, especially processes for generating, discussing, and resolving issues.

Commentary. A review of the research literature and professional practice should be able to inform different ways to deal with power dynamics—how to ensure all contribute as intended by inclusion in representation, such as how to structure discussions, when to use open versus anonymous voting, etc. There may be different group dynamics and methods to produce a group report when there is more or less agreement about fundamental issues. It would have to be decided how best to make such information available to the committees.

References

Asterisked documents are in the set provided by the Governing Board for this review.

Achieve. (n.d.). Next Generation Science Standards: Development Overview. Webpage: https://www.nextgenscience.org/development-overview

Achieve. (n.d.). Next Generation Science Standards. Developing the Standards. Webpage: https://www.nextgenscience.org/developing-standards/developing-standards

ACT. (2020 a). ACT National curriculum survey 2020. Available: http://www.act.org/content/dam/act/unsecured/documents/NCS_Report_Web_2020.pdf

ACT. (2020 b). ACT Technical manual. Version 2020.1. Available: https://www.act.org/content/dam/act/unsecured/documents/ACT_Technical_Manual.pdf

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (Eds.). (2014). *Standards for educational and psychological testing*. American Educational Research Association.

College Board. (2015). Test specifications for the redesigned SAT. Available: https://collegereadiness.collegeboard.org/pdf/test-specifications-redesigned-sat-1.pdf

ETS. (2015). *ETS Standards for quality and fairness*. Available: https://www.ets.org/s/about/pdf/standards.pdf

*Martineau, J.; Dadey, N., & Marion, S. (2018). *Literature review on developing or revising assessment frameworks to support a transition from paper-based to digitally based assessment*.

Mullis, I. V. S., & Martin, M. O. (Eds.). (2017). *TIMSS 2019 Assessment Frameworks*. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: http://timssandpirls.bc.edu/timss2019/frameworks/

Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. L., & Fishbein, B. (2020). *TIMSS 2019 International Results in Mathematics and Science*. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: https://timssandpirls.bc.edu/timss2019/international-results/

*National Assessment of Educational Progress Authorization Act of 2002. P.L. 107-279. Available: https://www.nagb.gov/about-naep/the-naep-law.html

National Assessment Governing Board. (November 17, 2018). National Assessment Governing Board Developing student achievement levels for the National Assessment of Educational Progress policy statement. Available: https://www.nagb.gov/content/dam/nagb/en/documents/policies/ALS-revised-policy-statement-11-17-18.pdf

*National Assessment Governing Board. (March 3, 2018). National Assessment Governing Board Framework development policy statement. Available: https://www.nagb.gov/content/dam/nagb/en/documents/policies/framework-development.pdf

*National Assessment Governing Board. (May 11, 2018). Update of the National Assessment of Educational Progress (NAEP) frameworks for mathematics, reading and other subjects. RFP# 91995918R0002. Attachment A: Statement of work.

*National Assessment Governing Board. (n.d.). Summary chart of framework updates.

National Governors Association Center for Best Practices, Council of Chief State School Officers. 2010. Common Core State Standards. National Governors Association Center for Best Practices, Council of Chief State School Officers, Washington D.C.

National Research Council. (2012). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. Committee on a Conceptual Framework for New K-12 Science Education Standards. Board on Science Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Organization for Economic Cooperation and Development. (2019). "What is PISA?", in *PISA 2018 Assessment and Analytical Framework*, OECD Publishing, Paris. Available: https://doi.org/10.1787/2c7c311d-en

Organization for Economic Cooperation and Development. (n.d.). PISA Governing Board. Webpage: http://www.oecd.org/pisa/contacts/pisagoverningboard.htm

U.S. Department of Education. (July 2018). A state's guide to the U.S. Department of Education's assessment peer review process. Available: https://www2.ed.gov/admins/lead/account/saa/assessmentpeerreview.pdf

*WestEd. (May 19, 2006). Final report: NAEP Science assessment 2009.

