Meeting of
May 13-14, 2021

Via Zoom

REVISED AGENDA

Thursday, May 13: 12:30 – 5:30 pm (EDT)

12:30 – 12:45 pm	**Welcome**
	Approval of May 2021 Agenda
	Approval of March 2021 Minutes
	Haley Barbour, Chair
12:45 – 1:15 pm	**Executive Director's Update**
	Lesley Muldoon, Executive Director
1:15 – 2:15 pm	**Independent Review Time**
2:15 – 2:30 pm	***Break***
2:30 – 4:30 pm	**NAEP Reading Framework Policy Discussion**
	Dana Boyd, Chair, Assessment Development Committee
	Mark Miller, Vice Chair, Assessment Development Committee
4:30 – 4:45 pm	***Break***
4:45 – 5:15 pm	**National Center for Education Statistics Update**
	James Lynn Woodworth, Commissioner
5:15 – 5:30 pm	**Resolution in Honor of Michael Casserly, Executive Director, Council of the Great City Schools**
	Haley Barbour

Friday, May 14: 12:30 – 5:30 pm (EDT)

12:30 – 1:45 pm	**Results from the 2019 NAEP Science Assessment (CLOSED)** *Grady Wilburn, National Center for Education Statistics*
1:45 – 2:00 pm	***Break***
2:00 – 3:50 pm	**NAEP Budget and Assessment Schedule (CLOSED)** *Peggy Carr, Associate Commissioner, National Center for Education Statistics* *Lesley Muldoon*
3:50 – 4:00 pm	***Break***
4:00 – 5:30 pm	**Next Generation NAEP (CLOSED)** *Peggy Carr*

2021 - 2022 QUARTERLY BOARD MEETING DATES AND LOCATIONS

August 5 - 7, 2021	**TBD**
November 18 - 20, 2021	**TBD**
March 3 - 5, 2022	**TBD**
May 12 - 14, 2022	**TBD**
August 4 - 6, 2022	**TBD**
November 17 - 19 , 2022	**TBD**

Governing Board Members
2020 - 2021

Honorable Haley Barbour, Chair
BGR Group, Founding Partner
Yazoo City, Mississippi

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

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

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

Gregory J. Cizek
Guy B. Phillips
Distinguished Professor of Educational
 Measurement and Evaluation
University of North Carolina
Chapel Hill, North Carolina

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

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

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

Paul Gasparini
Secondary School Principal
Jamesville-DeWitt High School
DeWitt, New York

Honorable James E. Geringer
Former Governor of Wyoming
Cheyenne, Wyoming

Eric Hanushek
Hanna Senior Fellow
Hoover Institution
Stanford, California

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

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

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

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

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

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

Julia Rafal-Baer
Chief Operating Officer
Chiefs for Change
Cranston, Rhode Island

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

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

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

Representative Mark White
Tennessee House of Representatives
Nashville, Tennessee

Grover J. "Russ" Whitehurst
Professor Emeritus
Stony Brook University
Fort Myers, Florida

Carey M. Wright
State Superintendent
Mississippi Department of Education
Jackson, Mississippi

Ex-officio Member
Mark Schneider
Director
Institute of Education Sciences

National Assessment Governing Board

Committee Structure
(2020-2021)

Assessment Development Committee
Chair	Dana Boyd
Vice Chair	Mark Miller
	Christine Cunningham
	Frank Edelblut
	Patrick Kelly
	Reginald McGregor
	Nardi Routten
	Michelle Blair (Staff)

**Committee on Standards,
  Design and Methodology**
Chair	Greg Cizek
Vice Chair	Carey Wright
	Jim Geringer
	Rick Hanushek
	Suzanne Lane
	Alice Peisch
	Julia Rafal-Baer
	Russ Whitehurst
	Sharyn Rosenberg (Staff)

Executive Committee
Chair	Haley Barbour
Vice Chair	Alice Peisch
	Dana Boyd
	Greg Cizek
	Jim Geringer
	Tonya Matthews
	Mark Miller
	Beverly Perdue
	Marty West
	Carey Wright
	Matt Stern (Staff)

Reporting and Dissemination Committee
Chair	Tonya Matthews
Vice Chair	Marty West
	Alberto Carvalho
	Tyler Cramer
	Paul Gasparini
	Beverly Perdue
	Ron Reynolds
	Mark White
	Laura LoGerfo (Staff)

Nominations Committee
Chair	Jim Geringer
	Dana Boyd
	Tyler Cramer
	Paul Gasparini
	Tonya Matthews
	Reginald McGregor
	Mark Miller
	Alice Peisch
	Tessa Regis (Staff)
	Lisa Stooksberry (Staff)

National Assessment Governing Board

Meeting of March 4-5, 2021
Virtual

OFFICIAL SUMMARY OF GOVERNING BOARD ACTIONS

Complete Transcript Available

National Assessment Governing Board Members Present

Haley Barbour, Chair
Alice Peisch, Vice Chair
Dana Boyd
Alberto Carvalho
Gregory Cizek
Tyler Cramer
Christine Cunningham
Frank Edelblut
Paul Gasparini
Jim Geringer
Eric Hanushek
Patrick Kelly
Suzanne Lane
Tonya Matthews
Reginald McGregor
Mark Miller
Julia Rafal-Baer
Ron Reynolds
Nardi Routten
Martin West
Mark White
Grover Whitehurst
Carey Wright
Mark Schneider (ex-officio)

Governing Board Members Absent

Bev Perdue

National Assessment Governing Board Staff

Lesley Muldoon, Executive Director
Lisa Stooksberry, Deputy Executive Director
Michelle Blair

Stephaan Harris
Donnetta Kennedy
Laura LoGerfo
Munira Mwalimu
Tessa Regis
Sharyn Rosenberg
Angela Scott
Matt Stern
Anthony White

National Center for Education Statistics (NCES)

Lynn Woodworth, Commissioner
Peggy Carr, Associate Commissioner
Gina Broxterman
Samantha Burg
Jing Chen
Jamie Deaton
Alison Deigan
Enis Dogan
Patricia Etienne
Elvira Germino Hausken
Eunice Greer
Shawn Kline
Daniel McGrath
Nadia McLaughlin
Taslima Rahman
Emmanuel Sikali
Holly Spurlock
Sheila Thompson
William Tirre
Ebony Walton
Bill Ward
Grady Wilburn

American Institutes for Research (AIR)

Rebecca Bates
George Bohrnstedt
Markus Broer
Jack Buckley
Christina Davis
Kim Gattis
Courtney Gross
Cadelle Hemphill
Angelica Herrera

Kimberly Imel
Young Yee Kim
Sami Kitmitto
Gabrielle Merken
Jasmine Park
Amy Rathbun
Terry Salinger

Chief State School Officers, CCSSO

Kirsten Carr
Fen Chou
Kathleen Lyons
Scott Norton

CRP, Inc.

Shamai Carter
Arnold Goldstein
Subin Hona
Anthony Velez
Edward Wofford

Educational Testing Service (ETS)

Randy Bennett
Jay Campbell
Gloria Dion
Patricia Donahue
Amy Dresher
Kadriye Ercikan
Gary Feng
Robert Finnegan
Michael Friesenhahn
Janel Gill
Helena Jia
Irwin Kirsch
Cara Laitusis
Daniel McCaffrey
Rupal Patel
Hilary Persky
Emilie Pooler
Shannon Richards
Sarah Rodgers
Lisa Ward
Nancy Waters

Karen Wixson

Hager Sharp

James Elias
David Hoff
Joanne Lim

The Hatcher Group

Jenny Beard
Sami Ghani
Robert Johnston
Zoey Lichtenheld
David Loewenberg
Alexandra Sanfuentes
Devin Simpson
Jenna Tomasello

Human Resources Research Organization (HumRRO)

Monica Gribben
Hillary Michaels
Anne Woods

Management Strategies

Harrison Moore

Optimal Solutions Group

Imer Arnautovic
Brian Cramer
Charlotte Notaras

Pearson

Scott Becker
Cristina Everett
Cindy Flockhart
Emily Hilligoss
Abigail Keller-Dombrock
Lillian Moore
Eric Moyer
Noemi Nolter
Paula Rios

Pat Stearns
Tammy Visco
Cathy White
Llana Williams

Westat

Chris Averett
Greg Binzer
Lauren Byrne
Laura Egan
Mike Fassbach
Zully Hilton
Lisa Rodriguez
Rick Rogers
Keith Rust

WestEd

Georgia Garcia
Cynthia Greenleaf
Mira-Lisa Katz
Andrew Latham
Mark Loveland
Sonya Powers
Matt Rudoff
Megan Schneider
Steve Schneider
Sarah Warner

Other Attendees/Speakers

Rabab Abdulghani, Umm Al-Qura University
Kim Ackermann, TX Education Agency
Tammie Adams, U.S. Department of Education
Deb Adkins, NWEA
Sarah Aguirre, Northside Independent School District
Melissa Ahlgrim, OK State Department of Education
Abdullah Ahmed, King Saud University
Maisaa Alahmadi, Education and Training Evaluation Commission (ETEC)
Adbullah Alamri, Department of Education at Taif
Annette Allen, U.S. Department of Education
Eileen Allen, NY State Education Department
Ahmed Alfakih, Albaha University
Mohammed Alghamdi, ETEC
Hind Alharbi, MOE

Khaleel Alharbi, ETEC
Hassan Alhaythami, Umm Al-Qura University
Ahmed Almakrob, PSAU
Karima Almazroui, Abu Dhabi Department of Education and Knowledge
Hadeel Almubarak, Alqimam International School
Bigad Alotaibi, Dar Aluloom
Khalid Alsobhi, Hope Collaborative
Nasser Alresaini, DU
Shakir Alshareef, Retired Educator, Ministry of Education
Noura Alshehri, Ministry of Education
Fahad Alzahrani, Ministry of Education
Wendy Amelotte, Warwick Public Schools
Aama Amin, KSU
Connie Anderson, Grandmaloutunes
Judith Anderson, U.S. Department of Education
Marit Andrews, NM Public Education Department
Kara Arundel, Industry Drive
Diana Arya, University of California, Santa Barbara
Lori Assaf, TX Education Agency
Francesco Avvisati, Organisation for Economic Co-operation Development
Melissa Babcock, OH Department of Education
Ellen Bailey, Utah State Board of Education
Vickie Baker, WV Department of Education
John Ballen, Core Knowledge
Glynis Barber, Coppin State University
Toni-Ann Barone, Baldwin Union Free School District
Jill Barshay, The Hechinger Report
Shaun Bates, MO Department of Elementary and Secondary Education
Angela Battaglia, UT State Board of Education
Mark Bauerlein, First Things
Melissa Beck, MS Department of Education
Jenn Bell-Ellwanger, Data Quality Campaign
Shelley Beard, OH Department of Education
Renee Behring, Houghton Mifflin Harcourt
Rebecca Bennett, MA Department of Elementary and Secondary Education
Molly Berger, WA Office of Superintendent of Public Instruction
Teresa Berndt, SD Department of Education
Julie Bertram, Lexia Learning
Reeda Betts, AL State Department of Education
Linda Bevilacqua, Core Knowledge Foundation
Ken Bigger, Chicago Literacy Alliance
Krystal Bishop, Southern Adventist University
Rolf Blank, STEM K–12 Research
Pamela Bonds, Chicago Public Schools
Catherine Boomer, OK State Department of Education
Robin Boone, NY State Education Department

Rachel Bradshaw, TN Department of Education
Latosha Branch, VA Department of Education
Lori Bresnahan, Danbury Elementary, Newfound Regional School District
Melissa Brown, Curriculum Associates
Emily Bryans, NY State Education Department
Nancy Brynelson, California State University
Kymyona Burk, ExcelinEd
Abby Burke, NE Department of Education
Linda Burrows, AZ Department of Education
Melody Bushley, VA Department of Education
Sara Byrd, Sumter School District
Ruth Caillouet, LA Department of Education
Anne Cannon, Department of Defense Education Activity
Heather Casey, Rider University
Michael Casserly, Council of the Great City Schools (CGCS)
Jill Castek, University of Arizona
Gina Cervetti, University of Michigan
Caroline Chigbo, Enugu State Universal Basic Education Board, Nigeria
Ashlina Chin, Reading is Fundamental
Roberta Ching, California State University
Byeong-Young Cho, University of Pittsburgh
Julie Clark, Utah State Board of Education
David Coffey, NY State Education Department
Michael Cohen, CenterPoint Education
Julie Coiro, University of Rhode Island
Amy Conley, Fortuna High School
Elizabeth Conners, Dedham Country Day School
Dea Conrad-Curry, Partner in Education
Catherine Coons, NY State Education Department
Bill Cordes, U.S. Department of Education
April Crawford, Utah Health
Rachel Crowley, Kaufman Independent School District
Nicole Daniel, Frederick Smith Secondary School
Barbara Davidson, StandardsWork Inc.
Barbara Davis
Elisabeth Davis, AL State Department of Education
Elizabeth Davis, Odell Education
Tom Deeter, IA Department of Education
Danielle Dennis, University of Rhode Island
Clara DeSorbo, NY State Education Department
Colin Dingler, ACT
Janice Dole, University of Utah
Donna Dubey, NH Department of Education
Nell Duke, University of Michigan
Morgan Dunton, ME Department of Education
Richard Duran, University of California, Santa Barbara

Samantha Durrance, Southern Regional Education Board
Wendy Dury-Samson, NY State Education Department
Kari Eakins, WY Department of Education
Ginger Earl, Dallas Baptist University
Katie Eckelmann, North East Independent School District
Christopher Edley, Jr., Berkeley Law and Opportunity Institute
Kristin Edwards, Pemberton Township Schools
Veda Edwards, U.S. Department of Education
Rachel Eggleston, University of Michigan
Christine Elegante, Utah State Board of Education
Jeremy Ellis, MO Department of Elementary and Secondary Education
Amy Endo, Houghton Mifflin Harcourt
Christina Erland, Core Knowledge Foundation
Charmane Espejo
Marianne Farnsworth, Utah State Board of Education
Amr Fayed, Fursan Aljazira Private Schools
Matthew Ferguson, SC Education Oversight Committee
Chester Finn, Thomas B. Fordham Institute
Jack Fletcher, University of Houston
Meochia Ford, Sumter School District
James Forester, U.S. Department of Education
Elena Forzani, Boston University
Brian Frain, Rockhurst University
Laconduas Freeman, Flat Shoals Elementary
Mimi Fuhrman, ACT
Rebecca Gagnon
Susanna Gall, University of the West Indies, Cave Hill Campus
Laura Goadrich, AR Department of Education
Joycelyn Gooding, University of the West Indies
Jonathon Gonzales, MD State Department of Education
Lori Govenettio, Iroquois Central School District
Alisha Green, Chester County School District
Lana Green, MD State Department of Education
Peg Griffin
Becky Goetzinger, National Center for Families Learning
Jonathon Gonzales, MD State Department of Education
John Guthrie, University of Maryland
Joy Hakim, Stories For Thinkers
Robin Hall, CGCS
Shelley Hamel, WY Department of Education
Whitney Hamilton, KY Department of Education
Laura Hancock
Patty Hansen, Sumter School District
Stephanie Hansen, SD Department of Education
Raymond Hart, CGCS
Donna Hawkins, Johnston County Schools

Chris Hayes, Washoe County School District
Lowry Hemphill, SERP Institute
Tracy Herman, KY Department of Education
Ken Hermens, Smarter Balanced Assessment Consortium
Anita Hernandez, New Mexico State University
Ann Herrmann, MD State Department of Education
Rachel Hesprich
Elizabeth Hess, OH Department of Education
Maggie Hicks, AL State Department of Education
Kathleen Hinchman, Syracuse University
Eric Hirsch, Core Knowledge Foundation
John Hirsh, Georgetown University
Andrew Ho, Harvard University
Latasha Holt, University of Louisiana at Lafayette
Krista Hotelling, Utah State Board of Education
Christy Hovanetz, Foundation for Excellence in Education
George Hruby, University of Kentucky
Gerunda Hughes, Howard University
Mona Humaid, United Arab Emirates University
Abdulsamad Humaidan, Southern Illinois University Carbondale
Nikki Ingram, Kershaw County School District
Carol Jago, University of California Los Angeles (UCLA)
Debbie Jameson, MO Department of Elementary and Secondary Education
Jaclyn Janaszak, Living with Letters
Megan Jensen, Literacy Design Collaborative
Rucker Johnson, University of California, Berkeley
Cathy Jones, AL State Department of Education
Cathy Jones-Stork, Palmetto State Literacy Association
Kathleen Judy, LA Department of Education
Abdulrahman Kamal, Ministry of Education
Loretta Kane, Berkeley City College
Suhayb Kattan, Western Michigan University
Steven Katz, NY State Education Department
Michael Kieffer, New York University
Alissa Kilpatrick, LA Department of Education
Emily Kimpton, Sumter School District
Rainey Knight, SC Education Oversight Committee
Jim Kohlmoos, EDGE Consulting Partners
Nancy Kolodziej, Tennessee Technological University
Andrew Kolstad, P20 Strategies LLC
Beth LaDuca, OR Department of Education
David Laird, TN Department of Education
Emily Leute, VT Agency of Education
Regina Lewis, ME Department of Education
Tamara Lewis, MD State Department of Education
David Liben, SAP

Sue Livingston, LaGuardia Community College – The City of University of New York
Brian Lloyd, MI Department of Education
Megan Lopez, Utah State Board of Education
William Lorie, National Center for the Improvement of Educational Assessment
Tina Love, U.S. Department of Education
Shelley Loving-Ryder, VA Department of Education
Phyllis Lynch, RI Department of Education
Banks Lyons, TN Department of Education
Kathleen Maher-Baker, MD State Department of Education
Scott Marion, National Center for the Improvement of Educational Assessment
Katina Marshall, College Board
Bren Martin, United Nations Association of the United States of America
Erin Maughan, National Association of School Nurses (NASN)
Margaret McKeown, University of Pittsburgh
Kristen McKinney, MO Department of Elementary and Secondary Education
Ebony McKiver, NE Department of Education
Charlene McKnight, Sumter School District
Joann McRell, KS State Department of Education
Heidi Mills, University of South Carolina
Jennifer Moone, Reading is Fundamental
Linda Moreno, VT Agency of Education
Jessica Morton, AL State Department of Education
Nichole Mosser, MI Department of Education
Raina Moulian, AK Department of Education and Early Development
Lelsie Mugan, NWEA
Kristen Munger, State University of New York at Oswego
Stephen Murphy, Cognia
Ellen Muscato, Douglas County School District
Juliana Musselman, TN Department of Education
Suzanne Naiman, Sarasota County Schools
Abdu Nashrey, Ministry of Education
Deborah Neisuler, Curriculum Associates
Yi-Chieh Newton, Florida Center for Reading Research
Susanne Nobles, ReadWorks
Jill Nogueras, VA Department of Education
Eileen Oboler, Literary Consulting Services, LLC
Colleen OBrien, RI Department of Education
Philip Olsen, WI Department of Public Instruction
Kathy Padgett, AL State Department of Education
Susan Palmiter, MI Department of Education
Cindy Parker
Ashley Patterson, Penn State University
Marissa Payzant, NE Department of Education
P. David Pearson, University of California, Berkeley
Sarah Pennington, Montana State University
Marianne Perie, Measurement in Practice, LLC

Nancy Perkins, MD State Department of Education
Jodi Pilgrim, University of Mary Hardin-Baylor
Susan Pimentel, Student Achievement Partners
Kelly Pizani, LA Department of Education
Neerja Punjabi, Peel District School Board
Lori Pusateri-Lane, WY Department of Education
Erica Queen, Harford Community College
Gwen Quinn, Lower Merion School District
María Guadalupe Ramírez-Silva, Dallas Independent School District
Carol Rasowsky, Learner's Journey
Sara Ratner, University of Sydney
Jennifer Ray, Round Rock Independent School District
Melanie Reaves, Montana State University Billings
Nicole Renner, Carnegie Learning
Patrick Riccards, Driving Force Institute
John Richard, OH Department of Education
Susannah Richards, Eastern Connecticut State University
Katherine Ringley, VA Department of Education
Eddie Rivers, U.S. Department of Education
Tina Roberts, OR Department of Education
Spa Robinson
Cecilia Roe, MD State Department of Education
Alicia Ross, Blue Ridge High School
Rebecca Rundlett, SC Department of Education
John Sabatini, Global, Integrated, Scenario-based Assessments (GISA)
Hailah Saleh, Ministry of Education
Danielle Saucier, ME Department of Education
Renee Savoie, CT State Department of Education
Petra Schatz, HI Department of Education
Angela Schroden, Stephanie Harvey Consulting
Sarah Schwartz, Education Week
Pamela Seastrand, Curriculum Associates
Vincent Segalini, Odell Education
Frank Serafini, Arizona State University
Maria B. Serpa, Lesley University
Lisa Scroggs, MO Department of Elementary and Secondary Education
Llewellyn Shealy, SC Department of Education
Amy Sheridan, DE Department of Education
Courtney Shimek, West Virginia University
Valerie Shinas, Lesley University, Graduate School of Education
Bonnie Short, Alabama Reading Initiative
Gina Sierzega, Lehigh University
Debra Silimeo, Silimeo Group
Adrienne Simmons, GA Department of Education
Nancy Sinotte, Warwick Public Schools
Paulson Skerrit, University of the West Indies

Michael Slattery, Huntington Ingalls Industries, Inc.
Candace Slobodnik, Crossland High School, Prince George's County Public Schools
Kelly Smith, ACT
Patriann Smith, University of South Florida
Adrienne Snow, Enfield Public Schools
Peggy Sorensen, OH Department of Education
Katherine Stahl, New York University
Shauna Stephanchick, STEP Up Consulting
David Steinerd, Johns Hopkins University
Jason Stephenson, OK State Department of Education
Mark Stephenson, KS State Department of Education
Alice Marie Stevens, Hobbs Municipal Schools
Pamela Stewart, DTRE
Christy Talbot, American Educational Research Association
Beth Tarasawa, NWEA
Katherine Tarca, MA Department of Education
Tara Tencza, Scotch Plains-Fanwood Public Schools
Lesley Thomas, East New York Middle School of Excellence
Elizabeth Tomev, WI Department of Public Instruction
Carolyn Trombe, NY State Education Department
Carolyn Turner, OH Department of Education
Katie Uelk, UChicago Impact, Urban Education Institute, University of Chicago
Anand Vaishnav, Education First
Sheila Valencia, University of Washington
Shannon Varley, Keystone Oaks School District
Sheri Vasinda, OK State University
Heather Villalobos Pavia, CO Department of Education
Deb Wade, OK State Department of Education
Stefanie Wager, OER Project
Kate Walsh, National Council on Teacher Quality
Sue Ward, ACT
Naomi Watkins, Utah State Board of Education
Natalie Wexler
Whitney Whealdon, Learning Tapestry
Katy Wiggs, South Carolina Virtual Charter School
Liz Williams, Utah State Board of Education
Arlette Willis, University of Illinois, Urbana Champaign
Maja Wilson, WA Office of Superintendent of Public Instruction
Crystal Wise, University of Illinois at Chicago
Angela Woodard, U.S. Department of Education
Roberta Woods, U.S. Department of Education
Kristen Wynn, MS Department of Education
Cindy Ziker, Ziker Research

Opening Remarks

Haley Barbour, Chair, called the session to order at 12:17 p.m. and welcomed attendees to the March 4, 2021, National Assessment Governing Board (Governing Board) meeting held by webinar.

Approval of March 2021 Agenda

Barbour requested a motion for approval of the March 2021 agenda. A motion to accept the agenda was made by Mark White and seconded by Alice Peisch. No discussion ensued and the motion passed unanimously.

Approval of November 2020 Board Meeting Minutes

Barbour requested a motion for approval of the minutes of the November 2020 Governing Board meeting. Carey Wright made a motion to approve the November 2020 minutes and Mark Miller seconded the motion. There was no discussion and the motion passed unanimously.

Action: 2019 NAEP Science Release Plan

Reporting and Dissemination Committee Chair Tonya Matthews presented a summary of the 2019 NAEP Science release plan. Tyler Cramer made a motion for the Board to approve the 2019 NAEP Science release plan. The motion was seconded. There was no discussion and all members approved.

Institute of Education Sciences Update

Mark Schneider, Institute of Education Sciences (IES) Director, spoke on two topics: (1) the School Survey Dashboard and (2) IES studies and how they relate to the Governing Board. First, a Presidential Executive Order requires IES to survey schools on the impact of COVID-19. The monthly School Survey Dashboard provides valuable data about online, in-person, and hybrid learning. Monthly reports will begin at the end of March and continue through June.

Beginning in August, IES will administer an expanded school survey on a monthly or bimonthly basis. This future school survey will include a broader range of questions about school conditions related to COVID-19 and recovery. Schneider appreciated using NAEP sample information to deploy the survey. The next school year will bring significant changes to schools, and Schneider hopes the survey will yield critical information on teaching, learning, and general school conditions across the nation.

Second, Schneider discussed three strands of a new study commissioned by IES and directed by the National Academies. The first strand reviews research topics that the National Center for Education Research and the National Center for Special Education Research fund. Each year, these agencies review more than 900 grant applications. The IES' proposal asked the National Academies to rethink the structure for research topics using a 10-year perspective.

The second strand of the National Academies' study focuses on staffing and budget issues for NCES. The proposal asked if NCES products are serving the nation and the taxpayers effectively. Schneider summarized the goal of this strand as answering the question, "what is the goal of the statistical collections that NCES does, and what's the best way of answering or getting those data?"

The third strand studies the technology of NAEP—is it optimal, and if not, how can NAEP administration improve? Schneider reported how NAEP is using automated scoring for reading and mathematics, which is a step forward from the past approach. However, Schneider challenged NCES to consider researching more cutting-edge assessment techniques, such as automated item generation. He acknowledged item generation might not be feasible at the current time but urged the Governing Board to be visionary since more efficient assessments could save time and money. NAEP's per student cost is estimated at more than three times the per student cost of the ACT and SAT even though NAEP does not generate individual student information. Schneider urged the National Academies to balance a review of existing technologies with costs.

Schneider's final comment called attention to the students performing below the NAEP Basic achievement level; he emphasized that the lowest-performing students are regressing, not progressing on NAEP. Many assessments focus on what these students cannot do, however, discussion must center on what these students know and can do. He suggested multi-stage or computer adaptive individual-level testing to uncover more information about the lowest-performing students. He cautioned this testing must be developed and deployed in a cost-effective manner.

Schneider answered several questions from Governing Board members about the School Survey Dashboard. School administrators or state analysts, rather than students, are responding to the survey which focuses on different types of learning delivery methods and which student subgroups are involved in each. Frank Edelblut asked how the data will be used, especially considering rapidly changing classroom environments. Schneider conceded that classroom dynamics are fluid but reassured the Governing Board that results would not be oversimplified. He reiterated data are collected monthly, which reflects the evolving situations in schools. Edelblut followed up with a question about the future of NAEP testing. Although NAEP does not report on each student, it provides a good benchmark and results are correlated with state summative results. Is there a way for NAEP to statistically meet the requirements of state summative tests in a coordinated effort between federal and state agencies? Schneider replied there are mapping methods that can be used to equate state test scores with NAEP, however, a more complete overhaul of the system would involve changing existing legislation. Additionally, there are differences between the purpose of NAEP and those of student-level state assessments.

Lynn Woodworth, NCES Commissioner, reminded Board members that legislation currently prohibits NAEP from being used to evaluate any school. Suzanne Lane asked if the survey is

collecting information on teacher professional development, absenteeism by method of instruction, and teacher strategies for engaging students. The survey only asks for absenteeism data. Schneider cautioned against adding too many questions since response rates for lengthy surveys are not high. Peisch worried about capturing information about extended absenteeism, essentially students who have dropped out, and Woodworth assured her enrollment data by race and subgroup would help capture this important information.

Executive Director's Update

Governing Board Executive Director Lesley Muldoon began by introducing Matt Stern, who recently joined the Governing Board staff. Stern serves as Assistant Director for Policy and Intergovernmental Affairs; his duties include monitoring federal legislation and budget recommendations, meeting with key partners and stakeholders of the Board, and reviewing policy issues for consideration by the Board. Prior to joining the staff, Stern was a K–12 policy advisor for the Senate's Health, Education, Labor, and Pensions (HELP) Committee and a former middle school teacher. Stern said he is looking forward to his new role.

Muldoon acknowledged the unprecedented challenges to learning that COVID presented including disruption of the NAEP assessments. However, Muldoon looked to opportunity, dubbing this year "the interstitial year," which allows time for the Governing Board to focus on broader strategic priorities.

With vaccination rates expanding and schools reopening, Muldoon expressed hope for a return to normalcy in education, with a goal to ensure a successful 2022 NAEP administration. The two NAEP-related surveys currently in the field should yield helpful information about student learning. Many are concerned with the persistent patterns of students falling behind in the last year, particularly students of color, low-income students, and students with disabilities. Researchers have indicated the disproportionate impact of COVID-19 on these populations, although no state or national assessment can corroborate these findings yet. Muldoon shared information highlighting how insights from NAEP can help inform educational recovery and improve student learning, especially for those students who are falling behind.

NCES has reported on plans to build several capabilities for the next generation of NAEP's digitally based assessments (DBAs). When COVID-19 closed schools in March 2020, there was pressure to remotely administer assessments. However, the security and logistics for contactless delivery methods require planning and research and consideration by the Governing Board of substantial policy implications.

Muldoon reported on activities related to Strategic Vision 2025. The Strategic Vision is an important organizing framework for Governing Board priorities, guiding the Board's work and facilitating attaining their goals to inform, engage, and innovate.

Muldoon concluded her update with a list of priorities for 2021, including: (a) review the current Science Framework, (b) release the 2019 Science assessment results for grades 4, 8, and 12, (c) monitor NAEP appropriations, (d) procure a new Technical Support contract to allow special research studies, and (e) redesign the website to better serve the public.

Reflections on Recommendations from the National Academies' Committee on Developing Indicators of Educational Equity

Matthews opened the session on developing indicators of educational equity. She introduced the first panelist, Rucker Johnson, the Chancellor's Professor of Public Policy in the Goldman School of Public Policy at the University of California, Berkeley. Johnson affirmed the importance of NAEP data and claimed that without NAEP, researchers cannot measure and track learning improvements and deficits related to school resources. Johnson provided extensive data on the big picture of inequity in education. Using Sean Reardon's Standardized Measure of Test Performance methodology, Johnson used NAEP as a benchmark to convert school and student test scores to grade-level equivalents at the national level, thus allowing comparisons of district per-pupil spending and achievement.

School district comparisons across the nation set the context for Johnson's examination of the impact of California's Local Control Funding Formula (LCFF), which included a significant investment of $18 billion in public schools from 2013 to 2019. The LCFF money was distributed using a progressive funding formula with implications for educational equity and ways in which interventions can narrow gaps. He compared data on how much LCFF funding districts received and how they spent the funds by achievement test scores and high school graduation rates. Johnson provided grade- and subject-level graphs illustrating spending trajectories before and after the LCFF was implemented. Using difference-in-difference estimates and controlling for student achievement, Johnson showed increases in student performance for low-income, high-funding districts. A $1,000 increase in per-pupil funding during a 3-year period from grade 6 to grade 8 was equivalent to significant math score improvements of .23 standard deviations in eighth grade, on par with a full year of learning. Results were similar across grades and subjects. Johnson described how investments in pre-kindergarten and transitional kindergarten translated to stronger outcomes in subsequent grades. Johnson's research highlights the importance of how districts spend funds to such student outcomes as high school graduation rates. He commented on the impact of the pandemic and distance learning, suggesting many of the gains will be reversed. He suggests California's inequity-oriented framework for resource allocation may be a model for national investments in education.

Gerunda Hughes, a member of the NAEP Validity Studies (NVS) Panel and Professor Emerita at Howard University, spoke about the role of NAEP as an indicator of educational (in)equity. Hughes suggested that not only does NAEP serve as an indicator of inequity but also NAEP can be infused with more equitable design, questions, and reporting.

In *The Journal of Negro Education* in 1995 Edmund Gordon stated "[e]quity speaks to fairness and social justice and the acknowledgement of differences. It references the differential or (un)equal distribution of resources or inputs for the purpose of meeting a specific need to address a particular purpose or outcome." Hughes explained that equity speaks to fairness, social justice, and unequal resources. It refers to the unequal distribution of resources so that individuals receive what they need to achieve an outcome, compared to equality where everyone receives the same resources.

Hughes pointed out the legislation states NAEP should (a) provide fair and accurate measurement of student academic achievement, (b) report trends, (c) be administered to a representative student sample, and (d) collect and report data on groups including race/ethnicity, socioeconomic status (SES), gender, and disability in a valid and reliable manner. She asserts the law allows collection of other student grouping data not listed in the legislation. She suggests additional variables to address (a) societal, (b) socioeconomic, (c) cultural, (d) familial, (e) programmatic, (f) staffing, (g) instructional, (h) linguistic, and (i) assessment inequities of our educational system. She outlined an NVS Panel proposal to make NAEP a more equitable assessment.

Hughes defined equitable educational assessments as fair and accurate measures with valid interpretations and uses. Equitable assessments should be aligned and validated with their specified interpretations and intended uses of results. Hughes provided an example from R. L. Thorndike's *Applied Psychometrics* textbook to illustrate the importance of validity of inferences and use of assessment results.

Hughes highlighted where an equity lens can be applied to NAEP, namely in: (a) sampling, (b) assessment design and development, (c) administration, (d) accommodations, (e) data analysis and reporting, (f) reporting and interpretations, and (g) use of results. She concluded with the five "E's" of equitable educational assessment: empathy, engagement, equity, evaluation, and equality.

Christopher Edley, the Honorable William H. Orrick, Jr., Distinguished Professor of Law at the University of California, Berkeley Law School, chaired the National Academies' Committee on Developing Indicators of Educational Equity. The Committee recommended developing indicators of educational equity in seven domains. Edley focused his presentation on one domain--elementary and secondary school learning, and three recommendations for indicators: (1) engagement in schooling; (2) performance in coursework; and (3) performance on tests. Committee members identified constructs to measure these indicators, e.g., engagement in schooling can include attendance or absenteeism and academic engagement. Future work needs to define the constructs.

Edley explained that the next steps are to use scientific research to build on existing systems of data to measure and collect the indicators. In some cases, research and development are needed. For example, there is no consensus on how to measure effective teaching. The committee

suggested proxies for indicators such as years of teaching experience. Edley emphasized the unique opportunity the education community currently faces—building a system of equity indicators while equity is a trending topic in the public consciousness. Edley admitted that some tailoring for subgroups of special interest and relevance (e.g., Native American/Alaska Native) may be needed, but there should be a core set of indicators with comparability across jurisdictions.

Some states are interested in implementing the committee's recommendations. Edley hopes the Governing Board will adopt a resolution commending the committee's effort to help elevate the importance of the initiative and to raise funds to support next steps. He offered an ambitious suggestion of a possible statutory expansion of the Governing Board to include overseeing a national system of educational equity indicators or to serve as an institutional home of the indicators.

Matthews facilitated Board member questions for the panelists. Jim Geringer asked Johnson about the most effective use of local funding. Few studies have the statistical power to answer this question. School resources matter and how money is spent matters. Educators need to think beyond *what* is working to *how* it is working and align interventions to school settings. Additional data such as out-of-school activities are needed.

Woodworth clarified several issues raised by panelists. The legal requirement for NAEP sampling does not preclude oversampling. NCES is field testing a new SES indicator with selected states. NCES staff is permitted to conduct secondary analyses, however, they are constrained by funding and staffing limitations.

Gregory Cizek emphasized that Johnson's presentation was too important for the limited time available. Edelblut questioned the dependencies of some of the equity indicators, e.g., absenteeism and success in class. If students do not attend class, they are less likely to succeed in school. Is there any weighting of the indicators? Edley stated weighting is a policy or political concern and not something the committee addressed.

Recess

The March 4, 2021, Governing Board meeting recessed at 3:05 p.m. and reconvened at 3:16 p.m.

Reading Comprehension in Large-Scale Assessment: A Symposium

Patrick Kelly moderated a panel of experts at the symposium for reading comprehension in large-scale assessment. Board members heard viewpoints of scholars as well as leaders of state, national, and international assessment programs. After general remarks on the role of background knowledge in reading comprehension, speakers focused on student background knowledge and its role in testing environments.

Dan Willingham, University of Virginia, began by acknowledging "expertise" studies, where students are given two passages of equivalent difficulty, with one passage on a randomly assigned topic and another passage for which the student has expressed interest in the topic. These studies consistently show students are much better at comprehending text on topics where they are highly familiar with the topic at hand. Another family of studies administers a background knowledge assessment and then examines the correlation between background knowledge scores and reading test scores. Willingham reported that the findings from this second set of studies support the idea that people with broad background knowledge consistently perform better on reading comprehension tests.

Willingham then described how psycholinguists think about comprehension on three levels, with knowledge a common factor across all. The first level is focused on making meaning within a sentence. Willingham discussed the difficulty of understanding a sentence and how, when words are rearranged in a sentence, the meaning changes because the syntax changes. Willingham argued that even when syntax is correctly applied many sentences are still ambiguous. Therefore, to understand the meaning of an ambiguous sentence, a student must depend on supplemental background knowledge, something many people take for granted since background knowledge is implicitly applied and resolves the confusion. The second level of comprehension is making meaning across multiple sentences. Willingham presented a series of sentences where the reader would need to infer some detail to understand the full sequence of events and make connections between the different actions to fully understand the meaning of the passage. Willingham stressed that all readers are constantly replacing omitted information with information from memory and noted that these replacements are central to the process of reading comprehension. For instance, he noted that providing just-in-time information, such as pop-up notes, would be problematic because it would replicate some of these inferencing skills that are deeply engrained in reading comprehension, i.e., doing it for students rather than having students do it themselves. Willingham concluded by stating that background knowledge is central to what reading is about and that attempting to mitigate background knowledge in assessments was not advisable.

Kelly thanked Willingham. For more context on how background knowledge has emerged as a centerpiece for the NAEP Reading Framework update, Kelly described the timeline of the framework update, beginning with an initial review in 2018 through present day revisions.

Gina Cervetti, University of Michigan, presented on behalf of the Framework Development Panel for the ongoing NAEP Reading Framework Update. She expanded on Kelly's timeline by describing in greater detail how background knowledge is currently addressed in the NAEP Reading Framework and assessment. She stated there are many kinds of knowledge that play important roles in reading comprehension and test scores may not reflect comprehension ability, but instead differences in knowledge regarding different topics. For example, some students may obtain knowledge through exposure to curriculum standards or experiences and will be better equipped to answer assessment items on selected topics. Cervetti reported that the current NAEP Reading Assessment includes two strategies to address variations in background knowledge, text selection and support features, which includes pop-up notes and introductions to passages. She illustrated the support features with respect to two passages.

Next, Cervetti compared how knowledge is addressed in the current NAEP Reading Framework and proposed framework update. While text selections remain similar, a shift in the proposed framework refers to the two support features (pop-up notes and introductions) as knowledge-based universal design elements (UDEs), bringing NAEP design in line with contemporary research on assessments. Under the framework update, Cervetti explained that introductions may be somewhat more elaborate and may also include brief videos, images, or audio recordings to provide topical information on topics that are likely to be unfamiliar. After demonstrating a passage example that included images, pop-up notes, and audio, Cervetti stated that these knowledge-based UDEs increase the validity of interpretations from the assessment and improve the ecological validity of the assessment by reflecting how people use supplemental resources to read texts on unfamiliar topics.

Ina Mullis, Executive Director of the Trends in International Mathematics and Science Study (TIMSS) and Progress in International Reading Literacy Study (PIRLS) International Study Center, spoke on minimizing the impact of background knowledge in the context of international assessments. She related PIRLS to NAEP, indicating they measure some similar domains and sample student populations rather than individuals, but PIRLS tests in different countries instead of states. Mullis added PIRLS is a fourth grade only assessment given every 5 years and measures linguistic skills and comprehension strategies. While recognizing that background knowledge is part of reading comprehension, PIRLS works to reduce the need for and impact of prior knowledge. The PIRLS framework includes two purposes for reading and four purposes for comprehension. The assessment includes 18 reading passages and five online informational texts, known as ePIRLS, that simulate internet reading. The passages and texts represent a wide range of content and settings, with background knowledge spread randomly throughout the assessment using a counterbalanced design across passages, tasks, and students. Passages reflect authentic reading experiences, and texts that depend on culture-specific knowledge are usually excluded. She clarified texts may introduce new information or knowledge, but it must be presented in a manner that can be easily understood by test takers unfamiliar with the topic, eliminating the need for pop-up text windows, for example. Additionally, to avoid creating advantages or disadvantages, items are passage dependent. That is, students do not need outside information to understand and answer items; they only need to read the text. Mullis noted that the advantage provided by background knowledge is regarded as bias. To reduce bias, a committee ensures texts and items avoid topics that favor specific ethnicities, geographical location, cultures, and gender. Given that half of the assessment's items are constructed response, PIRLS scoring guides are developed to meet important aspects of the assessment and define the responses as evidence of reading comprehension from the text. Mullis concluded by noting that there is no scoring advantage for students who display extra background knowledge in their responses.

Andreas Schleicher, Director for Education and Skills, and Special Advisor on Education Policy to the Secretary-General at the Organisation for Economic Co-operation and Development (OECD), described how the Programme for International Student Assessment (PISA) addresses background knowledge. He indicated PISA has adopted a contemporary definition of literacy that extends beyond understanding text. For example, PISA places a lot of emphasis on students'

ability to navigate ambiguity, assess the quality and credibility of information, and corroborate information. Background knowledge plays an important role in performing these tasks. However, other factors also contribute to performance, including motivation for the reading task, cognitive ability, as well as engagement and familiarity with the topic. Text factors are also important in assessing literacy, such as complexity of the text and what information students are asked to process. Schleicher indicated background knowledge affects each of the factors, but not always in a beneficial way. For example, students may draw from experiences in ways that are misleading relative to an assessment task. In any case, Schleicher noted that if students are able to answer test questions on the basis of background knowledge alone, then this would raise a fairness issue. PISA is administered in more than 100 countries and educational systems, providing a laboratory of cultural context to compare and contrast how the same item functions across different students in different settings. Like PIRLS, PISA uses the diversity of social and cultural context among countries to minimize bias and reduce the impact of prior knowledge, screening items for appropriateness for all students. Unlike PIRLS, PISA's authentic reading passages include cultural information as a way to have more authentic reading tasks. For example, introductory information would provide all the information a student needs to respond to items. No additional background knowledge is required. If a student has background knowledge of a concept, it would not provide an advantage because reading the passage is still required to answer the items. Pop-up notes and animations also appear in PISA, and tasks are purpose-driven.