*WestEd. (December 6, 2010). Final report: Technology and engineering literacy assessment for the 2014 National Assessment of Educational Progress.

*WestEd. (December 6, 2019). Design document: 2025 NAEP mathematics framework update.

WestEd. (draft). (February 12, 2021). Process report: 2025 NAEP Mathematics framework update.

Appendix A: Additional Documents Reviewed

Asterisked documents are in the set provided by Governing Board for this review. Double-asterisked documents are those consulted during the PCAP review.

*Boyer, M. (August 2020). Enhancing the meaning and utility of NAEP results through linking with other assessments.

**Council of Ministers of Education, Canada (CMEC). (1997). *Common Framework of Science Learning Outcomes K to 12*. Available: https://archive.org/details/commonframework00coun

**Council of Ministers of Education, Canada (CMEC). (2008). PCAP-13 2007: Report on the Assessment of 13-Year-Olds in Reading, Mathematics, and Science. (T.Y. 2007). Retrieved from: https://www.cmec.ca/Publications/Lists/Publications/Attachments/124/PCAP2007-Report.en.pdf

**Council of Ministers of Education, Canada (CMEC). (2011). PCAP-2010 Report on the Pan-Canadian Assessment of Mathematics, Science, and Reading. (T.Y. 2010). Retrieved from: https://www.cmec.ca/Publications/Lists/Publications/Attachments/274/PCAP2010.pdf

**Council of Ministers of Education, Canada (CMEC). (2014). PCAP 2013 Report on the Pan-Canadian Assessment of Science, Reading, and Mathematics. (T.Y. 2013). Retrieved from: http://cmec.ca/Publications/Lists/Publications/Attachments/337/PCAP-2013-Public-Report-EN.pdf

**Council of Ministers of Education, Canada (CMEC). (2015). Agreed memorandum on a council of ministers of education, Canada. Retrieved from: https://www.cmec.ca/Publications/Lists/Publications/Attachments/360/2016-08-17-Agreed-Memorandum-EN.pdf

**Council of Ministers of Education, Canada (CMEC). (2016). Pan-Canadian assessment program 2016 assessment framework. Retrieved from: https://www.cmec.ca/docs/pcap/pcap2016/PCAP-2016-Assessment-Framework.pdf

**Council of Ministers of Education, Canada (CMEC). (2018). PCAP 2016 Report on the Pan-Canadian Assessment of Reading, Mathematics, and Science. (T.Y. 2016). Retrieved from: https://www.cmec.ca/Publications/Lists/Publications/Attachments/381/PCAP-2016-Public-Report-EN.pdf

**Council of Ministers of Education, Canada (CMEC). (2019). PCAP 2016 Technical Report. (T.Y. 2016). Retrieved from: https://www.cmec.ca/Publications/Lists/Publications/Attachments/394/PCAP%202016%20Technical%20Report_FINAL_EN.pdf

**Council of Ministers of Education, Canada (CMEC). (2020). Pan-Canadian assessment program 2019 assessment framework. Retrieved from: https://www.cmec.ca/docs/pcap/pcap2019/PCAP-2019-Assessment-Framework-EN.pdf

**Council of Ministers of Education, Canada (CMEC). (n.d. a). Pan-Canadian assessment program (PCAP) mathematics assessment framework. (T.Y. 2010). Retrieved from: https://www.cmec.ca/docs/pcap/pcap2013/Math-Framework-April-2013-EN.pdf

**Council of Ministers of Education, Canada (CMEC). (n.d. b). Pan-Canadian assessment program (PCAP) reading assessment framework. (T.Y. 2007). Retrieved from: https://www.cmec.ca/docs/pcap/pcap2007/Reading-Framework-EN-Apri2013.pdf

**Council of Ministers of Education, Canada (CMEC). (n.d. c). Pan-Canadian assessment program (PCAP) science assessment framework. (T.Y. 2013). Retrieved from: https://www.cmec.ca/docs/pcap/pcap2013/PCAP-2013-Science-Assessment-Framework-EN.pdf

**Council of Ministers of Education, Canada (CMEC). (n.d. d). Pan-Canadian assessment program website: https://www.cmec.ca/240/Pan-Canadian_Assessment_Program_(PCAP).html

*Dickinson, E.R., Gribben, M., Schultz, S.R., Spratto, E., & Woods, A. (February 2, 2020). Comparative analysis of the NAEP science framework and state science standards: Technical report.