Schleicher indicated that the impact of background knowledge is one of the most difficult things to measure and shared some methods PISA uses to address the issue. Concerns about group level comparisons are addressed with a model in which performance comparisons within a country are based on every item, but linking to international scales is based only on the items for which there is no item-by-country statistical interaction. Furthermore, as an additional check, each country chooses a set of items that best represents country-specific knowledge, and their data are then rescaled using only those preferred items. When countries are compared based on these rescaled scores, results show no significant influence on the rankings of countries. Finally, PISA collects student contextual information on reading strategies. The assessments also collect metacognition data on reading, asking students about self-efficacy on different reading tasks and motivation, which are correlated with performance. PISA examines these relationships in different groups.

Kelly facilitated a short question and answer session with the international assessment panelists. Edelblut asked Mullis about the absence of UDEs in reading passages. Mullis clarified although passages are text rich and diverse, they are written in a manner that a student does not need background information. If a student can read the passage, it contains all the knowledge needed to answer the corresponding items. She added passages go through multiple stages of review. Cizek asked how much should background knowledge be controlled for in an assessment. He said if he encountered a word he did not know when he was reading, he would look it up; he supported this practice as a reading skill to be learned, but not pushed on students in an assessment. He asked what concerns speakers had about attempting to control for background knowledge in NAEP assessments, as proposed in the NAEP Reading Framework update. Mullis agreed that spoon-feeding information to students while they are reading is not authentic reading

and believes this strategy could be distracting, leading to adverse effects on reader comprehension. Schleicher argued for the need to control background information in a way that one can measure its impact differentially based on content knowledge, reading strategies, social background, and other factors. Geringer stated that he struggles with differentiating between assimilation and comprehension, though he believes background knowledge is important for reading comprehension. He drew parallels to other subjects like mathematics and physics and general problem solving where prior knowledge drives performance. Schleicher agreed it is a challenge to control for all background knowledge in an assessment. He asserted that the focus of standardized assessment is to exclude assessment tasks that clearly favor (or disfavor) certain groups, e.g., in a specific geographic area or cultural context. From the perspective of PISA, there is no problem if the assessment task is likely to be unfamiliar to all students.

Martin West invited Willingham to comment on the presentations from Mullis and Schleicher, citing fundamental differences in background knowledge definitions. What Willingham characterizes as essential, Mullis views as bias, and West wants to learn more about the differences. Willingham clarified that key differences include his belief that background knowledge does not need to be solved in terms of fairness because background knowledge is an integral part of reading comprehension; it is problematic to attempt to separate background knowledge from comprehension. Willingham posited that background knowledge should be part of the construct of any reading comprehension assessment.

Following the international assessment discussion, Kelly invited additional guest speakers to provide presentations on the role of background knowledge on several U.S. large-scale assessments. He noted that these assessments produce scores for individual students, while NAEP does not.

Jenna Chiasson, Louisiana Department of Education, discussed the state's innovative assessment pilot. The state uses an instructional review process, making it easier for school systems to adopt high quality instructional materials. Accordingly, Chiasson reported that seventy-five percent of Louisiana school systems are using the same English language arts curriculum which integrates social studies content, and this widespread adoption provides a unique opportunity to connect curriculum and assessments. The innovative pilot program uses several brief assessments rather than an end-of-year exam and includes a reading and writing assessment that is sequenced with knowledge-rich curriculum that measures student ability to understand and build knowledge from reading and then express that knowledge and understanding in writing. She indicated initial feedback and data have been positive, citing higher levels of engagement and time spent on reading and writing assessment tasks. Data also show historically disadvantaged students perform better on the pilot assessments than on Louisiana's traditional assessments. Chiasson noted that school districts preserve local control by selecting which books are used in instruction and which assessment students take. Chiasson compared the traditional assessment—the Louisiana Educational Assessment Program or LEAP—with the innovative pilot assessment. On LEAP, students engage in cold reads, a random selection of texts which are purposely unrelated to anything students have studied. Conversely, the pilot uses warm and hot reads, wherein warm reads involve passages that are topically related to what students have studied but have not

encountered in school and hot reads relate to actual passages and familiar texts students have encountered in school. Unlike the yearly essay on the traditional exam, the pilot provides the opportunity for students to write essays on a more frequent basis. Chiasson shared a prompt from the end-of-year pilot test, which consisted of an essay question that required students to use and extend the knowledge they gained from their English classes to synthesize texts from multiple sources and integrate their thoughts into a cohesive writing piece.

Rachel Kachchaf, Smarter Balanced Assessment Consortium, presented information on their approach to reading and background knowledge, beginning with a short overview of the Smarter Balanced assessment design before delving into their process for selecting passages. Kachchaf noted that the Smarter Balanced assessment development process leverages educator expertise and judgment to both select the passages and write assessment items. She explained that this educator involvement provides representation across a variety of backgrounds, certifications, experiences, and geographic locations. Passage selections are drawn from existing texts that are identified by educators and aligned to standards. Each passage is reviewed to adhere to bias and sensitivity guidelines as well as language complexity guidelines. From there, the passage undergoes an iterative review process by teams of educators in a holistic evaluation that includes quantitative and qualitative measures. Test items are then developed for each passage in a process that is also iterative. As with passages, items are reviewed multiple times for content, bias and sensitivity, and fairness, and the overall the process aligns with evidence-centered design. She offered an example of items that discussed playing at the beach or in snow—even though not all students encounter these experiences, they learn about them in school and can apply contextual knowledge to answer the items. Footnotes, introductory information through context-setting statements, and embedded glossaries provide additional support in the assessment on as-needed basis. In response to a clarification question, Kachchaf clarified that Smarter Balanced assessment items do not relate to the context-setting introductory statements.

John Sabatini, University of Memphis, presented on behalf of the Global, Integrated, Scenario-based Assessments (GISA). Unlike other assessments, GISA was formed using a federal grant and was part of an initiative by the Institute for Education Sciences (IES), which used teams to develop instruction and assessment. Sabatini led the K–12 assessment project. Based on a literature review, it was determined that some reading comprehension assessment constructs need to evolve to meet the demands of the 21st century. Sabatini reported that GISA work is compatible with NAEP and PISA and incorporates similar scenario-based items. He shared examples of how scenario-based tasks address background knowledge and are important for higher order comprehension. GISA uses natural language processing (NLP) techniques to identify topical knowledge and vocabulary related to the passage and task that students experience in the assessment. Sabatini described two of these techniques. In one, students decide if a word belongs to a topic or not and receives immediate feedback in the assessment. Although this tests current knowledge, it also activates prior knowledge that students bring to the assessment. Another technique embeds items from released NAEP assessment items in science and history in the beginning of the GISA reading test. These NAEP items are related to the passage students would experience in the GISA assessment. Sabatini summarized that both of

these techniques allowed the GISA reading comprehension assessment to measure topical knowledge before students engage with passages and test questions.

Referring to the presentations from the international and U.S. assessment leaders, Kelly asked Cervetti to summarize potential implications for the future of the NAEP Reading Assessment. Cervetti highlighted the abundant research and understanding about the importance of knowledge in reading comprehension, including knowledge of text and text genres, knowledge about how syntax works to create meaning, knowledge of the world, and knowledge about specific topics. Researchers have carefully documented the consistent and robust impact of topic knowledge on a reader's ability to respond to questions that require bridging inferences, forming connections within texts, making global inferences like understanding concepts or themes, and even recalling information from the text. Topic knowledge most likely impacts all processes described in the comprehension targets for the NAEP Reading Assessment.

Given that topic knowledge varies widely from one reader to another, Cervetti asserted that this presents several challenges for all reading comprehension assessments. First, because the passages that students encounter will always intersect with specific topics, the assessment scores may reflect students' knowledge of the topic at hand more than they reflect their comprehension ability. Second, topic knowledge is sometimes systematically distributed by group characteristics, such as the state, region, community, or culture in which students reside. To illustrate examples of this systematic difference, Cervetti discussed two states with different sequencing and pacing in their science curriculum, where one state addresses the science of light and sound waves in grade 3 and another state addresses that topic in grade 5 or higher. For an informational passage about light and sound waves on a grade 4 reading assessment, this topical familiarity could then contribute to students scoring higher in the state where students already encountered that topic through their state's science curriculum. Cervetti summarized that in this and other cases, the assessment challenge is that differences in reading comprehension performance detected across groups may be more related to topic knowledge than reading comprehension ability. All of the assessment programs leaders who presented recognize these assessment challenges and therefore attempt to mitigate the impact of topic knowledge to produce better estimates of the types of reasoning that students can do with text, and these types of student reasoning are similar to the comprehension targets that are the focus of the NAEP Reading Assessment. Cervetti highlighted that situating students in authentic reading was also a principle applied in several of the assessment programs described.

Cervetti provided a summary of strategies noted by the different assessment leaders to mitigate the impact of topic knowledge, especially where this knowledge might provide advantages to certain groups. For passage selection, assessment strategies included: choosing familiar texts because of shared curriculum; choosing unfamiliar authentic texts and providing supporting information in the assessment; avoiding texts that rely on culturally specific information or technical knowledge; grouping and sequencing thematically related texts allowing students to build knowledge as they read in the assessment; and ensuring that texts are engaging. For item development, strategies included: developing items that are text-dependent rather than knowledge-dependent; reviewing items for bias; and avoiding constructed response items based

on background knowledge. For other assessment features, several assessment programs use text introductions that include both topic information and purposes for reading, as well as pop-up definitions or footnotes. For reporting features, one program, GISA, considered knowledge in the scoring and interpretation by measuring readers' knowledge and using this understanding to support more expansive interpretations of assessment results.

Cervetti indicated that only some of these assessment strategies could be applied to NAEP because NAEP is prohibited from making an explicit curriculum connection, such as is done in Louisiana, for instance. Cervetti added all features proposed in the NAEP Reading Framework Update are also part of the current NAEP assessment. The framework update, however, includes a more robust approach to addressing differences in students' topical knowledge, given that it influences student performance and is not addressed in the comprehension targets or achievement level descriptions in the current NAEP framework. The assessment strategies proposed in the updated NAEP Reading Framework address the issue of topic knowledge in different but complementary ways to the international and U.S. assessments presented here. For instance, the updated framework: (a) increases the likelihood that students will have encountered at least some texts included in the assessment, (b) supports readers' engagement by focusing readers' attention on the most important information, and (c) provides introductions that address pivotal gaps in topic knowledge on an as needed basis, e.g., many passages and assessment blocks will not require these introductions. Cervetti acknowledged that there is no Universal Design element that can eradicate the influence of topic knowledge on comprehension entirely, but mitigation attempts are important for more fair and valid assessments.

Julia Rafal-Baer thanked Cervetti for the clarity of her remarks and, recognizing Chiasson, praised Louisiana's leadership in assessment and instruction. Rafal-Baer stated that assessments often drive instruction. Because of this, she is concerned that the NAEP Reading Assessment might send a signal that it is not important for students to build their knowledge about different topics. Rafal-Baer then noted questions about the impacts of Universal Design elements on students' testing experiences and asserted that more information was needed before she could comfortably support recommendations being proposed for the updated NAEP Reading Assessment.

Eric Hanushek noted the fundamental disagreement between how Willingham and Cervetti characterize the background knowledge issue for assessment, asking what is the legitimate adjustment to make to the NAEP Reading Assessment given the disagreement. Hanushek wants NAEP to do a better job of predicting future performance in careers and college and worries that adjustments might be made that make NAEP less predictive of these important outcomes.

Schleicher agreed the issue is not to eliminate the influence of background knowledge, but rather to eliminate bias at the group level. He added that the exemplars presented by Sabatini for GISA demonstrate that it is possible for assessments to statistically model the effects of topical knowledge and topical familiarity. Sabatini agreed with Schleicher that the assessment goal is less about reducing or eliminating the differences, but more about making sure we understand the source of the test score differences. Kachchaf added that it is important to ensure that experts

who know students well, such as educators, are making the determinations about which topics students may not be familiar with. Smarter Balanced relies on educators to indicate when additional contextual information is needed.

Lane asked whether Universal Design features, such as pop-ups, videos, or introductions, are themselves sources of construct irrelevance that may impede performance or decrease students' motivation by mandating them to do more to navigate the assessment. Cervetti described a NAEP special study conducted with 3,000 students which examined some of the features being discussed. Results of the study indicated students earned higher scores on passages with the features. Lane then asked if the study kept track of students who used the UDEs, which ones were used, and the relationship between use and performance. Kelly commented that process data could be a useful resource for answering these sorts of questions.

Paul Gasparini wondered if the preferred item analysis done for PISA results could inform NAEP Reading Framework discussions. Schleicher responded that these item analyses showed it was possible to achieve an equally unfair set of tasks across different cultural, linguistic, and national contexts. The objective is not to eliminate those influences but rather to account for them and make them visible. Schleicher also agreed with using process data to provide information about how students respond to items.

Dana Boyd referred to Hughes' presentation in the prior session on equity. Hughes specifically cited the importance of empathy, engagement, equity, evaluation, and equality for educational assessments, and how these concepts connect with the types of inferences that are drawn from assessment results. She asked how the Governing Board might work to prevent the mischaracterizing sorts of inferences that often implicate students of color, while also increasing equity and equality for our nation's students. Cervetti remarked that equity was of central importance in the development of update recommendations for the NAEP Reading Framework and assessment. Chiasson noted that equity is one of the drivers for Louisiana's innovative assessment pilot, and she is encouraged by pilot data which show higher levels of engagement from the students who are not reading on grade level and are encountering unfamiliar topics at the same time. Being familiar with the topics of passages seems to make students feel more empowered and engaged to perform at their best in the assessment pilot program.

West asked why the routinely performed differential item functioning (DIF) analyses are not sufficient for addressing the concerns Cervetti raised about group differences. He also asked if topic knowledge is not explicitly mentioned in the NAEP framework, then should the framework definition of reading be revised to include mastery of a diverse array of content knowledge. Cervetti said that assessments can be designed for equity so that DIF analyses are less likely to identify problems. Willingham responded he believes the current NAEP framework is not realistic about what reading is. Given that the test seems to prioritize broad yet shallow knowledge, he asserted that the framework should acknowledge this priority.

Cizek thanked Sabatini for clarifying what seems to be the defining issue, which is: does the Governing Board want to control for background knowledge in reporting a reading score or do we want to help explain reading performance because of background knowledge?

Reginald McGregor, referring to his work in industry, has found that the author of a report determined whether the report could be understood, and sometimes the likelihood of a document being understood was based on how things were being translated across international teams. In all cases, however, McGregor noted that it is important for workers to be able to review various reading materials and make sense of them. Based on McGregor's comment, Sabatini noted that maybe communication should be a stronger emphasis for future NAEP Reading Assessments.

Kelly thanked everyone for their presentations, comments, and discussions and for taking a deep dive into the issues surrounding background information in large-scale assessments. He also thanked the wider audience that attended this public session.

The meeting adjourned at 5:34 p.m. for the day.

Nominations for Board Terms Beginning October 1, 2021 (CLOSED)

Under the provisions of exemptions 2 and 6 of § 552b (c) of Title 5 U.S.C., the National Assessment Governing Board convened in closed session on Friday, March 5, 2021 from 12:00 to 12:30 p.m. to receive a briefing from Jim Geringer, Chair of the Nominations Committee, for Board terms that begin October 1, 2021.

Geringer noted that for the 2021 cycle, there are six vacancies in the following categories:

- Elementary School Principal
- General Public Representative
- Governor (Democrat)
- Governor (Republican)
- Local School Board Member
- Testing and Measurement Expert

For terms beginning on October 1, 2021, there are incumbents in the following three categories: Elementary School Principal, General Public Representative, and Governor (Democrat). There are no incumbents for the other three categories. For the category of Local School Board Member, these candidates were approved by the Board in March 2020.

Geringer reviewed the 2021 nominations process and timeline, which began during summer 2020. The final slate of candidates will be submitted to the Secretary of Education in April/May 2021, once commitment letters are received from the finalists.

Geringer reviewed the slate of finalists for terms that will begin on October 1, 2021. He presented information about the nominations received by number of applicants, gender, race/ethnicity, and geographical representation. The final slate of candidates was described, along with a listing of proposed finalists, for the categories of Elementary School Principal,

General Public Representative, Local School Board Member, and Testing and Measurement Expert. Geringer noted that nominations for the two Governor positions are made by the National Governors Association.

Board members engaged in discussion on the recommendations for the final slates of candidates for submission to the Secretary of Education.

NAEP Budget and Assessment Schedule (CLOSED)

Under the provisions of exemption 9(B) of § 552b of Title 5 U.S.C., on March 5, 2021, the Governing Board met in closed session from 12:35 p.m. to 1:50 p.m. to receive a briefing on the NAEP budget and assessment schedule from Peggy Carr, Associate Commissioner, NCES.

Carr provided updated projections on the current budget through 2024, noting potential implications for the NAEP Assessment Schedule for Board consideration. In addition, with the transition to the next generation eNAEP delivery platform, Carr outlined the need for special studies to investigate how changes to the content and/or administration of NAEP may affect trend and validity.

In May, the Board will receive two updates, one on the eNAEP transition and the other on the budget. The Board will use this information to identify next steps in upholding its policy priorities of utility, frequency, and efficiency.

Following a break, the meeting resumed in open session at 1:55 p.m.

Action: Updated NAEP Assessment Schedule

Barbour reminded Board members that after the November 2020 Governing Board meeting, the Board submitted letters to Congress that supported postponing the 2021 administration to 2022.

Congress then passed the Consolidated Appropriations Act of 2021, which included a waiver for 2021 NAEP administration. Through this waiver, Congress acknowledged the operational limitations of conducting NAEP in 2021 during COVID-19, allowing the Governing Board and NCES to postpone the legislatively mandated 2021 administration of NAEP Mathematics and Reading in grades 4 and 8.

To reflect these waiver provisions, the Board took action to update the NAEP Assessment Schedule. Alberto Carvalho made a motion that the Board approve the updated NAEP Assessment Schedule. Matthews seconded the motion. There was no discussion and the motion was unanimously approved. The approved NAEP Assessment Schedule is appended to these minutes.

Action: Nominations for Board Terms Beginning on October 1, 2021

Geringer made a motion that the Board approve the slate of Governing Board nominees for terms beginning on October 1, 2021 as presented earlier in closed session. Mark White seconded the motion. A brief discussion included confirmation that the Department's Office of General

Counsel vets potential appointees for conflicts of interest. With two members abstaining, the motion was approved unanimously.

State and TUDA Task Force Updates

Barbour welcomed partners from the Council of Chief State School Officers (CCSSO) and the Council of the Great City Schools (CGCS) to provide State and Trial Urban District Assessment (TUDA) task force updates. Barbour described the task force collaborations as extremely valuable to improve current NAEP processes and shape the future of NAEP. He introduced Shelly Loving-Ryder, Chair of the State Policy Task Force.

Loving-Ryder updated the Board members on the accomplishments and plans of the State Policy Task Force, a collaboration between the Governing Board and the CCSSO. The task force comprises individuals serving in a variety of roles from a diverse group of states. Typically, the task force discusses myriad topics, however, discussions during the past year focused on COVID-19 impacts. The task force provided the Governing Board with insights from the field about instructional modalities in schools and the impact on NAEP administration. They discussed communication strategies related to rescheduling NAEP from 2021 to 2022, particularly to emphasize that the delay was for logistical and operational reasons and does not reflect the importance of NAEP. Loving-Ryder thanked the Board and NCES for their thoughtful deliberations on the impact of COVID-19 and the decision to delay NAEP.

Loving-Ryder indicated the task force was briefed on the NAEP Reading Framework. She commented on the influence NAEP has on state standards and assessments. The task force looks forward to hearing more about the framework, especially how background knowledge is addressed.

The task force appreciated the Board's original strategic vision for its simplicity of focus. They are pleased with the addition of a third pillar, *engage,* to supplement *innovate* and *inform* in the 2025 Strategic Vision, because it is difficult to inform or innovate if there is no engagement. In closing, Loving-Ryder noted the task force received briefings on the report card releases. She introduced Scott Norton, Deputy for Programs at CCSSO.

Norton presented an update on the landscape of state assessments. In March 2020, the U.S. Department of Education waived state assessment requirements. In a letter to state education agencies on 2021 testing, the U.S. Department of Education emphasized the importance of assessment for understanding the impact of COVID-19 on student learning and as such, will not issue blanket assessment waivers for 2021. However, the Department offered some flexibility in state assessment requirements, such as waiving the 95 percent participation requirement and allowing shortened tests, remote administration, and extended administration windows. The U.S. Department of Education suggested students should not enter schools only to take a state assessment. Based on a recent CCSSO survey, most states plan to administer state assessments to as many students as possible in 2021; however, challenges persist. Several states have applied, or are expected to apply, for a waiver.

CCSSO staff and members are discussing the future of state assessment. Changes may be on the horizon, in part because of inconsistency in quality and in different types of assessment (e.g., classroom, district, state). These assessments typically do not belong to a single integrated system and some components may not align well to standards.

Norton offered some ideas for how states might meet these challenges. First, CCSSO is advocating for balanced assessment systems. In *Knowing What Students Know*, published in 2001, the National Research Council defined a balanced assessment system as "when the various types of assessments in the system are coherently linked through a clear specification of learning targets, they comprehensively provide multiple sources of evidence to support educational decision making, and they continuously document student progress over time." Norton acknowledged state and district partnerships are needed to create balanced systems, because most assessment occurs at the district, school, and classroom levels rather than the state level. CCSSO suggests a balanced assessment system should include (a) classroom-level formative assessments, (b) district-level interim or benchmark assessments, and (c) state-level summative or end-of-year assessments. Norton suggested "right sizing" expectations about summative assessments which currently tend to overshadow other assessments.

Muldoon facilitated questions from Board members for the State Policy Task Force representatives. Gasparini asked Norton about using assessments such as the New York State Regents Exams to inform instruction. Norton acknowledged that summative assessments are not as helpful for informing instruction as they are for end-of-year evaluation purposes and accountability. Geringer asked Norton to clarify his comment about coordinating state and district assessments. Norton meant that states and districts need to share information about assessments in a coordinated way.

Cizek asked about possible inequities when in-person administration is required for some assessments (e.g., WIDA Consortium). Norton reported that approximately 30 states planned to have remote students come to school to take assessments in-person but will likely reconsider their options given the recent guidance from the U.S. Department of Education. Lane asked about the potential uses and misuses of assessment data collected from remote unproctored testing conditions. According to a survey conducted about two months ago, five states were considering remote assessment administration. The number of states considering remote administration has dropped since then. For one, the District of Columbia has since applied for a waiver. Norton reported there may be one state considering remote proctoring as a way of monitoring testing. However, most states are not ready to offer remote testing. Lane added that testing under remote conditions, with or without a proctor, could lead to equity issues.

Michael Casserly began the TUDA Task Force update by noting he will step down as the Executive Director of CGCS at the end of June 2021, after serving for more than 44 years. He stated that it has been his honor to work alongside the Governing Board and averred that he is most proud of initiating the TUDA program. Casserly thanked everyone and turned the

presentation to Ray Hart who updated the Board members on the accomplishments and plans of the TUDA Policy Task Force.

The task force of 10 district leaders provides district perspectives and feedback to the Governing Board. Hart stated the TUDA is invaluable for CGCS members as it allows them to gauge their performance against their peers. Policy-focused discussions during the past year considered (a) the value of participating in NAEP, (b) the NAEP assessment schedule and participation during COVID-19 school disruptions, (c) the Reading Framework update, (d) adding contextual questions, and (e) communicating NAEP to the public.

In conversations about NAEP 2021, district leaders shared concerns about not having a representative sample and inherent bias in assessing only students attending school in person. CGCS provided feedback to the Governing Board on the NAEP administration schedule.

The CGCS supported proposed revisions to the NAEP Reading Framework, particularly inclusion of socio-cultural understanding of learning and development, incorporating science and social studies texts, and increasing the use of digital modalities. These changes will create a more fair and relevant assessment. Hart described the framework revisions as a sea change and useful for others to follow. Also, CGCS members appreciated new naming conventions such as comprehension targets replacing cognitive targets. Members support many of the proposed ideas for scaffolding for accessibility.

The Council discussed adding questions to student and teacher surveys to collect information about learning experiences and level of parental support to students during the pandemic. Hart suggested a partnership with the Council, Governing Board, and NCES communication teams to develop joint communication campaigns before and after the NAEP 2022 releases. The task force recommends extended communications to provide context for understanding NAEP results, especially the influence of school disruptions on student achievement.

On behalf of Barbour, Peisch acknowledged the tremendous contributions of Mike Casserly to NAEP and the Governing Board. Members echoed her thoughts and extended their thanks to Casserly and for the work of the Council. Muldoon applauded the work of Casserly and the Council in the progress that TUDA districts have made since the program started. Carvalho added his thanks for Casserly's advocacy and remarked on Casserly's friendship to public education which has "elevated the national landscape of opportunity for kids and educators." Carr and Woodworth thanked Casserly on behalf of NCES for being a partner of NAEP since its beginning. Carr described NAEP as a three-legged stool, with one of the legs being the CGCS. Woodworth added thanks for the tremendous support from Casserly and Hart in collecting and disseminating district data.

Muldoon facilitated questions from Governing Board members. Cramer asked if districts find contextual questions useful when comparing results with other districts. Hart responded they have not discussed specific questions, but they would like to add pandemic-related questions for

NAEP 2022. Cramer followed up by asking whether questions about student continuity would be useful for districts. Casserly commented they think this is an important topic. In addition, the Council is working on a study using NAEP contextual data from urban school districts. They want to understand whether results simply reflect demographics.

Hanushek asked if there are plans to expand the number of participating districts. Casserly noted that additional cities are interested in joining the TUDA, if and when there are funds available to support expanding the program.

Muldoon asked the panelists what they anticipate schools will look like for the 2021–2022 school year and how NAEP results might be used. Norton responded he thinks more students will return to in-person learning. The goal is for learning to be back on track, regular state assessment administration, and "normal" NAEP administration. Casserly reported that 51 of 77 districts are at least partly open. School leaders are eager to expand opening their buildings to more students this spring with only a handful remaining closed for in-person instruction for the remainder of this school year. He expects most districts to be open at the beginning of the next school year; however, not everyone will be back in person. Districts are likely to offer virtual and hybrid learning, and some parents and students will choose these modes. It is important to recognize these choices to prevent introducing bias into the NAEP sample.

Gasparini directed his earlier question about tying NAEP results to instructional practice to the TUDA Policy Task Force. There is only an indirect link because NAEP does not have school or classroom results. Casserly suggested NAEP results are useful for informing curriculum standards and instructional strategies and identifying subgroups needing additional support. Carvalho added the TUDA reports are very helpful by providing comparisons of participating districts. Leaders share information with each other to learn what others are doing when they show improvement and good performance. Hart suggested using the NAEP questions tool to understand differences between strong responses and how students in a specific school are likely to respond. This information can be used to inform curriculum and professional development.

The March 5, 2021, Governing Board meeting recessed at 3:05 p.m. and reconvened at 3:30 p.m.

NAEP Reading Framework Policy Discussion

The Assessment Development Committee (ADC) invited three scholars and leaders from the NAEP Reading Framework Visioning and Development Panels to lead a discussion about the latest draft of the NAEP Reading Framework. Dana Boyd, ADC Chair, thanked everyone for the work done thus far and provided the Governing Board with a progress report and summary of recent events in updating the framework. In July 2020, the draft framework was posted for public comment, and the Governing Board received feedback from a wide range of stakeholders. The framework contractor and Development Panel reviewed the public comments and shared a revision plan for incorporating the feedback. At the November 2020 Governing Board meeting, the panel received additional feedback from Board members regarding proposed revisions. Based

on this feedback from public comment and the Governing Board, the Panel submitted an updated draft of the framework to the Governing Board. The purpose of the current presentation is to give Board members an opportunity to provide final guidance before the May 2021 Governing Board meeting. Boyd reminded members the current framework was developed for paper-based assessments in 2004; as such, the Panel undertook an ambitious task of determining and assembling recommendations for updating the framework. The ADC's oversight of NAEP framework processes ensures a comprehensive, inclusive, and deliberate process that reflects research and standards in the field.

Boyd thanked the members of the Visioning and Development Panels, who are tasked with upholding the highest standards when developing the framework and incorporating the revisions. Boyd informed the Governing Board they would hear a high-level summary of the latest NAEP Reading Framework update, followed by a policy discussion on areas requiring additional deliberation and debate. Addressing concerns raised at the last meeting, Boyd stated that NCES has indicated that maintaining trend should be possible with careful planning and a gradual item development strategy. She introduced ADC and Panel members to present and address questions and comments.

P. David Pearson, Chair of the Reading Framework Visioning and Developing Panels, presented on behalf of the panels and WestEd. Pearson noted how advances in reading research, changes to state standards, and an increasingly digital world necessitated updates to the NAEP Reading Framework. Advances in research include differences in the knowledge and abilities needed to read and comprehend different types of text in various disciplines. Pearson highlighted the similarities and differences between the current framework and the proposed framework update. The updates reflect feedback obtained from the public and Governing Board members. Reading is defined as a complex process shaped by student, social, and cultural influences. Pearson reported that minor revisions to the current framework's definition add more context to the process of comprehension. He noted that the updated definition is more specific about the knowledge and tools the reader brings to the table and also about the sub-processes that constitute comprehension. Pearson said that a new comprehension target, Use and Apply, was introduced to the NAEP Reading Framework to reflect what assessments require students to do (i.e., comprehend the material and apply it to the task). This additional target is warranted by new research and state standards as well as prevalent practices in state and international assessments.

Visioning Panel member (and former Governing Board member) Susan Pimentel provided an update on the importance of using disciplinary context as a reporting feature. For the updated framework, new disaggregated scales for reading in science and social studies were added. These additions reflect the shift to disciplinary context. The latest research shows differences in the knowledge and skills required to read text in different disciplines; state standards also reflect this research. The added discipline areas supported by public comments were already part of the broad definitions for reading in literature, science, and social studies. These broad definitions also overlap with the current NAEP Reading Assessment item pool. The Panel believes the new

subscales will deepen insights for NAEP reporting, moving beyond the generic reporting on informational text and will enable educators to draw more precise inferences about student achievement. Pimentel provided the new disciplinary text definitions and shared related examples.

Pimentel next spoke on updates to purpose-driven assessments. In the revised framework, before starting to read a passage, students will see a purpose for reading the passage and what they will be asked to do. The panel proposes three levels of purpose: (a) broad purpose, either to read and develop understanding, or to read to solve a problem, (b) block specific purpose, to guide reading in the entire 30-minute block, and (c) task specific purpose, offered for each text students encounter. The purposes provide context, increase student engagement, and allow students to demonstrate comprehension. Pimentel noted that purposes also add ecological validity by more closely mirroring the type of reading that students do outside the assessment context.

Visioning and Development Panel member Gina Cervetti stated that the purpose of the NAEP Reading Assessment is to provide a valid measure of reading comprehension across a diverse range of test takers. To help accomplish this purpose, the NAEP Reading Framework update employs the principles of Universal Design for Assessment. In response to public comment and Board feedback, Cervetti reported that the Panel re-conceptualized assessment scaffolds to align with Universal Design for Assessment. Accordingly, in the framework update, a Universal Design Element (UDE) is defined as a design element that helps students access, organize, and express ideas in order to accomplish complex tasks. Similar to how these features appear in the current NAEP Reading Assessment, all students will have access to all UDEs. Cervetti summarized that UDEs allow NAEP to administer more rigorous and more complex comprehension tasks in short blocks, and UDEs do not provide answers to comprehension questions. Cervetti listed the three types of UDEs in the NAEP Reading Framework update: (a) motivational, (b) task-based, and (c) knowledge-based. Motivational UDEs are embedded into reading activities to encourage and support reader interest and engagement, especially when the reader encounters more complex or challenging reading passages. Task-based UDEs include directions for progressing through the assessment or a graphic organizer to record information. Knowledge-based UDEs supply a minimal amount of information about specific non-assessed concepts, topics, or vocabulary. Cervetti stated that these UDEs ensure NAEP Reading Assessment scores reflect differences in comprehension ability rather than differences in topic knowledge – topic knowledge is directly addressed in other NAEP assessments, such as in science, civics, and U.S. history.

Cervetti focused the next part of her presentation on providing the Governing Board more information regarding knowledge-based UDEs, and how they would appear in a NAEP Reading Assessment. Two kinds of knowledge-based UDEs are part of the current assessment and the Framework Development Panel is recommending that these features continue: (a) pop-up notes that provide brief explanations of words and phrases; and (b) passage introductions that provide information about the topic of the text, where that information is critical for comprehension

and may not be known by all readers. In the framework update, some of these features would be more elaborated, such as using video, audio, or photos, in passage introductions. To provide more context, Cervetti used the example of a passage involving the mention of a talent show. For the assessment, UDEs would enable students to listen to violin music before answering questions on the topic or see a pop-up definition of a potentially unfamiliar term such as "talent show." Cervetti also noted a released NAEP Reading Assessment block that used a pop-up definition for a term in a literary passage from a Turkish folktale.

Cervetti described key parameters for the development of the knowledge-based UDEs. She noted that they are not designed to equate students' topic knowledge, which is impossible. Most importantly, these UDEs may provide information that enables readers to reason with the text as intended, but they are not designed to address everyday information. There are differences between knowledge inherent in reading comprehension and that which is not. Knowledge inherent in reading includes text structures like cause-and-effect, story structure, or language structure. NAEP measures the ability to use these types of knowledge, which ultimately leads to comprehension that can be measured by the targets.

Pearson reflected on the March 4 symposium and its focus on the topic knowledge issue. He noted that the public comment draft of the NAEP Reading Framework included potential ways to measure students' topic knowledge, as done in the GISA assessment. In response to public comment and prior to Board feedback, those potential measures were removed and hence do not appear in the latest draft of the framework update. However, given the Board's comments in the symposium, Pearson suggested that NAEP should pursue a special study to carefully examine the influence of background knowledge on NAEP Reading Assessment performance. Finally, he reminded the Governing Board that many of the features being discussed for the framework update are part of the operational NAEP Reading Assessment and have been for many years. This holds true to the Board's charge to the Visioning and Development Panels to provide update recommendations that are evolutionary in nature.

Miller opened the floor for discussion related to the updates. After the Board expressed support for the disciplinary contexts recommended in the framework update, Suzanne Lane asked for evidence UDEs work for those who need it and do not increase performance for those who do not. Pearson noted that the purpose of UDEs is to provide support for all students, and Eunice Greer, NCES, referred to a study Cervetti spoke about during the March 4 symposium. Using scenario-based tasks with UDEs, the study created discrete versions of the tasks without design elements. Results showed the effect of scenario-based task format positively affected student performance across all achievement levels. Lane followed up to ask if some students were helped more than others, but Greer replied the study was too small to provide that level of information. However, she agreed it is important to know and perhaps another study is needed to examine impact of individual Universal Design features.

Hanushek asked about trend, and how the Board can be sure that trend will be maintained. After listening to the symposium presentations, he concluded if background knowledge is important

then there are only two avenues to be pursued—reduce it as much as possible or revisit it after the assessment to explain differences. He asked: will the Governing Board be able to maintain and assess trend if the framework changes the measurement as opposed to trying to use background information to explain differences in trends and levels? Greer referenced the document NCES provided to the Governing Board that describes the process of rolling out new content and evaluating it in comparison with past content and items. Data suggest if the assessment is moving forward with enough of the same passage blocks, this increases the likelihood that trend will be maintained. She added it is an empirical question of whether trend can be maintained, and this will need to be carefully evaluated based on the incremental approach described in the NCES document. Greer reminded Hanushek the assessment is not adding new content areas because there are already passages in the current NAEP Reading Assessment that relate to science and social studies; the framework update specifies that, instead of aggregating these assessment blocks under one informational subscale, these assessment blocks will now be disaggregated with one subscore to address reading in science and another subscore to address reading in social studies. In terms of the UDEs, 13 of the 15 are not new and will not affect trend; their influence is already well documented. Additionally, the assessment already has UDEs that address background knowledge. Greer stated that NCES will carefully evaluate the new UDE features.

Cervetti clarified that adding UDEs motivate and engage students, and without these features there is a risk that lower performing students or those without relevant background knowledge will be less engaged and not able to fully participate in the assessment. She said it is important to address these issues in the design phase because they cannot necessarily be identified in assessment results. Pimentel added that the issue being discussed is not about helping lower performing students do better. She said the point is to make the test fair. She gave an example of a student who does not know about a topic but is a good reader. UDEs put students on a fair footing, while not providing the answer.

Nardi Routten asserted that knowledge-based UDEs are not about "spoon-feeding students." She gave an example of the term cricket, which has at least three different meanings: an insect, a phone, and a sport. If the reading passage is about the sport, a short video would be important to ensuring that the inaccurate conception of the term does not contaminate measurement from the reading test items that are addressing the comprehension targets of the NAEP Reading Assessment.

Cizek referred to Lane's previous inquiry. He emphasized that it is not encouraging if a design element is helpful across all achievement levels. He used Braille as an example. He argued that if Braille were added to all assessments and all students showed improvement, that would show a testing problem, since it should only help the visually impaired. He thinks this is a serious issue that should not be relegated to a special study.

Cizek asked two questions: (a) did the Framework Development Panel rely on Universal Design for instruction or Universal Design for assessment? (b) did the Panel provide any guidance on

how to determine when to gloss terms in a passage and the extent to which this should be done? Pearson responded that the perfect amount of information to add is unknown, but the amount recommended is cursory and provides only fundamental awareness of a topic for a student without prior knowledge. NCES uses various panels of experts, including educators, to oversee text selection, item design, and block design and make professional judgments regarding which blocks deserve and need UDEs and at what level of detail or specificity. He added that some passages currently have a short introductory text, but given the availability of digital media the framework update proposes that NCES evaluate if different types of multimedia would be useful.

Cizek clarified that he sought to understand if students who lack prior content knowledge are helped by the supports and students who had prior knowledge were not. He added that he would be in favor of a study on this topic. Peggy Carr reported that NCES has conducted many studies on accommodations and universal design features. She noted that these studies are very expensive and need to be developed in a very scientific way to discern true evidence. Further, Carr stated that very few of these accommodations and features have had the level of study now being requested by the Board. However, Carr added that after routine analyses prompt removal of problematic items from the operational NAEP assessment, the matrix design of NAEP means that any noise in the measurement of student performance will be randomly distributed across all student groups.

Although not part of the framework, Kelly suggested a need to measure the background knowledge students bring to the assessment to determine the impact on reading comprehension. He noted that the limited use of UDEs in an assessment adds context and supports authenticity of the assessment. He added that process data could also be a resource for looking at the impact of background knowledge.

Peisch initially thought that UDEs would help students who otherwise would not do well on the assessments and this might mask their true ability. However, it seems that in the discussion today, the framework update recommendation is for UDEs to support all students because whether a student is high achieving or low achieving, there are topics that not everyone knows or is familiar with. Pimentel confirmed that Peisch was correct in her interpretation of the update recommendation.

Edelblut stated that it was important to ensure that the assessment itself is not contributing to the different disparities illuminated by NAEP results. He asked Cizek to comment on this interpretation of the recommendations for UDEs. Cizek responded that there seems to be no conclusive evidence to indicate that knowledge-based UDEs are, in fact, assisting the students who need them.

Ron Reynolds said the framework update recommendations appear to be based upon a hypothesis that some portion of the variability and ability in reading can be attributed to a feature missing from the assessment instrument. He asked whether anyone could provide an estimated magnitude of this problem, i.e., if the proposed UDEs are in place, how much of a gain does

NAEP expect to observe in those unfairly measured by the current assessment? Carr answered it was difficult to determine; a study would need to be run to randomly assign students and to identify statistically significant differences.

Matthews reminded Governing Board members to keep in mind that equity is not an accommodation. She posited that these framework recommendations challenge NAEP and the Governing Board to be less complicit in the institutionalized disadvantages embedded into educational systems and assessments. She expressed that bravery is required to deinstitutionalize these disadvantages, and that she was proud of the Board for engaging in this thoughtful work.

Russ Whitehurst posited that equity is largely a characteristic of environments and opportunities, not a characteristic of an assessment. He commented on the length and accessibility of the framework draft and noted that it was written in academic language. He asserted that equity was never explicitly defined in the draft, though it is mentioned. He reasoned that equity in the framework draft was being conceptualized as an effort to support fairness by giving every student who takes NAEP the opportunity to generate a score that is not affected by their differential access to a national culture that young people should be socialized into in order to qualify for various jobs and to be successful in other settings. Whitehurst expressed concern that providing supports in an assessment context will be detrimental in the long term, because in the business world, standards must be met without support. He argued that all students should have equal opportunities in educational experiences, to the extent possible. He worried that some of the framework recommendations will make the assessment less rigorous, which would ultimately hurt those it was designed to help.

McGregor commented that, as a Board member, he receives various NAEP reports and materials referencing psychometric concepts. He added that, although he might not know much about psychometrics, he is an engineer, and if there was a technical report on jet propulsion, he could comprehend that report while a psychometrician could not. He said that while both the psychometrician and the engineer are smart, they are smart on different subjects. He reasoned that these are the types of topic knowledge disparities that are natural occurrences across students as well. McGregor asserted that UDEs do not provide an advantage; instead, they allow for stronger universal access to the assessment.

Hanushek asked for information about the range of studies that are anticipated to deal with the issues from this framework update. Pearson summarized the research and development conducted by NCES as part of typical item development procedures and noted that special studies will be listed in the Assessment and Item Specifications document, which is a companion to the framework update that has not yet been drafted.

In closing, Boyd outlined next steps for the reading framework. The panel will use feedback gathered from today's meeting to make final revisions in April 2021.
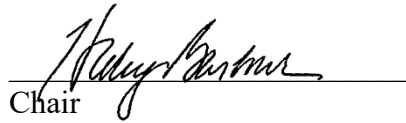
Concluding Remarks and Next Steps

Peisch thanked the panelists and Board members for a productive meeting with challenging topics requiring decisions. The May quarterly Board meeting is expected to be conducted virtually; times for that meeting will be provided soon. During the next meeting, the Board plans to (a) take action on the Reading Framework, (b) receive a full briefing on the 2019 NAEP Science results, and (c) continue work related to the Strategic Vision.

Meeting Adjourned

Board Vice Chair Peisch requested a motion to adjourn. Gasparini made a motion to adjourn; West seconded the motion. The motion was approved unanimously, and the meeting adjourned at 5:28 p.m.

I certify to the accuracy of the minutes.

_____   April 27, 2021
Chair   Date

National Assessment Governing Board

Executive Committee Meeting

Report of March 3, 2021

CLOSED SESSION

Executive Committee Members: Haley Barbour (Chair), Alice Peisch (Vice Chair), Dana Boyd, Gregory Cizek, Jim Geringer, Mark Miller, Martin West, Carey Wright.

Executive Committee Members Absent: Bev Perdue, Tonya Matthews

National Assessment Governing Board Members: Tyler Cramer, Christine Cunningham, Paul Gasparini, Reginald McGregor, Ron Reynolds, Mark Schneider (ex-officio).

National Assessment Governing Board Staff: Michelle Blair, Stephaan Harris, Donnetta Kennedy, Laura LoGerfo, Lesley Muldoon, Munira Mwalimu, Tessa Regis, Sharyn Rosenberg, Angela Scott, Matthew Stern, Lisa Stooksberry, Anthony White.

National Center for Education Statistics Staff: Peggy Carr, Pat Etienne, Dan McGrath, Holly Spurlock, James Lynn Woodworth, Alison Deigan, Bill Ward, Brian Cramer, Ebony Walton, Enis Dogan, Eunice Greer, Gina Broxterman, Grady Wilburn, Jing Chen, Nadia McLaughlin, Samantha Burg, Shawn Kline, Taslima Rahman, William Tirre, James Deaton,.

U.S. Department of Education Staff: Judith Anderson, James Forester.

The Executive Committee met in closed session from 11:00 a.m. to 12:30 p.m. to discuss the NAEP technology platform, the budget, and assessment schedule.

The closed session was called to order by Chair Haley Barbour at 11:00 a.m.

These discussions were conducted in closed session because the disclosure of cost data would significantly impede implementation of contract awards. Therefore, this discussion is protected by exemption 9(B) of section 552b(C) of Title 5 U.S.C.

Barbour reminded members of the confidential nature of the discussions before turning to Peggy Carr, Associate Commissioner, National Center for Education Statistics (NCES). Carr led a presentation on the Next Generation of NAEP: Planning for the Future. Carr communicated that NCES is thinking about the future of NAEP, a transition to an upgraded Next Generation NAEP administration platform, potential implications for the assessment schedule, and how to reduce costs for the program. NCES has three priorities in mind for updates to NAEP administration:

(1) online; (2) device-agnostic; and (3) contactless administration. Carr also provided a briefing on the budget.

Lesley Muldoon, Executive Director, then facilitated a discussion on potential implications for the NAEP Assessment Schedule in the short- and long-term. Muldoon also reminded the committee of the need to update the NAEP Assessment Schedule to comply with congressional action taken in December 2020.
The session concluded at 12:25 p.m.

OPEN SESSION

Executive Committee Members: Haley Barbour (Chair), Alice Peisch (Vice Chair), Dana Boyd, Gregory Cizek, Jim Geringer, Mark Miller, Martin West.

Executive Committee Members Absent: Tonya Matthews, Bev Perdue, Carey Wright.

National Assessment Governing Board Members: Tyler Cramer, Paul Gasparini, Reginald McGregor, Ron Reynolds.

National Assessment Governing Board Staff: Michelle Blair, Stephaan Harris, Laura LoGerfo, Lesley Muldoon, Munira Mwalimu, Sharyn Rosenberg, Angela Scott, Matthew Stern, Lisa Stooksberry, Anthony White.

National Center for Education Statistics Staff: James Lynn Woodworth, Peggy, Carr, Gina Broxterman, Jing Chen, Brian Cramer, Enis Dogan, James Deaton, Pat Etienne, Eunice Greer, Dan McGrath, Nadia McLaughlin, Holly Spurlock.

Contractors: American Institutes for Research (AIR): Jack Buckley, Kim Gattis, Young Yee Kim, Sami Kitmitto. Education First: Anand Vaishnav. Educational Testing Service (ETS): Jay Campbell, Gloria Dion, Amy Dresher, Emilie Pooler, Nancy Waters, Karen Wixson. CRP: Arnold Goldstein, Subin Hona, Edward Wofford, Anthony Velez. Hager Sharp: David Hoff, Joanne Lim, Debra Silimeo. The Hatcher Group: Devin Simpson, Jenny Beard, Alexandra Sanfuentes. Westat: Chris Averett, Lauren Byrne, Lisa Rodriguez.

U.S. Department of Education Staff: Judith Anderson, James Forester.

Others: Chester Finn.

Following the closed session, the Executive Committee reconvened in open session from 12:30 to 1:00 p.m. to discuss the status and next steps for (1) updating the Assessment Schedule and (2) carrying out Strategic Vision 2025.

Chair Haley Barbour opened with remarks about the actions taken at the November 2020 Board meeting to recommend to Congress that administration of NAEP be postponed from 2021 to 2022. Barbour called for a motion to recommend action by the full Board to update the NAEP Assessment Schedule. The motion was made by Jim Geringer and seconded by Vice Chair Alice

Peisch. There was no further discussion. The Executive Committee made a unanimous recommendation that was adopted for consideration by the full Board.

Barbour introduced Lesley Muldoon, Executive Director, and Lisa Stooksberry, Deputy Executive Director, to provide an update on Strategic Vision 2025. Muldoon shared the progress of draft work plans and proposed accomplishments for the next year. Muldoon presented a timeline of staff efforts to date, indicating that staff will provide committee-level progress reports at each quarterly meeting and an annual report every November. Muldoon then turned to Stooksberry to lead the presentation on accomplishments and priorities led by standing committees.

Stooksberry stated that COSDAM is responsible for two priorities: linking studies and achievement levels. Stooksberry signaled there are three accomplishments that COSDAM is working towards for linking studies and a working group for achievement levels. Stooksberry asked Greg Cizek to talk more about COSDAM priorities for the year.

Cizek reported that during the COSDAM meeting there was discussion about linking studies and the need to focus on linking studies that have policy-relevant goals. Cizek stated that a main outcome for this effort is to establish a formal mechanism for collaborating with the Reporting and Dissemination Committee (R&D). Cizek noted that COSDAM members Eric Hanushek and Julia Rafael-Baer stepped forward to lead this effort. Cizek said the goal is to better understand external data sets that can be mined to identify policy relevant sources to use to make recommendations. Cizek also provided an overview of the achievement levels work plan and advocated for collaboration with R&D to create an "interpretive guide" to communicate with influential people and the public about best practices, as well as appropriate and inappropriate uses of NAEP data and achievement level reporting. Cizek asked that the Governing Board staff think about how the Board can push that forward as a cross-committee effort.

Next, Stooksberry stated that the Assessment Development Committee (ADC) has been working on three proposed accomplishments and turned to Dana Boyd, Chair, and Mark Miller, Vice Chair, for an update. Boyd noted that the proposed accomplishments include initiating adjustments to the framework development process, creating a plan for updating remaining frameworks, and completing the science framework review. Boyd indicated that Greg Cizek and Cary Wright will be joining the ADC meeting to further cross-committee collaboration.

Stooksberry then asked Marty West, Vice Chair of R&D, to speak about the Committee's priorities. West mentioned that in its recent meeting R&D discussed the release plan for science assessment results later this year. Following on the heels of a Committee session focused on how socioeconomic status and income could be linked to NAEP data, the Committee noted the importance of having a similar panel discussion at a future Board meeting. West noted the importance of providing such data to researchers, and the Committee is thinking about not only how to make existing data useful but how to expand it moving forward to put increase NAEP's relevance and utility among stakeholders.

Stooksberry then described the Executive Committee-led priority related to the Assessment Schedule. The first accomplishment for 2021 is alignment of the assessment schedule to the

congressional waiver, the second is creating plan for additional state and TUDA assessments in the future (as currently reflected in the Board's approved assessment schedule), and the third is establishing policy priorities to inform next generation eNAEP transition. Stooksberry mentioned there will be a plenary session in May on the next generation eNAEP transition.

Stooksberry concluded by thanking Angela Scott for leading the Strategic Vision among Board staff and recognizing all staff for their contributions to this effort. Stooksberry asked if Board members had any comments, questions, or responses.
Marty West asked about the possibility of expanding state level reporting to other subject areas (other than the required reading and mathematics assessments) and mentioned it would be useful to know how much of an appetite there is for participation from the states.

Mark Miller thanked staff – especially Angela Scott – for moving forward with Strategic Vision. Chair Barbour also thanked Lesley Muldoon, Lisa Stooksberry, and staff.

Chair Barbour called for a motion to adjourn. Gregory Cizek made the motion. Vice Chair Peisch seconded the motion.

The meeting adjourned at 12:58 p.m.

I certify the accuracy of these minutes.

_____   April 15, 2021
Haley Barbour, Chair   Date

National Assessment of Educational Progress
Schedule of Assessments
Approved March 5, 2021

The *National Assessment of Educational Progress (NAEP) Authorization Act* established the National Assessment Governing Board to set policy for NAEP, including determining the schedule of assessments. (P.L. 107-279)

Year	Subject	National Levels Assessed	State Grades Assessed	TUDA Grades Assessed
2020	Long-term Trend*	9-year-olds 13-year-olds		
2021				
2022	Reading Mathematics Civics U.S. History Long-term Trend*	4, 8 4, 8 8 8 17-year-olds	4, 8 4, 8	4, 8 4, 8
2023				
2024	Reading Mathematics Science Technology and Engineering Literacy Transcript Studies	4, 8, 12 4, 8, 12 8 8	4, 8 4, 8	4, 8 4, 8
2025	Long-term Trend	~		
2026	**READING** **MATHEMATICS** Civics U.S. History	4, 8 4, 8 8 8	4, 8 4, 8	4, 8 4, 8
2027				
2028	Reading Mathematics **SCIENCE** Technology and Engineering Literacy Transcript Studies	4, 8, 12 4, 8, 12 4, 8 8	4, 8, 12 4, 8, 12 4, 8 8	4, 8 4, 8 4, 8
2029	Long-term Trend	~		
2030	Reading Mathematics **CIVICS** **U.S. HISTORY** **WRITING**	4, 8 4, 8 4, 8, 12 4, 8, 12 4, 8, 12	4, 8 4, 8 8 4, 8, 12	4, 8 4, 8 4, 8

NOTES:
* Long-term Trend (LTT) assessment not administered by computer until 2024. All other assessments will be digitally based.
~ LTT assessments sample students at ages 9, 13, and 17 and are conducted in reading and mathematics.
BOLD ALL CAPS subjects indicate the assessment year in which a new or updated framework is implemented, if needed.

National Assessment Governing Board

Committee on Standards, Design and Methodology

Report of March 2, 2021

COSDAM Members: Gregory Cizek (Chair), Carey Wright (Vice Chair), Jim Geringer, Eric Hanushek, Suzanne Lane, Alice Peisch, Julia Rafal-Baer, and Russ Whitehurst.

Other Governing Board Members: Dana Boyd and Mark Miller.

Governing Board Staff: Executive Director Lesley Muldoon, Deputy Executive Director Lisa Stooksberry, Stephaan Harris, Laura LoGerfo, Munira Mwalimu, Sharyn Rosenberg, Angela Scott, and Matt Stern.

NCES Staff: Commissioner James (Lynn) Woodworth, Associate Commissioner Peggy Carr, Gina Broxterman, Jing Chen, Brian Cramer, Enis Dogan, Pat Etienne, Eunice Greer, Daniel McGrath, Nadia McLaughlin, Holly Spurlock, Bill Tirre, and Grady Wilburn.

Other Attendees: American Institutes for Research: George Bohrnstedt, Markus Broer, Kim Gattis, Cadelle Hemphill, Sakiko Ikima, Young Yee Kim, and Sami Kitmitto. CRP: Shamai Carter, Arnold Goldstein, and Anthony Velez. Education First: Anand Vaishnav. Educational Testing Service: Jay Campbell, Gloria Dion, Amy Dresher, Gary Feng, Helena Jia, Hilary Persky, and Karen Wixson. Hager Sharp: David Hoff and Joanne Lim. The Hatcher Group: Devin Simpson. Pearson: Jennifer Galindo, Eric Moyer, and Cathy White. Westat: Lauren Bryne and Keith Rust. WestEd: Sonya Powers. Other: Chester Finn.

Welcome and Overview of Agenda

Chair Gregory Cizek called the meeting to order at 2:02 p.m. ET and asked all COSDAM members to briefly introduce themselves. He noted that ADC Chair Dana Boyd and Vice Chair Mark Miller would be joining the meeting for the brief discussion on reviewing framework processes.

Review and Revision of Mathematics and Reading Achievement Level Descriptions

Cizek began with a brief explanation of achievement level descriptions (ALDs). At the most general level, NAEP has three achievement level policy definitions: *NAEP Basic*, *NAEP Proficient,* and *NAEP Advanced*. The achievement level descriptions translate these general policy definitions into specific expectations for a given subject and grade assessed by NAEP

that are more informative about what students at each achievement level should know and be able to do. He explained that the ALDs provide important validity evidence for the NAEP achievement levels, and that the upcoming study to be conducted by Pearson will show us whether students within a given achievement level can actually do the things that the ALDs claim they should be able to do.

Sharyn Rosenberg provided a brief update on the current status of the study, which was also summarized in the advance materials. When this work was conceptualized, the panel meetings were intended to be conducted in person; however, the COVID-19 pandemic made it impossible to do so during the first half of 2021. A proposal to conduct the meetings remotely was discussed by COSDAM during the December 2020 meeting; concerns were expressed related to data security and panelist engagement. Shortly after this discussion, the NAEP program received a Congressional waiver to move the next administration of the mathematics and reading assessments from 2021 to 2022, which meant that there was an additional year before results from this study would be needed for NAEP reporting. This development, along with the approval and plans for distributing the first vaccines for COVID-19, led to a decision by staff and COSDAM leadership to extend the project schedule for this work to allow for the possibility of in-person panel meetings in late 2021 and early 2022. A contract modification is in progress; the status of this work, including an updated Design Document, will be discussed at the May COSDAM meeting.

Rick Hanushek asked whether there should be an achievement level for below NAEP Basic; Cizek responded that the current Board policy does not treat this category as an official achievement level but that this issue is related to the next topic on the agenda.

Below the NAEP Basic Achievement Level

Cizek noted that it is important to better understand what students below NAEP Basic know and can do, but that having an official achievement level is not necessarily the only or best way to do this. He explained that the range of performance in this category spans from zero to just below NAEP Basic.

Rosenberg stated that at the direction of COSDAM leadership, Board staff commissioned a paper to describe how state and international assessments handle the lowest category of achievement. This paper will be completed in approximately one month and can serve as a resource for future Committee discussion on this topic. As a subcontract to the Human Resources Research Organization (HumRRO), the paper is being prepared by Karla Egan of Edmetric. The paper will look at how many state assessments have a Below Basic achievement level, what the nature of that achievement level description is, how it compares to descriptions of the other levels, and potential pros and cons of a Below Basic achievement level for NAEP.

Suzanne Lane noted that many states have very coarse descriptions for Below Basic (or whatever the lowest category is called) that are mostly in terms of limitations but that New York does an exceptionally good job of describing what students in the lowest category can do. Cizek closed by noting that the consideration of a Below Basic achievement level has serious

design implications for NAEP, including making sure that there are sufficient items towards the bottom of the scale that can be used to measure and describe what the lowest performing students know and can do. Decisions about the number of achievement levels cannot be made in isolation from operational considerations related to test development, design, and administration. Peggy Carr agreed with Cizek and noted that in most cases, NAEP has very few items at the bottom of the scale. Julia Rafal-Baer expressed concern with not having enough items in this range at this point in time given the prediction of lower student performance in the wake of the COVID-19 pandemic; the need for more items and better information about student performance at the lower end of the scale is greater than ever before.

Proposed Strategic Vision Activities

Cizek explained that the brief discussion on this topic is a preview of two other agenda items, those related to NAEP linking studies and the Achievement Levels Work Plan. He noted that Rosenberg prepared a short presentation to orient Committee members to those topics.

Rosenberg noted that COSDAM members brainstormed potential Strategic Vision activities during the December meeting. Since that time, staff have been developing potential work plans to implement each of the eight priorities. The next step is for each committee to discuss proposed year one goals for the Strategic Vision priorities that they are leading, recognizing that much of the work will occur in cross-committee groups. Cizek will be sharing key takeaways from this discussion with the Executive Committee at their meeting the following day.

There was no additional Committee discussion on Strategic Vision activities at this time.

Framework Development Processes

Cizek transitioned to the topic of framework development and noted that he and Carey Wright had some initial discussions with Dana Boyd and Mark Miller regarding cross-committee work on potential enhancements to the Board policy on framework development. He stated that this work is intended to be distinct from the update of the NAEP Reading Framework. Cizek noted that framework development is clearly in the purview of the Assessment Development Committee but that collaboration seems desirable as COSDAM members also have interest and expertise in this area. Two papers have been commissioned to serve as a resource for future discussions on this topic: former Governing Board Executive Director Cornelia Orr is synthesizing historical information on NAEP framework development and the Center for Assessment (under subcontract to HumRRO) is describing how NAEP framework development processes relate to other assessments and best practices.

Cizek welcomed Boyd and Miller and invited them to address the Committee. Boyd expressed appreciation for the opportunity to collaborate with COSDAM to inform future framework development processes. Miller reiterated a commitment to continuous improvement following the revision to the Board's framework development policy in 2018.

There was no additional Committee discussion on framework development processes.

Next Steps for NAEP Linking Studies

Cizek began by noting that linking studies have been a part of Board conversations on the Strategic Vision. He explained that this topic is not one of his own most important priorities but that he recognizes the value in informing the public and providing context for what NAEP results mean as they relate to other important indicators of student achievement.

Cizek noted that the advance materials contain information about several existing studies, but that not all studies necessarily have findings that are policy-relevant. Therefore, he sees a need to identify policy-relevant findings from existing studies and to determine how to best synthesize, leverage, and communicate those findings. Following those steps, the Board could identify policy-relevant goals that could be addressed through additional studies and create a plan for prioritizing studies to accomplish those goals. He cited the need to work closely with the Reporting and Dissemination Committee (R&D) and NCES. Cizek asked the Committee for feedback on the proposed next steps.

Jim Geringer stated that his definition of policy-relevant is to use NAEP to inform education leaders and policymakers about what the field of education is doing right and where there is room for improvement. But he acknowledged that relevance to policy could be defined in a variety of other ways. He added that consistency in the results between NAEP and other assessments can provide affirmation that NAEP is measuring something relevant, even if the purposes of the assessments differ somewhat. Cizek responded that he does not view linking studies as providing validity evidence for NAEP given the variety of purposes among the various assessments but that linking studies can provide relevant and useful information about how NAEP fits into a constellation of other assessments.

Hanushek noted that he conceives of there being two types of linking studies: 1) validation studies that compare NAEP to other assessments, and 2) the relevance of NAEP to important real-world outcomes and indicators, e.g., college attendance and employment. Cizek noted that given Hanushek's depth of understanding and interest in this topic, he may wish to be part of a subset of COSDAM members that can begin discussions with R&D members to help move this work forward.

Rafal-Baer agreed with Hanushek's framing of two linking purposes and stated that the prediction of employment outcomes is particularly important. She is concerned about recent trends of learning loss and of decreased enrollment in community college. Rafal-Baer suggested that the Board may want to consider proposing changes to the NAEP legislation to allow NAEP to link to some data sources that are currently prohibited.

Russ Whitehurst underscored the importance of predictive outcomes and proposed that such information should be used as an external anchor to inform framework development by focusing on the content that is most predictive of future outcomes. Several COSDAM members countered that there is important NAEP content that may not be predictive but should still be assessed, and that predictive validity is not the most important criterion for NAEP framework

development given the intended standards-referenced interpretations and intended uses of NAEP scores.

Carey Wright and Alice Peisch both discussed the use of NAEP to inform policy decisions in their states, as NAEP has been the driver of a lot of reform work in both Mississippi and Massachusetts.

Cizek closed the discussion by inviting Hanushek and Rafal-Baer to be the two COSDAM members who might work with identified R&D members and Board staff to identify policy-relevant findings from existing linking studies and discuss how this work can be highlighted in ways that are actionable to policymakers. He thanked them for agreeing to do so. He suggested that COSDAM receive a brief update on the status of this work at the next meeting, based on initial conversations of those involved in the effort.

Status of the Achievement Levels Work Plan

Cizek explained that he led the development of the Achievement Levels Work Plan that the Board adopted last year. The plan describes the activities that the Board plans to undertake to respond to the recommendations in the evaluation of NAEP achievement levels conducted by the National Academies of Sciences, Engineering, and Medicine. The ultimate goal for this work is to lead to the removal of the trial status of the achievement levels.

Cizek described the purpose of this session as providing a status update on the implementation of the planned activities. He briefly reviewed the status of each activity, which was also explained in the advance materials. He noted that the COVID-19 pandemic precluded some of the activities from happening on the timeline originally envisioned. Proposed next steps are: 1) Monitor progress and provide input on the studies to review and revise ALDs; 2) Determine how the communication of existing studies and prioritization of new studies can provide context for how the NAEP achievement levels relate to other external indicators; and 3) Set up a contract to accomplish remaining activities that have not been started.

Geringer raised the question of the intended uses of NAEP. Cizek responded that COSDAM has had several discussions with the R&D Committee about the intended meaning of NAEP and intended uses of NAEP. Last year the Board adopted a statement to articulate the intended meaning of NAEP, but additional work is needed to further flesh out and then communicate appropriate and inappropriate interpretations and uses of NAEP. Rosenberg noted that a forthcoming contract is intended to support the remaining activities.

Wrap Up

In closing, Cizek noted that he and Wright will attend the brief ADC discussion on framework development processes. He acknowledged that there are many follow up activities based on the meeting discussion, and that an additional item that COSDAM may need to discuss throughout this year with NCES is related to potential technical challenges for the 2022 NAEP

administrations. Committee members briefly discussed the interpretation of the 2022 results in light of the COVID-19 pandemic and the importance of the contextual questionnaires.

Cizek adjourned the meeting at 3:52 pm ET.

I certify the accuracy of these minutes.

_____
Gregory Cizek, Chair

<u>April 16, 2021</u>
Date

Reporting and Dissemination Committee Meeting

March 1, 2021

10:00 am - 12:15 pm

Closed Session

Under the provisions of exemption 9(B) of § 552b of Title 5 U.S.C., on March 1, 2021, the Governing Board met in closed session from 10:00 am to 10:45 am to receive a briefing on *embargoed results of the 2019 National Assessment of Educational Progress in Science*

10:00 am – 10:45 am   *Attendance for closed session*

Reporting and Dissemination Committee Members: Chair Tonya Matthews, Vice Chair Marty West, Alberto Carvalho, Tyler Cramer, Paul Gasparini, Ron Reynolds, Mark White

Reporting and Dissemination Committee Members Absent: Governor Bev Perdue

Governing Board Members: Dana Boyd, Christine Cunningham, Eric Hanushek, Patrick Kelly, Reginald McGregor, Mark Miller, Alice Peisch, Carey Wright

Governing Board Staff: Laura LoGerfo, Michelle Blair, Stephaan Harris, Donnetta Kennedy, Lesley Muldoon, Munira Mwalimu, Sharyn Rosenberg, Angela Scott, Matt Stern, Lisa Stooksberry, Anthony White

National Center for Education Statistics Staff: Peggy Carr, Brian Cramer, Pat Etienne, Jasmine Fletcher, Daniel McGrath, Holly Spurlock, Ebony Walton, Grady Wilburn

Contractors: AIR: Kim Gattis, Young Yee Kim, Sami Kitmitto; CRP: Shamai Carter, Anthony Velez; ETS: Marc Berger, Jay Campbell, Amy Dresher, Robert Finnegan, Cassandra Malcom, Lisa Ward; Hager Sharp: James Elias, David Hoff, Joanne Lim; The Hatcher Group: Jenna Tomasello; Optimal Solutions Group: Imer Arnautovic, Charlotte Notaras; Pearson: Scott Becker; Silimeo Group: Debra Silimeo; Westat: Chris Averett

---

Chair Tonya Matthews called the Reporting and Dissemination Committee meeting to order at 10:00 am on Monday, March 1, 2021. The meeting's first session offered a preliminary preview of the 2019 results from the National Assessment of Educational Progress (NAEP) Science assessment. Because these results will not be released until mid- to late May 2021, this session was closed to the public. Only Governing Board members, Board staff, staff from the National Center for Education Statistics (NCES) and their contractors attended. The Reporting and Dissemination (R&D) Committee extended an invitation to all Board members to join the meeting and learn the results; several accepted this invitation to participate. Grady Wilburn presented and explained the results, after which he fielded questions from the committee members. The closed session ended on time at 10:45 am in accordance with the Federal Register notice.

10:45 am - 12:15 pm Attendance for open sessions

Reporting and Dissemination Committee Members: Chair Tonya Matthews, Vice Chair Marty West, Alberto Carvalho, Tyler Cramer, Paul Gasparini, Ron Reynolds, Mark White

Reporting and Dissemination Committee Members Absent: Governor Bev Perdue

Governing Board Members: Dana Boyd, Eric Hanushek, Alice Peisch

Governing Board Staff: Laura LoGerfo, Michelle Blair, Stephaan Harris, Donnetta Kennedy, Lesley Muldoon, Munira Mwalimu, Sharyn Rosenberg, Angela Scott, Matt Stern, Lisa Stooksberry

National Center for Education Statistics Staff: Peggy Carr, Jing Chen, James Deaton, Pat Etienne, Daniel McGrath, Nadia McLaughlin, Holly Spurlock, Ebony Walton, William Ward

Contractors: <u>AIR</u>: George Bohrnstedt, Markus Broer, Kim Gattis, Cadelle Hemphill, Young Yee Kim, Sami Kitmitto; <u>CRP</u>: Shamai Carter, Arnold Goldstein, Anthony Velez, Edward Wofford; <u>ETS</u>: Jonas Bertling, Jay Campbell, Gloria Dion, Amy Dresher, Robert Finnegan, Paul Jewsbury, Hilary Persky, Courtney Sibley, Yan Wang, Lisa Ward, Ryan Whorton, Karen Wixson; <u>Hager Sharp</u>: James Elias, David Hoff, Joanne Lim; <u>The Hatcher Group</u>: Robert Johnston, Devin Simpson; <u>Optimal Solutions Group</u>: Imer Arnautovic; <u>Silimeo Group</u>: Debra Silimeo; <u>Westat</u>: Chris Averett, Lauren Byrne, Jason Nicholas

Other: Rolf Blank; <u>Wayne State University</u>: Latitia Watkins

---

Draft Release Plan for the 2019 NAEP Science Results

The R&D Committee reconvened in open session at 10:50 am to review the proposed plan for releasing the 2019 NAEP Science results. Laura LoGerfo, the Governing Board's assistant director for reporting and analysis, explained the plan to the committee members. Dan McGrath, director of reporting for NAEP at NCES, requested two changes to the draft plan. First, McGrath sought less specificity for which data NCES Commissioner Lynn Woodworth would present at the release event. The draft plan recommended that Commissioner Woodworth share results from the Trends in International Mathematics and Science Study, which was released in December 2020. However, NCES staff requested that the data Commissioner Woodworth will present not be named until a later date.

Second, McGrath requested that any video produced by the Governing Board to explain the content of the three subscales on the NAEP Science assessment be shown separately from the presentation of the results. This reflects NCES' desire to distinguish sharply between the Governing Board's work and that of NCES. The committee agreed to amend the plans to accommodate these requests.

Tyler Cramer praised the exceptional organization of the last two virtual release events and noted that videos explaining each subscale could be posted and disseminated via social media easily. Matthews inquired if there were any differences in uptakes between videos posted to social media and video clips from releases. The Hatcher team responded that they would investigate that query.

Marty West observed that the "short and sweet" approach taken for the October release of the Grade 12 NAEP results in Reading and in Mathematics succeeded and suggested that this same strategy be applied to the Science release event. The Board could invite science-focused stakeholders to pre-record questions for Board members who would answer them during the event. Christine Cunningham, a science curriculum expert on the Governing Board, could participate along with another current or former Board member expert in science content.

The committee agreed that messaging should focus on the importance of science education, given the pandemic and efforts to develop and distribute vaccines. The NAEP Science assessment occurred before COVID-19 impacted the nation and the world, yet the relevance of an assessment measuring knowledge and skills in life sciences, earth sciences, and physical sciences (e.g., chemistry) remains resoundingly consequential.

Alberto Carvalho commended the summary graphs Grady Wilburn shared in the closed session and recommended the release plan focus on addressing two critical questions: (1) What is the audience learning; (2) Is there some causality the audience will or should infer? Thus, the event should guide participants and stakeholders in interpreting the results easily yet correctly. To

facilitate the interpretation of complicated, nuanced results, the staff should excise chunks of the release event for dissemination more broadly after the event itself.

LoGerfo acknowledged that the plan would reflect the amendments suggested by the committee. Tyler Cramer made a motion to approve the release plan and present it to the full Governing Board for approval on Thursday, March 4th; Marty West seconded the motion, and the motion passed unanimously.

Understanding Socioeconomic Status and NAEP

Vice Chair Marty West convened several panelists to discuss the measurement and interpretation of socioeconomic status in NAEP. West introduced the topic by noting that the committee has long bemoaned NAEP's reliance on the increasingly convoluted indicator of student eligibility for free and reduced-price lunch to capture socioeconomic status (SES). Similarly, the idiosyncratic items about socioeconomic status on the student contextual questionnaire seem weak in comparison to those in other data collections. The Reporting and Dissemination Committee is not alone in their dissatisfaction as members of the Committee on Standards, Design and Methodology as well as the Assessment Development Committee also find these items lacking.

This session intended first to present approaches for improving the measure of SES underway by the NAEP team and second to offer alternative measurement methods. To the first intention, William Ward of NCES and Markus Broer of the American Institutes for Research (a NAEP contractor) shared insights into research and development work within NAEP. To the second intention, Thomas Kane of the Harvard Graduate School of Education and Rick Hanushek of Stanford University's Hoover Institution (and Governing Board member) lent the committee members insights from their research.

Ward explained the fundamental assumptions which underlie NAEP's construct of SES. A measure of SES for the NAEP program must be useful, relevant to educational outcomes, and work in similar ways across grades 4, 8, and 12. The measure should comprise existing variables, so that any new iteration of SES can be applied to previous data to chart critical trends. To conduct such trend analyses, NAEP needs to measure the same construct over time so that changes in scores reflect changes in what students know and can do and not changes in the variable or construct. This criterion imposes a strict limitation, but panelists did offer some tentative solutions to this constraint.

Ward reminded the committee how NAEP currently captures SES, which is through eligibility for the National School Lunch Program (NSLP). This strategy suffers from variable reliability across grades, large within-category differences, and changes in eligibility across time, among

other issues. Ideally, SES comprises three sources of information: (1) parental educational attainment; (2) parental occupational status; and (3) family income. The expert committee which advises NAEP on collecting contextual data recommended measuring school-level SES and neighborhood-level SES, then combining those with student-level SES to build an expanded measure of SES.

However, challenges emerge. Specific items from the contextual questionnaire are vulnerable to change over time, such as outdated references to brand-name technology like Nooks or to items that are so prevalent they no longer distinguish SES categories. Other challenges reside in missing or erroneous data, which derive from a wide range of reasons, from states opting out of the student questionnaire to students' lack of knowledge on particular items, e.g., about a quarter of fourth-graders do not know their parents' education.

Markus Broer then described his investigations into a new and improved measure of SES with extant NAEP data so as to analyze trends from 2003. His measure includes:

- number of books at home;
- students' eligibility for NSLP;
- percent of students eligible for NSLP at school the student attends; and
- parents' highest level of education.

Broer finds that NSLP retains its value, validity, and power to explain variance in assessment performance, despite the aforementioned issues with the variable. At grade 4, the fourth component of the index--parents' highest level of education--is excluded, due to inaccurate reporting or missing data. When analyzed, Broer's index explains more variance in NAEP scores than NSLP alone and than measures of SES employed by large-scale assessment programs like the [Trends in International Mathematics and Science Study](#) (TIMSS). This index also shows expected correlations with other measures and explains achievement at the national level, at the state level, at the TUDA level.

To present an alternative approach, Professor Tom Kane shared his results from modeling the relationship between income and achievement in the NAEP data. Kane noted disagreements in the field about how achievement gaps by SES have changed over time, with Sean Reardon at Stanford revealing a significant widening in the achievement gap and Rick Hanushek and others seeing a flattening or narrowing of the gap. Kane challenged the foundation of those disagreements by arguing that SES does not serve as a good proxy measure for income and advocated for using income alone.

Kane enumerated weaknesses in different methods to capture income through NAEP, such as imputing income from students' race, maternal education, state, and urbanicity. Other attempts include matching school locations to neighborhood mean income from Census data. But this method is vulnerable to inaccuracy from increasing prevalence of school choice, and the vast

majority of variance in parental income lies within schools, not between schools. Kane's third alternative approach -- of adding a parent questionnaire for a subsample of students and schools -- would require changes to several laws.

The strongest, most valid alternative approach to measuring income on NAEP, posited Kane, is by linking NAEP data to Census data. Kane reassured the committee that this approach would safeguard student privacy. Specifically, students would enter their addresses into the secure system provided by NAEP. The device would match the students' address to a neighborhood, for which the mean income would be drawn from Census data. The students' address would be deleted from the device or platform after the match is made and only the neighborhood mean income level would be retained and assigned to students' records.

Following Kane's conclusion, Rick Hanushek presented his thoughts on measuring income and SES. Hanushek and his colleagues have combined outcomes and SES measures from several sources, including Long-Term Trend NAEP, Main NAEP, TIMSS, and the Programme for International Student Assessment (PISA) to find that achievement gaps by SES have declined over the last fifty years.

Hanushek averred that there is no standard or accepted measure of SES. As such, measuring SES gaps over time presents even greater difficulties than maintaining trend in NAEP. Given that, and the challenges explained by Bill Ward, factoring parental education together with items in the home represents the best strategy to capturing SES. Hanushek admitted the sources of uncertainty in NAEP, such as accuracy in students' replies, the burden on students, and shifted meanings of constructs over time.

To illustrate his points, Hanushek showed how the construct "items in your home" quickly becomes outdated. In 1990, the NAEP questionnaire asked students if they have encyclopedias in their homes and whether their families regularly receive deliveries of newspapers and magazines. Even a NAEP respondent in 2003 would find a question about encyclopedias amusing at best and confusing at worst. He then walked the committee members through graphs showing the relationship between SES and achievement across different datasets.

These illuminating presentations provoked thoughtful questions from the committee members. West invited members to send LoGerfo any additional questions that the panelists could answer via email and/or follow-up conversations. He then summarized the conversation succinctly. Approaches taken by Broer and Hanushek share conceptual similarities--based on information taken from the NAEP student contextual questionnaire--but differ in methodological approach, i.e., constructing an additive index or an index through principal components analysis or using NAEP alone or in combination with other data. Kane's approach differed from those completely and shifted focus away from SES to income and away from continuing historical precedent to innovating a different approach for the future.

West then posed the first query to Ward, questioning why NCES assumes that the SES measure must behave the same across grade levels and rely only on existing questions. Does this lead to an assumption that questions about SES, such as household items, must be understood by both twelfth-graders and fourth-graders? Currently, researchers struggle to use NAEP data and express dissatisfaction with the SES measure. How much should what NAEP does now and in the future be constrained by retaining connections to the past?

Ward objected to West's premise and suggested that NCES should become more innovative in how questions are phrased so that all students can answer questionnaire items accurately. For example, NAEP is currently exploring how to ask fourth-graders about their family structures, which is often complicated and tricky to capture in a survey. Ward claimed it is incumbent upon NAEP to gather accurate information from students more effectively. However, West pointed out an intrinsic contradiction in Ward's reply; developing new questions nullifies the claim that variables now must be compatible with variables in the past.

West asked Kane if he ever compared the percentage of variance in achievement explained by income only with that explained by SES. Kane has not yet conducted this comparison but predicted that income would capture considerably more variance than SES. Kane also clarified that he can impute income using school locations through Census data dating back to 1990, which would allow trend calculations. Only the student-level neighborhood income measure would be new.

Kane concluded his response by beseeching the Board to help sort out the muddled picture of SES and achievement. Non-researchers could grasp achievement patterns more easily if NAEP used an index. But, given the variety of SES indices, there is no consensus on the "right" index. Indeed, the "right" SES measure likely differs with the question being asked. Some research shows the gaps widening; others show the gaps narrowing. All of this research shows wide variation in results with SES, so using only income could clarify the issue.

Time expired, and Matthews concluded the session with an enthusiastic thanks to West and to the panelists. She warned that this conversation was intended to foment more robust and longer conversations in the future. With that cliffhanger, Matthews adjourned the meeting at 12:16 pm.

As promised, R&D Committee members did post questions in the chat and through subsequent emails. Paul Gasparini wondered if NAEP release events highlight differences by SES and how policymakers use this information. Tyler Cramer sought explication on the interoperability of NAEP data with external data. Matthews highlighted three points for the committee to pursue further:

1. Within-school SES differences
2. Maternal outcomes as indicators
3. Acknowledging student understanding (or lack thereof) to academic descriptions of socio-economic status

Ron Reynolds sent Rick Hanushek a question: *Could you please unpack the comparison you drew between operationalizing SES and maintaining NAEP trend (…only more difficult)? I suspect it involves tradeoffs between validity and reliability...*

Hanushek replied: With the SES trends in scores, we want to look at gaps between students at different points in the underlying SES distribution, e.g., students in the bottom SES quartile versus students in the top SES quartile. But family SES is estimated from survey background questions that change over time. Thus, there is a recurring question of whether we are measuring SES in the same way over time -- and thus whether any NAEP score differences reflect how we are measuring SES or how well kids at different points in the SES distribution are performing.

Ron Reynolds corresponded with Tom Kane about what sources of income his measure includes. Kane responded: The Census and Current Population Survey questionnaires ask about a series of sources of income individually for each person above the age of 15 in the family/household: wages, self-employment, interest/dividends, Social Security, Supplemental Security Income, public assistance, retirement/pensions, VA payments, unemployment insurance, child support, alimony. Then, "total income" is just the arithmetic total of the individual items respondents reported for all members of the family/household. It does <u>not</u> include non-money income—such as SNAP (Food Stamps), Medicaid or housing subsidies. It also does not include Earned Income Tax Credits.

The previous research (which...finds that gaps are widening) compiles studies using different types of income measures, including questions where a parent or student is simply asked "What is the total family income?" and respondents are given categories from which to choose.

R&D Committee leadership expressed thanks to the panelists for their time during and after the meeting to clarify their positions and address questions.

I certify the accuracy of these minutes.

_____   April 15, 2021
Tonya Matthews, Chair   Date

RELEASE PLAN FOR THE
NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS (NAEP)

The Nation's Report Card: 2019 Science

The national results of the 2019 National Assessment of Educational Progress (NAEP) Science assessment for grades 4, 8, and 12 will be released to the public in May 2021. Typically, results from these assessments are released a year after administration, however, the shift to digital-based assessment required additional quality control processes and statistical checks. The release will be held virtually to comply with public health norms in response to the COVID-19 crisis. The event will be webcast live for a national audience and last approximately 70-75 minutes.

OVERVIEW

The event will begin with a welcome, followed by an introduction by Board member Christine Cunningham, a professor of education and engineering, who works to make engineering and science more relevant and accessible, especially for populations underrepresented and underserved in engineering and science.

A video produced by the Governing Board will introduce the three Science assessment subscales by showing how students engage in the study of life sciences in both extraordinary and ordinary ways. These ways will connect to elements seen in the NAEP Science assessment framework. For example, when schools closed in March 2020, parents found videos online to instruct their children on proper hand-washing techniques to combat COVID-19, to lead their children through science experiments with baking soda, and to make slime. Students participated in backyard bio blitzes while others invented innovative ways to address the Flint water crisis or discovered a novel small molecule that could lead to a cure for COVID-19.

Then focus will shift to data presentations by both the Commissioner and Associate Commissioner of the National Center for Education Statistics (NCES). The Commissioner will share recent highlights of science data from NCES. After which, the Associate Commissioner will release and present the 2019 NAEP Science results for the nation's fourth-, eighth-, and twelfth-grade students, providing an overview of the national data and illuminating national

trends. Associate Commissioner Carr will share highlights of results from subscales of the 2019 NAEP Science assessment and provide summary slides, after which a question-and-answer session will proceed. As with the release for the 2019 NAEP Reading and Mathematics results, grade 12, Governing Board staff will collaborate with NCES staff to select, direct, and ask the questions.

Once the data portion of the event concludes, we will replicate the approach taken for the release of the Grade 12 NAEP data, with pre-recorded questions from stakeholders and answers provided in real time by Governing Board members and/or alumni.

DATE AND LOCATION

The release event will occur in mid- to late May via virtual platform. The Chair of the Reporting and Dissemination Committee will set the release date, in accordance with Governing Board policy, in collaboration with the National Center for Education Statistics, and following Committee acceptance of the final report card.

ACTIVITIES BEFORE THE RELEASE

In the weeks before the release event, the Governing Board will launch a social media campaign to build interest in the release, with special focus on stakeholders involved in science, tagging influencers in this field and former Board members prominent in science education. The Board's website will dedicate a webpage to release events.

Shortly before the release, NCES will host a call for members of the media, during which NCES will present highlights and answer questions. NCES will oversee an embargoed website with results available to stakeholders approved for access by NCES, including Congressional staff and media. The goal of these activities is to provide a comprehensive overview of the findings, to deepen understanding of the results, and to help ensure accurate reporting to the public.

REPORT RELEASE

The Commissioner of the National Center for Education Statistics will release the report card on the NAEP website—at 12:01am the day of the release event. The Governing Board press release, the full and abridged versions of the 2019 NAEP Science Assessment Frameworks, and related materials will be posted on the Board's web site. The site will feature links to social networking sites and multimedia material related to the event.

CENTRAL MESSAGES

Activities before and after the release, as well as the release itself, will promote several messages. First, data from NAEP illuminate critical gaps in students' knowledge and skills within the three science domains assessed by NAEP. By focusing on what content is challenging in these domains and for which students, actions to bolster student knowledge and skills may be more directed and effective. Second, science knowledge and skills do not dwell only among the elite echelons of academia and famous science fairs; everyone can and should participate in the study and practice of science. Science education allows students to understand the world in which they live and learn to apply scientific principles to their lives. Third, international assessments and other NCES data offer helpful information and context to interpret the NAEP results.

ACTIVITIES AFTER THE RELEASE

The Governing Board's communications contractor will work with Board staff to coordinate additional post-release communications efforts to target communities and audiences. The subscale videos will be publicized on social media. The goal of these activities is to extend the life of the results and provide value and relevance to stakeholders.

National Assessment Governing Board

Nominations Committee
(Closed Session)

Report of February 22, 2021

Nominations Committee Members: Governor Jim Geringer (Chair), Tyler Cramer, Paul Gasparini, Tonya Matthews, Reginald McGregor, Mark Miller, Alice Peisch.

Board Member Absent: Dana Boyd.

Board Staff: Donnetta Kennedy, Munira Mwalimu, Tessa Regis, Lisa Stooksberry.

Under the provisions of exemptions 2 and 6 of § 552b (c) of Title 5 U.S.C., the National Assessment Governing Board's Nominations Committee met in closed session virtually on Monday, February 22, 2021 from 5:30 p.m. to 6:30 p.m. to review, discuss, and take action on finalists for Board terms beginning October 1, 2021.