*Jago, C. (March 2009). A history of NAEP assessment frameworks.

*Johnston, W.T., Stephens, M., & Ratway, B. (September 2018). Review of state curricular standards in mathematics: Supplemental exhibits.

**Kelly, D.L., Centurino, V.A.S., Martin, M.O., & Mullis, I.V.S. (Eds.) (2020). *TIMSS 2019 Encyclopedia: Education Policy and Curriculum in Mathematics and Science*. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: https://timssandpirls.bc.edu/timss2019/encyclopedia/

**Rostamanian, A. (2020). Canada. In Kelly, D.L., Centurino, V.A.S., Martin, M.O., & Mullis, I.V.S. (Eds.), *TIMSS 2019 Encyclopedia: Education Policy and Curriculum in Mathematics and Science*. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: https://timssandpirls.bc.edu/timss2019/encyclopedia/

*Stephens, M., Johnston, W.T., & Ratway, B. (December 2018). Review of state curricular standards in mathematics: Methodological narrative.

**Wendt, H., Schulz-Heidorf, K., Walter, D., & Steffensky, M. (2020). Germany. In Kelly, D.L., Centurino, V.A.S., Martin, M.O., & Mullis, I.V.S. (Eds.), *TIMSS 2019 Encyclopedia: Education Policy and Curriculum in Mathematics and Science*. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: https://timssandpirls.bc.edu/timss2019/encyclopedia/

Appendix B:
Review of Framework Processes in the Pan-Canadian Assessment Program (PCAP)

Relevance of PCAP

According to the *TIMSS 2019 Encyclopedia: Education Policy and Curriculum in Mathematics and Science*, the U.S. is not the only participating country without a national mathematics or science curriculum. Other countries without national curricula in these subjects in grade 4 include Belgium (Flemish), Bosnia and Herzegovina, Canada, and Germany (Kelly et al., 2020, Introduction p. 7). Among these four countries, only Germany has national education standards that are binding across the primary divisions of the country. In general, each of Germany's 16 federal states, however, has a different curriculum aligned to those standards (Wendt et al., 2020, Germany p. 1).

In this list of countries without national curricula, only the U.S. and Canada have a national assessment, and in Canada, it is only at grade 8. This assessment, known as the Pan-Canadian Assessment Program (PCAP), assesses student achievement in reading, mathematics, and science. Like NAEP, participation in PCAP is based on random sample selection (Rostamanian, 2020, Canada p. 8).

Assessment Frameworks

The Council of Ministers of Education, Canada (CMEC) oversees PCAP. Documentation on this assessment program is available on the CMEC website (CMEC, n.d. d). The first administration of PCAP was in 2007, following a CMEC directive that "a new pan-Canadian assessment program was needed to reflect changes in curriculum, integrate the increased jurisdictional emphasis on international assessments, and allow for the testing of the core subjects of mathematics, reading, and science." (CMEC, n.d. d). PCAP has been administered every third year since 2007.

CMEC provides a PCAP assessment framework document for each of these administrations. These documents each describe one or more of four frameworks in the PCAP programs (reading, mathematics, science, and questionnaire). In the most recent assessment framework published (for 2019, CMEC, 2020), there is a chapter dedicated to each of the four frameworks. Each of these chapters includes a description of its subject framework, variously characterized as a "working definition" (mathematics), "definition" and "organization of the domain" (science), "definition" following a "theoretical background" (reading), and "description" followed by "core questions" (questionnaire).