Governor Geringer welcomed members and provided a preview of the agenda. He described the timeline that began in summer 2020 with the call for nominations, noting that there are four vacancies. Three categories are part of the 2021 cycle:

- Elementary School Principal
- General Public Representative
- Testing and Measurement Expert

A fourth category, Local School Board Member, was not filled in 2020. The finalists in this category will be presented to The Honorable Miguel Cardona, Secretary of Education, for 2021 appointment along with finalists in the other three categories.
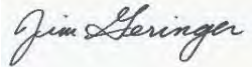
Governor Geringer summarized activities undertaken for the 2021 nominations process. He described the number of nominations received and provided an overview of candidate demographics. Tonya Matthews credited staff member Stephaan Harris for outreach conducted during the 2021 campaign that yielded an increase in diverse applicants. Governor Geringer reminded Committee members that all applicants' ratings were discussed during a conference call on January 27, 2021.

Committee members briefly discussed finalists by category and made suggestions for the closed plenary session to be held Friday, March 5, 2021. Geringer described next steps once the Board takes action on the final slate of candidates to be presented to the Secretary.

Governor Geringer asked for a motion to approve the Nomination Committee's recommendations on the final slate of candidates for the 2021 Board vacancies, to be submitted to the Board for discussion and action. The motion was made by Tyler Cramer, seconded by Mark Miller, and approved unanimously.

The meeting adjourned at 6:30 p.m.

I certify the accuracy of these minutes.

_____   April 15, 2021
Jim Geringer, Chair   Date

On Thursday, May 13th, Governing Board members will convene in small groups to discuss recommendations which surfaced during a session at the March 2021 meeting entitled *Reflections on Recommendations from the National Academies' Committee on Developing Indicators of Educational Equity*. That session featured three presenters who shared considerations for equity within NAEP.

A concise summary of the three presentations follows, after which the goals for the May 2021 discussion are described.

March 2021 Quarterly Meeting Summary: Equity Session

At the March 2021 quarterly meeting of the National Assessment Governing Board, members heard from three experts about considering equity within the context of assessment generally and the National Assessment of Educational Progress (NAEP) specifically.

- **Rucker Johnson**, the Chancellor's Professor of Public Policy in the Goldman School of Public Policy at the University of California, Berkeley. Johnson affirmed the importance of NAEP data and claimed that without NAEP, researchers cannot measure and track learning improvements and deficits related to school resources.
  - Johnson used NAEP as a benchmark to convert school and student test scores to grade-level equivalents at the national level, thus allowing comparisons of district per-pupil spending and achievement.
  - Johnson described how investments in pre-kindergarten and transitional kindergarten translated to stronger outcomes in subsequent grades. Johnson's research highlights the importance of how districts spend funds to such student outcomes as high school graduation rates.

- **Gerunda Hughes**, a member of the NAEP Validity Studies (NVS) Panel and Professor Emerita at Howard University, spoke about the role of NAEP as an indicator of educational (in)equity. Hughes suggested that NAEP can be infused with more equitable design, questions, and reporting.
  - Hughes explained that equity speaks to fairness, social justice, and the unequal distribution of resources so that individuals receive what they need to achieve an outcome, compared to equality where everyone receives the same resources. Equitable assessments should be aligned and validated with their specified interpretations and intended uses of results.
  - Hughes suggested collecting and/or reporting student grouping data not specifically listed in the legislation, such as additional variables to address (a) societal, (b) socioeconomic, (c) cultural, (d) familial, (e) programmatic, (f) staffing, (g) instructional, (h) linguistic, and (i) assessment inequities in the educational system.
  - Hughes highlighted where an equity lens can be applied to NAEP, namely in: (a) sampling, (b) assessment design and development, (c) administration, (d)

accommodations, (e) data analysis and reporting, (f) reporting and interpretations, and (g) use of results. She concluded with the five "E's" of equitable educational assessment: empathy, engagement, equity, evaluation, and equality.

- **Christopher Edley**, the Honorable William H. Orrick, Jr., Distinguished Professor of Law at the University of California, Berkeley Law School, chaired the National Academies' Committee on Developing Indicators of Educational Equity. The Committee recommended developing indicators of educational equity in seven domains. Edley focused his presentation on one domain—elementary and secondary school learning— and three recommendations for indicators: (1) engagement in schooling; (2) performance in coursework; and (3) performance on tests.
  - Committee members identified constructs to measure these indicators, e.g., engagement in schooling can include attendance or absenteeism. Future work needs to define the constructs. Some tailoring of the indicators for subgroups of special interest may be needed, but there should be a core set of indicators with comparability across jurisdictions.
  - The next steps are to build on existing data to measure and collect the indicators. In some cases, research and development are needed. NAEP is identified as a possible indicator of "disparities in performance on tests."
  - Edley asked the Board to consider "how NAEP and related data should be used to provide context and how NAEP-related data should be used within the Equity Indicators System."
  - Edley hopes the Governing Board will adopt a resolution commending the committee's effort to elevate the importance of the initiative and to raise funds to support next steps.
  - He also offered an ambitious suggestion to expand the Governing Board's statute to include overseeing a national system of educational equity indicators or to serve as an institutional home for the indicators.

A robust, yet very brief question-and-answer session followed the presentations. Highlights from the short discussion focused on urging educators to think beyond _what_ factors in education work to _how_ they work, on aligning interventions to school settings, and calling for additional data to capture students' educational experiences more fully, with a focus on subgroups.

In response to a few of the recommendations, Lynn Woodworth, Commissioner of the National Center for Education Statistics (NCES), shared that no law precludes NAEP from oversampling student groups. NCES is field testing a new SES indicator with selected states. NCES staff is permitted to conduct secondary analyses, however they are constrained by funding and staffing limitations.

Given the brevity of the discussion, staff deemed a subsequent conversation critical to deliberating upon the recommendations.

May 2021 Quarterly Meeting

The virtual meeting approach thus far has prevented small groups of members convening for activities beyond committee meetings. Given the content of the panelists' recommendations, however, small groups seem more amenable than a plenary session to facilitate deliberations on the recommendations' merits. Thus, members will exit the main meeting to meet in separate Zoom 'rooms' on Thursday, May 13th for small group deliberations.

Three goals drive these small group discussions:

1. What are the Board's goals for addressing equity through NAEP?
2. Which of the recommendations by Hughes and Edley warrant further discussion and/or pursuit?
3. Of those tagged for further effort, how should the Board prioritize the recommended activities?

Every Governing Board member will be assigned to a small group. The small groups will meet for approximately an hour. Each will be led by a Governing Board member who will offer a few questions to prompt conversation. A staff member will take notes. Observers may listen to the discussion, but only Governing Board members will participate in the discussion.

This May meeting will *not* include a reporting of the conversations. Instead, staff will write summaries of each group's conversation, which will be required reading prior to the August board meeting when a plenary session will address the topic.

To prepare for this session, we encourage you to read through: (1) Gerunda Hughes' PowerPoint slides; (2) a brief report from Edley that explicates in summary form the work of the National Academies' committee, next steps, and partnerships for the effort.

The Role of NAEP as an Indicator of Educational (In)Equity

Gerunda B. Hughes

Professor Emerita, Howard University

National Assessment Governing Board (NAGB) Meeting – March 4, 2021

NAEP as an Indicator of Equity in Education

- The National Academies of Sciences, Engineering, and Medicine recently released a report titled, *Monitoring Educational Equity*.

- The report acknowledges that disparities in educational attainment among different population groups have characterized the United States throughout its history and proposes to establish and implement a system of indicators of educational equity.

- In the report, NAEP is identified as a possible indicator of "disparities in performance on tests".

What is Equity?

Equity speaks to fairness and social justice and the acknowledgement of differences. It references the differential or (un)equal distribution of resources or inputs for the purpose of meeting a specific need to address a particular purpose or outcome.

Gordon, E.G. (Summer, 1995). Toward an equitable system of educational assessment. *The Journal of Negro Education, 64(3), pp.360-372.*

Provisions of "The NAEP Law"

- **Purpose** – "…to provide, in a timely manner, a *fair and accurate measurement of student academic achievement* and *reporting of trends* in such achievement in reading, mathematics, and other subject matter as specified in this section."

- **Measurement and Reporting** –"The Commissioner of Education Statistics …shall-

    A. Use a *random sampling process* which is consistent with relevant, widely accepted professional assessment standards and that produces data that are *representative on a national and regional basis*;

    B. Conduct a national assessment and *collect and report assessment data*, including achievement data trends, *in a valid and reliable manner* on student academic achievement in public and private…schools…

    G. Include information on special groups, including, whenever feasible, *information collected, cross tabulated, compared, and reported by race, ethnicity, socioeconomic status, gender, disability, and limited English proficiency…*

Inequities in Public Education

- Societal inequity

- Socioeconomic inequity

- Cultural inequity

- Familial inequity

- Programmatic inequity

- Staffing inequity

- Instructional inequity

- Linguistic inequity

- Assessment inequity

Source: https://www.edglossary.org/equity

Attributes of Equitable Educational Assessments

- ***Fair*** – Fair assessments are sensitive to the characteristics of different groups being assessed and thereby, where appropriate, employ equitable strategies in the design, development, and delivery of the assessment and in the reporting, interpretation, and uses of assessment results.

- ***Accurate measurement*** – Accurate measurement occurs when measurement error is minimized for all groups of test-takers.

- ***Valid interpretations and uses*** – Equitable measures are aligned and validated with their interpretations and uses of assessment results.

Validity of Inferences and Uses of Assessment Results

"There are several levels of inference that can be made from a test. Consider a reading comprehension test built on several passages drawn by an appropriate random sampling procedure from the articles appearing in *Reader's Digest*. Comprehension of each passage is tested by a set of multiple-choice questions, and [the] score is the number of correct answers chosen."

A low group mean score on this test might lead to any of the following inferences. Given each inference, how might the assessment results be used?

1. The individuals in this group have a low level of understanding of these passages.
2. The individuals in this group will have difficulty in understanding the contents of *Reader's Digest*.
3. The individuals in this group are, in general, poor readers.
4. The individuals in this group are not likely to do well in college.

Adapted from: Thorndike, R.L. (1982). *Applied Psychometrics.* Boston: Houghton Mifflin Company

Equity in Educational Assessments

Stages of Test/Testing Process

- Purpose
- Sampling
- Design & Development
  - Content, Item formats
- Administration
  - Accessibility, Accommodations, Standardization
    - Mode, Timing, Language, etc.
- Scoring
- Analysis of Data
  - "Mirror/Thermometer", "X-Ray", "MRI", "Biopsy"
- Reporting and Interpretation of Test Results
- Use of Test Results

Characteristics of Test-Takers

- English Learners
- Economically Disadvantaged Students (SES)
- Culturally Diverse Students
- Students with Disabilities
- Gender
- Racially/Ethnically Diverse Students

Equity in NAEP: Sampling

- ## Identifying the Assessment Population
  - Oversample for some subgroups of interest – While "The NAEP Law" calls for *representative sampling*, it may be necessary to oversample for some groups such as Native Americans or ethnic groups whose presence in the general population has shown significant changes over time.
  - For years, eligibility for the National School Lunch Program (NSLP) was used as a proxy for socioeconomic status (SES); however, the validity of NSLP eligibility was shown to decrease over time. Therefore, an expert panel issued recommendations to NCES on how to improve the measurement of SES for NAEP.

- ## Developing the item pool
  - Ensure there are enough of items at all points along the score scale in order to accurately measure the achievement of all student groups being assessed.

- ## Sampling of Subject Matter Content
  - What will be the bases for determining the subject matter content emphases on NAEP? The Content Frameworks only or something else?

Source: Improving the Measurement of Socioeconomic Status for the National Assessment of Educational Progress: A Theoretical Foundation (November 2012).

Equity in NAEP: Assessment Design & Development

- Selecting reading passages

  - Select or create reading passages on a variety of topics to measure reading comprehension with the intended purpose of *maximizing student engagement* across the assessment population. Reading passages can be standardized by ensuring they have the same or nearly the same reading load or length. Allow students to choose which passage or passages they will read.

  - Focus on Students: Standardize the Level of Engagement in a reading passage.

Equity in NAEP: Administration

- Minimize Mode Effects
  - Choose administration modes that minimize measurement error.

- Minimize Device Effects
  - Allow students to take the assessment on devices with which they are familiar.
  - Focus on Students: Standardize Level of familiarity with the device on which they will take the test.

- Digitally Based Assessments (DBAs)
  - Color Contrast, Zooming, Text-to-Speech, Equator Editor, Calculator

Equity in NAEP: Accommodations

- Employ Universal Design features such as:

- Extra time

- Large print

- Language enhancements

- Different item presentation designs

- Multiple-choice
- Short Answer
- Matching
- Fill in the Blank
- Essay
- On-the-Job Evaluations

- Performance
- True-False
- Interviews
- Observations
- Projects

Equity in NAEP: Analysis of Data & Reporting

- "The NAEP Law" states that the Commissioner of Education Statistics shall report *meaningful/useful* statistics for each of the various subgroups of interest.

- Continue to make NAEP data available for conducting secondary analyses with a host of analytic tools such as the NAEP survey questionnaires.
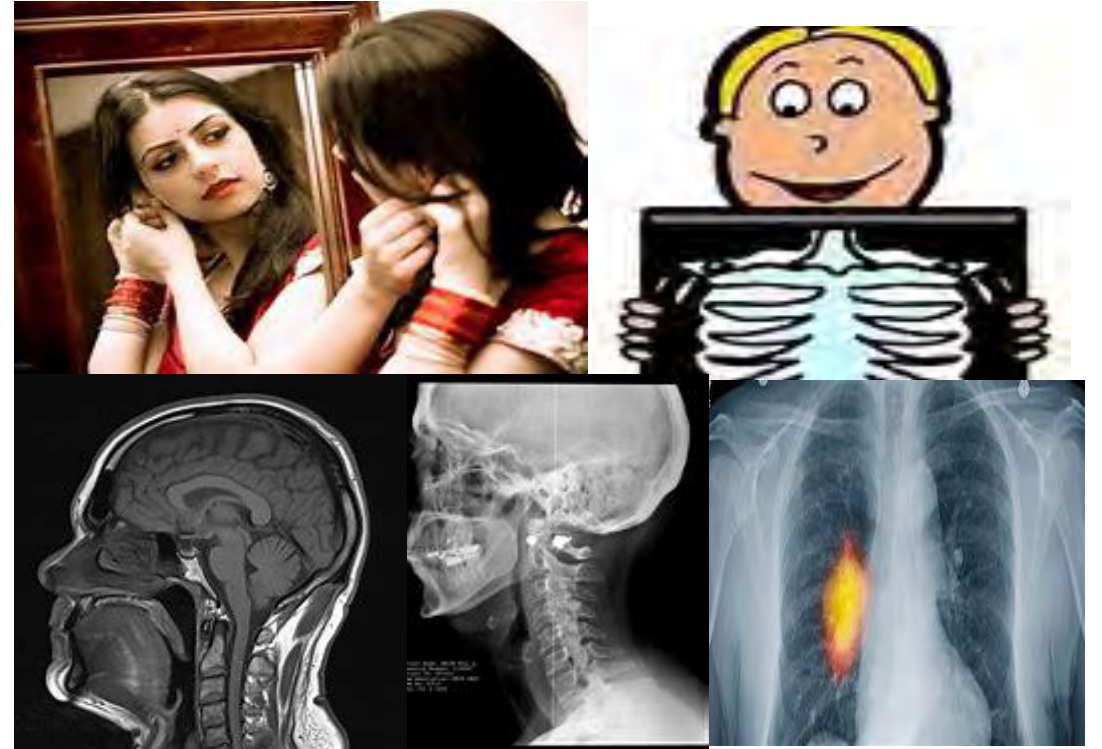
Equity in NAEP: Reporting and Interpretations

- Evaluate the validity and reliability of interpretations and inferences about subgroup performances and comparisons that are made from NAEP data and reports.
  - Highlight *all* comparisons in academic achievement between subgroups defined by race, ethnicity, not just the white-black or white-Hispanic performance gaps. Share the Asian-white gaps as well.

Equity in NAEP: Use of NAEP Results

- Report to a variety of audiences that may be able to use NAEP statistics to improve educational and social equity in a variety of contexts.
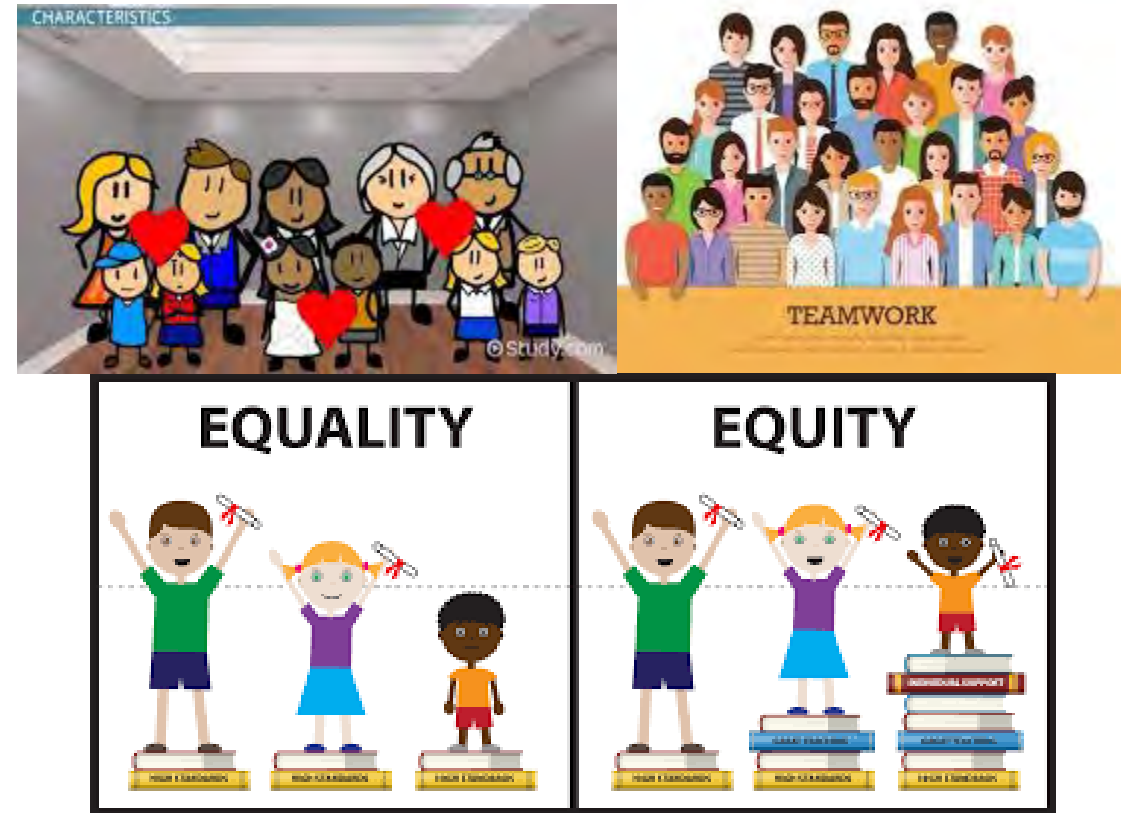
Call to Action for Equity

- The current enthusiasm for "equity" in various aspects of our educational and social environments exists side-by-side with long established, fully functional, institutionalized caste systems in the United States based on race, ethnicity, gender, socioeconomic status, and English language proficiency.

- Not so long ago, the enthusiasm was about "equality". The equality outcome remains illusive. And yet, "equity" is likely to be more difficult to achieve because it requires those who have the resources, the power, and control to share/use these prized commodities with those who need them, but whom they may perceive as "others".

The 5 E's of Equitable Educational Assessment

- Empathy
- Engagement
- Equity
- Evaluation
- Equality

References

Gordon, E.W. (1996). Toward an equitable system of education assessment. *Journal of Negro Education*, 64(3), 360-372.

Improving the Measurement of Socioeconomic Status for the National Assessment of Educational Progress: A Theoretical Foundation. (2012). Washington, D.C.: National Center for Education Statistics.

National Academies of Sciences, Engineering, and Medicine. (2019). *Monitoring educational equity*. The National Academies Press. https://doi.org.10.17226/25389.

Wilkerson, I. (2020). *Caste: The origins of our discontents*. Random House.

Thorndike, R.L. (1982). *Applied Psychometrics*. Boston: Houghton Mifflin.

EDUCATIONAL EQUITY INDICATORS PROJECT (EIP)

—A Foundation for Narrowing Opportunity and Outcome Disparities—

> "Equity" is the absence of significant disparities between student subgroups—not individuals—in opportunities or outcomes. It requires a group-level fit between opportunities and needs. There must be adequate efforts to mitigate the effects on outcomes of structural disadvantages and adversity that disproportionately affect different student groups. Equity is not equality. Inequity need not be unlawful discrimination.[1]

Why this project? Why now?
Because now is different, but will prove fleeting.

Introduction. Racial reckoning and the reduction of inequality require increased educational equity. To that end, the National Academies of Sciences, Engineering, and Medicine (NAS)[1] has recommended a framework for a "national" system of K-12 equity indicators in its report, *Monitoring Educational Equity*.[2] Building on those research-driven recommendations, this proposed Equity Indicators Project (EIP) will improve the evidence available to policymakers, administrators, and the public at all levels, creating durable indicator systems to monitor and compare equity in participating jurisdictions. The project term will be 30 months.

We will augment the national NAS design to include early childhood and postsecondary disparities, while inviting jurisdictions to augment the national system with additional indicators and comparison subgroups. We will focus on four early-adopter states and districts within them. We will also provide technical assistance to states eager to improve their monitoring of equity now, without waiting for the four state pilots.

The process for creating these indicator systems must be aimed at broad consensus and include insiders, outsiders, and researchers at the state and local tables.

Appendices

A. NAS Indicators Framework

B. Workflows

C. Notional Timeline

D. Project Principals, Other Research Partners, and Advisors

E. AERA Role as a Principal Partner

F. Adversity, Context, and Whole Child Equity

G. Budget

H. Project Principals, Other Research Partners, and Advisors

---

[1] *Monitoring Educational Equity* (2019). NAS is the recent merger of the National Academy of Sciences, the National Academy of Engineering, the Institute of Medicine, and the National Research Council—NAS, NAE, IOM, and NRC. We use the familiar acronym, rather than NASEM.

[2] The NAS identified seven domains, such as "educational attainment" or "access to quality curricula and instruction". It recommended 16 indicators distributed across those. For each indicator there are 1-4 specific variables to be defined and measured (constructs), such as "on-time high school graduation rate" or "teachers' years of experience". See *Appendix A*.

Opportunity Institute, U.C. Berkeley Law, & Partners

Context. Perhaps tribalism and racial hierarchy are in our biological and social DNA. Even so, in this American moment, we can make progress against racial hierarchy—both rhetorical *and programmatic*. Inequality is a trending topic in public discourse. COVID-19 incidence, response, and recovery planning have stoked concern, along with criminal justice. Biden has short-listed racial justice, along with the pandemic, the economy, climate change, and America's global standing. Leaders in politics and business have promised progress. Battling racial disparities in educational opportunities and outcomes is essential to honoring our best dreams. To guide our battle, we need indicators that identify critical problems, illuminate promising strategies, measure successes, and continually renew righteous urgency.

We have several premises. ●Evidence should be fundamental, along with passion. Science should be more powerful than intuition and politics whenever possible, although all three are necessary to decisionmaking. ●Indicators must be useful to policymakers, advocates, and the public. ●Building consensus and designing the indicator system must include researchers, administrators, policymakers, and teachers. However, the process must also include outsiders—such as representative community leaders and advocacy groups. ●The indicator system must be sustainable[3], because large-scale change requires consensus in understanding the problem, identifying what is important, maintain funding, consistent leadership, and reporting to consumers. ●Some states are eager to start building equity indicator systems. They will not wait for pilot projects. They deserve support soon because they, too, sense the moment.

NAS Indicators, and Local Augmentation. The NAS framework provides a core set of 16 indicators, but with the expectation that state and local jurisdictions will add more based on their particular concerns. The committee did not recommend a core set of subgroups for comparisons (crosstabs), considering those to be policy and political judgments rather than research conclusions.[4] An early goal of this EIP will be to identify a consensus set of subpopulations—starting from the ESSA and CRDC obligations—again with the expectation that jurisdictions will define additional groups based upon their local salience.

A major aspect of EIP is the consensus, technical engineering of the constructs/measures that make up each indicator. What should be the precise and comparable definition of high school GPA, and how can it be accurately collected? What is the best instrument for assessing social-emotional readiness for kindergarten? Without agreed details, implementation and comparability are impossible.

For each of the 16 indicators there are 1-4 recommended constructs (measures). For example, *domain **B***

---

[3] We mean "sustainable" with regard to funding, leadership, and political support.

[4] Congress specified four in ESSA. The department' s Civil Rights Data Collection system requires reporting on XX.

Opportunity Institute, U.C. Berkeley Law, & Partners

Diagram of Concepts

Domains

Seven Indicator Domains

A. Kindergarten Readiness

B. *K-12 Learning*

C. Educational Attainment

D. Extent of Racial, Ethnic, and Economic Segregation

E. Access to Quality Early Learning Programs

F. Access to Quality Curricula and Instruction

G. Access to Supportive School and Classroom Environments

Indicators

3 of 16+ Indicators:

3. Engagement & Schooling

4. Performance in coursework

5. Performance on tests

Constructs / Measures for Each Indicator

- Attendance/ Absenteeism
- Academic Engagement

- Success in classes
- Accumulating credits (being on track to graduate)
- Grades, GPA

- Achievement in reading, math, and science
- Learning growth in reading . . . .

Engineering Technical Details

? ?
? ?
? ?
? ?

---

is K-12 Learning and Engagement. It has 3 *indicators*, which collectively are further defined by 7 recommended *constructs*. EIP will take the critical, technical, next step of refining each construct with a precise, consensus definition. Then in each state, we will determine whether a construct requires existing, modified, or new data sources. See the **Diagram of Concepts,** above.

For another example, indicator number 10, Disparities in Access to Effective Teaching, has 3 constructs to be measured: Teachers' years of experience; Teachers' credentials/ certification; and Diversity in the teaching force. For the most part, these are proxies for a *direct* measure of effectiveness—for which there is no expert consensus. Moreover, even the proxy constructs themselves lack uniform definitions to allow comparability. Credentialing standards vary.

Three final notes. First, equity indicator systems will differ from the myriad existing dashboards and accountability schemes because those are not focused on research-driven variables covering both outcomes *and* opportunities, and do not specifically probe disparities that are critical for equity. Second, the system points to the drivers of the disparities in order to identify needed changes in policies and spending, not just the end-of-the-pipeline results of underlying decisions. Third, several recommended indicators/constructs—NAS plus augmentations—require additional research and consensus, which should shape an R&D agenda. Indicators vary in degree of difficulty, sometimes for technical reasons but often because there is insufficient research or consensus. Therefore, continuous improvement is vital and requires guidance by a coordinating body. This body might be akin to the prestigious National Assessment Governing Board (NAGB), created by Congress to oversee the gold-standard National Assessment of Educational Progress (NAEP). This role includes advancing an R&D agenda.

*

Domains

NAS Educational Equity Indicators

A
1. Academic readiness

2. Self-regulation and attention skills

B
3. Engagement and schooling

4. Performance in coursework

5. Performance on tests

6. On-time graduation

C
7. Postsecondary readiness

D
8. Students' exposure to racial, ethnic, and economic segregation

E
9. Access to and participation in high-quality pre-k programs

10. Access to effective teaching

F
11. Access to and enrollment in rigorous coursework

12. Curricular breadth

13. Access to high-quality academic supports

14. School climate

G
15. Non-exclusionary discipline practices

16. Nonacademic supports for student success

- New domain—2-3 indicators for Early Childhood

- New domain—2-3 indicators for Post-Secondary

---

*

Five EIP Workflows. The programmatic activity can be grouped into five workflows, not strictly sequential. For more description, *see Appendix B*.

A. *Engage* four states, including at least one urban and one rural district in each. Engage to build consensus on the NAS recommended framework plus state & local augmentation, tailoring the indicator system to local concerns using additional indicators and subgroups. This requires state and local discussion "tables" with (a) insiders, (b) outsiders, and (c) researchers. The tables are both politically and substantively crucial.

We will try to achieve comparability across jurisdictions for the NAS indicators, and maximize use of existing data sources and other "report card" efforts, including those mandated by federal statutes and regulations.

B. *Technical assistance* and multi-district collaboration for jurisdictions that do not want to wait for a beta-tested system. To deliver TA, we will enlist several NGOs and experts from other jurisdictions.

C. *Beta-test* the initial 4-state indicator systems with current and modified data. Publish both technical and general audience white papers with context, early findings, guides to interpretation. (*Cf*., the cluster of NAEP dissemination efforts, but far less ambitious.)

D. *Expand participation*: Set the stage for broadening agreement on and implementation of the equity indicator system: insiders and outsiders in other states; national associations; and stakeholders beyond education. Engage Congress, the Biden Administration, national associations, and national NGOs.

E. *Communication strategies and tools* to support EIP Partners, state and local leaders, and national and state/local advocacy organizations. As the work begins in CA, parents, funders, and leaders will need to be able to talk about the work and bring stakeholders at all levels along. The Data Quality Campaign will develop language and resources to communicate why these indicators matter to multiple audiences and provide communications support and advice throughout the duration of EIP. DQC will support storytelling so that partners, funders, and policy leaders can understand the challenges and opportunities involved in successful engagement and implementation. This will lay a foundation for follow-on work by engaging with policy and education leaders as well as state-based advocacy partners.

Foundational Work for the Project. EIP builds on three pathbreaking bodies of work, each of which involved two or more EIP principal partners.

- A study committee of the National Academies of Sciences, Engineering, and Medicine (NAS) recommended a framework, *Monitoring Educational Equity* (2019). The committee, chaired by Christopher Edley, Jr., marshalled the research literature to select equity indicators, and narrowed scores of possible measures to a parsimonious, policy-relevant, and practical few. See *Appendix A*. Governments, advocates, and researchers have advanced other equity "dashboards" which are generally consistent with the NAS framework or are too vague to be helpful in deciding policy or allocating resources.

- Another project, *The Science of Learning and Development*[5] published two important peer-reviewed journal articles. *One article* (2018) presented the consensus of several prominent researchers on principles that have emerged from recent developments in brain and human development research. The science principles explain the neurobiology that connects student learning with chronic stress, adverse childhood experiences, and the child's context generally[6]. It then identifies some research-based, ameliorative interventions. The *second article* (2018) details the implications of this recent brain science for education practice. The authors of these articles included two of the EIP Principals, Linda Darling-Hammond (LPI) and David Osher (AIR).

This and related research informed some of the more novel NAS analysis. See details in *Appendix F: Adversity, Context, and Whole Child Equity*. Note, specifically:

---

[5] The six organizational partners for the first two phases of SoLD were: the Opportunity Institute (lead), Christopher Edley; Turnaround for Children, Pamela Cantor; Learning Policy Institute (Linda Darling Hammond); American Institutes for Research, David Osher; Education Counsel, Bethany Little; and Harvard Graduate School of Education, Todd Rose.

[6] For a brief description of adverse childhood experiences (ACEs), *see https://www.samhsa.gov/child-trauma/recognizing-and-treating-child-traumatic-stress.* (Substance Abuse and Mental Health Services, U.S. Dept. of Health & Human Services, 2020)

CONCLUSION 3-1: The circumstances in which students live affect their academic engagement, progress, and attainment in important ways. If narrowing disparities in student outcomes is an imperative, schools cannot shirk the challenges arising from context. Neither can they confront these challenges on their own. Contextual factors that bear on learning range from food and housing insecurity to exposure to violence, unsafe neighborhoods, and adverse childhood experiences to exposure to environmental toxins. Children also differ in their individual responses to stress. Addressing student needs, in light of their life circumstances, requires a wide variety of resources. It is a responsibility that needs to be shared by schools, school systems, other agencies serving children and families, and nongovernmental community organizations.

RECOMMENDATION 4: Indicators are needed to document the existence and effectiveness of integrated, cross-agency services.

- The *Getting Down to Facts II* project of Policy Analysis for California Education (PACE, 2018) commissioned 36 peer-reviewed technical studies and 19 derivative research briefs by leading researchers. Like a similar project a decade earlier, this was an empirical, state-of-the-state assessment of California's P-12 system. Studies addressed many salient policy questions and highlighted significant equity gaps, but stopped short of operationalizing the research into actionable indicators. Now, our EIP will take what was learned from that earlier enterprise and expand it so it can be used by educators and policymakers to inform and enable a narrowing of disparities. Like GDTF I and II, the EIP will develop the research agenda and commission a small set of synthesis studies (because of budget constraints). We will suggest a broader set studies for a separate, follow-on effort.

- Congress chartered the *Equity and Excellence Commission*, focused on K-12. Its unanimous report, *For Each & Every Child*[7] (2013), divided the policy reform landscape into five domains: (1) Finance and Efficiency; (2) Teaching, Leading and Learning; (3) Early Childhood; (4) Meeting the needs of children in high-poverty communities; and (5) Governance and Accountability. It agreed on 62 recommendations across those domains. The commission was co-chaired by Dean Christopher Edley, Jr. of UC Berkeley Law School, and Professor Mariano-Florentino Cuéllar of Stanford Law School.[8]

Partners & Advisors. The *Principal Partners* will: (a) collaboratively guide EIP as a whole; (b) conduct or manage specific parts of the workflows (A-E); and (c) receive funds through EIP[9] or directly from funders. *Appendix C* has the full, current list of Principal Partners, Partner Researchers, and Advisors. Excepting the few noted there, all have committed to participate, conditional on funding. We will secure formal commitments and descriptions of roles when the funding picture is clearer.

---

[7] U.S. Department of Education, *For Each and Every Child—A Strategy for Education Equity and Excellence* (2013).

[8] Professor Cuellar, now Associate Justice on the California Supreme Court, became co-chair when Reed Hastings, CEO of Netflix, had to resign.

[9] Practically, this means Opportunity Institute will contract Principal with

Principal Project Partners (3/4/20)	
Opportunity Inst.; UC Berkeley Law	American Education Research Assn
Policy Analysis for CA Education	Data Quality Campaign
The Learning Policy Institute	American Institutes for Research
Educational Oppty Project at Stanford	

*

What would success look like? Agreement in several jurisdictions on detailed specifications for an indicator system, supported solidly by insiders and outsiders. Successful beta-testing, yielding results comparable across early-adopter jurisdictions. Broad preliminary support in the research, practitioner, and advocacy communities. Indications of support on Capitol Hill. Passing of the baton to another entity, e.g., the federal government, National Governors Association, or the Council of Chief State School Officers. And more hope for of each and every child.

***

Opportunity Institute, U.C. Berkeley Law, & Partners

SUMMARY

TABLE S-1 Proposed Indicators of Educational Equity

DOMAIN	INDICATORS	CONSTRUCTS TO MEASURE
A Kindergarten Readiness	1 Disparities in Academic Readiness	Reading/literacy skills Numeracy/math skills
	2 Disparities in Self-Regulation and Attention Skills	Self-regulation skills Attention skills
B K–12 Learning and Engagement	3 Disparities in Engagement in Schooling	Attendance/absenteeism Academic engagement
	4 Disparities in Performance in Coursework	Success in classes Accumulating credits (being on track to graduate) Grades, GPA
	5 Disparities in Performance on Tests	Achievement in reading, math, and science Learning growth in reading, math, and science achievement
C Educational Attainment	6 Disparities in On-Time Graduation	On-time graduation
	7 Disparities in Postsecondary Readiness	Enrollment in college, entry into the workforce, enlistment in the military
D Extent of Racial, Ethnic, and Economic Segregation	8 Disparities in Students' Exposure to Racial, Ethnic, and Economic Segregation	Concentration of poverty in schools Racial segregation within and across schools
E Equitable Access to High-Quality Early Learning Programs	9 Disparities in Access to and Participation in High-Quality Pre-K Programs	Availability of licensed pre-K programs Participation in licensed pre-K programs

TABLE S-1 Continued

DOMAIN	INDICATORS	CONSTRUCTS TO MEASURE
F Equitable Access to High-Quality Curricula and Instruction	10 Disparities in Access to Effective Teaching	Teachers' years of experience
		Teachers' credentials, certification
		Racial and ethnic diversity of the teaching force
	11 Disparities in Access to and Enrollment in Rigorous Coursework	Availability and enrollment in advanced, rigorous course work
		Availability and enrollment in advanced placement, international baccalaureate, and dual enrollment programs
		Availability and enrollment in gifted and talented programs
	12 Disparities in Curricular Breadth	Availability and enrollment in coursework in the arts, social sciences, sciences, and technology
	13 Disparities in Access to High-Quality Academic Supports	Access to and participation in formalized systems of tutoring or other types of academic supports, including special education services and services for English learners
G Equitable Access to Supportive School and Classroom Environments	14 Disparities in School Climate	Perceptions of safety, academic support, academically focused culture, and teacher-student trust
	15 Disparities in Nonexclusionary Discipline Practices	Out-of-school suspensions and expulsions
	16 Disparities in Nonacademic Supports for Student Success	Supports for emotional, behavioral, mental, and physical health

TO:	National Assessment Governing Board Members
FROM:	Governing Board Staff
DATE:	May 3, 2021
RE:	Status Report and Next Steps for NAEP Reading Framework update

This memo repeats updates sent to Board members on April 23, 2021.

Background

The Governing Board has been undertaking updates to the NAEP Reading Framework for nearly two years. This process started with a unanimous March 2019 Board action to initiate an update to the framework, which provided specific guidance to a panel of subject-matter experts and NAEP stakeholders. An initial draft of the proposed framework update was released for public comment in June 2020. Using extensive public comment, engagement with stakeholders, and several Board policy discussions on the framework, the draft was revised and submitted for discussion at the March 2021 Board meeting.

Current Status

Several points have emerged in the Board's continued discussions on the Reading Framework update:

1) Board members are all committed to maintaining the level of quality and rigor that has long characterized the NAEP assessments.

2) While the initial charge to the panel may not have explicitly asked them to prioritize the stable reporting of trend, the Board is now prioritizing trend.

3) Consensus is emerging among Board members on several aspects of the proposed framework update (see Table 1 at the end of this memo). This includes proposed updates in the following areas:
   a. Comprehension Targets
   b. Disciplinary Contexts
   c. Purposes
   d. Text Types
   e. Text Source
   f. Text Format
   g. Text Complexity
   h. Language Structures and Vocabulary
   i. Reporting

4) There are two important areas where the Board has not reached consensus, which relate to the role of background knowledge (specifically, topical familiarity) in reading comprehension. These include:
   a. **Definition of reading:** There has been some concern that the framework updates change the construct that is being measured. Further Board discussion will attempt to clarify the construct in the existing framework (2004) and the construct as defined in the updated framework.

   b. **Universal Design Elements (UDEs):** Essentially, the question is about the degree of change in the framework update regarding UDEs and how they will be used in the

assessment. The Assessment Development Committee (ADC) is proposing to remove new knowledge-based UDEs that have to do with video, audio, and photo introductions to passages, i.e., not adding new knowledge-based elements, but retaining the ones already on the assessment such as written introductions. (See attached the latest revision of the framework update reflecting this removal.[1])

An optional Board webinar was held on April 30, 2021, to address Board member questions about how the current NAEP Reading Assessment relates to the latest draft of the framework. Presenters included NCES and representatives of the framework's Technical Advisory Committee, and more than half of the webinar allowed for Board members to discuss and ask additional questions.

The April 30 webinar was recorded for members who could not attend, and **the Executive Committee requests that all Board members either attend live or view the recording before the May 13 Reading Framework plenary session.**

Documentation in Response to Questions

Given some of the questions that have been raised so far, staff are providing the attached materials:

1) *Revised Draft NAEP Reading Framework Update* (updated 4/21/21), which appears in track changes. It reflects the removal of multimedia knowledge-based Universal Design Elements and clarifying edits relative to the draft reviewed at the March Board meeting, e.g., reducing academic language, removing redundancies, and improving graphics. (Please contact Michelle Blair to receive a copy that shows in track changes *all edits* to the framework since the public comment version was released last June.)

2) *NCES overview of current NAEP Reading Assessment* (updated since March 2021 meeting), which now includes additional information about the NAEP Reading Assessment over time (from paper to digital) and revised assessment development plans to further bolster the likelihood of maintaining trend.

3) *Redacted NCES report on evaluation of universal design elements in Scenario-Based Tasks (SBTs)* (new resource; embargoed); this research has been referred to in several Board discussions, and key findings are summarized in the *NCES overview of the current NAEP Reading Assessment* (listed above). As secure material, ***this report can only be discussed in closed session***. It was provided under separate cover to Board members.

4) *NCES Questions and Answers on the NAEP Reading Framework update* (new resource), which responds to questions asked by several Board members.

5) *Guidance from Framework Technical Advisory Committee* (new resource), which was provided to the Framework Development Panel after the March 2021 Board meeting on key questions regarding topic knowledge in reading comprehension assessment constructs and Universal Design Elements.

---

[1] Multi-media introductions are removed from the framework update and will be proposed as an area for special study. Special studies are listed as part of the Assessment and Item Specifications, a separate document that is reviewed after a framework is adopted and specifically written for NCES implementation purposes.

6) *Results from a survey conducted by CCSSO* (new resource), which gathered more detailed information on states' approaches to topic knowledge.

Next Steps

The Executive Committee would like to enable full and productive deliberations on the Reading Framework at the May 2021 Quarterly Board Meeting. So, it is critical that all Board members review the above materials and raise any questions with ADC or staff so that all issues can be addressed to the extent possible.

TABLE 1: Similarities and Differences Between the 2009–2019 and 2026 NAEP Reading Frameworks

Green rows are areas of agreement, i.e., no remaining concerns have been raised by Board members.

Yellow rows are areas under discussion.

	Current Framework and Assessment	2026 Framework Update
Definition	Reading is an active and complex process that involves: ● Understanding written text. ● Developing and interpreting meaning. ● Using meaning as appropriate to type of text, purpose, and situation.	Reading comprehension is making meaning with text, a complex cognitive process shaped by students' social and cultural influences. To comprehend, readers: ● Engage with text in print and multimodal forms; ● Employ personal resources that include foundational reading skills, language, knowledge, and motivations; ● Extract, construct, integrate, critique, and apply meaning in activities across a range of contexts.
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

	Current Framework and Assessment	2026 Framework Update
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, 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

	Current Framework and Assessment	2026 Framework Update

		• or experts, or serve as an audience
		• Knowledge-based UDEs
		   – Text providing brief topic previews
		   – Pop-up notes for definitions of words or phrases
		   – Resetting by providing correct response to answered questions
		
		[Differences compared with current framework/assessment are listed in **bold** above; all others are already part of the assessment]

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 • 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

Reading Framework for the 2026 National Assessment of Educational Progress

*** *April 21, 2021 Draft* ***

National Assessment Governing Board
U.S. Department of Education

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*
Associate 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

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

Fen Chou
Program Director, Standards, Assessment,
and Accountability

Scott Norton
Deputy Executive Director, Programs

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.

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. After the tutorial, students engage with two assessment blocks, each including one or more texts and approximately 10 questions. Texts may include images, graphics, or even a short video, and assessment items include both selected response and constructed response formats. The digital platform allows for a greater variety of formats, including selecting key words or sentences in a passage, dragging and dropping responses to complete a sequence or chart, completing a matrix or grid, and selecting more than one correct response. Hybrid items combine selected and constructed responses.

When students finish answering assessment questions, they participate in a digital survey, answering both general and reading-related questions. Student surveys collect demographic data and students' perceptions about access to technology and their reading habits and experiences in school, home, and the community. Together, the assessment blocks and survey take roughly 90 minutes. Teachers and administrators also complete surveys. Data collected as students navigate the digital assessment can provide valuable information about how students process texts and information during the assessment. For example, process data can reveal the time students take to read texts and respond to questions, how often they return to the text as they answer questions, and their use of optional digital tools.

While maintaining the essential structure and purpose of previous paper-and-pencil assessments, the development and implementation of digitally based assessments is key in maintaining NAEP's position as a leader in large-scale assessment.

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 address advances in research in reading. 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 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 ~~the~~ 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). 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:

(1) 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

(2) 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).

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 and fairness 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 guidelines, 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

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~~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 ~~watershed Rand~~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.

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 ~~imbued with~~that involves socially and culturally specific characteristics and practices~~.~~ (Bronfenbrenner & Morris, 2006; Lee, 2016b, 2020; National Academy of Sciences, 2018; Zelazo, 2013).

Drawing from previous frameworks and these research understandings, this updated NAEP Reading Framework attends to four key features of 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 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 students' social and cultural influences. 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 contexts.

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 communication practices and the application of reading to tasks that address wide-ranging purposes and contexts.

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 address the sociocultural aspects of the cognitive processes known as 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, motivations, prior experiences, agency, opportunities to learn); and

- Increase the precision of inferences about student reading achievement in the U.S.

Overview of the Updated NAEP Reading Framework's Key Components

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 scientific research in human development and learning, including reading comprehension. 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. 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. ~~The 2026 NAEP Reading Assessment invites students to read texts and respond 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

6

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 contexts. Texts, too, are sampled to address purposes within disciplines, affordances offered by digital and multimodal formats, and text complexity criteria for each tested grade. 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.

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 to emphasize equity, rigor, precision, and validity. 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 (1) socioeconomic status within race/ethnicity and (2) former English (ELs) learners, in addition to current ELs and non-ELs, in order to describe student performance in more precise and detailed ways.

The framework also proposes to measure an expanded set of contextual variables, as is current practice, via student, teacher, and administrator questionnaires and by expanded the increased use of digital process data to provide further precision and explanation ofmore 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.

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 to explicitly acknowledge the sociocognitive processes of reading. Reading comprehension is defined as making meaning with text and four key features are highlighted—contexts, readers, texts, and activities.

- Emphasis on three additional, research-based concepts: (1) how social and cultural experiences shape learning and development; (2) how reading varies across disciplines; and (3) 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	Digital texts ● Static – non-moving print, graphics, or images on screen

	Current Framework and Assessment	2026 Framework Update
	● Dynamic – navigation across modes (print, video, other) or nonlinear locations (hypertext link)	● Expanded use of dynamic formats – navigation across modes (print, video, 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
		  – Student exemplars as mentor texts
		  – 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, videos, or photographs providing brief topic previews
		  – Pop-up notes for definitions of words or phrases
		  – 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)	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	Disaggregation by all existing categories, adding ● Disciplinary contexts ● Socioeconomic status within race/ethnicity ● Former English learners (ELs) as well as current ELs and non-ELs
	Data collected from student, teacher, and administrator questionnaires on contextual variables of interest	Data collected from student, teacher, and administrator questionnaires on expanded set of contextual variables
	Some data collected from students' test taking behaviors (process data) in digital administrations	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.

The 2026 NAEP Reading Framework recommends updates necessary to deliver assessments that are relevant, fair, 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, ~~traces~~presents the definition's origins in policy and scholarship on reading comprehension, and ~~culminates in~~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 students' social and cultural experiences. 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.

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 ~~and situated within particular settings and practices~~ (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 read by themselves and with others; silently or orally; and lightly for a general impression or closely to prepare for a debate.

  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 ~~frameworks of~~ knowledge ~~(e.g., a birthday party)~~ 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.

Texts. Texts are artifacts generated by authors to communicate their ideas. 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 ~~of~~ 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*, 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.

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~~construction integration (C-I) ~~Models~~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 explicitly cited in the 2009-2019 Framework. Related to the features in the 2026 Definition of Reading Comprehension, the RAND model posited that reader, text, and activity reside in a sociocultural context, describing how "the identities and capacities of readers, the texts that are available and valued, and the activities in which readers are engaged with those texts are all influenced by, and in some cases determined by, the sociocultural context" (pp. 11-12). The 2009-2019 Framework also introduced the centrality of "using meaning as appropriate to type of text, purpose, and situation" (p. 3). The 2026 NAEP Reading Assessment will continue NAEP's longstanding focus on reading comprehension, rather than foundational skills or writing.

Updating the NAEP Reading Framework

The 2026 NAEP Reading Framework is updated to reflect three research-based developments that help to ensure that the NAEP Reading Assessment is a precise, fair, and accurate measure of reading comprehension. The first is how sociocultural experiences shape learning and development, including the learning and development of reading comprehension and, consequently, its assessment. The second is how reading varies across disciplines. The third ~~development is~~regards the ~~increasing~~ use of digital and multimodal texts.

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 ~~learners'~~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)

This perspective builds on longstanding understandings from scholarship in psychology and education. Over 30 years ago, Cronbach (1990) predicted that the psychology of individuals would have to take into account the highly contextualized framing of learning implied by Bronfrenbrenner's (1979) ecological approach. He noted that to fully understand individual development, psychologists and educators would have to engage in systematic analysis of the interactions among the attributes of students and the characteristics of the settings in which their learning is fostered and assessed. For many engaged in the study of assessment, a perspective that accounts for contextual facets of the assessment space is needed to assess more complex constructs. One of these complex constructs is reading comprehension, which can be assessed with greater relevance, precision, fairness, and validity by better reflecting contemporary understandings about the nature of the process.

A second update in the 2026 NAEP Reading Framework is the recognition of 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 account more precisely for how well U.S. students comprehend what they read in texts and situations that more closely approximate reading practices in today's society. By building on past frameworks and research traditions while embracing more recent developments in assessment, NAEP ~~honors its mission of~~will continue to both ~~leading~~lead and ~~reflecting~~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. ~~Five~~To that end, five additional ~~sets of new or~~ updated ~~assessment~~ components are ~~introduced~~also presented: disciplinary contexts, purposes, texts, universal design elements, 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~~. It also allows~~ and the natural variation that readers encounter in reading in home, school, community, and workplace settings. In this way, NAEP ~~assessment to account~~accounts for a wide range of factors that influence reading comprehension~~,~~ ~~mitigating~~ and mitigates potential bias that might result from a narrower operationalization of reading comprehension. ~~That is, building planned variation into every facet of the assessment provides opportunities for readers with varied backgrounds to find connections to their knowledge and experiences. Although it continues to be the case that students read the same texts and complete the same tasks and that their responses are evaluated in the same way, these assessment components help to create a more equitable standardized assessment.~~

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 these texts. Comprehension Targets are used in NAEP to generate the questions, ~~or test~~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. ~~One~~An additional target, *Use and Apply,* ~~is an update that reflects~~has been added to reflect the importance of applying comprehension to new situations.

~~Although different, the~~Each comprehension ~~targets involve~~target involves inferences that readers tend to find more or less challenging in general~~, items~~. 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 ~~the~~each comprehension ~~targets~~target.

Locate and Recall. The first ~~set of~~Comprehension ~~Targets~~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 ~~targets~~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 ~~set of~~Comprehension ~~Targets~~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. The readerReaders 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 set of Comprehension TargetsTarget, 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, motivations, and perceptions of self. 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 every aspect of readers' engagement with text and that influence how readers respond to and learn from the experience of reading. As a result of this principle, the 2026 NAEP Reading Assessment contextualizes almost every component of reading comprehension. 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, authors 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 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 ~~forays into~~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 Design 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 for their use.. 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 ~~should~~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 ~~they~~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. ~~On line~~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 SBAC̶Smarter 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 (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—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.

Task-based UDEs. Task-based UDEs are designed to clarify requirements and guide readers in their use of available resources; they.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 ~~or mentor texts~~ 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 ~~, not readers' comprehension skill, but instead their~~ 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 assessment will be better able to use the processes described in the comprehension targets as they read and respond to questions. For instance, in comprehending a text called Patagonia Glaciers, a reader who happens to have knowledge about glaciers is likely to be better able to successfully answer the comprehension questions than one who might be a skilled reader but has no relevant topic knowledge. Knowledge-based UDEs for the 2026 NAEP Reading assessment expand the use of brief passage introductions that offer topic previews ~~in the form of brief text, videos, or photographs.~~ . The 2026 assessment continues using vocabulary pop-ups to offer on-demand definitions of untested vocabulary. Such knowledge-based UDEs, will help to address this long-standing potential source of bias in assessment, resulting in more accurate measurement of text comprehension across readers (Johnston, 1981). Within the 2026 NAEP Reading Assessment, knowledge-based UDEs, as well as task-based and motivational UDEs, are not necessarily a part of every assessment block. Additionally, many current NAEP blocks do not include knowledge-based UDEs, and tasks that do not include knowledge-based UDEs will continue to be a part of the 2026 NAEP Reading Assessment.

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. 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 ~~all readers possess the same~~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.

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 achievement to reflect these data, collectively called contextual variables. These include race/ethnicity, English language proficiency, socio-economic status, and region of the country~~, and, for special NAEP initiatives, large cities and districts.~~.. 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' habits and attitudes (e.g., How much do students like school~~; how~~? How often do they read for pleasure, go to the library, and/or read or write on a digital device?).

NAEP collects data to gain insight into contextual variables 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' perceptions of the texts they read, their 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 insights into the knowledge, interest, motivation, engagement, habits, attitudes, language competence, skills, and strategies that students bring to their reading comprehension. 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. ~~Therefore, additional information is important for contextualizing and better understanding the circumstances in which low-performing readers learn.~~

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 explanatory variables associated with reading performance more visible. Instead of portraying student groups as unitary and homogeneous, this approach will yield a more nuanced and complete measure to better understand reading disparities as the result of a ~~complexity~~complex of factors. (For more information about ~~reporting~~how contextual variables are reported, see Chapter 4.)

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 differencedifferences in a richer context (Guthrie & Humenick, 2004; Guthrie & Klauda, 2015).

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, SESsocioeconomic 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 performancesperformance.

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.

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" or peers, 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.
~~*Disciplinary Contexts Purposes*~~				

28

	Features of the NAEP Definition of Reading Comprehension			
Assessment Components	Contexts	Readers	Texts	Activities
Texts *y*	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.
~~Disciplinary Texts~~ *~~Digital Texts~~* ~~Text Complexity~~				
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	Gather information	Gather information about	Gather information	Gather information

	Features of the NAEP Definition of Reading Comprehension			
Assessment Components	Contexts	Readers	Texts	Activities
Questionnaire Items	about the contexts of readers' lives and experiences in and out of school.	demographics, motivation, and in- and out-of-school reading practices.	about the amount and kinds of texts that readers encounter in and out of school settings.	about reading activities that readers commonly engage in at school and outside of school.
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 ~~using~~leveraging digital assessment features, including item response formats, Universal Design Elements (UDEs), and process data, ~~and item formats~~ in line with principles of validity, fairness, and inclusivity (Thompson, Johnstone, & Thurlow, 2002). 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.

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.

Designating Disciplinary Context

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. 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.

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.

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. (see Appendix C). 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 parts of the story and respond to items situated in threetwo purpose-driven tasks. Because test-takers encounter additional texts and items in different parts of the block, more specific purposes may be given to situate their work on particular comprehension items in the context of each new text. Note, in this example, each additional text is an excerpt from the same story.

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**

Introduction

You will read a book called *Hana Hashimoto, Sixth Violin* by Chieri Uegaki. After you read the story, you will participate in a small discussion group with three classmates to understand how Hana changes over the course of the story as a result of events involving her family.

To prepare for the book discussion, you will read the story while completing three tasks.

1. Identify important events in the story and consider what these events say about the characters.
2. Learn more about Hana and other important characters from their words, feelings, and actions in the story.
3. Use your understanding of the characters to predict what might happen after the story ends.

Your teacher for this project will be Mr. Beadle

You will work with three classmates in your discussion group: Brian, Diana, and Michael

Welcome

You will read the story, *Hana Hashimoto, Sixth Violin*, by Chieri Uegaki and Qin Leng to prepare for a book discussion.

First, you will learn about **important events** in the story and **characters' thoughts, feelings, and actions.**

Then, you will **write about what the main character, Hana, is like as a person so that you are ready to discuss the book** with three peers.

Your teacher for this project will be Mr. Obas:

You will work with three classmates in your discussion group:

Gia   Gabe   Luisa

NEXT

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.

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 ~~coherence~~cohesion*. Are the texts well-written and ~~considerate,~~ 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 See Exhibit ~~1~~4 in Appendix A. 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).

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.

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.

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 25 in Appendix A shows important textualtext 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, coherencecohesion, 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 common text structures and text features that are found in common across disciplinary contexts (see Exhibit 36 in Appendix A). 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.

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 ~~complex textual~~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.

Developing Comprehension Items

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 4 in Appendix A):~~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 ~~textual~~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 ~~texts will~~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 ~~ask~~be answerable by readers ~~to draw~~only drawing upon text-independent domain knowledge, topic knowledge, knowledge of technical vocabulary or idiomatic expressions, or conceptual or domain knowledge in particular subject areas. Knowledge-based UDEs are ~~therefore~~ incorporated into given blocks to maximize students' ability to engage with the content that is being tested. Thus, knowledge-based UDEs are designed to reduce the noise associated with knowledge sources not being assessed in a given block and also provide orientations to the topical knowledge addressed in the text(s).

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.

~~Exhibit 4 in Appendix A~~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.

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 (see Exhibit 58 in Appendix A). 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 language known to be part of students' everyday oral proficiency, rare words of limited application across grade-appropriate texts, discipline-specific concepts, 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

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 accurate interpretations of test results, thereby increasing the ~~potential~~ validity of scores across the diversity of test takers (c.f., Mislevy, 2016; Thompson et al., 2002).

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 valid 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.

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

41

	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 unfairly advantage these students, resulting 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).
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" (see page 6)."

Exhibit 3.6 Alignment of the 2026 NAEP Reading Assessment ~~With~~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 ~~limited~~ 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 ~~(e.g., a term not assumed to be possessed by *all* students).~~. 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 that is easily understandable regardless of a student's background (e.g., experience, knowledge, language use, interest) 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 be presumed to be able to read and understand 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."

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. 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 and; 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

Chicken Forum Released Item #3

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 ~~two~~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 ~~SBACS~~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. 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). 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 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 NAEP Reading Assessment. Additional details are provided in the item specifications. Some examples of UDEs are presented in the next sections. 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 ~~(in this case, a word bank) is provided for,~~ all students ~~to enable them to select from an assortment of character traits and select the one most in line with~~ have an equivalent opportunity to focus more of their ~~reasoning about the main character based on her actions in the story. More~~ time and attention on the use and apply construct to be measured, rather than ~~one~~ on trying to generate a character trait word ~~choice could be an acceptable answer, but some selections are better than others, and the appropriateness of any word is linked to the reader's ability to provide a reasonable justification for their choice.~~. 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 ~~new~~ item in the relatively short period of time allotted by the assessment. ~~Such~~This clarity of expectations also maximizes the likelihood that readers ~~are able to~~will cognitively engage with complex NAEP-designed reading experiences within the short time frame allotted to each block.

   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 ~~is designed to~~

decreasedecreases construct-irrelevance by providing a set of words from which test-takers can *select*, rather than *generate*, a relevant character trait. That is, theThe 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.

Exhibit 3.11. A Grade 4 IntegrateUse and InterpretApply 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 it as part of their constructed response.

In the story, the author writes, "Hana swallowed her nerves like medicine and leaned toward the microphone. She would just do her best."

What do you think the author is trying to tell the reader about what kind of person Hana is? Choose a character trait from the word bank to describe Hana's personality. Then, use the box to explain what kind of person Hana is and what makes you think this.

WORD BANK	
helpful	curious
brave	proud
smart	nervous
afraid	confident
forgetful	determined

Great job! Now you will use what you have learned about Hana to write about **what Hana is like as a person** so you are ready to discuss with your peers. **Use your chart to help you.**

Part A. Select a **character trait** from the word bank that best describes Hana.

WORD BANK	
helpful	curious
brave	proud
smart	nervous
afraid	confident
forgetful	determined

Hana Hashimoto Story

Completed Chart

Part B. Explain how Hana can be described using the character trait you selected in Part A. Be sure to use evidence from your chart about **Hana's thoughts, feelings, and actions.**

DONE

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 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 3015 second video of children playing the violin string instruments serves to pique students' interest in the topic of athe reading passage. The passage is about a girl who learnsenters a talent show contest to perform the violin she has just learned how to play the violin. 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 3015 second video clip of students playing stringed instruments for the Grade 4 textshort 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, non-biased items" as well as "precisely defined constructs" (Thompson et

al., 2002, p. 10) by stimulating prior interest and motivation and thus removing some construct-irrelevant variance for students who might come to an assessment task with no prior interest in the topic or activity that is the focus of the assessment itemblock.

Motivational UDEs may also 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).

Exhibit 3.13. Teacher and student task characters remind the reader of the task goal <u>for the second task</u>.

As you work in your literature circles, remember the goal is to think about how Hannah grows and changes over the course of the story.

Ok, so how do you think we should get started?

I collected all of our notes about Hana into one notepad.

Thanks Diana! That definitely helps to have all of the ideas in one place.

OUR NOTES

Note 1	In the beginning, Hana's brothers made fun of her.
Note 2	But, Hana kept practicing anyways.
Note 3	When she stepped on stage, Hana was nervous.
Note 4	Hana remembered her Ojiichan and played what she knew.
Note 5	Hana's family loved it, and Hana decided to play again next year.

Which sentences best describe Hana's personality at the beginning, middle, and end of the story. Click and drag your answers from the notepad into the appropriate boxes below.

Beginning   Middle   End

Now, to prepare for the discussion, you will write about **what Hana is like as a person.**

Here are some of my notes about **Hana. Can you add some more?** Be sure to use specific details from the story about her thoughts, feelings, and actions.

Hana Hashimoto, Sixth Violin
By Chieri Uegaki & Qin Leng

Perhaps next year Hana would be able to perform one of Ojiichan's favorite pieces. But for now, Hana played a little melody she had been practicing, remembered from night lit by dancing fireflies. She imagined that the notes would drift out through the window, past the bright rabbit moon and beyond, and Ojiichan would hear them and smile.

OUR NOTES

Note 1: Hana's brothers made fun of her. She practiced anyway. The text says, "Hana practiced every day."

Note 2: When Hana gets on stage, she is feeling nervous. The texts says, "Hana swallowed her nerves like medicine."

Note 3:

Note 4:

NEXT

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 short videos, images,written texts only, unless video, image, or a previewother kinds of specific concepts addressed in theintroductions 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 particular readers. DeterminationA 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 translation,translations—and simulations of interesting or challenging phenomena) can provide visual and multimedia cues to support readers' understanding of unfamiliar vocabulary or challenging concepts.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 ~~Assessed and Not~~to Be Assessed in the 2026 NAEP Reading Assessment

Knowledge ~~Not Intentionally~~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
~~Knowledge of:~~ • ~~text structures (descriptive, causal, compare and contrast, problem-solution, etc.)~~ • ~~vocabulary and language structures~~ • ~~genres and rhetorical structures~~ • ~~authors' craft~~ ~~That enables students to demonstrate their ability to:~~ ~~use~~Students' Ability to: • Recall specific text information • Use text features to derive meaning • ~~discern authors' rhetorical strategies and purposes~~ • ~~draw~~Draw inferences based on information in text • ~~synthesize information across text or multiple texts~~ • ~~analyze~~Integrate information within and across texts • ~~critically evaluate~~Analyze information presented in text • Analyze authors' rhetorical strategies and purposes • Evaluate sources of information in text • ~~use~~Use and apply ~~knowledge~~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 ~~why~~how Hana's brothers ~~flee~~first respond to her decision to play the ~~house every day~~violin in the talent show, based on ~~other character's words and~~ their actions and words (Hana's brothers ~~cover their ears~~"nearly fell out of a tree" and ~~complain about the "horrible noise" from Hana's violin practicing).~~they tell her, "you'll be a disaster!"). Since the whole story is 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. 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 ~~helps to~~like this one orient readers to the topic of the text ~~in an effort to~~, 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 ~~inference~~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.

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),

- 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 capacity of NAEP results.

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, and disability status (, i.e., supported by an individualized educational program), and English learner status.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 equity, rigor, precision, and 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 factors that can expand 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.

To achieve broader equity goals— with particular attention to providing within an integrated system that provides more nuanced reports and useful data to key stakeholders on research-based contextual variables focused on opportunities to learn—. 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. Reconceptualize reporting and contextual variables as an integrated system to explain student performance in ways that make the data collected more useful for policy makers and educators. Provide information on research-based contextual variables (derived from demographic, questionnaire, and process data) focused on opportunities to learn.

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. 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. 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.

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 and explanatory capacity of NAEP through reporting by disciplinary contexts; disaggregating results within demographic categories; and expanding reporting categories for English learners.

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, as measured by student eligibility for the National School Lunch Program; 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.

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 as it does in the U.S. 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. 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 (e.g., high quality teaching, rigorous curriculum, community-based institutional structures such as libraries) may serve as an underlying driver of achievement.

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.

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. ~~To the extent possible, information is also gathered from non-NAEP sources such as state, district, or school records to minimize the burden on those asked to complete the 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* (e.g., students' self-perceptions about engagement and motivation, knowledge, self-efficacy, agency, effort, and interest) and *environmental characteristics* (students' perceptions about facets of home, community, or school settings, including their perceptions about classrooms, sense of belonging, and support).

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 ~~an~~the outcome of interest, ~~such as~~ reading comprehension scores on NAEP. The 2026 NAEP Reading Framework's emphasis on the cultural assets of individuals and 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 taking into account students' differential engagement with reading and their access to home and community resources such as libraries, tutoring, and out-of-school programs, the expanded contextual variable data are intended to help contextualize and explain students' differential performance on the NAEP Reading Assessment.

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, ~~2016~~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). ~~From the multitude of potential contextual variables, the 2026 NAEP Reading Framework expands the questionnaires by adding a manageable, selected set of research-based and malleable factors. By taking into account students' differential engagement with reading and their access to home and community resources such as libraries, tutoring, and out-of-school programs, the expanded contextual variable data may help contextualize and explain students' differential performance on the NAEP Reading Assessment, and thereby support policymakers and stakeholders in identifying potential means to shift policy and education practice to better serve our nation's students. Guided by the latest research, the 2026 NAEP Reading Assessment includes contextual variables that seek to capture both reader characteristics and environmental characteristics.~~

With the aim of supporting policymakers and stakeholders to understand student performance and craft effective policy and education practice, 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) predict performance on the outcome measure of reading comprehension; 2) be malleable (that is, it can be influenced in school and community settings); and 3) avoid unnecessary or inappropriate intrusions into the lives of students' families. 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). 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). 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, sense of self-efficacy, 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

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 (OTL), 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:

Perceptions of School and Community Resources

1. **School social support** refers to the extent to which students perceive that their teachers and peers believe they 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 perceive themselves to be 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 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).

Perceptions of Teacher, Instructional, and Classroom Supports

1. **Teacher support for reading engagement** refers to the extent to which students perceive their teacher(s) as providing 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 perceive their teacher(s) to support their interests, self-efficacy, and reading goals (Wigfield & Wentzel, 2007).

3. **Teacher support for students' background experiences** refers to the students' perceptions 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 perceive 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 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. Consideration of such factors is consistent with research on the importance of social and emotional well-being to learning (~~Damasio 1995; Markus and Kitayama 1991; Weiner 1985~~Durlak et al., 2015; Elias, 2019; Guthrie~~, J. T.,~~ & 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~~, Roderick~~ et al~~.,~~., 2012), and approaches that build on and engage students' out-of-school identities and interests to make learning meaningful and relevant (Katz~~, Brynelson & Edlund,~~ et al., 2019; Shin~~, Daly & Vera,~~ et al., 2007).

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.

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

Enhancing NAEP's Explanatory Reporting Capacity

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. 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 as a function of malleable factors, either for reader characteristics (e.g., student motivation) or environmental characteristics (e.g., opportunity to learn factors), 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. Implementing the changes summarized below can mitigate potential misinterpretations and increase the usefulness of NAEP data.

1. **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.

2. **Revise Questionnaires**. To increase the capacity to examine the impact of contextual variables related to readers and their environments, 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).

3. **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.

4. **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.

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

6. **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,

statistically reliable reports beyond the standard ones offered by the Nation's Report Card.

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.

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.

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. 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.

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.

~~Glossary terms placeholder~~

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.

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

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).

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.

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

For All Grade Levels

~~Exhibit 1.~~

Principle: The percentage of different text formats (static or dynamic) and modalities (print, sound, image, and multimodal) should reflect their distribution in the population of texts that students encounter in and out of school at different grade levels.
• As dynamic and multimodal texts increase in our society and schools, NAEP should aim to keep pace with those shifts.
 • Current NAEP: 80% print, 20% other modalities

Exhibit 1 provides guidance to developers about sampling different kinds of texts (where texts include multimodal forms of representation). The underlying assumption in the exhibit is that there exists a continuum of forms of representation. That continuum is bounded at the one end by more static, print texts and at the other end by a complex and variable range of text types, features, and purposes. The exhibit provides advice about sampling for the present (80/20 static/dynamic and multimodal) and the future (to reflect the distributions in school and society).

Exhibit 2. Illustrative Examples of Texts and Other Media Across Single Static and Dynamic Texts and Multilayered Digital Text Environments

SINGLE STATIC TEXT	SINGLE DYNAMIC TEXT
Examples of single static genres and forms of continuous prose, non-continuous prose, and everyday reading materials from which designers might sample as readers read to engage in literature, science, or social studies and history are found in Exhibit 2 in this appendix.	*Nonlinear text* Single text with hyperlinks that only connect to ideas within the same document; may also contain one or more dynamic media elements *Dynamic media* • Dynamic image • Video • Podcast • Digital poster • Infographic • Interactive timeline • Interactive chart or graph • Data visualization • Blog • Simulation

MULTILAYERED DIGITAL TEXT ENVIRONMENT

- Augmented reality text
- Blog
- Database
- Digital creation/composition tool
- Dynamic simulation
- Email
- Interactive model

- Google document or Google folder
- Role play simulation
- Search engine
- Social media (e.g., Facebook, Instagram, Twitter)
- Threaded discussion
- Webpage or website

Exhibit 2 provides examples of the types of texts/media that designers should consider for the three text environments (single static, single dynamic, and multilayered digital) in NAEP blocks.

Exhibit 3. Commissioned Texts: Parameters and Constraints

Guidelines for Using Commissioned Texts
The following guidelines seek to provide clarity about the circumstances under which commissioned texts might be used and the criteria with which developers should use such commissioned texts: • Rare, never to exceed more than 5-10% of all texts included in NAEP at any grade level; 5% limit at 12th grade unless permission issues are encountered • Only used when an appropriate authentic text cannot be located to include within a text set for a block, but never as an "anchor" text for a block • Authored by writers within the discipline in which the block is situated and using specific criteria to meet strict guidance re: form and purpose • Vetted for accuracy, authenticity, and appropriateness by experts in the discipline, NCES's text selection panel, and the ADC • No items asking students to evaluate source credibility of such commissioned texts will be used • Will meet the same complexity and other criteria for text selection as all texts for NAEP Reading

Exhibit 3 summarizes the guidelines that developers will use to determine if, when, and how texts will be commissioned to meet particular needs that cannot be met by sampling already published (i.e., authentic) texts.

Exhibit 4. Passage Lengths for Grades 4, 8, and 12

Grade	Range of Passage Lengths (Number of Words)
4	200-800
8	400-1,000
12	500-1,500

Exhibit 2.Exhibit 4 provides ranges for the total number of words in the text(s) within a given block. This total might be distributed across 1-4 texts depending on the broad purpose (Reading to Develop Understanding or Reading to Solve a Problem) of a block.

<strikethrough>Exhibit 5.</strikethrough> **Exhibit 5. Typical Text Elements Across Disciplinary Contexts**

Context	Genres and Text Types	Discourse, Language Structures, and Text Elements
Literature	**Fiction** **(Short stories, novels, plays)** • Myths, legends, and fables • ~~Short stories~~ • Coming of age stories • ~~Novels~~ • ~~Dramas~~ • ~~Poetic traditions~~ • Satires • Science fiction • ~~Satires~~ • Magical realism • ~~Biographies~~ • Fantasy • ~~Memoirs~~ • Comic books • Graphic novels • Manga • Fanfiction **Poetry** • Haiku, sonnet, ballad, dirge, epic, etc. **Related Nonfiction** • Memoirs • (Auto)biographies • Literary analyses • ~~Literature reviews~~Reviews and recommendations • Author profiles ~~and biographies~~	• Plot ~~and~~ types • Character types • Narrative elements (character ~~structures~~, setting, plot, conflict, rising action, climax, resolution) • Figurative language (symbolism, imagery, simile, metaphor, personification, satire) • Point of view • ~~Dialogue~~ • Theme • Soliloquy, dialogue, and monologue • Diction~~ and~~, word choice • Repetition, exaggeration • ~~Exaggeration~~ • ~~Theme and message~~ • Flashback • Foreshadowing • Mood, tone, irony, paradox, and sarcasm • Visual and graphical elements such as illustrations and photographs • Multimodal elements such as narrative soundscapes • Description~~, exposition~~ • Narrative and ~~narrative elements and~~expository text structures
Science	• ~~Reports~~ • Science reports • Press releases • ~~News briefs~~ • Science news and features • Science magazine articles • Reference materials and field guides • Discovery narratives~~, biographies,~~	• Linguistic frames and signals for organizing arguments, comparisons, sequences and/or causal chains • Abstraction and nominalization (e.g., use of technical terms like transpiration to represent a sequence of events in an explanation~~ sequence))~~

	- Biographies and first-person accounts - Blogs and other forms of public engagement in science - Science websites, such as those of universities, federal and state agencies, formal research groups, hospitals, etc. - Raw data - Bench notes and science journals - JournalProcedures - Published research articles - Personal communications	- Epistemological qualificationEmbedded definitions (science specific words explained in the text) - Science-specific definitions for polysemous words (e.g., heat, energy) - Qualification of claims: may, probably, indicates, suggests, etc. - Spatial (place, location) and temporal indicators (era, time, sequence, and tense) - Linguistic and numeric indicators of magnitude and scale - Visual and graphical elements such as charts, tables, graphs, equations, diagrams, schematics, models, photographs, digital scans and images - Multimodal elements such as simulations orsimulation, time lapse photography and animations
Social Studies	- Primary, secondary, - Historical and tertiary text traditions (mainly in history) - Primary:contemporary documents such as newspaper articles, editorials, political cartoons, broadsides, blogs, census data, diaries, letters, speeches, inventories and records of sale, advertisements, archival documents, cultural artifacts - Secondary: interpretive - Biographies and autobiographies - Historical and contemporary photographs and video - Data (tables, charts, graphs, infographics) conveying information such as demographic, employment and education levels, voter registration and turnout statistics, Gross Domestic Product and other economic measurements, etc.	- Linguistic frames and signals for organizing arguments, comparisons, and/or causal chains - Lexical expressions that mark chronology or argument - Abstraction and nominalization (e.g., to develop a chain of reasonings across events and happenings, e.g., this stance of brinkmanship...) - Rhetorical markers of persuasion - Lexical expressions that mark chronology or argument - Historical expressions and ideologicalterminology - Ideological markers of language and rhetorical devices (word choices, emotional appeals, hyperbole) - Visual and graphical elements such as maps, timelines, political cartoons, photographs - Multimodal elements such as digital stories, procedural texts, public service announcements

• Interpretive explanations ~~of~~or arguments about historical, social, and cultural phenomena and trends. • Procedural texts, public service announcements	• Event models (how historical events are described) • Spatial (place, location) and temporal indicators (era, time, sequence, and tense)

Note: Many text types and elements are common across disciplines. All texts should include information about their sources and authors. In general, NAEP applies a standard of accuracy and trustworthiness to the texts it selects, especially in matters of scientific inquiry. For certain tasks, however, it is necessary to use texts with questionable, or at least different, levels of accuracy and trustworthiness if the purpose of a block, or a task within a block, is to engage students in analysis and critique of texts. It is even more likely that NAEP will employ texts that represent different perspectives on an issue when students are asked to compare the multiple perspectives that texts/authors bring to a social or scientific issue.

Exhibit 5 provides a list of the text types and elements that test developers will consider as they sample texts within the three disciplinary contexts of literature, science, and social studies. Examples are provided for both broad organizational structures (genre and text type) and highly specific features that define the nature and flow of discourse at more specific levels of text (sections, paragraphs, sentences, and even words). While it is impossible in NAEP to represent the entire range, these elements define the portfolio of possibilities that developers will consult in selecting specific texts, making sure that a range of broad organizational structures and specific features are represented in the sample for each discipline and each grade level.

Exhibit 6. (Note: Many text types and elements are common across disciplines.)

Text Structures and Features Within and Across Single Static and DynamicTexts and Multilayered Digital Text Environments

SINGLE STATIC TEXT	SINGLE DYNAMIC TEXT
~~Textual~~Text **structures** are comparable to those in a printed format for texts designed to inform, entertain and/or persuade. ~~Textual~~Text **features** may include visual media elements in a single text comparable to those in a printed format that convey meaning through primarily static words, numbers, and/or visual graphics, such as those in a still photograph, diagram, or table.	~~Textual~~Text **structures** include one or more **nonlinear** elements (e.g., hypermedia or hyperlinks) for readers to quickly move from one location or mode to another, but still *within the same text* (e.g., a navigational menu at the top of a document). ~~Textual~~Text **features** include one or more **multimodal** elements (words, moving images, animations, color, music and sound) embedded into a single text or other media element

~~COMPLEX TEXTUAL~~MULTILAYERED DIGITAL TEXT ENVIRONMENT

~~Text~~In multilayered digital text environments (Cho & Afflerbach, 2017), text **structures** may include one or more static or dynamic texts, with a strong likelihood of nonlinear elements both within a text (e.g., hypermedia or hyperlinks) that may lead to another text (e.g., another webpage within the same website or another webpage on a different website). **Text features** may include linked texts may contain either related or conflicting textual ideas. Multimodal elements (words, moving images, animations, color, music and sound) may appear in any or all texts.

Note: Ideas within each cell are likely to change and expand as new kinds of texts and technologies continue to emerge.

Exhibit 46 describes the possible relationships among important factors in shaping the distribution of texts, especially now that many of the texts within NAEP will bring digital affordances along with those of print texts. It provides an overview for developers about what they should expect in blocks built in accordance with the 2026 NAEP Reading Framework. Ideas within each cell are likely to change and expand as new kinds of texts and technologies continue to emerge.

Exhibit 7. Distribution of Cognitive Comprehension Targets Across Grade Level and ~~Blocks~~Broad Purposes

~~Grade Level and Comprehension Targets~~Rules of Thumb
· The distribution of items for the comprehension targets should be monitored at the pool level (across the two broad purposes—Reading to Develop Understanding and Reading to Solve a Problem) at each grade level · All Comprehension Targets are employed at each grade level. · All Comprehension Targets require students to consult the text in order to select or construct responses. What changes across targets (from Locate and Recall, to Interpret and Integrate, to Analyze and Evaluate, to Use and Apply) is the sophistication of the text-based reasoning and the inferences involved. · Moving across grades, the proportion of higher-level Comprehension Targets increases · RDU blocks, by definition, do not require the application of ideas to a new task. Hence the bulk of Use and Apply items will be in RSP blocks; however, NAEP should be open to the possibility that an RDU block might merit an item based on the Use and Apply Comprehension Target.

Grade	Combined Block Pool: both Reading to Develop Understanding and Reading to Solve a Problem Blocks (% Target Ranges per Block)
Grade 4	
Locate and Recall	15 - 30 ~~40~~%
Integrate and Interpret	10 - 30 ~~40~~%

90

Analyze and Evaluate	10 - 20%
Use and Apply	~~10~~0 - 20%
Grade	**Combined Block Pool: both Reading to Develop Understanding and Reading to Solve a Problem Blocks (% Target Ranges per Block)**
Grade 8	
Locate and Recall	10 - 20%
Integrate and Interpret	20 - 30%
Analyze and Evaluate	20 - 30%
Use and Apply	~~10~~0 - 20%
Grade 12	
Locate and Recall	10 - 20%
Integrate and Interpret	~~20~~25 - 30%
Analyze and Evaluate	~~30 - 40~~25- 35%

Use and Apply	~~200~~ - 30%

Exhibit 7 provides both the principles and ranges anticipated for the distribution of items for each comprehension target within blocks developed for each broad purpose (RDU and RSP) at grades 4, 8, and 12. Because item development is so greatly influenced by the affordances of the texts selected, the ranges for item types will vary from block to block, even within each broad purpose. Hence, as with previous frameworks, NAEP monitors the range of comprehension targets by looking at the total distribution across all of the blocks within a grade level for each disciplinary context.

Exhibit 8. Inclusion and Exclusion Criteria for Connected Language and Vocabulary

Language Structures & Vocabulary Included / Excluded ~~From~~from Testing	Criteria
Included	• Words and language structures that appear across numerous texts, either across literary texts (e.g., *despise, benevolent*) or across social studies and natural sciences texts (e.g., *resolution, commit*) • Words or phrases necessary for understanding at least a local part of the context linked to central ideas in the passage • Words and language structures found in grade-appropriate texts • Words that label generally familiar and broadly understood concepts, even though the words themselves may not be familiar to younger learners (e.g., *timid*). • Words that include word parts (roots and affixes) useful to acquire and figure out the meaning of unfamiliar words (e.g., *disregard, counterargument*).

	• Language that expresses logical relations between ideas (e.g., phrases that include connecting words such as *although, in contrast*) • Expressions that refer to characters, events, or ideas previously introduced in the passage (e.g., *those alliances, this phenomenon*)
Excluded	• Rare words of limited application across grade-appropriate texts and discipline-specific concepts (e.g., fiduciary, photosynthesis) • Idiomatic expressions (e.g., spill the beans, up in the air) • Words and language structures that are already likely to be part of students' oral proficiency at a specific grade level.

*Note: A total of 30 percent of items in any assessment block will assess passage-relevant Language Structures and Vocabulary knowledge while concurrently measuring a specific comprehension process.

Exhibit 6.Exhibit 8 describes the types of words and structures that developers may and may not include when developing the set of vocabulary items for a given block. Vocabulary items are doubly categorized: (a) by the language structures and features in this table; and, (b) by the comprehension targets. In terms of reporting, scores on vocabulary items are aggregated with other comprehension items to create an overall comprehension block score for each student.

Principle and Provisional Distribution Targets for Sampling Assessment Design Elements: Text Formats and Modes

For All Grade Levels
Principle: The percentage of different text formats (static or dynamic) and modalities (print, sound, image, and multimodal) should reflect their distribution in the population of texts that students encounter in and out of school at different grade levels.   • As dynamic and multimodal texts increase in our society and schools, NAEP should aim to keep pace with those shifts.   • Current NAEP: 80% print, 20% other modalities

Exhibit 7. Range of Design Features for Assessment Components with which Students Might Engage in A Block

Assessment Component	More constrained and conventional assessment features		More complex, dynamic, and iterative assessment features
Block	Less involved specific reading purposes that focus students' attention on a theme, question, or problem to be explored during the block (e.g., consider how a character changes throughout a story). Not, all tasks within the block necessarily work directly toward this theme.	⟷	More involved specific reading purposes paired with an essential inquiry question or problem to be examined (e.g., using an author interview, nonfiction texts, and a fiction story based on real issues, consider why an author includes characters with different perspectives despite the author's own perspective on the issue stated during the interview). All tasks within the block will help readers work towards this theme, question, or problem.
Role of readers	Reader is less constrained (assigned less of a role) by specific reading purposes that contextualize expectations for how to engage with provided texts and tasks.	⟷	Reader is more constrained by specific reading purposes and role expectations about how to engage with provided texts and tasks. Readers may be assigned (or choose to take on) particular roles, and their role may be more specified, particularly in relation to reading purpose(s) and expected outcome(s).
Tasks	*Inter-relatedness:* Purpose-driven tasks are situated in line with context norms but tasks are more loosely structured with less probability of readers moving back and forth across tasks; less need for resetting. *Culminating elements:* Less involved culminating task that loosely addresses the question/problem; not a major driver of the block.	⟷	*Inter-relatedness:* Purpose-driven tasks are situated in line with context norms but tasks are more tightly structured so that one task builds on the previous; more probability that tasks are interdependent; more need for resetting. *Culminating elements:* More involved culminating task at the end of an activity that directly addresses the question or problem; major driver of the block.

Assessment Component	More constrained and conventional assessment features		More complex, dynamic, and iterative assessment features
Assessment Component	More constrained and conventional assessment features		More complex, dynamic, and iterative assessment features
Texts	*Number*: 1-3 related texts; excerpts rather than entire texts from some texts may be included rather than in their entirety *Dynamism*: More static texts with minimal dynamic features *Linearity:* Fewer nonlinear structures to navigate within or across texts; less variation in structures across texts *Features*: Texts include a narrower range of features and fewer types of media. *Perspectives:* Less variation in content, purposes, perspectives across texts.	⟷	*Number*: 2-4 interconnected texts (or excerpts from longer texts); readers may be asked to choose only some to engage with in line with task purposes *Dynamism*: More texts with dynamic or multimodal text features *Linearity:* More nonlinear structures to navigate within or across texts; more variation in structures across texts *Features*: Texts include a wider range of features and more types of media *Perspectives:* More variation in content and a wider range of purposes and perspectives across texts.
Universal Design Elements (UDEs)	Less complex reading purposes that may involve UDEs for knowledge or motivation but lesser need for task-based UDEs.	⟷	More complex and inter-related reading purposes that may involve UDEs for knowledge or motivation but greater need for task-based UDEs.