The 2019 PCAP framework document has a 6-page introduction to the PCAP, its contrast with classroom assessments, its languages and modes of administration, reporting aspects, and monitoring role. The document closes with a 3-page chapter on assessment design, briefly covering scale characteristics, administration time, numbers of booklets, descriptions of item types (selected response and constructed response), and item release schedules.

The framework document from the 2016 cycle of PCAP contains much of the same information. Although PCAP assessed students on all three subjects starting in 2007, the frameworks for a given content area do not appear prior to the year it was first a "primary" domain for PCAP (2007 for reading, 2010 for mathematics, and 2013 for science). The framework documents for those years, moreover, cover only the framework of the "primary" domain. Thus, the text for the reading framework first appears in 2007, then again, with some updates and variations in the

2016 assessment frameworks document and again (with some changes) in the document for 2019.

Key Aspects of Framework Processes

These documents, together with information on the PCAP section of the CMEC website, as well as public and technical reports published through the 2016 cycle (except for 2007, which does not have a technical report), are collectively called the "program documentation" here. Program documentation describes some of the processes for developing the PCAP frameworks. They leave some aspects of framework processes unaddressed.

Authority and/or Legislative Mandate

There is no legislative mandate for the administration of PCAP. Authority over the program is exercised by the CMEC, whose members are the provincial/territorial education ministers of Canada. CMEC is governed by a memorandum; this agreement does not explicitly address standards, curriculum, instruction, or assessments among its objectives or duties. The CMEC memorandum, however, lists that the Council "may conduct and support research and cross-jurisdictional assessments." (CMEC, 2015, p. 2)

There is no readily available official agreement currently governing the PCAP program. The first PCAP public report (CMEC, 2008) indicates that CMEC convened an August 2003 PCAP working group which commissioned a "concept paper […] that would elaborate on issues of structure, development planning, operations, and reporting" (p. 2) The report does not cite this concept paper. The report states, however, that the working group used it to define the PCAP, a definition followed by six brief bulleted statements addressing (among other aspects) assessed domains, population, frequency, basis ("the commonality of all current juristictional [sic] curricular outcomes across Canada", p.2).

Descriptions of Framework Derivation Process

None of the PCAP sources offer a description of how a person or group derived the current frameworks .

Intended Relationship to Academic Standards or Curricula of the Assessed Population

Sources indicate that the PCAP frameworks are informed by the curricular goals/objectives/outcomes of the participating provinces/territories. Each content area framework and public report either states or implies that the PCAP frameworks cover what is common across participants' curricular goals/objectives/outcomes.

Role of Curricula/Content Standards of the Assessed Population

Each content area framework indicates it is informed by one or two of three kinds of external sources. The first kind, addressed by all three frameworks, concerns the curricula of the participating provinces/territories. The mathematics and science frameworks each reference reviews, authored by CMEC and not published, comparing the curricula of that content area, across Canada. The reading framework implies that a review was conducted, but only refers the reader to official jurisdictional websites for updated curricula.

Role of Education Research in the Content Area

The second kind of external source concerns education research in the content area. For the reading framework, it is "current research findings and best practices in the field of literacy development and the learning of reading." (n.d. b, p. 1). The original reading framework (from the cycle 2007 assessment) does not cite one specific document that summarizes the relevant education research, but instead provides the author's (or authors') own view(s) about the domain of reading, citing several other sources, primarily in reading/literacy theory. The domain description section of the reading framework chapter of the cycle 2016 assessment framework document (CMEC, 2016) is a significantly expanded or updated version of the cycle 2007 reading framework, with more research sources cited, including some published after the original framework. The corresponding section of the reading framework chapter in the cycle 2019 assessment framework document (CMEC, 2020) is mostly unchanged from the cycle 2016 document.

Neither the mathematics nor the science frameworks indicates that it is directly informed by education research in the respective content area. (They may be indirectly informed by research, however, through other frameworks consulted.)

Role of Other Frameworks

We identified a third kind of source informing assessment frameworks: Other frameworks for curricula or assessments.