Exhibit 8. Illustrative Examples of Texts and Other Media Across Single Static and Dynamic Texts and Complex Textual Environments

SINGLE STATIC TEXT	SINGLE DYNAMIC TEXT
Examples of single static genres and forms of continuous prose, non-continuous prose, and everyday reading materials from which designers might sample as readers read to engage in literature, science, or social studies and history are found in Exhibit XXX.	*Nonlinear text* Single text with hyperlinks that only connect to ideas within the same document; may also contain one or more dynamic media elements *Dynamic media* • Dynamic image • Video • Podcast • Digital poster • Infographic • Interactive timeline • Interactive chart or graph • Data visualization • Blog • Simulation

COMPLEX TEXTUAL ENVIRONMENT

96

- Augmented reality text
- Blog
- Database
- Digital creation/composition tool
- Dynamic simulation
- Email
- Interactive model

- Google document or Google folder
- Role play simulation
- Search engine
- Social media (e.g., Facebook, Instagram, Twitter)
- Threaded discussion
- Webpage or website

The NAEP Reading achievement level descriptions (ALDs) articulate specific expectations of student performance in reading at grades 4, 8 and 12. Like other subject-specific ALDs, the NAEP Reading ALDs presented in this appendix translate the generic NAEP policy definitions into grade- and subject-specific descriptions of performance.

NAEP Policy Definitions

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

Range ALDs

This Framework presents range ALDs for NAEP Reading. For each achievement level, the corresponding range ALD details observable evidence of student achievement. In many cases, range ALDs also illustrate "changes" in skills across achievement levels, portraying an increasingly sophisticated grasp of the material from one achievement level (and from one grade level) to the next. Achievement levels are also cumulative, meaning each ALD in each grade includes all the reading achievement expectations identified in all the lower achievement levels and grade levels.

Range ALDs should not be confused with reporting ALDs. The fundamental difference between the two is straightforward; range ALDs communicate **expectations**, and reporting ALDs convey **results**. In other words, range ALDs are **conceptually driven**, based on the model of reading and the Assessment Construct in the NAEP framework. They answer the question, given what we know about the development of reading, what should students be able to do at different grade and achievement levels when responding to different combinations of texts and tasks? By contrast, reporting ALDs are **empirically driven**, based on **actual** performance of students who have taken NAEP. They answer the question, given the distribution of NAEP performance, what can students at different grade and achievement levels do when responding to various combinations of texts and tasks?

The ~~2025~~2026 NAEP Reading Framework does not provide reporting ALDs; those are constructed using empirical data during a later stage in the NAEP cycle, i.e., a live administration of the NAEP Reading Assessment. Further detail about the development of the reporting ALDs for NAEP is provided in the Governing Board's policy statement on achievement level setting.

~~**Multiple Disciplinary Contexts for Reading**~~

Organizational Features and Structures of the Reading Construct: Contexts, Purposes, Comprehension Targets, and Text Complexity

The ALDs in this appendix are structured to mirror the presentation of the reading construct provided in the Framework narrative. The primary organizational structure in the Framework narrative is the disciplinary context. Whereas the prior (2009) NAEP Reading Framework identified two reading contexts (literary and informational) this ~~2025~~2026 Framework has identified three (science, social studies, and reading). In the ALDs below, all three disciplinary contexts are described within each performance level.

~~Connections to the Sociocultural Model of Reading~~

Comprehension Targets and Text Complexity

Over the course of the NAEP Reading Assessment, students will engage with texts of various discourse structures and an appropriate grade-level range of text complexity. While reading these texts within an assessment block, students will complete varied reading comprehension activities that include specific purposes, tasks, processes, and consequences. The reader, per his or her achievement level, will employ various knowledge types to accomplish the assessment's reading comprehension activities. In doing so, the reader will demonstrate achievement relative to four *comprehension targets:* (1) Locate and Recall; (2) Integrate and Interpret; (3) Analyze and Evaluate; and (4) Use and Apply. Students at each achievement level are expected to meet the demands of each comprehension target. However, as the complexity of texts increases on a given reading assessment, students, on average, are expected to demonstrate less competency with skills associated with higher-level comprehension targets, such as Use and Apply.

Broad and Specific Reading Purposes

~~According to the sociocultural model, reading~~ Reading activities in an assessment block are situated within not only a disciplinary context but also a broad reading purpose. ~~This section describes~~Each assessment block is designated as having one of two *broad* purposes: Reading to Develop Understanding or Reading to Solve a Problem. Reading to Develop Understanding (RDU) blocks ask students to *read and comprehend deeply* (analyzing, inferencing, interpreting, and critiquing) in or across disciplinary contexts. By contrast, Reading to Solve a Problem (RSP) blocks ask students to demonstrate understanding across multiple texts and related perspectives in order to solve a problem. Reading to Solve a Problem activities do involve comprehending text, but in the service of a specific action or product, such as a classroom presentation.

Both RDU and RSP blocks also have *specific* purposes with reader roles that shape how and why readers engage with the ~~mapping of~~tasks, texts, and items in each block. Unlike the broad purposes, these specific purposes are applicable only to the texts in a given task in the assessment block. The purpose-driven statements will reflect the contexts and scenarios in which reading ~~purposes~~in the real world occurs. The subsections below describe how specific reading purposes map to disciplinary contexts.

Literary Texts. People engage in reading literature for the following purposes:
- To understand human experience
- To entertain themselves and others
- To reflect on and solve personal and social dilemmas
- To appreciate and use authors' craft to develop interpretations

In school, students read, create, and discuss literary texts such as poems, short stories, chapter books, novels, and films. Outside of school, students participate in book clubs, create fan fiction and book reviews, follow and discuss authors, dramatize literary works with animation and music, and more. NAEP simulates these Contexts of Reading to Engage in Literature by providing test takers with activities to respond to literary and everyday texts like those read in and outside of school.

Science Texts. People engage in reading science for the following purposes:
- To understand natural and material phenomena
- To design solutions to problems
- To explore and discuss issues and ideas
- To consider impacts on themselves and society

In school, students read, create, and discuss science texts such as explanations, investigations, journal articles, trade books, and more. They design solutions to engineering challenges, use diagrams and flow charts, and follow step-by-step procedures to investigate scientific phenomena. Outside of school, students engage in reading science when participating in games, cooking, and crafts, and reading and viewing science and health news. NAEP simulates these Contexts of Reading to Engage in Science by providing test taskers with activities to respond to science and everyday texts like those read in and outside of school.

Social Studies Texts. People engage in reading social studies for the following purposes~~, among them these~~:
- To understand past events and how they may impact the present
- To explore and discuss issues and ideas
- To understand human motivation, perception, and ethics
- To advocate for change for themselves and society

In school, students read social studies texts such as primary and secondary source documents, historical narratives in textbooks, case studies, current events, <u>maps, data,</u> court cases, and more. They read, create, and discuss memoirs, timelines, and biographies. Outside of school, people engage in reading history and social studies when participating in trivia games, crafts, civic activities, community discussions, self-help, and community service. NAEP simulates these contexts of reading to engage in social studies by providing test tasks with activities to respond to history/social studies and everyday texts like those read in and outside of school.

NAEP Reading Achievement Levels: Grade 4

NAEP Basic

Fourth-grade students performing at the *NAEP Basic* level should be able to locate specific pieces of information, identify relationships between explicitly stated pieces of information, make simple inferences and interpretations ~~within~~in static, dynamic, and ~~between~~multimodal texts, create summaries, and show understanding of vocabulary in the disciplinary contexts.

When engaged in reading literary texts such as fiction, drama, film, poetry, and literary nonfiction, fourth-grade readers performing at the *NAEP Basic* level should be able to use textual evidence as support to identify or determine literary elements such as character point of view, theme or central message, problem, and setting. Readers should be able to explain how a text's illustrations contribute to what is conveyed by the text, explain the differences between poems, drama, and prose, and show understanding of vocabulary and simple figurative language. Readers ~~can~~should be able to produce a simple summary of a text and continue the narration of an incomplete story to a conclusion of their making.

When engaged in reading science texts such as exposition (including literary nonfiction), argumentation, and procedural texts (including investigations), fourth-grade readers performing at the *NAEP Basic* level should be able to use textual evidence as support to determine the main idea and how it is supported by key details, determine and interpret an author's point of view or purpose, and distinguish between fact and opinion. Readers should be able to interpret and integrate information presented in a text visually, quantitatively, and orally, analyze specific results of a simple multistep procedure, and show understanding of academic and domain-specific vocabulary. Readers ~~can~~should be able to apply simpler ideas acquired through reading to solve a new problem.

When engaged in reading social studies texts such as exposition (including literary nonfiction), argumentation, and documents of historical and literary significance, fourth-grade readers performing at the *NAEP Basic* level should be able to determine the main idea and how it is supported by key details, determine and interpret an author's point of view or purpose, and distinguish between fact and opinion. Readers should be able to describe the overall structure of a text and compare and contrast explicit information found in a firsthand and secondhand account of the same event or topic. Readers ~~can~~should be able to produce a simple summary of a text and integrate information from lower complexity sources to produce a new text of informational or argumentative purpose.

NAEP Proficient
Fourth-grade students performing at the *NAEP Proficient* level should be able to make more complex inferences and interpretations, reconcile inconsistencies within and across ~~a text or~~static, dynamic, and multimodal texts, and explain how an author uses reasons and evidence to support particular points in a text.

When engaged in reading literary texts such as fiction, drama, film, poetry, and literary nonfiction, fourth-grade readers performing at the *NAEP Proficient* level should be able to use textual evidence as support to describe in depth character, setting, and plot, and to explain how a theme or central message is conveyed through details in a text. Readers should be able to analyze how a printed version of a text relates to its multimedia version and show understanding of nuances in word meaning. Readers ~~can~~should be able to produce a detailed summary of a text and rewrite a story from a different character's perspective.

When engaged in reading science texts such as exposition (including literary nonfiction), argumentation, and procedural texts (including investigations), fourth-grade readers performing at the *NAEP Proficient* level should be able to use textual evidence as support to explain events, procedures, ideas, and concepts based on specific information in and across texts. Readers should

be able to make predictions and to interpret an author's point of view or purpose, including in reference to a procedure or experiment and in comparison to another text's author. Readers should be able to develop a new procedure or experiment based on knowledge acquired from information gained from reading texts.

When engaged in reading social studies texts such as exposition (including literary nonfiction), argumentation, and documents of historical and literary significance, fourth-grade readers performing at the *NAEP Proficient* level should be able to use textual evidence as support to explain events, procedures, ideas, and concepts based on specific information in and across texts. Readers should be able to explain how information presented in a text visually, quantitatively, and orally contributes to an understanding of a text. Readers should be able to produce a detailed summary of a text and adopt the persona of a historical figure when producing a new text of informational or argumentative purpose.

NAEP Advanced

Fourth-grade students performing at the *NAEP Advanced* level should be able to make complex inferences and to support their interpretations, conclusions, and their judgments based upon evidence within and across static, dynamic, and multimodal texts.

When engaged in reading literary texts such as fiction, drama, film, poetry, and literary nonfiction, fourth-grade readers performing at the *NAEP Advanced* level should be able to use textual evidence as support to explain character motivation and behavior and how characters interact with setting and plot. Readers should be able to evaluate how characters or themes resonate with society and their personal lives. Readers should be able to apply knowledge acquired about author's craft to produce a literary work evidencing their understanding.

When engaged in reading science texts such as exposition (including literary nonfiction), argumentation, and procedural texts (including investigations), fourth-grade readers performing at the *NAEP Advanced* level should be able to determine the significance of information and arguments made in a text. Readers should be able to make predictions and to interpret an author's point of view or purpose and to argue for or against a particular interpretation.

When engaged in reading social studies texts such as exposition (including literary nonfiction), argumentation, and documents of historical and literary significance, fourth-grade readers performing at the *NAEP Advanced* level should be able to determine the significance of information and arguments made in a text. Readers should be able to make predictions and to interpret an author's point of view or purpose and to argue for or against a particular interpretation. Readers should be able to use acquired knowledge about a topic, conduct brief research, and produce a historical document, such as a political cartoon or a personal bill of rights.

NAEP Reading Achievement Levels: Grade 8

NAEP Basic

Eighth-grade students performing at the *NAEP Basic* level should be able to find information in static, dynamic, and multimodal texts, make simple inferences and

interpretations within and between texts, make predictions, create objective summaries, analyze word choice, and show understanding of vocabulary in the disciplinary contexts.

When engaged in reading literary texts such as fiction, drama, film, poetry, and literary nonfiction, eighth-grade readers performing at the *NAEP Basic* level should be able to use textual evidence as support to determine theme or central idea and aspects of character, setting, and plot. They should be able to compare basic literary attributes of two or more texts and make judgments about how each author presents events. Readers show understanding of vocabulary and figurative language. They ~~can~~should be able to develop a simple objective summary of a text and produce an argumentative text that prosecutes or defends the actions of a character by using evidence from the reading text.

When engaged in reading science texts such as exposition (including literary nonfiction), argumentation, and procedural texts (including experiments), eighth-grade readers performing at the *NAEP Basic* level should be able to use textual evidence as support to determine the central ideas and conclusions of a text and explain how a text makes connections among and distinctions between individuals, ideas, and/or events. Readers should be able to integrate quantitative or technical information expressed in words in a text with a version of that information expressed visually (e.g., in a flowchart, diagram, model, graph, or table), show understanding of how to follow precisely a multistep procedure of an experiment, and show understanding of academic and domain-specific vocabulary, key terms, and symbols. Readers ~~can~~should be able to apply simpler ideas acquired through reading to solve a new problem.

When engaged in reading social studies texts such as exposition (including literary nonfiction), argumentation, and documents of historical and literary significance, eighth-grade readers performing at the *NAEP Basic* level should be able to determine the central ideas, determine and interpret an author's point of view or purpose, and distinguish between fact, opinion, and reasoned judgment in a text. Readers should be able to identify key steps in a text's description of a process related to social studies (e.g., how a bill becomes law). Readers ~~can~~should be able to produce a simple objective summary of a text and integrate information from multiple sources to produce a new text of informational or argumentative purpose.

NAEP Proficient

Eighth-grade students performing at the *NAEP Proficient* level should be able to make more complex inferences and interpretations, form explanations and generalizations, generate alternatives, and apply new ideas acquired through reading to a new problem or context~~.~~ when reading static, dynamic, and multimodal texts. Students should be able to use text-based evidence to support arguments and conclusions.

When engaged in reading literary texts such as fiction, drama, film, poetry, and literary nonfiction, eighth-grade readers performing at the *NAEP Proficient* level should be able analyze the development of the theme or central idea over the course of a text and how particular lines of dialogue or incidents in a text propel, the action, provoke a decision, or reveal aspects of character. Readers should be able to analyze how a printed version of a text relates to its multimedia version and how text structure contributes to meaning and style. They ~~can~~should be able to analyze how word choice impacts a text's meaning and tone. Readers ~~can~~should be able

to develop a detailed objective summary of a text and produce an informational text that analyzes how different authors developed a similar theme or central idea.

When engaged in reading science texts such as exposition (including literary nonfiction), argumentation, and procedural texts (including experiments), eighth-grade readers performing at the *NAEP Proficient* level should be able to use textual evidence as support to analyze the specific results of a multistep procedure based on explanations in the text, analyze how the author acknowledges and responds to conflicting evidence and/or viewpoints, and analyze how two or more texts provide conflicting information on the same topic, identifying where the texts disagree on matters of fact or interpretation. Readers should be able to compare and contrast information gained from experiments, simulations, video, or multimedia sources with that gained from reading a text on the same topic. Readers should be able to generate an alternative procedure or experiment based on knowledge acquired from information gained from reading texts.

When engaged in reading social studies texts such as exposition (including literary nonfiction), argumentation, and documents of historical and literary significance, eighth-grade readers performing at the *NAEP Proficient* level should be able to use textual evidence as support to explain how a text makes connections among and distinctions between individuals, ideas, and/or events (e.g., through comparisons, analogies, or categories). Readers should be able to analyze the relationship between a primary and secondary source on the same topic and analyze how two or more texts provide conflicting information on the same topic, identifying where the texts disagree on matters of fact or interpretation. They should be able to analyze the structure an author uses to organize a text and develop a detailed objective summary of a text. Readers canshould be able to produce an argumentative text that proposes a form of social action based on knowledge acquired and opinions formed from the reading texts.

NAEP Advanced

Eighth-grade students performing at the *NAEP Advanced* level should be able to make complex inferences and to support their interpretations, conclusions, and their judgments based upon evidence within and across static, dynamic, and multimodal texts. Students should be able to evaluate the relevance and strength of evidence to support an author's claims.

When engaged in reading literary texts such as fiction, drama, film, poetry, and literary nonfiction, eighth-grade readers performing at the *NAEP Advanced* level should be able to use textual evidence as support to analyze how multiple literary elements in a text relate to each other and to analyze points of view of and between character(s) and the reader/audience. Readers should be able to analyze how a modern text draws on themes, patterns of events, or character types from myths or traditional stories, and then evaluate how these elements resonate with society and their personal lives. Readers should be able to produce a literary text that adapts elements of a myth into a contemporary retelling based upon the reader's personal experience.

When engaged in reading science texts such as exposition (including literary nonfiction), argumentation, and procedural texts (including experiments), eighth-grade readers performing at the *NAEP Advanced* level should be able to analyze the development of the central idea over the course of the text. They should be able to delineate and evaluate the argument, claims, and

reasoning in a text, including whether the evidence is relevant and sufficient to support the claims. Readers ~~can~~should be able to produce a new argumentative or informative text that synthesizes information from a range of sources to demonstrate a coherent understanding of a process, phenomenon, or concept.

When engaged in reading social studies texts such as exposition (including literary nonfiction), argumentation, and documents of historical and literary significance, eighth-grade readers performing at the *NAEP Advanced* level should be able to analyze the development of the central idea over the course of the text and analyze how the author acknowledges and responds to conflicting evidence and/or viewpoints. Readers should be able to delineate and evaluate the argument, claims, and reasoning in a text, including whether the evidence is relevant and sufficient to support the claims. They ~~can~~should be able to produce an informative text that traces and connects various factors (e.g., economic and societal) by incorporating acquired knowledge through reading multiple sources and conducting brief research.

NAEP Reading Achievement Levels: Grade 12

NAEP Basic

Twelfth-grade students performing at the *NAEP Basic* level should be able to find information in static, dynamic, and multimodal texts, make inferences and interpretations within and between texts, make predictions, create objective summaries, analyze word choice, and show understanding of vocabulary in the disciplinary contexts.

When engaged in reading literary texts such as fiction, drama, film, poetry, and literary nonfiction, twelfth-grade readers performing at the *NAEP Basic* level should be able to use textual evidence as support to analyze the development of the theme or central idea over the course of a text and to analyze points of view of and between character(s) and the reader/audience. They should be able to compare literary attributes of two or more texts and make judgments about how each author presents events. Readers show understanding of vocabulary and figurative language. They ~~can~~should be able to develop an objective summary of a text and produce an informational text that applies a common theme or central idea culled from multiple texts to a current societal issue.

When engaged in reading science texts such as exposition (including literary nonfiction), argumentation, and procedural texts (including experiments), twelfth-grade readers performing at the *NAEP Basic* level should be able to use textual evidence as support to analyze the specific results of a multistep procedure based on explanations in the text, explain how specific individuals, ideas, and/or events interact and develop over the course of a text, and analyze how the text structures information or ideas into categories or hierarchies. Readers should be able to compare and contrast findings presented in a text to those from other sources and show understanding of general academic and domain-specific vocabulary, key terms, and symbols. Readers should be able to generate an alternative procedure or experiment based on knowledge acquired from information gained from reading texts.

When engaged in reading social studies texts such as exposition (including literary nonfiction), argumentation, and documents of historical and literary significance, twelfth-grade readers performing at the *NAEP Basic* level should be able to explain how specific

individuals, ideas, and/or events interact and develop over the course of a text, determine and interpret an author's point of view or purpose, and distinguish between fact, opinion, and reasoned judgment in a text. Readers should be able to show understanding of general academic and domain-specific vocabulary and of figurative language and be able to develop an objective summary of a text by paraphrasing its complex concepts and information. They ~~can~~should be able to integrate information from multiple sources to produce a new text of informational or argumentative purpose.

NAEP Proficient

Twelfth-grade students performing at the *NAEP Proficient* level should be able to make more complex inferences and interpretations, form explanations and generalizations, generate alternatives, and apply new ideas acquired through reading to a new problem or context~~.~~ when reading static, dynamic, and multimodal texts. Students should be able to use text-based evidence to support arguments and conclusions.

When engaged in reading literary texts such as fiction, drama, film, poetry, and literary nonfiction, twelfth-grade readers performing at the *NAEP Proficient* level should be able to analyze how two or more themes or central ideas interact and build on one another to produce a complex account over the course of the text. Readers should be able to analyze how text structure contributes to meaning and style. They ~~can~~should be able to analyze how word choice impacts a text's meaning and tone. Readers ~~can~~should be able to develop a detailed objective summary of a text and produce a new text of literary purpose based on an archetypal conflict discovered in the reading texts.

When engaged in reading science texts such as exposition (including literary nonfiction), argumentation, and procedural texts (including experiments), twelfth-grade readers performing at the *NAEP Proficient* level should be able to use textual evidence as support to analyze an author's point of view or purpose, including in providing an explanation, describing a procedure, or discussing an experiment, identifying important issues that remain unresolved. Readers should be able to integrate and evaluate multiple sources of information presented in diverse media or formats (visually or in words) in order to address a question or solve a problem. Readers ~~can~~should be able to produce a new argumentative or informative text that synthesizes information from a range of sources to demonstrate a coherent understanding of a process, phenomenon, or concept.

When engaged in reading social studies texts such as exposition (including literary nonfiction), argumentation, and documents of historical and literary significance, twelfth-grade readers performing at the *NAEP Proficient* level should be able to use textual evidence as support to analyze how the central ideas interact and build on one another to produce a complex account. They should be able to analyze the themes, purposes, and rhetorical features of foundational U.S. documents and evaluate the effectiveness of the structure in the text's exposition or argument. They should be able to develop a detailed objective summary of a text. Readers ~~can~~should be able to evaluate multiple sources of information presented in different media or formats (visually or in words) in order to produce an argumentative text with evidence to structure and support a judgment.

NAEP Advanced

Twelfth-grade students performing at the *NAEP Advanced* level should be able to make complex inferences and to support their interpretations, conclusions, and their judgments based upon evidence within and across static, dynamic, and multimodal texts. Students should be able to use an understanding of legal and ethical principles to develop a text or presentation on a matter of social debate.

When engaged in reading literary texts such as fiction, drama, film, poetry, and literary nonfiction, twelfth-grade readers performing at the *NAEP Advanced* level should be able to use textual evidence as support to analyze and evaluate multiple interpretations of text (e.g., multimedia versions of a text) to the source text. Readers ~~can~~should be able to use acquired knowledge to produce an informational text analyzing how elements of an era's poetry (e.g., Romanticism's celebration of nature; rejection of industrialization) are evidenced in the work of one or more poets.

When engaged in reading science texts such as exposition (including literary nonfiction), argumentation, and procedural texts (including experiments), twelfth-grade readers performing at the *NAEP Advanced* level should be able to delineate and evaluate the argument, claims, and reasoning in a text, and evaluate the hypotheses, data, analysis, and conclusions in a text. They should be able to explain how style and content contribute to the power, persuasiveness, or beauty of the text. Readers ~~can~~should be able to produce a new argumentative or informative text that utilizes an understanding of legal and ethical principles to address a scientific matter of debate (e.g., uses of genetic databases).

When engaged in reading social studies texts such as exposition (including literary nonfiction), argumentation, and documents of historical and literary significance, twelfth-grade readers performing at the *NAEP Advanced* level should be able to delineate and evaluate argument, claims, and reasoning in a text. They should be able to explain how style and content contribute to the power, persuasiveness, or beauty of the text. Readers ~~can~~should be able to produce a new argumentative or informative text that utilizes an understanding of legal and ethical principles to address a societal matter of debate (e.g., indigenous peoples' land rights).

———

This ~~last section presents two hypothetical examples~~appendix is provided to describe design considerations, based on the principles outlined in the framework, that assessment developers might weigh as they develop blocks. Each design decision requires tradeoffs, and assessment developers must consider which tradeoffs to make and why. Such decisions are guided by the components of the assessment—the disciplinary context, broad purpose, tasks and texts, and comprehension targets. Moreover, developers must consider whether and how different design features (item response formats, UDEs, and process data) will be used so that a broad array of features are included, in purposeful ways, across the multiple blocks that are sampled.

Employing the 2026 NAEP Reading Assessment ~~blocks. The first example illustrates a~~**Framework Principles: Assessment Components**

The 2026 NAEP Reading Assessment Framework describes three areas of design considerations about which developers will make decisions: the block components (disciplinary context, broad reading purpose, specific reading purpose, and reader role); the task components (tasks, texts, and items); and the design features (item response formats, UDEs, and process data). See Exhibit 1 for an illustration of how these areas relate to one another.

It is important to note that developers do not necessarily make decisions about these three areas in this order; rather, some of these decisions might be iterative and mutually informative. For example, in developing a literature block for a certain grade level, the developer might first choose a text and broad reading purpose and then determine the reader's role and a specific purpose appropriate to the text. Thus, the areas are only used to illustrate the relationship of these considerations to one another and how students might experience the block.

First, students learn what disciplinary context and broad purpose they are working in, and then they learn the specific purpose and their role. Second, students are given a text or texts to read and tasks to work on as they read that text. As students engage with the texts and tasks, they complete comprehension items, which are situated within the tasks, as illustrated in Exhibit 1. Third, design features such as item formats, UDEs, and process data are used to leverage the digital assessment environment to measure, as precisely as possible, how well students perform on the blocks. The relationships among all of these features of the assessment are synergistic. The disciplinary context and broad reading purpose drive the specific reading purpose, reader role, selection of texts, and the tasks; all of which, in turn, inform the comprehension items. Items are created in relation to item response formats, as different formats are used to collect different kinds of information. Similarly, all assessment components inform the use of UDEs because UDEs are used to help ensure that all students can gain access to the tasks required of them to complete the assessment and that the assessment measures students' reading comprehension of the texts and not something else (e.g., how well they can read or follow test directions). In this manner, a well-integrated block results, with all of the parts working in tandem.

Exhibit 1 illustrates the assessment components and their relationship to one another. Each block defines a disciplinary context, broad purpose, block-specific purpose, and reader role. Each block also outlines 2-3 tasks, which are explicitly stated to the reader and which might include sub-tasks, for readers to complete as they read one or more texts. For each task, there might be one or more comprehension items. UDEs are only employed as needed to bolster construct validity and ensure better measurement of the reading comprehension construct. Similarly, process data are only collected in places where developers think it might be useful for understanding why students perform the way that they do or for informing revision or future research and development.

As developers develop a block, they make decisions about each of the components described in Exhibit 1. In the following section, we describe some of the different considerations developers might think about as they make decisions about the assessment components illustrated.

Exhibit 1. Design Components of a 2026 NAEP Reading Assessment Block
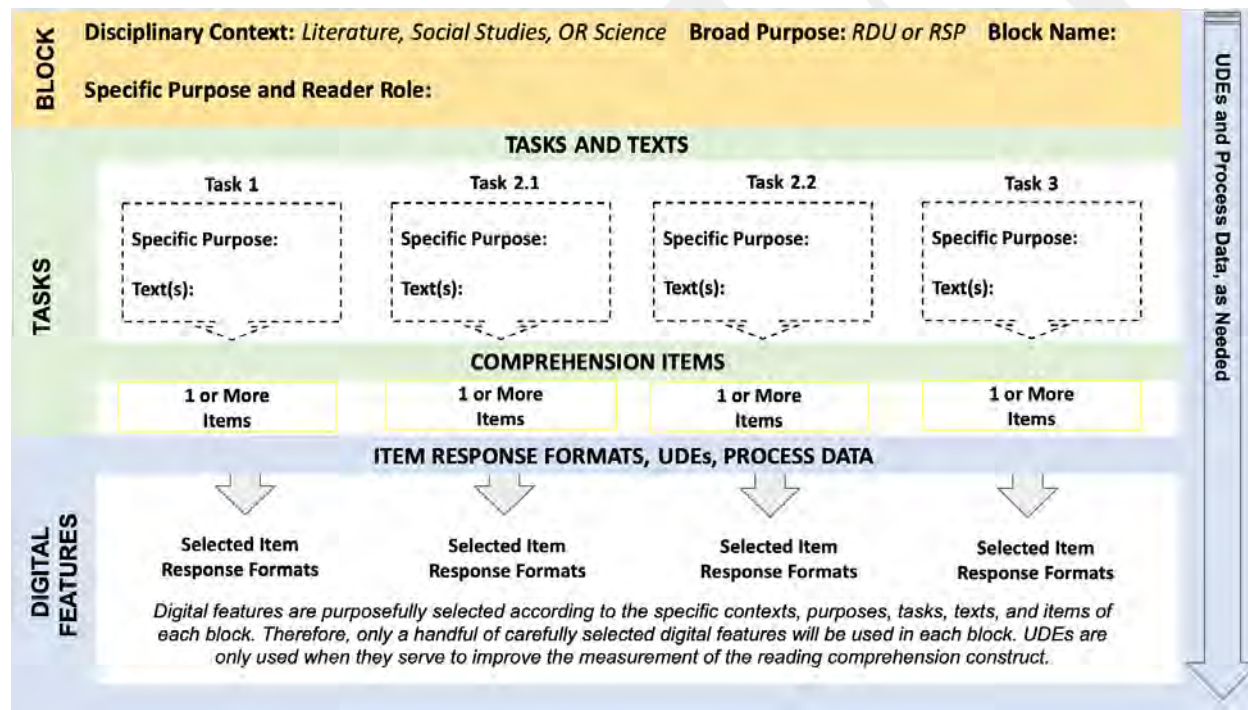
Considering the Range of Variations Within Assessment Components and Across a Block

When blocks are developed in accordance with the 2026 NAEP Reading Framework, the expectation, as outlined in Chapters 2 and 3, is that any of the components in a block (i.e., rows in the exhibit) can vary along a continuum, as depicted in Exhibit 2. That is, some blocks are more likely to include static texts and less cumulative tasks, items, and/or UDEs from one item to the next (left of center on the continuum), while other blocks are more likely to include dynamic/multilayered texts and more cumulative tasks, items, and/or UDEs from one item to the next (right of center on the continuum).

Exhibit 2 illustrates the continuum of design features from which developers might choose for each assessment component in the testing block. Note that within a given block, one component may have features that fall more on the left end of the continuum while features of

110

another component fall more on the right. Further, the complexity of different design features, and therefore of assessment components, may vary within a task. For example, for one task/text, the features might be less complex, but for a second task/text, they might be more complex. Or, for a single task/text, the purpose might be straightforward but the UDEs might be more complex. In all blocks, formats and features will continue to provide opportunities for readers to engage with an array of texts and tasks made possible in the digital platform used for all NAEP assessments.

Exhibit 2. Continuum of Variation in Features of Assessment Components Within a Block

Assessment Component	Less Dynamic and Cumulative Across Content and Format		More Dynamic and Cumulative Across Content and Format
Specific Reading Purposes	Purposes allow readers to focus attention on developing a deep understanding of a theme, question, or issue to be explored during the block. Not all tasks or items within the block necessarily work directly toward this theme, and there are opportunities for items to be less related to the specific purpose.	⟷	Purposes are paired with an essential inquiry question or problem to be examined throughout the task. All tasks and items within the block help readers work towards this theme, question, or problem.
Reader Role	Fewer parameters are specified for the reader's role. The reader is placed in a situation that provides fewer pieces of information about how to engage with the provided tasks and texts. The reader might be placed within a situation that contextualizes expectations for how to engage with provided texts and tasks. However, this situation provides less information about that role.	⟷	More parameters are specified for the reader's role within the block. The reader is placed in a situation that provides multiple pieces of information about how to engage with the provided tasks and texts. Readers may be assigned a particular role, and their role may be more specified, particularly in relation to reading purpose(s) and expected outcome(s).
Tasks	Purpose-driven tasks and items are situated in line with disciplinary context, but tasks are less related to one another	⟷	Purpose-driven tasks are situated in line with disciplinary context but tasks are more tightly structured so that one task builds on the previous;

	with less probability of readers moving back and forth across items within tasks; less need for resetting. Less involved culminating task, or no culminating task. Task not necessarily a determinant of all items in block.		more probability that tasks are interdependent; may have more need for resetting. More involved culminating task at the end of an activity that directly addresses the question or problem; major driver of the block.
Texts	*Number*: 1-3 topically related texts; excerpts may be included. *Dynamism*: More static texts with minimal dynamic features. *Linearity:* Fewer nonlinear structures to navigate within or across texts; less variation in structures across texts. *Features*: Texts include a narrower range of features and fewer types of media.	⟷	*Number*: 2-4 topically related and interconnected texts may be included. Readers may be asked to choose only some texts to engage with and in line with task purposes. *Dynamism*: More texts with dynamic and/or or multimodal text features. *Linearity:* More nonlinear structures to navigate within or across texts; more variation in structures across texts. *Features*: Texts include a wider range of features and more types of media.
Items	Items are less connected to the overall specific reading purpose for the block and there are more opportunities for items to be related, but less connected, to this specific purpose and to the related tasks; Less dynamic item formats to support less complex tasks and items.	⟷	Items are more connected to the overall specific reading purpose for the block. There are more opportunities for items to be more directly related to the specific reading purpose for the block and to the related tasks; More dynamic item formats to support more complex/multilayered tasks and items.
Universal Design Elements (UDEs)	Fewer cumulative reading purposes that may require UDEs for knowledge or motivation and potentially	⟷	More cumulative reading purposes that may require UDEs for knowledge or motivation and potentially greater need for task-based UDEs.

	lesser need for task-based UDEs.		
Process Data	Potentially fewer locations where process data involving reading actions could provide additional information about comprehension performance; sources may include, but not be limited to, timing data, navigation data (use of look back buttons), and use of varied item response formats.	⟷	Potentially more locations where process data involving reading actions could provide additional information about comprehension performance; sources might include, but not be limited to, timing data, more complex navigational practices across multiple sources and/or use of more dynamic item response formats.

Specific Guidelines for Block Development

Despite the range of variations in assessment components described above, as developers consider the different decisions they must make when designing a block, it is useful to keep the following points in mind:

1. Students deserve to know the tasks that lie ahead of them in the block. Guidance in the form of task-based UDEs is essential.
   a. Both purpose and reader role need to be made apparent at the outset of a block.
   b. Students should be reminded of purpose and role at the outset of each specific task within a block.
2. Since directions can be a source of construct irrelevant variance, they should always be conveyed in as accessible and straightforward a register as possible.
3. There is always a button available to allow students to listen to directions (or listen and read at the same time).
4. Just as expectations that students will be able to handle more complex text across the grades, so the expectations that they will be able to handle more complex guidance and activities also increases.
5. Cognitive labs, block tryouts, and pilot testing should ultimately guide NAEP in determining the optimal balance among these principles, especially when they come into conflict with one another. The experience in GISA and in the current 2019 operational NAEP SBT blocks offer an existence proof that these guidance features are manageable by 4th, 8th, and 12th graders. When these sorts of guidance features were included along with other UDEs in the 2017 special study, the enhanced blocks provided an overall comprehension performance advantage and resulted in higher motivational ratings by students, especially in the earlier grades. NAEP needs to monitor these matters with great vigilance.

Block Sketches

Sketches of three different blocks are provided to illustrate a range activity within assessment blocks that students might encounter when they participate in the 2026 NAEP

113

Reading Assessment. To accomplish this goal, the Appendix offers three hypothetical sketches of blocks (showing only a sampling of items from each) that might be developed using the components (from Chapter 2) and the design principles (from Chapter 3) of the 2026 NAEP Reading Framework. Importantly, these sketches are designed to exemplify key concepts from the framework and do not represent blocks or items that will be used on future NAEP assessments. Tasks presented with multiple sample items are provided to help readers of the framework envision how theoretical ideas in the framework might guide assessment design. However, these sketches do not represent fully expectations for enacting the NAEP style guide and other test specifications.

The first example (labeled *Hana* because it is built upon a short story text entitled *Hana Hashimoto, Sixth Violin* by Chieri Uegaki and Qin Leng) illustrates a block developed for the broad purpose of Reading to Develop Understanding (RDU~~) block, and the second example illustrates a~~). The second example (labeled *Hill District* because it is built upon a set of activities surrounding an authentic civic issue in the Hill District neighborhood of Pittsburgh, PA) illustrates a block developed for the broad purpose of Reading to Solve a Problem (RSP~~) block.~~

~~The first example outlines components in a Reading to Develop Understanding ().~~ And the third (labeled *EB White* because it is built upon a pair of texts, one *about* and one *by* the author E. B. White) illustrates a second, but more traditional, RDU~~) block in which fourth graders read to engage with texts in a literature context. In this block, fourth grade~~ block. Referring to the underlying continuum of variation for assessment components within blocks as detailed in Exhibit 2 above, these three block sketches are situated on three hypothetical points along that continuum, as illustrated in Exhibit 3.

Exhibit 3. Underlying Continuum of Variation in Assessment Components in the Block Design for E.B. White, Hana, and Hill District Block Sketches

Less Dynamic and Cumulative   More Dynamic and Cumulative

EB White   **Hana**   **Hill District**
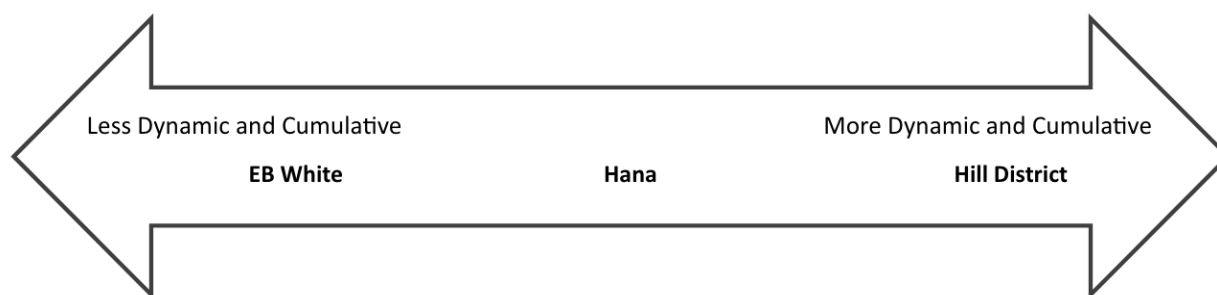
An overview of the three block sketches. As suggested, *Hana* exemplifies what features of assessment components in RDU blocks might look like at the center of the continuum. In this block, grade 4 readers preview a short video of young children playing in an orchestra ~~and then~~ (a motivational UDE) to pique their interest in playing the violin, the topic of the text. Then, they read and interpret story excerpts from the short story, Hana Hashimoto, ~~*Sixth Violin,*~~ by Chieri Uegaki ~~as they~~ in preparation for a book discussion with three peers. First, students are asked to read to develop an understanding of the ~~main~~ characters, key events, and author's craft ~~and~~ . Second, they apply their insights to ~~predict events beyond the story.~~ describe what Hana is like as

114

a person. so that they are ready to contribute to the discussion.

~~The second example illustrates what eighth~~ The *Hill District* block includes features of assessment components more characteristic of those toward the right of the continuum that 12th graders might encounter in a ~~Reading to Solve a Problem (~~RSP~~)~~ block with texts situated in a social studies context. In this block, students engage in more ~~complex~~cumulative reading tasks that might include two to four more dynamic or multilayered texts and involve greater integration across texts and items, all of which contribute to a generative opportunity to use and apply meaning from ~~the text to solve a problem. While both assessment blocks include tasks, texts, items, and UDEs, differences in what readers experience illustrate just a sampling of the range of possible design features from which developers might choose in their creation of purpose-driven tasks embedded in any single block.~~ multiple texts to solve a problem.

~~**Grade 4 Assessment Block. Reading to Develop Understanding in a Literature Context: Hana Hashimoto, Sixth Violin**~~

EB White illustrates a second RDU block, but for an 8th grade literature context and with a more traditional look and feel than the *Hana* block. It retains many of the features students might encounter in commercially available standardized tests of reading comprehension, on state reading examinations, or on blocks characteristic of NAEP tasks developed from earlier frameworks. In fact, this example was created by using the two texts from a released 8th grade NAEP Block drawn from the 2011 NAEP Assessment.

When viewing these examples, it is important to keep in mind the following points:

- The purpose of these block sketches is to help readers of this 2026 Reading Framework develop an understanding of the range of comprehension activity and assessment components students might experience when they participate in the NAEP Reading Assessment.

- None of the examples is complete in the sense that all of the components and features are fully developed in the exact form in which they would appear on a finished test booklet. These examples are more like elaborated sketches that provide a preview of what each

block might look like, recognizing that not all of the actual items, UDEs, and other features are fully developed. Sometimes, for example, the type of UDE needed is specified but not actually provided (e.g., a particular word might make a plausible pop up vocabulary definition), or the type of comprehension item is indicated but not actually developed (e.g., an analyze/evaluate item is needed here to test students' understanding of the author's use of irony). In some cases (e.g., the Hill District block), two exemplars with different formats are provided to illustrate alternative ways to design task and item features in any particular block.

- While all three exemplar blocks include purposes, contexts, tasks, texts, items, and UDEs, differences in what readers experience illustrate just a sampling of the range of possible design features from which developers might choose in creating purpose-driven tasks embedded in any single block.

- Any given block, even a block that is situated toward one or the other end of the continuum (from Exhibit 7 in Appendix A), may have some features that lean more toward the center or even in the other direction. In other words, a given block might lean toward the traditional end of the continuum on texts (as does the Hana block) but toward the innovative end on item formats (as does Hana). The *EB White* block lends is otherwise classic RDU block, but lends itself to a Use/Apply culminating task (which is more characteristic of RSP blocks).

- The inclusion of the *EB White* exemplar has been included intentionally to reflect NAEP's commitment to maintain a healthy sample of tasks that feature print-based texts, RDU purposes, relatively few UDEs, and items that reflect the entire array of comprehension targets. As in all aspects of development, NAEP builds on its current strengths as it incorporates important developments in the nature of texts and tasks that students encounter in the ever-changing world of literacy.

Hana Hashimoto, Sixth Violin, Grade 4

The following example (not intended to be a complete block or to represent an actual NAEP Reading assessment) offers a sketch of what a Grade 4 Reading to Develop Understanding in a Literature Context block might look like. In the sketch, we walk through the assessment components described in the framework and illustrated in the block design visual (see Exhibit 4). These include the block components (context, purpose, grade level), the tasks (the tasks as well as the texts and items that students use to accomplish those tasks), and the digital features (item response formats, UDEs, and process data). In so doing, we describe how these components might be used by assessment developers when creating blocks to achieve some of the aims described in the framework.

Exhibit 4. Block Design for Hana

BLOCK

Disciplinary Context: *Literature, Grade 4* **Broad Purpose:** *Reading to Develop Understanding (RDU)* **Block Name:** *Hana*

Specific Purpose: Read to learn about what Hana is like as a person so you can participate in a book discussion with classmates
Reader Role: To work with three classmates to prepare for the book discussion

UDEs and Process Data, as Needed

TASKS

TASKS AND TEXTS

Task 1

Specific Purpose: Learn about important events in the story and characters' thoughts, feelings, and actions
Text(s): *Hana Hashimoto, Sixth Violin*

Task 2

Specific Purpose: Write about what Hana is like as a person so that you are ready to discuss the book with peers
Text(s): *Hana Hashimoto, Sixth Violin*

COMPREHENSION ITEMS

6-7 Items

6-7 Items

DIGITAL FEATURES

ITEM RESPONSE FORMATS, UDEs, PROCESS DATA

Selected Item Response Formats:
Single and multiple selection multiple choice; fill in the blank; short constructed response

Selected Item Response Formats:
Short constructed response; matching; zones; short constructed response

Task-based UDEs: Teacher and student task characters; Task-reminder; Word bank
Motivational UDEs: Teacher and student task characters; String Instruments Video

Block Components (Disciplinary* Context*, Purposes,* and *Tasks

Reader Role). This block is designed to assess how ~~4th-grade~~Grade 4 readers develop understanding within a single, print text ~~by forming an~~ in a literary context. In this block, readers identify important events in the story and analyze how characters' thoughts, feelings, and actions describe the kind of people they are. Then, readers use and apply what they have learned to form an overall interpretation ~~about a story's~~of the main character, Hana. They choose a character trait from a word bank and then ~~applying that understanding to consider what might happen after the story ends. More specifically, readers are invited to engage with a group of fourth-grade students (represented by task characters in the assessment) who are reading the text, *Hana Hashimoto, Sixth Violin*, by Chieri Uegaki. In this book, a young girl named Hana signs up to play the violin in her school's talent show after having had only three lessons. While many items give students opportunities to demonstrate their understanding and develop their thinking across the story, the texts (video and story) and items are relatively independent of one another. The test block also includes opportunities to develop understanding around other aspects of the story that may, or may not, contribute to that characterization. Throughout the block, readers are asked to activate and employ their personal, cultural, and literary knowledge and resources by drawing on textual evidence to make thoughtful interpretations of the text~~explain how Hana fits that character trait based on the thoughts, feelings, and actions they have already interpreted.

Specific Reading Purpose(s) and Reader Role. At the beginning of the assessment (see Exhibit ~~4~~5), readers are ~~invited~~told that they will read the story *Hana Hashimoto, Sixth Violin*, by Chieri Uegaki and Qin Leng. Then, they are introduced to the specific purpose and reader role of reading to participate in a small book discussion group ~~about the story *Hana Hashimoto, Sixth Violin*~~with ~~the~~ three ~~other 4th~~fourth grade ~~student~~ classmates (represented in the assessment by task characters~~. A~~ Gia, Gabe, and Luisa). They are also introduced to their teacher for the

project (represented by the ~~task character joins them to explain the discussion goal, which focuses on understanding how Hana grows and changes over the course of the story as a result~~ task character Mr. Obas).

Then, a **task-based UDE** in the form of ~~events involving her classmates and her family. To~~ two statements informs students what tasks will be expected of them. Here, students are told that, to prepare for the book discussion, ~~students are told~~ they will read ~~parts of~~ the story and ~~respond to items situated in three purpose-driven tasks to:~~ 1) ~~identify~~ learn about important events in the story and ~~consider what these events say about the characters; 2) learn more about Hana and other important characters from their words~~ characters' thoughts, feelings, and actions ~~in the story; and 3) apply their understanding of the characters in order to predict what might happen after the story ends.~~; and, 2) use what they have learned about Hana's to describe what she is like as a person. **Motivational UDEs** (here, student ~~task character classmates~~ and ~~a~~ teacher ~~task character; see also bottom of Exhibit 1~~ avatars and an introductory video) serve to ~~situate and~~ motivate readers to engage with the block.

Exhibit ~~1. Task-specific purposes~~5. Specific purpose, reader role, and ~~student~~ task characters serve to situate readers in a Grade 4 Reading to Develop Understanding block involving the short story *Hana Hashimoto, Sixth Violin* by Chieri Uegaki and Qin Leng

Introduction

You will read a book called *Hana Hashimoto, Sixth Violin* by Chieri Uegaki. After you read the story, you will participate in a small discussion group with three classmates to understand how Hana changes over the course of the story as a result of events involving her family.

To prepare for the book discussion, you will read the story while completing three tasks.

1. Identify important events in the story and consider what these events say about the characters.
2. Learn more about Hana and other important characters from their words, feelings, and actions in the story.
3. Use your understanding of the characters to predict what might happen after the story ends.

Your teacher for this project will be Mr. Beadle

You will work with three classmates in your discussion group: Brian, Diana, and Michael

Welcome

You will read the story, *Hana Hashimoto, Sixth Violin*, by Chieri Uegaki and Qin Leng to prepare for a book discussion.

First, you will learn about **important events** in the story and **characters' thoughts, feelings, and actions.**

Then, you will **write about what the main character, Hana, is like as a person so that you are ready to discuss the book** with three peers.

Your teacher for this project will be Mr. Obas:

You will work with three classmates in your discussion group:

Gia   Gabe   Luisa

NEXT

Next, test-takers ~~are invited to~~ view a ~~30~~15 second video of young children playing in an orchestra found at a website designed for young children ~~(https://www.wonderopolis.org/wonder/why-do-orchestras-need-so-many-people~~ (see Exhibit ~~A~~26). This short video is an example of a **motivational** ~~and knowledge-based~~ **UDE designed to introduce readers** to the sounds and emotions one might experience when playing in an orchestra~~, while providing minimal background knowledge to students who may be less familiar with stringed instruments such as the violin.~~.

Exhibit ~~2~~6**. The teacher task character and a** ~~pre-reading preview~~<u>motivational UDE in the form</u> **of a** ~~30~~15 **second video clip of students playing stringed instruments serve to pique** ~~students~~<u>students'</u> **interest** ~~and provide minimal background knowledge needed to make sense of the story~~

In this story, the main character Hana, decides to enter a talent show to perform the violin, a stringed instrument.
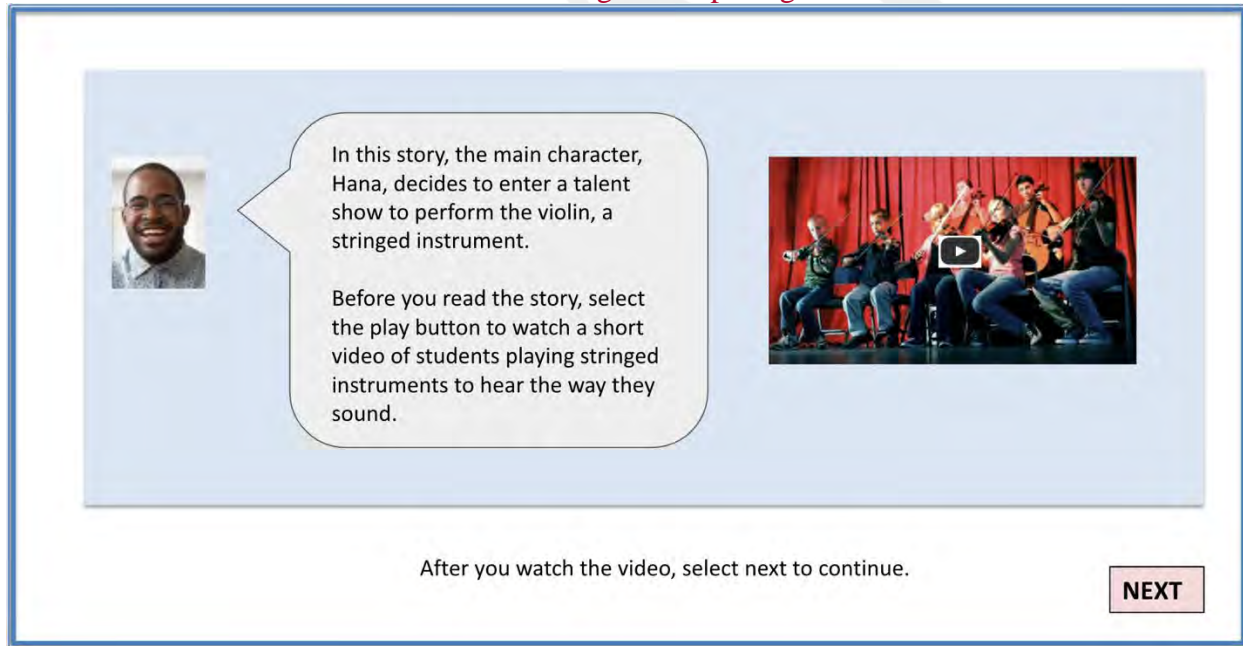
Before you read the story, clink the blue link to watch a short video of students playing stringed instruments, including the violin, to hear what it sounds like.

https://www.wonderopolis.org/wonder/why-do-orchestras-need-so-many-people

After you watch the video, click next to continue. | NEXT |

~~After learning about the three task-specific purposes in this literature block and viewing the video, readers engage with several passages from the book that contain important information about Hana and other minor characters. Through these passages~~

In this story, the main character, Hana, decides to enter a talent show to perform the violin, a stringed instrument.

Before you read the story, select the play button to watch a short video of students playing stringed instruments to hear the way they sound.

After you watch the video, select next to continue. | NEXT |

Task Components (Tasks, Text(s), and Items). <u>After viewing the video about string instruments, students then begin reading the story and working through the tasks.</u>

Tasks. <u>After students are asked to read the story, the teacher reminds them of the specific reading purpose for the block (to prepare for a discussion) as well as the students' first task as they prepare for this discussion: learning about the events and characters (see Exhibit 7). In this case, the task reminder for the first task stays on the screen until students are ready to do the second task. At that point, the teacher offers a reminder of the second task, which is to write</u>

about what Hana is like as a person. To do this, students are asked to use evidence from the story that they have already collected and interpreted on Hana's thoughts, feelings, and actions.

Text: Hana Hashimoto, Sixth Violin. In this story, a young girl named Hana signs up to play the violin in her school's talent show after having had only three lessons. Through the story, readers learn that Hana's desire to take lessons was inspired by a recent visit to Japan to see her Ojiichan, or grandfather, who plays the violin. They also learn that despite much teasing and doubting from her brothers, Hana practices and practices for the talent show, inviting everyone she can to be her audience. When it comes time to play her violin in the talent show, Hana is at first nervous and thinks to herself, "This is going to be a disaster." However, as she looks out at the audience, she sees her friends and family. Then, Hana recalls her Ojiichan telling her to do her best and decides that is what she will do. She plays some of the everyday sounds she recalls her grandfather playing for her (e.g., a mother crow calling her chicks"). At the end of her performance, Hana takes "a great big bow." That night, her family asks her to play more of her sounds. The story ends with Hana playing her violin to herself before she goes to sleep, imagining the notes drifting out through her window and to Ojiichan in Japan while the author hints that Hana will keep practicing so that she might perform again in next year's talent show.

In the digital assessment format, readers can scroll through the story as they read, and the items appear aside the text so that readers can easily refer to the text as they complete the comprehension items. At the Grade 4 level, some illustrations from the original source text might accompany the story, as they do here (see Exhibit 7).

Comprehension Items. The array of items provides students with opportunities to develop their thinking across the story and demonstrate their understanding. Throughout the block, readers are asked to draw on textual evidence to make thoughtful interpretations of the text. The text and items are suitably independent of one another so that a student's performance on one item does not impact their performance on another item. The test block also includes opportunities to develop understanding around aspects of the story that may, or may not, contribute to the final task. Generally, however, the items help students work towards the specific purpose of the block (in this case, preparing for a book discussion), as well as the goal of each task. Exhibits 7-13 illustrate items that help students accomplish the first task of learning about the events and characters. Exhibits 14-16 illustrate items that then help students accomplish the second task of using what they have learned about the characters' thoughts, feelings, and actions to characterize Hana, in particular, by writing about what she is like as a person (see Exhibits 14-16).

Item response types ~~would~~ vary from simple multiple choice to short answer or hybrid constructed response items to give readers different kinds of opportunities to demonstrate their understanding in the block. **Sample questions** at this point ~~may~~might, for example, include single selection multiple choice items to assess readers' ability to locate and recall important events and other details (see ~~Exhibit 3) as well as~~ Exhibits 7 and 8), short constructed-response items that include fill in the blank options (see Exhibit 9), multiple selection multiple choice items (see Exhibit 10), and longer short constructed response items that ask readers to interpret and integrate ~~character traits~~details about the character's thoughts, feelings, and actions into their understanding of the story (see Exhibit ~~4~~11).

Exhibit ~~3. Example of~~ **7. A Grade 4 RDU block illustrating** a **Locate and Recall** multiple choice~~, locate and recall~~ item ~~in~~. **The teacher reminds the reader of the specific purpose (to prepare for** a ~~Grade 4 RDU block~~

How do Hana's brothers first respond to her decision to play the violin in the talent show?

A ⭘ They are happy for Hana.   ⊖

B ⭘ They make fun of Hana.   ⊖

C ⭘ They promise to help Hana.   ⊖

D ⭘ They are angry with Hana.   ⊖

Exhibit 4. ~~Example of a short constructed-response interpret~~discussion) **and** ~~integrate item in a Grade 4 RDU block~~ **the first task (to learn about events and characters)**

At this point in the story, do you think Hana's brothers supported her decision to play in the talent show. Please explain your thinking with facts from the story.

To prepare for the discussion, first read the short story and learn about the events and characters.

Hana Hashimoto, Sixth Violin
By Chieri Uegaki & Qin Leng

When Hana Hashimoto announced that she had signed up for the talent show and that she would be playing the violin, her brothers nearly fell out of a tree.

"That's just loopy," said Kenji. "You're still a beginner."

"Stop kidding," said Koji. "You can barely play a note."

"It's a *talent* show, Hana."

"You'll be a disaster!"

Hana squared her shoulders and took her violin and bow inside, leaving her brothers laughing like monkeys in the tree.

What does Hana want to do for the talent show?

A ⭘ Sing a song   ⊖

B ⭘ Tell jokes   ⊖

C ⭘ Play the violin   ⊖

D ⭘ Climb a tree   ⊖

NEXT

Within the block, **knowledge-based UDEs** might include pop-up boxes providing a hint about the meaning of certain ~~domain specific words or general topics of a text~~vocabulary terms that are not tested (in this case, describing what a ~~talent~~talent show is, see Exhibit ~~5) so~~8). This provides readers ~~are provided the minimal~~with some background knowledge from which to make sense of the story ~~and engage~~. This allows readers to focus on engaging with items designed to measure, in this case, their ability to make inferences about characters from their thoughts, feelings, and actions in the story.

Exhibit 8. A Grade 4 Integrate and Interpret item illustrating a knowledge-based UDE in the form of a pop-up box defining the term "talent show." The blue pop-up box appears when a test-taker clicks on the underlined term. This example also illustrates a multiple choice integrate and interpret item

Exhibit 9. A Grade 4 Locate and Recall item illustrating a fill in the blank short constructed response

Exhibit 10. A Grade 4 Locate and Recall item illustrating a multiple selection multiple choice response format

In addition, **a look-back button (~~or~~a task-based UDE)** is embedded into ~~the~~items with excerpted ~~quote in this item; if~~text (see Exhibits 11 and 12). If readers ~~choose~~wish, they can click on the underlined quote to see exactly where the excerpted text is located in the context of the original story ~~passage~~ in the assessment space. Multiple choice and constructed response item formats are interspersed throughout the assessment.

Exhibit 5~~11~~. A Grade 4 ~~Integrate~~Analyze and ~~Interpret~~Evaluate item illustrating a task-based UDE in the form of a look-back button that refers ~~the reader~~readers to the relevant section of text within the story and a ~~knowledge-based UDE in the form of a pop-up box defining the term "talent show" for the text _Hana Hashimoto, Sixth Violin_ by Chieri Uegaki. The blue pop-up box appears when a test-taker clicks on the highlighted term.~~

The text says, "When Hana Hashimoto announced that she had signed up for the **talent show** ⟶ and that she would be playing the violin, her brothers nearly fell out of a tree...Hana practiced every day, just like Ojiichan. And every day, her brothers fled the house with covered ears, complaining about the horrible noise."

A **talent show** is a show in which different people perform a special skill.

Thinking about this part of the text, why do you think Hana's brothers flee the house every day?

    a.   They are angry with Hana
    b.   The noise of the violin bothers them
    c.   They have somewhere important to be
    d.   They like the way the violin sounds

As depicted in Exhibit 6, students could also be given a word bank (a task-based UDE) **from which to select relevant character traits when asked to describe the kind** **of person Hana is. Then, in an analyze and evaluate item with a hybrid**short **constructed response format**, students could be asked to use that word to **describe Hana's actions and then explain their thinking using evidence from** **the story.**

Hana Hashimoto, Sixth Violin
By Chieri Uegaki & Qin Leng

The day of the talent show arrived and the school auditorium thrummed with excitement. Backstage, Hana waited with a walloping heart. Finally, Hana heard the master of ceremonies call her name.

As Hana walked onto the stage, her violin tucked under her arm and bow gripped tight in her hand, an oceanic roar filled her ears.

Things seemed to be moving in slow motion, and for one dizzy moment, Hana thought, "Kenji and Koji were right. This is going to be a disaster." She wished she could turn into a grain of rice and disappear into a crack between the floorboards.

She could hardly see with the spotlight in her eyes. Yet, as Hana looked out into the audience, certain faces appeared to her, as if through a telescopic lens.

The story says, "She wished she could turn into a grain of rice and disappear into the crack between the floorboards."

What do you think the author is trying to tell the reader about how Hana is feeling? Use details from the story to explain your answer.

NEXT

~~Exhibit 6.~~ **A Grade 4 Analyze and Evaluate 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 it as part of their answer in the box.**

In the story, the author writes, "Hana swallowed her nerves like medicine and leaned toward the microphone. She would just do her best."

What do you think the author is trying to tell the reader about what kind of person Hana is? Choose a character trait from the word bank to describe Hana. Then, use the box to explain what makes you think this.

WORD BANK

helpful	curious
brave	proud
smart	nervous
afraid	confident
forgetful	determined

Students could also be given a timeline on which to drag and drop their responses about how the main character changes over the course of the story (see Exhibit 7). A collection of relevant and irrelevant notes about the character can be provided from which students can select the best answers. Once completed, students would then have access to this informational graphic as a writing support when answering the final Use and Apply item (see Exhibit 7).

Toward the end of the story, readers learn that when Hana is on stage, she first becomes nervous and doubts herself, but then imagines her Ojiichan telling her to do her best. Hana decides to play what she knows — the sound of a crow, lowing cows, her neighbor's cat, and rain on a paper umbrella. Her family loves her performance so much that later that evening, they ask her to play them more musical notes around the dinner table. The story ends when Hana recalls the numerous songs her Ojiichan shared with her and imagines what she might play in next year's talent show.

Exhibit 12. The items for the first task help students develop an understanding of the events and characters as in this Grade 4 Integrate and Interpret short constructed response item

Hana Hashimoto, Sixth Violin
By Chieri Uegaki & Qin Leng

She could see her brothers, melting into their seats. She saw her best friend, Jas, giving her two thumbs up. And there, her smiling mother, and her father, camera in hand.

Hana held a breath, then ballooned her cheeks before letting it out. With a *whoosh*, the roaring in her ears receded. Then, as everyone seemed to disappear beyond the light shining down on her like a moonbeam, she remembered.

"*Gambarunoyo, Hana-chan.*" Do your best, her grandfather had told her. Ojiichan would be cheering for her.

Hana swallowed her nerves like medicine and leaned toward the microphone. She would just do her best.

The story says that "Hana swallowed her nerves like medicine and leaned toward the microphone. She would just do her best."

Why do you think Hana decides to do her best? Explain your answer using details from the story.

NEXT

Exhibit 13. A Grade 4 Integrate and Interpret Item for the first task using a single response multiple choice format

Hana Hashimoto, Sixth Violin
By Chieri Uegaki & Qin Leng

"This is the sound of a mother crow calling her chicks," she said. She placed the violin under her chin, held her bow in position and played three raw, squawky notes.

As Hana continued to play all the special sounds she had practiced, the air around her came alive with buzzing bees...and lowing cows...and squeaking mice...and croaking frogs.

Finally, as the last sound effect trailed away, Hana tucked her bow and violin under her arm. "And that," she said to the audience, "is how I play the violin."

Then she took a great big bow.

Later, after dinner, Kenji surprised Hana by asking for an encore. "Make that funny cow sound again," he said. Then Koji said, "Make that crazy cat sound, too." So Hana did. And when her mother and father and brothers all laughed, she happily played her sounds again.

Which statement do you think is true about Hana's brothers at the **end** of the story?

A ○ They do not like her playing anymore.

B ○ They want to learn to play.

C ○ They have always liked her playing.

D ○ They do not like her playing but do now.

NEXT

The story ends when Hana recalls the songs her Ojiichan shared with her and imagines what she might play in next year's talent show. At this point, students are invited by the teacher to start the second task, which is to write what Hana is like as a person in preparation for the book discussion (see Exhibit 14).

One of the classmates (a task character in the assessment) acts as a **motivational UDE** to motivate the student to engage in collecting notes for the second task, as the classmate has already completed part of the activity. The task character also acts as a task-based UDE in reminding the student that they should use specific details from the story about Hana's thoughts, feelings, and actions. Once completed, students have access to the full set of notes, as these completed notes are transferred to the next item (see Exhibit 15).

Exhibit 14. Teacher and student task characters remind readers of the second task goal in this Integrate and Interpret item

In Exhibit 15, the other two classmates serve as **motivational and task-based UDEs** to engage students in the task while also reminding them to stay focused on the character's thoughts, feelings, and actions. The student's responses from the previous item are carried over to the next item as the completed notes, which also serves to motivate the student since they have already completed the work. These notes could also be "reset" if the student did not enter appropriate notes in the previous item so that the student's score on this item is not dependent on how they responded previously.

In Exhibit 15, the student is asked to move the notes from their notepad into the chart as they sort the notes into Hana's thoughts, feelings, and actions in preparation for writing about the kind of person she is. In the final task (see Exhibit 16), the student has access to this chart as a writing support when they answer the final use and apply item. Again, notes that are incorrect are reset so that the final item is not dependent on the way they responded to this one.

Exhibit 15. The student's responses from their completion of the previous item are carried over to the next item as the completed notes. A graphic organizer with drag and drop features offers students an efficient way to demonstrate their understanding of how the text conveys the character's thoughts, feelings, and actions in this Grade 4 Integrate and Interpret item

A longer constructed response item such as the example shown in Exhibit ~~8~~16 is designed to assess readers' ability to Use and Apply ~~their understanding to a new situation beyond~~understandings learned from the story ~~itself. In~~to form a characterization of Hana. As readers engage with this final part of the ~~assessment~~ block, ~~after listening to one of the student task characters orally~~the teacher invites them to use their chart (which they have access to) to write what Hana is like as a person in preparation for the discussion.

Then, as depicted in Exhibit 16, in a use and apply item with a hybrid constructed response format, students are given a word bank (a task-based UDE) from which to select a relevant character trait (these could be hot spots; in other words, when readers click on them, the word is highlighted and gets recorded as the student's answer to Part A) when asked to describe ~~how Hana reacted to her brothers' behavior earlier in the story, readers are invited to join the discussion group with three task character classmates and contribute their ideas.~~ the kind of person Hana is. Instead of spending time generating character trait words (which is not part of the construct this item aims to measure), the student can select from those provided. This allows the student to focus their limited time and cognitive resources on applying evidence from the text about Hana's thoughts, feelings, and actions to an analysis of the kind of person Hana is.

Exhibit 16. This final, two-part Use and Apply item illustrates the use of a task-based UDE in the form of a word bank of character traits as well as a hybrid item format where students select a choice and write about it. Students use what they have learned from the text about Hana as a person and apply that understanding to draw a conclusion about the kind of person she is.

Performance Evidence and Indicators

. When interpreting reading achievement from performance on the 2026 NAEP Reading Assessment, multiple indicators can be used to ~~situate and~~ explain what students are able to do. As indicated earlier in this chapter, each block would be classified with a primary disciplinary context, grade level, and broad purpose. Scores from the Hana Hashimoto, Sixth Violin block, then, ~~can demonstrate~~describe what Grade 4 students can do in a literature context as part of a Reading to Develop Understanding block. The block is designed to measure ~~their~~students' ability to develop their understanding ~~within~~of a single text and then apply that understanding ~~to~~in a simple culminating event (in this case, ~~making a prediction,~~describing the kind of person Hana is based on her thoughts, feelings, and actions in the story~~, about what will happen after the story ends).~~).

Test developers keep ~~an elaborate~~a detailed account of all decisions that go into classifying texts and generating items from comprehension targets in each block. This process enables NAEP to compile a description of what 4th graders (or sub-groups of 4th graders) can do in each disciplinary context as they engage with texts and test items, while also being encouraged to draw from and use the knowledge, skills, and experiences they bring to that reading context.
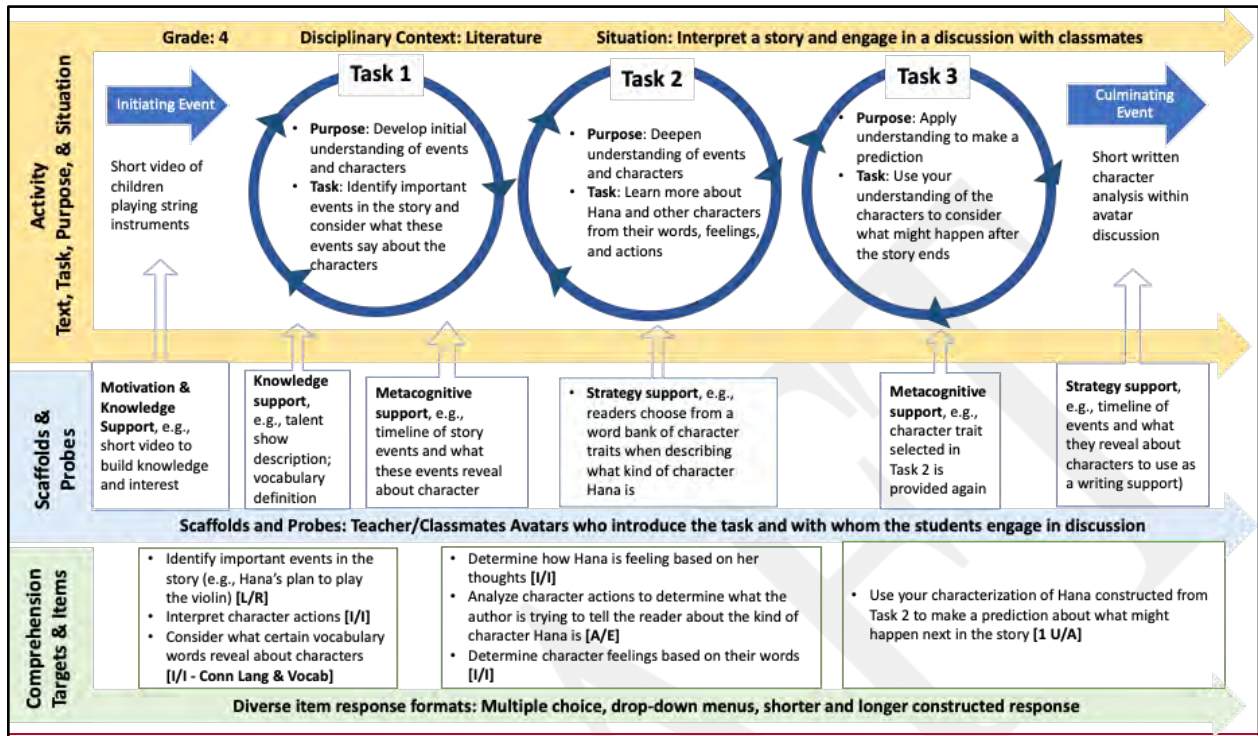
Hill District, *Grade* 8 Assessment Block: Reading to Solve a Problem in a Social Studies *12*

Block Components (Context

Context

, Purposes, and Reader Role). This block is designed to assess how 8th12th grade readers develop understanding across multiple texts in a social studies context by forming an interpretation of the perspectives of multiple community members linked to both current and historical events and then applying that understanding to solve a problem. (See Exhibit 17 for the block design and Exhibit 18 for the introduction to the block).

Exhibit 17. Block Design for Hill District Sketch

BLOCK

Disciplinary Context: *Social Studies, Grade 12* **Broad Purpose:** *Reading to Solve a Problem (RSP)* **Block Name:** *Hill District*
Specific Purpose: Read to form an interpretation of the perspectives of multiple community members linked to both current and historical events and then apply that understanding to synthesize and share their work with project leaders (student task characters)
Reader Role: To help project leaders compile and organize the perceived benefits and concerns about the park plan from different perspectives, including their own.

UDEs and Process Data, as Needed

TASKS AND TEXTS

TASKS

Task 1	Task 2.1	Task 2.2	Task 3
Specific Purpose: Read multiple documents to learn about the park and what different community members think about the plan. **Text(s):** two web pages and a set of four twitter comments	**Specific Purpose:** Learn about the history of Pittsburgh's Hill District and how that history is related to the parks' design. **Text(s):** search engine results and two web pages	**Specific Purpose:** Use a digital notepad to compile and organize benefits and concerns by perspective and provide supporting evidence from texts read previously	**Specific Purpose:** Choose two perspectives and summarize the views of different community members and share their own views so they can be added to the final report

COMPREHENSION ITEMS

3-5 Items	3-4 Items	3-4 Items	3 Items

ITEM RESPONSE FORMATS, UDEs, PROCESS DATA

DIGITAL FEATURES

Selected Item Response Formats: Single/multiple selection MC; short constructed response	**Selected Item Response Formats:** Single/multiple selection MC; zone; short constructed response	**Selected Item Response Formats:** Single and multiple selection MC; grid; short constructed response	**Selected Item Response Formats:** short constructed response; possible audio recording (in future)

Selected UDEs interwoven with item response format: Task characters, purpose setting statements, graphic organizers, and digital notepad

More specifically, readers are invited to engage with ~~a group of~~three students (represented by task characters in the assessment) who ~~are motivated to learn about a current civic project deeply rooted in their city's history: The City of Pittsburgh has recently announced~~ have been asked by the Mayor to compile and organize public reactions to an ambitious plan ~~for~~proposed by the City of Pittsburgh. Known as the "I-579 Cap Project," the plan involves the construction of an overpass park that reconnects the Hill District and Downtown. Park designers at a landscape architecture firm have created a proposed park design.

~~Park designers at a landscape architecture firm have created a proposed park design.~~
~~Test takers are asked to learn about this project by considering the role of a key aspect~~The tasks in this Reading to Solve a Problem block reflect design features that are more dynamic and cumulative in terms of content and format, as depicted toward the right side of the continuum in Exhibit 2. For example, readers are constrained by specific purposes and role expectations about how to engage with provided texts. The four tasks (and related sub-tasks) are tightly structured so that one task builds on the previous, such that readers are asked to learn more about the project goals and get a general sense of the public's comments before they are asked to gain a deeper understanding of the historical significance of the proposed park ~~design: the inclusion of a 13-year-old African-American girl named Keisha who appears on illustrated signs throughout the park. Park designers have proposed including signs of Keisha in many park locations to provide details about the African-American community's history in the Hill.~~ .

The test block also includes opportunities for students to engage with several interconnected digital texts (e.g., excerpts from social media, search engine results, and multimedia websites and online news articles) that represent the perspectives of different kinds of community members and cuts across issues of contemporary and historical relevance. Throughout the block, readers ~~engage with a collection of~~ XX ~~historical and contemporary multimodal texts to develop~~are asked to activate and employ their personal, cultural, and civics

knowledge and resources by drawing on textual evidence in multiple modes to make thoughtful interpretations and evaluations of the text. Of note, several UDEs and dynamically formatted items are designed to motivate and guide students through the series of challenging assessment tasks in a multilayered digital environment.

Specific Reading Purpose(s) and Reader Role. At the beginning of the assessment (see Exhibit 18), students learn that the city has recently unveiled the park plan to the public on its website and city residents have been invited to share their reactions on various social media. Students are also introduced to three high school aged task characters selected by the Mayor to help compile comments in preparation for a series of public working meetings (see Exhibit 19). In a school partnership with the city, the three high schoolers have invited other students to help them organize comments from different community members. This situation inspires the question/problem that guides readers' inquiry in the assessment block: How do different community members feel about the proposed park project and what interests inform their comments?

Exhibit 18. A social studies context and reader role serve to situate readers in a Grade 12 Reading to Solve A Problem block involving several interconnected digital texts

Introduction

The City of Pittsburgh recently announced an ambitious plan for the construction of a highway overpass park known as the "I-579 CAP Project" that reconnects the Hill District and Downtown.

The proposed park design was posted on the city website and community members have begun to share their reactions on various social media. To prepare for the city's next meeting, the Mayor has tasked a team of high school students to help organize the comments according to the varied interests of different community members.

It's a big task, and you have been invited to help.

Click next to learn more. **NEXT**

Your Task

You will work with three high school students who were selected by the mayor to lead the project:

Kai | Moises | Jasmine

To accomplish this goal, you will do four tasks:

1. Read multiple documents to learn about the park plan and what different community members think about the plan.

2. Learn about the history of Pittsburgh's Hill District and how that history is related to the park's design.

3. Describe some of the benefits and concerns about the park from different perspectives, or viewpoints, including your own.

4. Share your work with the student project leaders for a meeting with the Mayor.

NEXT

Task Components (Tasks, Text(s), and Items).

Tasks. To support their inquiry, students are told they will read multiple documents and respond to items situated in four purpose-driven tasks to: a) learn more about the proposed park plan and keep notes about what different community members think about the plan; b) learn about the history of Pittsburgh's Hill District and how that history is related to the park's design; c) synthesize some of the benefits and concerns about the park from different perspectives, including their own and d) share their work with the student project leaders for a meeting with the Mayor. Several task-based UDEs (e.g., graphic organizers and purpose setting statements) and motivational UDEs (three student avatars, a recent event, and an ~~understanding of~~ opportunity to express their own opinions about the ~~Hill District's history~~project) serve to guide and ~~then clarify~~motivate readers to engage with the ~~planned vision of~~block.

Texts. After learning about the four task-specific purposes in this social studies block, readers engage with a digital text set that contains important information and viewpoints related to the proposed park ~~to different~~plan. These include Twitter comments from community members ~~of the~~; a set of search engine results and pull-down menu items from a website; and text passages on websites about the project embedded with comments from Pittsburgh ~~community. While some of the items~~residents, photographs, a short video, and an artist's rendering of the park plan. With each new text, readers learn more about proposed features of the park plan that help to build their understanding of how different community members view the park's features from various perspectives and how the history of Pittsburgh's Hill District is relevant to the park's plan.

Comprehension Items. Item response types would vary from simple multiple choice to short answer or hybrid constructed response items to give ~~students~~readers different kinds of opportunities to demonstrate their understanding ~~and~~ in the block and apply that understanding to

136

solve the problem. While some items give students opportunities to demonstrate their understanding and develop thinking within a specific text, other items are designed to assess how readers ~~integrate ideas~~navigate and ~~perspectives~~make meaning across sources representing multiple ~~sources~~ and diverse perspectives.

After being asked to read text and watch a short video on a website about the park project (Exhibit 20), sample questions may, for example, include single or multiple response formats for multiple choice items that ask readers to locate and recall important details about the project from the passages and the video (Exhibits 21 and 22). Other questions might assess students' ability to integrate and interpret textual and visual information from an artist's rendering of the site improvement plan on a different website (see Exhibit 22). **Task-based UDEs** (e.g., one of three task characters) provide short prompts (shown at the top of Exhibits 20 and 23) designed to cue the reader about the steps they are completing as they read across different sources to solve the problem.

Exhibit 20. A Grade 12 RSP block illustrating the directions that readers are asked to follow as they engage with texts and items. The task character reminds the reader of the specific purpose and the first task

Exhibit 21. A Grade 12 Locate and Recall item illustrating a multiple-selection multiple choice response format

TRIB LIVE | SEA | Righting a Wrong | Next City

Pittsburgh City Council on Tuesday authorized about $32 million in spending for the Interstate 579 "cap" designed to link the Lower Hill District and Downtown via a three-acre park.

PennDOT will oversee the work expected to start by July and end in late 2021, according to the Sports and Exhibition Authority of Pittsburgh and Allegheny County.

Plans call for handicapped-accessible pedestrian pathways, an amphitheater, stage, lawns, landscaping, recreation areas, art and replacement of walkways in the area.

"The 'cap' project will be transformative for the Hill District by removing a physical barrier and re-establishing connectivity to centers of employment, education and services in Downtown Pittsburgh," according to the SEA website.

I-579, known as the Crosstown Expressway, runs between the Hill and Downtown. The cap will essentially serve as a large bridge deck made of 8-inch-thick, reinforced concrete slabs supported by beams and pillars.

It will complement the Penguins' long-awaited $450 million residential, retail and office redevelopment plan for the 28-acre former Civic Arena site.

The SEA in 2016 received a $19 million federal grant for the work and is kicking in an additional $5.2 million. The remaining funding is coming from the state.

Council unanimously approved the allocation without comment.

After learning about the park plan in the text and in the video, select **all** statements that are true of the Interstate I-579 Cap Project?

A ◯ The project is funded by the Pittsburgh City Council.

B ◯ The project will re-connect the lower Hill District and Downtown Pittsburgh.

C ◯ The project provides new green spaces for residents to exercise.

D ◯ The project will increase access to employment in Downtown Pittsburgh.

NEXT

Exhibit 22. A Grade 12 Locate and Recall item illustrating a single-selection multiple choice item response format

TRIB LIVE | SEA | Righting a Wrong | Next City

Pittsburgh City Council on Tuesday authorized about $32 million in spending for the Interstate 579 "cap" designed to link the Lower Hill District and Downtown via a three-acre park.

PennDOT will oversee the work expected to start by July and end in late 2021, according to the Sports and Exhibition Authority of Pittsburgh and Allegheny County.

Plans call for handicapped-accessible pedestrian pathways, an amphitheater, stage, lawns, landscaping, recreation areas, art and replacement of walkways in the area.

"The 'cap' project will be transformative for the Hill District by removing a physical barrier and re-establishing connectivity to centers of employment, education and services in Downtown Pittsburgh," according to the SEA website.

I-579, known as the Crosstown Expressway, runs between the Hill and Downtown. The cap will essentially serve as a large bridge deck made of 8-inch-thick, reinforced concrete slabs supported by beams and pillars.

It will complement the Penguins' long-awaited $450 million residential, retail and office redevelopment plan for the 28-acre former Civic Arena site.

The SEA in 2016 received a $19 million federal grant for the work and is kicking in an additional $5.2 million. The remaining funding is coming from the state.

Council unanimously approved the allocation without comment.

According to the article, which organization is funding a large part of the 'cap' project?

A ◯ Lower Hill District

B ◯ Penn DOT

C ◯ Cross Town Expressway

D ◯ Sports and Exhibition Authority

NEXT

Examples of short constructed-response items earlier in the block might ask readers to integrate and interpret information about how park designers plan to modify the city's use of natural resources to address environmental concerns (Exhibit 24). Later in the block, readers might be asked to integrate and interpret information in an online newspaper article about the historical significance of the park's design (Exhibit 25) or to analyze and evaluate the requests of some community members to include park features that honor the history of their neighborhood (Exhibit 26). Also depicted in Exhibit 26 is a **task-based UDE** in the form of a task character that serves to remind students of their reading purpose in the second task.

Exhibit 24. A Grade 12 RSP short constructed-response item that asks readers to integrate and interpret information about how park designers plan to address environmental concerns

Exhibit 25. A Grade 12 short constructed-response item with a look-back button (task-based UDE) that asks readers to integrate and interpret information in an online newspaper article about the historical significance of the park's design

Exhibit 26. A Grade 12 short constructed-response item that asks readers to integrate and interpret information on a web page with a look-back button (task-based UDE). The task character reminds readers of the specific purpose of the second task

Other potential items might ask readers to locate and evaluate the relevance of search engine results pertaining to the historical significance of some of the park's features (see Exhibit 27) or locate (navigate to) and then analyze information from a website's menu to evaluate the expertise of the group responsible for publishing information about the park project (see Exhibits 28 and 29 respectively). Both of these tasks and items can be designed to collect timing and navigation process data about the choices readers make as they navigate multilayered digital environments such as search engines and websites with menus.

Exhibit 27. **A Grade 12 selected response zone item designed to capture process data about which link is selected paired with a short constructed response scored item that asks readers to analyze and evaluate the relevance of their search engine choice**

TASK 2. Now, can you help us with some Internet research to find background information to share?

Directions: Read the list of Google Search Results. Choose the link most likely to have information about the history of Pittsburgh's Hill District and why this history is relevant to the park's plan. Then, use the box to explain why that link is the best choice. Use evidence from the search engine results to support your thinking.

https://www.bostonmagazine.com › news › 2017/04/05
A Park Beneath the I-93 Overpass Will Open in June - Boston
Apr 5, 2017 — A new park, called Ink Underground, is scheduled to open under the I-93 overpass in June. It will stretch eight acres and feature public art.

https://archive.triblive.com › local › pittsburgh-allegheny
Pittsburgh City Council OKs 'cap' park over Crosstown ...
Jul 2, 2018 — A project to build a 'cap' over Pittsburgh's Crosstown Boulevard and ... A proposed 'cap' linking the Lower Hill District and downtown Pittsburgh would include a city park and ... office redevelopment plan for the 28-acre former Civic Arena site. ... The Shrines of Pittsburgh

https://pittsburghpa.gov › dcp › hill-district
The Greater Hill District Master Plan | pittsburghpa.gov - City of ...
The Neighborhood. As one of Pittsburgh's earliest and largest neighborhoods, the Hill District continues to play an important role in the story of African-Americans ...

https://www.post-gazette.com › 2019/06/14 › stories
Righting a wrong: New park over I-579 to reconnect ...
Jun 14, 2019 — A $32 million effort to 'right a wrong' that occurred half a century ago gets its start Friday. A slew of politicians and stakeholders will gather in a parking lot near PPG Paints Arena to break ground on a three-acre park that will straddle Interstate 579/Crosstown Boulevard.

Your explanation:

NEXT

Exhibit 28. A Grade 12 item selected response zone item designed to capture process data about how readers navigate through hyperlinked web pages

TRIB LIVE | SEA | Righting a Wrong | Next City

Pittsburgh City Council on Tuesday authorized about $32 million in spending for the Interstate 579 "cap" designed to link the Lower Hill District and Downtown via a three-acre park.

PennDOT will oversee the work expected to start by July and end in late 2021, according to the Sports and Exhibition Authority of Pittsburgh and Allegheny County.