The domain description sections of the different versions of the PCAP reading frameworks (those in the cycle 2007, cycle 2016, and cycle 2019 framework documents) do not reference any such sources.

By contrast, the mathematics framework indicates that it is based on (the assessment frameworks for) the School Achievement Indicators Program (SAIP, which preceded PCAP), PISA and TIMSS. The documents indicate it has been guided by two National Council of Teachers of Mathematics (NCTM) documents: *Principles and Standards for School Mathematics* and *Curriculum Focal Points for Prekindergarten through Grade 8 Mathematics*. Although these different frameworks are described in the domain description section of the PCAP mathematics framework, their connection to the latter is not made explicit. That is, the PCAP mathematics framework does not report how its categories relate to the categories in these other frameworks.

The PCAP science framework also references the SAIP assessment framework and indicates it "takes into account findings from" PISA and TIMSS. (CMEC, n.d. c). However, the document seems to draw most heavily from another CMEC-authored framework, *Common Framework of Science Learning Outcomes K to 12* (CMEC, 1997).

Role of Professional Standards

The PCAP sources do not reference professional standards.

Sources for the Assessment Design

By "assessment design," we mean the way in which a domain description is made operational through weighting, test blueprints, item format decisions, and related specifications. The PCAP sources do not reference a process or other sources that inform the assessment design portion of the PCAP frameworks.

Authorship of Framework Documents

The first PCAP public report (from the 2007 cycle) indicates that in August 2003, a working group of "of experienced and knowledgeable representatives from several jurisdictions and including an external authority on measurement theory, large-scale assessment, and educational policy" (CMEC, 2008, p. 2) started the process of developing the assessment program. A "concept paper" (not cited) "would elaborate on issues of structure, development planning, operations, and reporting." (p. 2) The working group drew on this concept paper to "define" PCAP as follows:

"[PCAP will] be administered at regular intervals[,] be administered to students who are 13-year-olds at the start of the school year[,] be based on the commonality of all current juristictional [sic] curricular outcomes across Canada[,] assess reading, mathematics, and science[,] provide a major assessment of one domain with a minor concentration on the two other domains[, and] focus on reading as the major domain in the first administration in 2007. For each subject area, a thorough review of curricula, current assessment practices, and research literature was then undertaken and reports were written to indicate the common expectations among all jurisdictions." (p. 2)

The sources do not document the membership of this group, nor reference working groups or identify authors of the individual subject-area frameworks.

The cycle 2016 technical report references a working group and a specific contractor for updating the reading framework, but not the composition of the group.

Constituency Review Processes

Program documentation does not reference external or public review of frameworks.

Processes for Reviewing, Updating, and Revising Existing Frameworks

The cycle 2016 technical report indicates that the reading framework was updated for that assessment year. The text does not specify a process for arriving at a decision to review or update the framework. The description of the revision process is brief and does not document directives or parameters for the update nor consensus or constituency review processes. The document does not describe the specific changes made to the reading framework. (These changes, however, can be assessed through document comparison.)

Approval

PCAP program documentation does not reference a formal approval process for frameworks.

NAEP Science Framework

Under the leadership of the Assessment Development Committee (ADC), the Governing Board over the last three years has undertaken framework updates in mathematics and reading. As part of the Board's 2018 revision to the Framework Development Policy and the recently adopted Strategic Vision, the Board has set a goal of reviewing frameworks in a more proactive and timely manner.

Contributing to this goal, the ADC will soon lead a review of the NAEP Science Framework. Before the new Framework Policy (approved in 2018), the Board had undertaken only relatively minor updates to existing frameworks or the implementation of entirely new frameworks. Reflecting on the NAEP Mathematics and NAEP Reading Framework updates, the Committee recently discussed one potential process refinement to future framework reviews, which are the preliminary activities that take place to inform the ADC and the Board about whether or not a framework should be updated. ADC is considering holding a public comment period *in advance* of the framework review. The intent of this public comment collection would be to enable the Board to consider a wider array of perspectives as it makes the decision about whether or not to update a NAEP framework.

The ADC has agreed that it would be useful to have an earlier, comprehensive view of the issues in a given content area before a framework review begins. During its initial discussion, the ADC also noted that early public comment supports credibility for eventual Board decisions.

A draft of the public comment request is attached for the Committee's feedback.

Science Assessment Framework for the 2028 National Assessment of Educational Progress

AGENCY: National Assessment Governing Board, U.S. Department of Education.

ACTION: Notice of opportunity for public comment for the Science Assessment Framework for the 2028 National Assessment of Educational Progress (NAEP).

SUMMARY: The National Assessment Governing Board (Governing Board) is soliciting public comment for guidance in updating the Assessment Framework for the 2028 National Assessment of Educational Progress (NAEP) in Science.

The Governing Board is authorized to formulate policy guidelines for NAEP. Section 302 (e)(1)(c) of Public Law 107-279 s specifies that the Governing Board determines the content to be assessed for each NAEP Assessment. Each NAEP subject area assessment is guided by a framework that defines the scope of the domain to be measured by delineating the knowledge and skills to be tested at each grade and subject, the format of the assessment, and the achievement level definitions – guiding assessments that are valid, reliable, and reflective of widely accepted professional standards. The NAEP Science Assessment Framework was last revised in 2005. It is anticipated that the Governing Board will decide about the extent of revision needed to update the NAEP Science Assessment Framework at the National Assessment Governing Board quarterly meeting on March 3-5, 2022.

Public and private parties and organizations are invited to provide written comments and recommendations relative to the current framework, adopted in 2005. Comments should specifically address: (a) whether the NAEP Science Framework needs to be updated; (b)

if the framework needs to be updated, why is a revision needed; and (c) what should a revision to the framework include? This notice sets forth the review schedule and provides information for accessing additional materials that will be informative and useful for this review.

SUPPLEMENTARY INFORMATION:

Assessment and Item Specifications elaborate on the framework as guidance for item development conducted by the National Center for Education Statistics (NCES) and the NAEP assessment development contractor(s). The framework development and update process also produces recommendations for contextual variables, which supports NCES' development of the questionnaires administered to students, teachers, and schools to help the public understand the achievement results in each subject. By engaging NAEP's audiences, partners, and stakeholders in the panels that provide recommendations for NAEP frameworks and by seeking public comment, NAEP frameworks reflect content valued by the public as important to measure. Additional information on the Governing Board's work in developing NAEP Frameworks and Specifications can be found at https://www.nagb.gov/naep-frameworks/frameworks-overview.html.

All responses will be taken into consideration before finalizing the recommendations for the update of the NAEP Science Assessment Framework. Once finalized, recommendations will be used to guide a framework update process, if an update is needed for the 2028 NAEP Science Assessment.

Additional information (including the materials referenced below) can be found on the project website at https://www.nagb.gov.

Existing Science Framework for the NAEP

The existing framework (adopted in 2005) can be downloaded from the Governing Board website at https://www.nagb.gov/naep-frameworks/science.html.

Governing Board's Periodic Review and Updating of NAEP Frameworks

Governing Board policy articulates the Board's commitment to a comprehensive, inclusive, and deliberative process to determine and update the content and format of all NAEP assessments. For each NAEP assessment, this process results in a NAEP framework, outlining what is to be measured and how it will be measured. Periodically, the Governing Board reviews existing NAEP frameworks to determine if changes are warranted. Each NAEP framework development and update process considers a wide set of factors, including but not limited to reviews of recent research on teaching and learning, changes in state and local standards and assessments, and the latest perspectives on the nation's future needs and desirable levels of achievement.

In 2021, the Board is initiating a review of the NAEP Science Framework. The Governing Board's NAEP Science Framework review will use general public comment collected through this notice as well as expert commentary to determine whether a framework update is required and the type of updates that may be needed. Learn more about framework update processes at https://www.nagb.gov/content/dam/nagb/en/documents/naep/NAEP-Frameworks-FAQ_FINAL.pdf.