Plans call for handicapped-accessible pedestrian pathways, an amphitheater, stage, lawns, landscaping, recreation areas, art and replacement of walkways in the area.

"The 'cap' project will be transformative for the Hill District by removing a physical barrier and re-establishing connectivity to centers of employment, education and services in Downtown Pittsburgh," according to the SEA website.

I-579, known as the Crosstown Expressway, runs between the Hill and Downtown. The cap will essentially serve as a large bridge deck made of 8-inch-thick, reinforced concrete slabs supported by beams and pillars.

It will complement the Penguins' long-awaited $450 million residential, retail and office redevelopment plan for the 28-acre former Civic Arena site.

The SEA in 2016 received a $19 million federal grant for the work and is kicking in an additional $5.2 million. The remaining funding is coming from the state.

Council unanimously approved the allocation without comment.

Let's try to find out more about the Sports and Exhibition Authority (SEA) to see why they might be spending so much money to support the park project.

Directions: Select the link that will tell you more about the Sports and Exhibition Authority.

142

Exhibit 29. A Grade 12 critical online resource evaluation item that asks readers to analyze and evaluate the extent to which an organization has the appropriate qualifications to publish details about the proposed park plan on their website using a hybrid constructed response

Dynamic response items in the testing block can also be used to capture process data (e.g., how long students take to complete the item and the order of selections and answer changes) while assessing reading comprehension performance. The item in Exhibit 30, for example, asks readers to analyze and evaluate a small set of comments shared on social media in order to characterize the interests of different community members in relation to the proposed park plan. In this context, the drag-and-drop dynamic response format provides two additional functions; it serves as an alternative to writing each response as well as a **task-based UDE** to guide the language students use to classify comments into categories of accurately worded perspectives. This particular task-based UDE is also designed to introduce students to perspectives they will be asked to consider later in the testing block as part of the culminating Use and Apply task.

Exhibit 30. A Grade 12 dynamic response item that asks readers to analyze and evaluate four comments on social media. The drag-and-drop response format serves as an alternative to writing and also serves as a task-based UDE to guide students' classification of items into categories of accurately worded perspectives

As was noted in Chapter 3, NAEP should continue the trend of exploring the use of other interactive or dynamic response formats made possible with emerging digital tools. To that end, the next pair of items (Exhibits 31 and 32) serves to provide an illustrative example of how task-based UDEs might be used alternatively to compare how readers engage with comprehension items that use different types of response formats.

In both instances, readers are asked to categorize comments from community members about the park project and the intentional pairing of motivation and task-based UDEs serve to guide students and sustain their willingness to persist with multiple document inquiry tasks. Exhibit 31 applies a multiple-selection response format with a **task-based UDE (table) and motivational UDE (task character)** that serve to support readers as they engage in one particular item in the block. That is, the table is designed to first help readers focus their attention on relevant comments on the left side (rather than referring back to them in the original text) and then apply their understanding to the task at hand., match each comment with one or more specific benefits on the right.

At the beginning of the assessment block (see Exhibit XXX), students learn that the city has recently unveiled the park plan to the public on its website, and the plan is now open for public comment. City residents have posted comments and questions (depicted in a series of twitter posts).

Insert something here about the motivational UDE's in the authentic Twitter posts and the overpass knowledge-based UDE and the role they play in the early part of the block.

The four twitter posts are designed to inspire the question that guides readers' inquiry in the block: Why does Keisha matter to the city park project? As test takers proceed, they are introduced to this question (see Exhibit XXX) and the four task-specific purposes for engaging with the texts and comprehension items in the block: explore the background history of the Hill District, demonstrate an understanding of the texts they encounter, and craft an historically informed presentation for the general public that clarifies and illustrates Keisha's role in the park (e.g., representing and celebrating the history of the Hill).

Exhibit 10. Readers are Situated Within a Disciplinary Context and Broad Purpose in the Reading to Solve a Problem *Hill District* Block

Introduction

The City of Pittsburgh recently announced an ambitious plan for the construction of a highway **overpass** park that reconnects the Hill District and Downtown. The proposed park design has been posted on the city website. Read the public's comments to see what they are saying about the plan. Then, click next to continue.

A Cortland @cortland
Wow – this will be a great place to bring my kids! **#Hill District**

Jay Anderson @janders459
I don't understand why the city wants to spend their money on this park. Can someone explain? **#Hill District**

Pedro Carano @caranofamily
I like the idea of a park, but why should be it built on a highway overpass? **#PittCityPlanners** **#Hill District**

Ms. Peters @petersgrade8
I noticed in the park plan there were several signposts with a picture of a young girl named Keisha. Where can I read more about Keisha so I can talk with my students about how she fits in the planner's vision of the park? **#PS57**

NEXT

Your goal is to help the public understand the history of Pittsburg's Hill District and generate an answer to this question: **Why does Keisha matter to the park city project and what role does she play in fulfilling the park planner's vision?**

To accomplish this goal, you will:

1. Explore the background history of Pittsburg's Hill District to understand what has happened to "The Hill" from the 1940's to the present.
2. Examine different perspectives about the proposed park plan to determine what park visitors might consider confusing or concerning about the plan.
3. Apply your understanding of the issues to imagine how you would respond to the public's concerns.

You will work with Mr. Ramirez, a regional historian, and three teens: Brian, Diana, and Michael

Mr. Ramirez

NEXT

Tasks

Readers are asked to engage in purpose-driven tasks across multiple stages of reading (see Exhibit 4.18) to make sense of a focal problem, the historical context in which the problem is rooted, different perspectives on the problem, and the potential action in response to the problem. In the initial stage, students have opportunities to build background knowledge about the problem (i.e., people lack understanding of the Hill District and why Keisha matters). In the following stage, students will encounter multiple texts about the history of the Hill District, to help them explicitly understand ideas that might initially be confusing to park visitors. Topics are selected to help students build knowledge about various aspects of the Hill (e.g., vibrant cultures, thriving community businesses, discrimination, and segregation) to understand what it was like in the past and what has happened to the Hill from the 1940s to the present (e.g., urban renewal, demolition of the Hill, civil protests, civic arena and parking lot development). Students are supported in examining ideas from two different perspectives to help them to imagine a possible pathway to address the problem (e.g., how to clarify Keisha's role and why she is effectively positioned to fulfill the park planner's vision).

Texts

Readers are asked to comprehend and consult different forms of disciplinary texts and popular media texts. Historical texts may include both primary and secondary sources, such as historical photos and maps, archived black-and-white news articles, textbook-like written summaries, or visual timeline charts. Students may also be asked to read some online multimodal texts when learning about the problem and people's diverse opinions through news articles and website comments.

Readers carry out a series of historical reading tasks with specific purposes as they demonstrate the range of comprehension processes, such as those involved in close reading of a historical text, synthesizing within and across multiple texts, analyzing historical arguments

using textual evidence, employing historical frameworks such as social structures or historical patterns, evaluating historical interpretations, and demonstrating historical perspectives. These tasks and texts are also socially situated in that the purposes, processes, and consequences of reading are considered in relation to the challenges associated with urban development both locally, in Pittsburgh, and across the country.

When choosing texts, test developers take into consideration the length and level of complexity to ensure selected texts and related tasks are suitable for 8th graders completing the entire block in 20-40 minutes (e.g., passage length, structures, vocabulary, knowledge demands, motivational features).

Items

Comprehension processes are identified throughout the block and linked to an appropriate balance of items among the intended targets (Locate and Recall, Integrate and Interpret, Analyze and Evaluate, Use and Apply). Given that this is a Reading to Solve a Problem block, more attention might also be given to Use and Apply items (with less focus on Locate and Recall items), so that readers have time to fully develop and express their solution to the problem in a 40-minute timeframe. Item difficulties might increase throughout the block with variations in attention paid to unique text features and task demands as well as qualitative differences within each comprehension target category.

Universal Design Elements

As shown in Exhibit 4.19, the block design includes a range of digitally enhanced UDEs as readers comprehend texts, respond to items, and reflect on their performance. In the initial stage, an task character (a regional historian designated as a knowledge-based UDE) presents the reader with a primary purpose for reading; then, the reader (alongside task character classmates that represent motivational UDEs) is asked to decide how to conduct brief research to find out more about the history of Pittsburgh's Hill District and generate their claims and responses to the inquiry question.

We can insert a visual that illustrates what these task characters might look like with these dialogue prompts.

Task-based UDE's may include an image-based timetable that sequentially displays important local and national histories designed in the form of a graphic banner with pop-up notes. A list of keywords and relevant information offers a built-in knowledge support in the form of a searchable resource compilation (e.g., historical terms, specific names and places, civil rights movement). These task-based design elements (a graphic timetable and a searchable resource compilation) also serve as motivational UDE's in that they are designed to assist with organizing and analyzing information throughout the testing block while also helping to facilitate real-world connections and sustain 8th graders willingness to persist in this block's challenging collection of tasks.

We can insert a visual that illustrates what this timeline image might look like next to a list of keywords. Julie could draft an idea?

Diverse but intuitive response formats can be selected to facilitate reader engagement and reduce the cognitive memory load involved in expressing responses to test items designed to measure comprehension performance. Students are likely to benefit from embedded task guidance provided by task character guides and/or a graphical overview of block-specific reading

~~tasks to help monitor where they are and where they should focus their attention next to work toward the culminating task. Ultimately, decisions about UDEs should be specific to the block as test developers consider what is needed to fulfill the goal of obtaining comprehension scores that validly and fairly represent high-level comprehension processes in complex reading contexts.~~

In contrast, Exhibit 32 engages readers in a similar matching process, but for this item, a task character **(motivational UDE)** ask readers to move each comment into the appropriate cells of a table that is part of a retractable digital notepad **(task-based UDE** marked near a blue arrow to illustrate how it can be minimized and maximized on the screen as needed**)**; readers use the notepad to store, organize, and recall important details as they read across multiple sources to solve the problem at hand. Similar to how students engage in reading across multiple documents outside of a testing environment, the digital notepad enables students at several points in the testing block to click on the notepad (which makes the table appear) to add and organize details as they continue to learn more and build a deeper understanding about how different community members feel about the park project from their varied and diverse perspectives. Exhibit 33 illustrates how the same notepad could have been paired with a different item earlier in the task when students were reading on a different website.

Of course, as was also noted in Chapter 3, 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. Pairing the development of any innovative task-based UDEs with careful piloting efforts will ensure that design features yield their intended outcomes for as many students as possible.

Exhibit 31. A Grade 12 multiple-selection response grid item with a task-based UDE (table) and motivational UDE (task character) that serve to support readers as they engage in one particular item in the RSP block

Exhibit 32. A Grade 12 dynamic matching response grid item with a motivational UDE (task character) and task-based UDE (retractable digital notepad) that serve to support readers at multiple points in the RSP block as they read across multiple sources to solve the problem at hand

Exhibit 33. A Grade 12 dynamic matching response grid item with a task-based UDE (retractable digital notepad) that serves to support readers at another point in the RSP block as they read across multiple sources to solve the problem at hand

Culminating Task. Toward the end of the Reading-to-Solve-A-Problem task, the three task characters remind students they are close to accomplishing their goal. In the first part of the task (Exhibit 34), students are asked to use what they learned about what different community members think about the proposed park plan (as stored in their digital notepads) and apply that understanding to provide evidence-based descriptions of their benefits and concerns from a certain perspective in order to help the task characters submit their final report to the Mayor. By suggesting "this is a big task so can you help with two of the perspectives and then I'll find the other three?", the high-school aged avatars recognize the difficulty of the task and provide support, as a **motivational UDE,** while still asking students to demonstrate their ability to use and apply what they have learned about the views of different community members in preparation for the final report. Readers are also reminded that they have access to the four websites they read and their digital notepad **(task-based UDEs)** to help them accomplish this culminating task.

For the second part of the task, students are asked to share their own evidence-based views of the park proposal plan and the task characters promise to also include their opinions in their final report. This item serves to validate the student's own voice and agency as an important contributor to the group's final summary. Exhibit 35 illustrates how this item might look using a short-constructed response format, similar to those in existing NAEP assessment blocks, and Exhibit 36 is included to depict what an item might look like longer into the future, as NAEP continues to explore alternative response formats that offer authentic opportunities for students to choose their preferred response format (e.g., written or audio recording) to express their own opinions to the problem posed by this testing block. Again, pairing the development of these innovative features with new considerations for scoring and careful piloting efforts will ensure that design features yield their intended outcomes for as many students as possible while never unintentionally disadvantaging some populations of students.

Exhibit 34. This Use and Apply item with open-constructed response format illustrates the use of a task character (motivational UDE) that reminds students of their goal, recognizes the difficulty of the task, and provides support.

TASK 3. We are ready to summarize the views of different community members for our meeting with the Mayor. This is a big task, so can you help with two of the perspectives and then I'll finish the other three? Thank you!

Directions. Open your notepad to recall comments raised by different community members in the texts you read about the I-579 Cap Project. Think about how their comments reflect different perspectives. Then complete the items below. You can also select the notepad to view your notes or click the links on the left to view any of the sources you read.

Notepad

Sources

TRIB LIVE

SEA Website

Righting a Wrong Website

Next City Website

Part 1. Choose one perspective (safey, recreational, environmental, economic, or historical) and summarize briefly the benefits and concerns about the park proposal from that perspective.

Be sure to cite one at least one piece of evidence from the texts you read to support each benefit and concern you listed.

Part 2. Choose a second perspective (safety, recreational, environmental, economic, or historical) and summarize briefly the benefits and concerns about the park proposal from that perspective, using evidence from the texts.

We will include your written summary as part of our final report to the Mayor.

NEXT

Exhibit 35. This final Use and Apply item with open-constructed response format illustrates the use of a task character (motivational UDE) who reminds students they have accomplished their goal and validates the test-taker's role by inviting them to use what they learned and apply that understanding by sharing their own opinion.

Task 4. Now that you know more about the features of the park plan and the perspectives of Pittsburgh residents, you probably have your own opinion too! We'd love to include your own opinions in our final summary for the Mayor too.

Notepad

Directions. Imagine you lived in Pittsburgh and will attend the community meeting to express your views. Follow the directions to share your opinion. You can also select the notepad to view your notes or click the links on the left to view any of the sources you read.

Choose the perspective (recreational, environmental, economic, or historical) that best relates to your own interests in the CAP Project and summarize briefly what you think about the park proposal from that perspective. Support your thinking using evidence from the text.

Sources

TRIB LIVE

SEA Website

Righting a Wrong Website

Next City Website

Exhibit 36. This alternative format for the final Use and Apply item with open-constructed response format illustrates the use of motivational UDEs for two purposes: a task character who invites students' own opinion paired with an opportunity to choose their preferred format (text or audio) for expressing their opinion.

Performance Evidence and Indicators

. Scores from the Hill District block reveals what Grade 812 students can do when Reading to Solve a Problem in a social studies context. Ultimately, NAEP produces descriptions of what 8th12th graders (or sub-groups of 8th12th graders) can do in each disciplinary reading context. Thus, from students' participation in the Hill District block (and other assessment blocks designated as Reading to Solve a Problem in social studies contexts), it is possible to characterize how well eighth gradeGrade 12 students are able to comprehend and use multiple sources while engaging in social-studies inquiries involving a collection of relatively short but nonetheless complex multilayered digital texts and a range of digitally enhanced items and access tools.

Exhibit 12. Concept Sketch of a Reading to Solve a Problem Activity Block: Keisha Reconnects the Hill with Downtown in the City of Pittsburgh
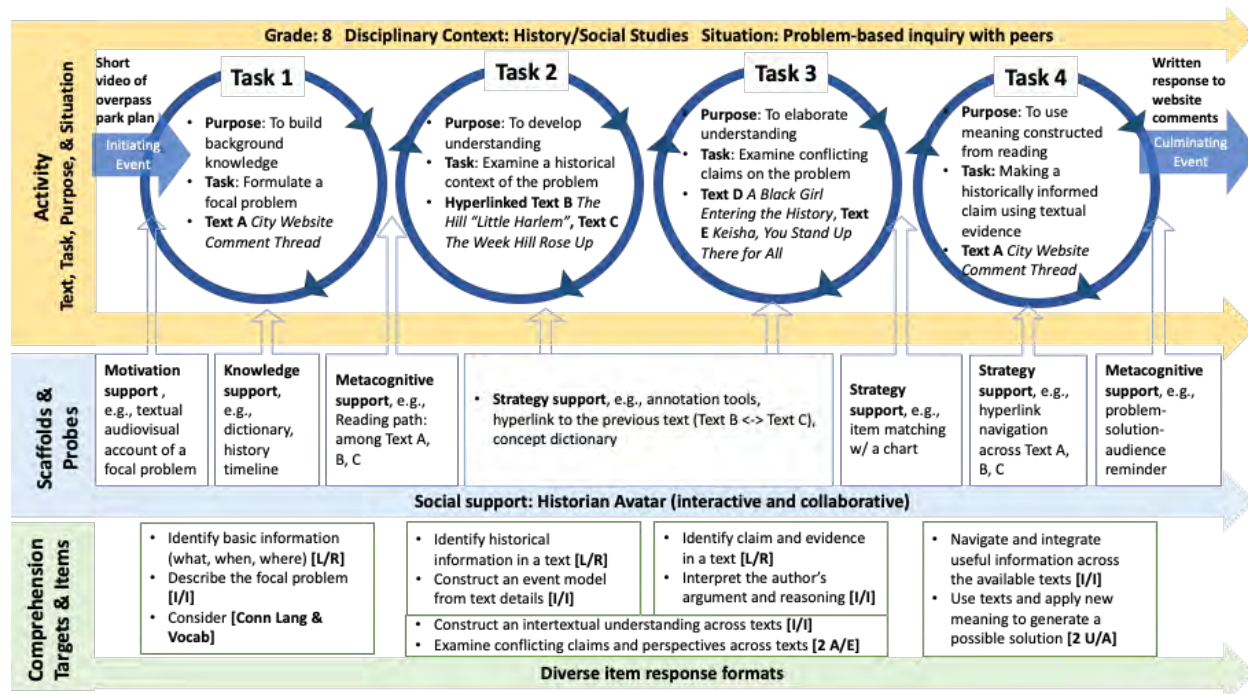
E.B. White

The last example offers a sketch of what a Grade 8 Reading to Develop Understanding in a Literature Context block might look like. This example illustrates what a block might look like if it occupied a space along the left end of the continuum portrayed in Exhibit 2. Here, students have more time to develop deep understanding of the texts. Tasks are relatively simple, and so fewer digital design features are needed to support the complexity of the task. When fully developed, this block should provide a good opportunity for students to demonstrate reading to develop understanding, by answering text-based questions that promote close reading of two texts as well as drawing inferences about how the ideas in the two texts inform one another.

Block Components (Disciplinary Context, Purposes, and Reader Role). In this example, students read and answer questions about two texts representing common literary genres: (a) a *biographical sketch* about the author E. B. White, and (b) a short human-interest *essay* by him. Some of the items will query the sketch, others will query the essay, and one item will require reasoning across the texts. These texts are a part of a NAEP released block that was used in the 2011 NAEP Assessment. The texts appear here (in Exhibits 46 and 47), as they did in that assessment.

At the outset, readers are provided a specific reading purpose and informed about the role (working on their own) they will be asked to assume during the block, composed of two common literary genres—a biographical sketch and a human-interest essay (see Exhibit 37).

Exhibit 37. Introduction to EB White

Introduction

You will read two texts: (1) a biographical sketch about the author E. B. White, most famous for writing *Charlotte's Web*, and (2) an essay that White wrote for The New Yorker magazine.

You will answer questions about each text. Then, you will explain how the description of E. B. White in DiConsiglio's biographical sketch applies or does not apply to the narrator of E.B. White's essay, *Twins*.

NEXT

Task Components: Tasks, Text(s), and Items). This E. B. White block has three tasks that include, 1) Reading and answering questions about the biographical sketch, *Not Just for Kids Anymore;* 2) Reading and answering question about the essay, *Twins*, and 3) Reasoning across the two texts to explain how what was learned in *Not Just for Kids Anymore* helps to understand E.B. White, the narrator of the essay, *Twins*. See exhibits See Exhibit 38, which shows task 1.

Exhibit 38. Introduction to the grade 8 EB White literature block

Task 1: Read the biographical sketch to learn about how DiConsiglio viewed E. B. White's career. Then answer the 4 questions about this text.

Meet the author: E. B. White, the author of children's classics *Charlotte's Web* and *Stuart Little*, was also a great essayist.

Not Just for Kids Anymore

By John DiConsiglio

"I have a lot of the cat in me," said author E. B. White, "and cats are not joiners."

Perhaps that is why White, one of the country's greatest writers, is so hard to label. His essays for *The New Yorker* appealed to an urbane crowd, but he is best remembered for his children's books. He loved the bustle of New York City, but was happy raising chickens on a Maine farm. And just when critics thought they had him pegged as a benign satirist, he'd write a biting condemnation of the dangers of technology.

Select the answer that shows why DiConsiglio thought it was hard to label E. B. White.

A ⃝ White was as happy in a crowd in New York City as on a farm in New England.

B ⃝ White was a great writer.

C ⃝ White was well-liked by many people.

D ⃝ White could write more than one type of prose.

NEXT

The comprehension items for Task 1 could help the reader develop understanding on

154

segments of the biographical sketch that focus on characteristics of White that might be useful in Task 3 (see Exhibit 39). Plausible segments for focus could be…

- The very first paragraph in which he compares himself to a cat.
- His adaptability (equally comfortable in NYC or Maine).
- Mood variation—benign satire to biting critique.
- The statement near the end suggesting that his essays matched his personality.
- The very last statement, suggesting that he was an eminently likeable character.

In terms of UDEs, note that there is a knowledge-based introductory UDE just before the title of the biographical sketch. Several relatively obscure terms are singled out as possible vocabulary pop-ups for a definition. No explicit motivational UDEs are provided.

Exhibit 39. Task 1 would involve additional items

Task 1: Read the biographical sketch to learn about how DiConsiglio viewed E. B. White's career. Then answer the 4 questions about this text.

……

Two additional items about DiConsiglio's biographical sketch would follow here
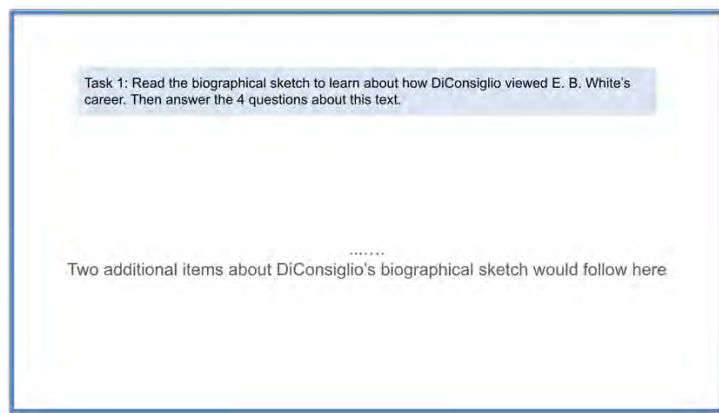
For Task 2, comprehension items should focus on the narrator White's statements that say something about his personality and attitudes toward the world around him (see Exhibits 40-42). Candidates for items include:

- Getting more than we bargained for and the sighting of the cow and her twins.
- White's characterization of the moose cow resentful of the onlookers
- The description of the mother and child as unaware of the special treat before their eyes
- The fawn's attempt to "hide" behind the leaf of the plant.
- One of several contrasts between the natural environment in a forest and the urban substitute of a zoo.

In terms of UDEs, similar to the biographical sketch there is a knowledge-based introductory UDE just before the title of the biographical sketch. Also several relatively obscure terms are singled out as possible vocabulary pop-ups for a definition. No explicitly motivational UDEs are provided.

Exhibit 40. Task 2 for the grade 8 EB White block illustrating an Integrate and Interpret item with a short constructed response item format

Task 2: Read E.B. White's essay in light of DiConsiglio's biographical sketch, making connections as you read. Then, answer the questions.

E. B. White was not only a great author for children, he was also the preeminent essayist of his time. This essay, written as a "Talk of the Town" piece for The New Yorker, provides a hint of his powers..

Twins

On a warm, miserable morning last week we went up to the Bronx Zoo to see the moose calf and to break in a new pair of black shoes. We encountered better luck than we had bargained for.

The cow moose and her young one were standing near the wall of the deer park below the monkey house, and in order to get a better view we strolled down to the lower end of the park, by the brook. The path there is not much traveled. As we approached the corner where the brook trickles under the wire fence, we noticed a red deer getting to her feet. Beside her, on legs that were just learning their business, was a spotted fawn, as small and perfect as a trinket seen through a reducing glass. They stood there, mother and child, under a gray beech whose trunk was engraved with dozens of hearts and initials. Stretched on the ground was another fawn, and we realized that the doe had just finished twinning. The second fawn was still wet, still unrisen. Here was a scene of rare sylvan splendor, in one of our five favorite boroughs, and we couldn't have asked for more. Even our new shoes seemed to be working out all right and weren't hurting much.

White begins the essay by suggesting that "*we encountered better luck than we had bargained for.*" What was the better luck?

Use the blank box below to type your response. Use evidence from the text to explain your thinking.

[NEXT]

Exhibit 41. Task 2 continues for the grade 8 EB White block illustrating an Analyze and Evaluate item with a multiple choice item response format

Task 2 (continued).

The doe was only a couple of feet from the wire, and we sat down on a rock at the edge of the footpath to see what sort of start young fawns get in the deep fastnesses of Mittel Bronx. The mother, mildly resentful of our presence and dazed from her labor, raised one forefoot and stamped primly. Then she lowered her head, picked up the afterbirth, and began dutifully to eat it, allowing it to swing crazily from her mouth, as though it were a bunch of withered beet greens. From the monkey house came the loud, insane hooting of some captious primate, filling the whole woodland with a wild hooroar. As we watched, the sun broke weakly through, brightened the rich red of the fawns, and kindled their white spots. Occasionally, a sightseer would appear and wander aimlessly by, but of all who passed none was aware that anything extraordinary had occurred. "Looka the kangaroos!" a child cried. And he and his mother stared sullenly at the deer and then walked on.

As the birth of the twins happened for all to see, how does the narrator contrast what he and his partner saw with what others saw?

A ◯ White and his partner saw something special but others saw only the ordinary. ⊖

B ◯ White and his partner saw a moose but others saw a kangaroo. ⊖

C ◯ White and his partner saw a monkey but others saw a the mother moose. ⊖

D ◯ They all saw the deer and walked on. ⊖

[NEXT]

156

Exhibit 42. Additional items accompany task 2

Task 2 (continued).

....
Three additional items about White's essay, *Twins*, would follow here

For Task 3, which was foreshadowed by the original block-specific purpose at the outset, both texts are involved. A task-based UDE, in the form of a partially completed note-taking chart (see Exhibits 43 and 44), might be provided to assist students in organizing their response to a final Use and Apply extended constructed response item (see Exhibit 45).

Exhibit 43. An Integrate and Interpret item illustrating a matching item response format

Task 3: Comparing ideas across the two passages

The final question (item 10) will require you to show how the ideas from *Not Just for Kids Anymore* apply to the narrator of the essay, *Twins*.
To prepare for that final item, fill out the chart below by moving phrases from the idea box into the the blank spaces in the chart.

1. Idea from *Not Just for Kids Anymore*	2. How this applies to the narrator of *Twins*
Cats are not joiners.	White and his companion hung back from the others who could see the moose.
He could adapt to many settings.	
	He was critical of the mother and child, who seemed not to appreciate the incredible good fortune of witnessing the twin birth.
He was comfortable on a rural farm with animals.	

Idea Box
- When at the zoo, the narrator was able to sit back and enjoy the birth of the twins.
- He showed great respect for the animals at the zoo.
- He is capable of biting criticism.
- He graduated from Cornell University.

NEXT

157

Exhibit 44. Integrate and Interpret item illustrating resetting of item responses from prior item

Task 3: Completed Chart: Comparing ideas across the two passages

No question to answer on this screen. Below you will see the chart from the previous page with the phrases from the idea box when they have been dragged into the right spaces in the chart. You can refer back to this chart when you complete the next (and last) item on this block.

1. Idea from *No Longer Just for Kids*	2. How this applies to the narrator of *Twins*
Cats are not joiners.	**White and his companion hung back from the others who could see the moose.**
He could adapt to many settings.	When at the zoo, he was able to sit back and enjoy the birth of the twins.
He was capable of biting criticism.	**He was critical of the mother and child, who seemed not to appreciate the incredible good fortune of witnessing the twin birth.**
He was comfortable on a rural farm with animals.	He showed great respect for the animals at the zoo.

Idea Box
- When at the zoo, the narrator was able to sit back and enjoy the birth of the twins.
- He showed great respect for the animals at the zoo.
- He was capable of biting criticism.
- He graduated from Cornell University.

NEXT

After completing the drag and drop task with the chart (Exhibit 43), students receive feedback about how the chart might best have been completed in Exhibit 44. The KB UDE, called resetting, is provided so that students do not carry misconceptions into the final item in Exhibit 45.

Task 3: Compare ideas across the two passages

For the final task, you will use ideas from the biographical sketch to support your thoughts about how ideas from DiConsiglio's biographical sketch apply to the narrator of *Twins*.

Use the underlined completed chart on the previous page or go back to either passage to get ideas to support your answer. Type your answer into the box below.

DONE

As suggested earlier, the EB White block sketch provides an example of what blocks might look in under the auspices of the 2026 when they are developed with an RDU Broad Purpose as the driving force in design. Blocks like these have long been a part of the NAEP Reading Assessment Portfolio and will continue to be included going forward. For the convenience of the reader, the full version of the two texts used for this block appear in Exhibits 46 and 47.

Exhibit 46. The First Text for the E. B. White Task: A Biographical Sketch. Meet the author: E. B. White, the author of children's classics Charlotte's Web and Stuart Little, was also a great essayist.

Not Just for Kids Anymore

"I have a lot of the cat in me," said author E. B. White, "and cats are not joiners."

Perhaps that is why White, one of the country's greatest writers, is so hard to label. His essays fo*r The New Yorker* appealed to an urbane crowd, but he is best remembered for his children's books. He loved the bustle of New York City, but was happy raising chickens on a Maine farm. And just when critics thought they had him pegged as a benign satirist, he'd write a biting condemnation of the

dangers of technology.
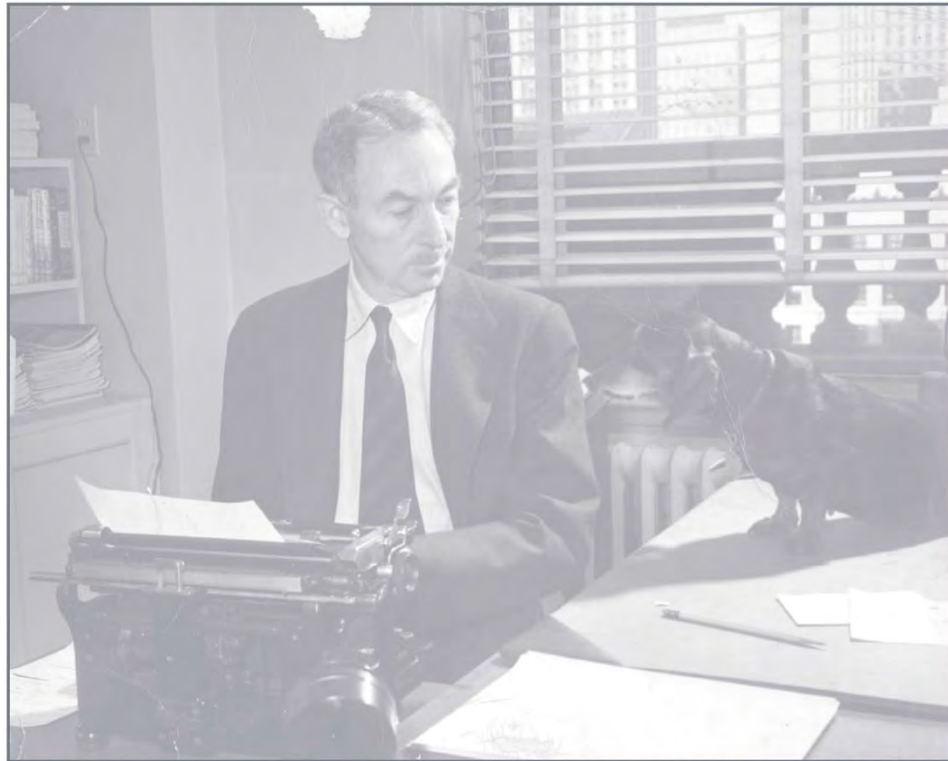
E. B. White and Minnie, his dachshund, at *The New Yorker* offices in the late 1940s.

The son of a piano manufacturer, Elwyn Brooks White was born in Mount Vernon, New York, in 1899. His family was prosperous, and White was raised with the mix of sophistication and common sense that would mark his writing. After graduation from Cornell University, White spent a year as a newspaper reporter in New York City, then decided to drive across the country with a friend in a Model T Ford. The trip gave White a lifetime of anecdotes, and spawned a legend or two. "When they ran out of money," White's friend, James Thurber, noted, "they played for their supper—and their gasoline—on a fascinating musical instrument that White had made out of some pieces of wire and an old shoe."

When White returned to New York City in the mid-1920s, he spent a few years bouncing between advertising jobs and unemployment before trying his hand again at writing Borrowing his brother's typewriter, he began pounding out sketches and poems. On a lark, he sent some essays to a fledgling magazine called *The New Yorker.* Since its founding in 1925, the magazine had struggled to find its niche, and White's work helped put *The New Yorker* on the map. His essays were funny and sophisticated; they spoke equally to socialites and cab drivers, professors and plumbers. Through his essays, which he wrote for nearly

50 years, White helped giv*e The Ne*w *Yorker* its voice and identity.

In 1945, already a leading literary figure, White embarked on his second career: writing children's books. He moved from New York to a farm in Maine, where he raised chickens and geese. Seeking a way to amuse his nieces and nephews, White started to write stories for them. "Children were always after me to tell them a story and I found I couldn't do it," he said. "So I had to get it down on paper."

A vivid dream about a mouselike character led to S*tuart Little*. Then, in 1952, White published *Charlotte's Web*. The book, which was inspired by White's own farm animals, is arguably the most famous children's story published in the 20th century.

By the time he died from Alzheimer's disease in 1985, White's essays had appeared in more college anthologies than those of any other writer. Many said his essays matched his personality: subtle without being simple, critical without being mean.

Indeed, one New *York Time*s critic wrote, "There are times reading an E. B. White book of essays when you think he must be the most likable man of letters alive. If you are some kind of writer yourself, you probably want to imitate him."

<div align="right">-By John DiConsiglio</div>

E. B. White was not only a great author for children, he was also the preeminent essayist of his time. This essay, written as a "Talk of the Town" piece for The New Yorker*, provides a hint of his powers.*

Twins

by E. B. White

On a warm, miserable morning last week we went up to the Bronx Zoo to see the moose calf and to break in a new pair of black shoes. We encountered better luck than we had bargained for.

The cow moose and her young one were standing near the wall of the deer park below the monkey house, and in order to get a better view we strolled down to the lower end of the park, by the brook. The path there is not much traveled. As we approached the corner where the brook trickles under the wire fence, we noticed a red deer getting to her feet. Beside her, on legs that were just learning their business, was a spotted fawn, as small and perfect as a trinket seen through a reducing glass. They stood there, mother and child, under a gray beech whose trunk was engraved with dozens of hearts and initials. Stretched on the ground was another fawn, and we realized that the doe had just finished twinning. The second fawn was still wet, still unrisen. Here was a scene of rare sylvan splendor, in one of our five favorite boroughs, and we couldn't have asked for more. Even our new shoes seemed to be working out all right and weren't hurting much.

The doe was only a couple of feet from the wire, and we sat down on a rock at the edge of the footpath to see what sort of start young fawns get in the deep fastnesses of Mittel Bronx.

The mother, mildly resentful of our presence and dazed from her labor, raised

162

one forefoot and stamped primly. Then she lowered her head, picked up the afterbirth, and began dutifully to eat it, allowing it to swing crazily from her mouth, as though it were a bunch of withered beet greens. From the monkey house came the loud, insane hooting of some captious primate, filling the whole woodland with a wild hooroar. As we watched, the sun broke weakly through, brightened the rich red of the fawns, and kindled their white spots. Occasionally, a sightseer would appear and wander aimlessly by, but of all who passed none was aware that anything extraordinary had occurred. "Looka the kangaroos!" a child cried. And he and his mother stared sullenly at the deer and then walked on.

In a few moments the second twin gathered all his legs and all his ingenuity and arose, to stand for the first time sniffing the mysteries of a park for captive deer. The doe, in recognition of his achievement, quit her other work and began to dry him, running her tongue against the grain and paying particular attention to the key points. Meanwhile the first fawn tiptoed toward the shallow brook, in little stops and goes, and started across. He paused midstream to make a slight contribution, as a child does in bathing. Then, while his mother watched, he continued across, gained the other side, selected a hiding place, and lay down under a skunk-cabbage leaf next to the fence, in perfect concealment, his legs folded neatly under him. Without actually going out of sight, he had managed to disappear completely in the shifting light and shade. From somewhere a long way off a twelve-o'clock whistle sounded. We hung around awhile, but he never budged. Before we left, we crossed the brook ourself, just outside the fence, knelt, reached through the wire, and tested the truth of what we had once heard: that you can scratch a new fawn between the ears without starting him. You can indeed.

Footnote

Sample items in the framework are being provided to exemplify key concepts in the framework and do not represent items that will be used on future NAEP assessments. These sample items may not represent accurately the full set of NAEP style guide and other test specifications. Tasks presented with multiple sample items are provided to help readers of the framework envision how theoretical ideas in the framework might guide assessment design, but they do not represent fully expectations for enacting the NAEP style guide and other test specifications.

Exhibit 1. Example of a Matching Selected Response Item for a Webpage Text from PISA's Rapa Nui Block

Exhibit 1, from PISA's Reading Literacy test for 15-year-olds, illustrates a matching item response format. After reading a webpage, students are asked to "drag and drop" the causes and effects offered at the bottom of the table into the appropriate places in the table.

Exhibit 2. Example of a Matching Selected Response Item from a Grade 12 PARCC Block

Choose **two** central ideas that are developed in the passage from "Biopiracy in India: The Case of the Aubergine." Drag each idea into one of the sections of the Venn diagram labeled Central Idea. Then, drag the quotation that illustrates the relationship between the two central ideas to the central section of the Venn diagram.

The government of India may be too disorganized to stand up to corporations like Monsanto.	Genetically-modified foods are currently controlled by a few large corporations.	Farmers in India have not been paid for the seeds and crops they have cultivated for generations.	Monsanto has not lived up to the pledges it has made in public statements.	Developments in agricultural technology should be used to help feed the people who are now going hungry.
"In response, the national biodiversity authority has announced its plans to prosecute Monsanto for carrying out this research without seeking its permission and the consent of hundreds of thousands of farmers who have cultivated these varieties for generations." (paragraph 5)	"More than half (53 per cent) of all genetically modified and organic seeds traded worldwide are owned by three multinational companies, according to the environmental group Greenpeace." (paragraph 12)

"This is all the more poignant in India, where thousands of debt-ridden farmers have in recent years resorted to taking their own lives to escape the misery of crop failure and financial ruin." (paragraph 12) | "The world's top ten agro-chemical companies own almost 75 per cent of all seeds globally." (paragraph 12) | | "In developing nations where farmers often rely on subsistence agriculture to eke out meager livings, the controversial and highly lucrative industry of genetic engineering is thrown into sharper relief against a backdrop of widespread poverty." (paragraph 12) |

Example 2, from a PARCC Grade 12 task, illustrates a matching format. Students are asked to "drag" the ideas into the venn diagram.

Exhibit 3. Example of a Zones Selected Response Item Format and the Use of Task Characters from ePIRLS' Mars Block

Exhibit 3, from an ePIRLS task for grade 4 students, illustrates a zones item format. The item asks students to "click on the website tab 'Rover Called Curiosity'." To do so, students must click on the tab of the webpage with the same title. This item also illustrates the use of task characters, or avatars. An animated icon of a teacher shows "Mr. Webster," and another one shows the "Student," who is the test taker.

Exhibit 4. Example of a Grid Selected Response Item from PISA's Rapa Nui Block

Exhibit 4, from PISA's Reading Literacy test for 15-year-olds, illustrates the use of a grid item response format to efficiently collect data about students' ability to analyze multiple fact/opinion statements.

Exhibit 5. Example of a Zones Item for an Internet Text from ePIRLS' "Elizabeth Blackwell" Block

Exhibit 5, from ePIRLS' assessment for grade 4 students, provides an example of the use of a zones item format. Here, students are asked to "Click on the link that is most likely" to have the requested information – in this case, "information about the life and achievements of Doctor Elizabeth Blackwell." This exhibit also illustrates the use of an Internet text in the form of a search engine results page.

Exhibit 6. Example of an In-line Choice Item from ePIRLS' Mars Block That Also Collects Process Data on Where Students Click on the Web Page

Exhibit 6, from ePIRLS' assessment for grade 4 students, asks students to use the digital diagram to answer questions by selecting responses from a drop-down menu (an in-line choice item). This item also collects **process data** of where on the graphic stimulus students click. While the clicks are not scored as items, they allow test makers to collect valuable information about why students might perform the way that they do. Such information can be useful for test development and also for outside researchers.

Exhibit 7. Example of a Short Constructed Response Item from PISA's Galapagos Islands Block

PISA

The Galapagos Islands
Question 5 / 7

Refer to the different webpages on the website on the right. Type your answer to the question.

The Conservation webpage cites two examples of programs that were undertaken to protect the giant tortoises.

What is the key difference in the approach taken between these two conservation programs?

THE GALAPAGOS ISLANDS – A NATURAL TREASURE

| About | Animals | Conservation | Volunteer |

Located 1000 kilometres west of the South American coast lie the Galapagos Islands - one of the most fascinating places in the world.

There are currently 95 indigenous species of animals that exist solely on the various islands of the archipelago. Many people travel to the Galapagos Islands to observe these special animals in their natural habitat. The islands are often referred to as a "living laboratory" because they offer scientists great research potential. Being near the equator, the islands receive ample sunshine, while the strong ocean currents provide cool breezes. Many plants and animals thrive in this environment. Tourists and scientists alike are fascinated by the animals who seem just as curious about humans as we are about them. Galapagos animals evolved for centuries without human interference or predation and consequently, when approached by humans, they don't show fear like most animals throughout the world. They often wander up to visitors! This behavior creates amazing photo opportunities, but it has made the animals very vulnerable.

Over the years, the ecosystem surrounding the Galapagos Islands has been threatened due to human activity on the islands. Damage to the ecosystem has had negative consequences on populations of many of the Galapagos animals. Thankfully, with the work of committed researchers, the ecosystem is slowly recovering.

The Galapagos Islands

South America
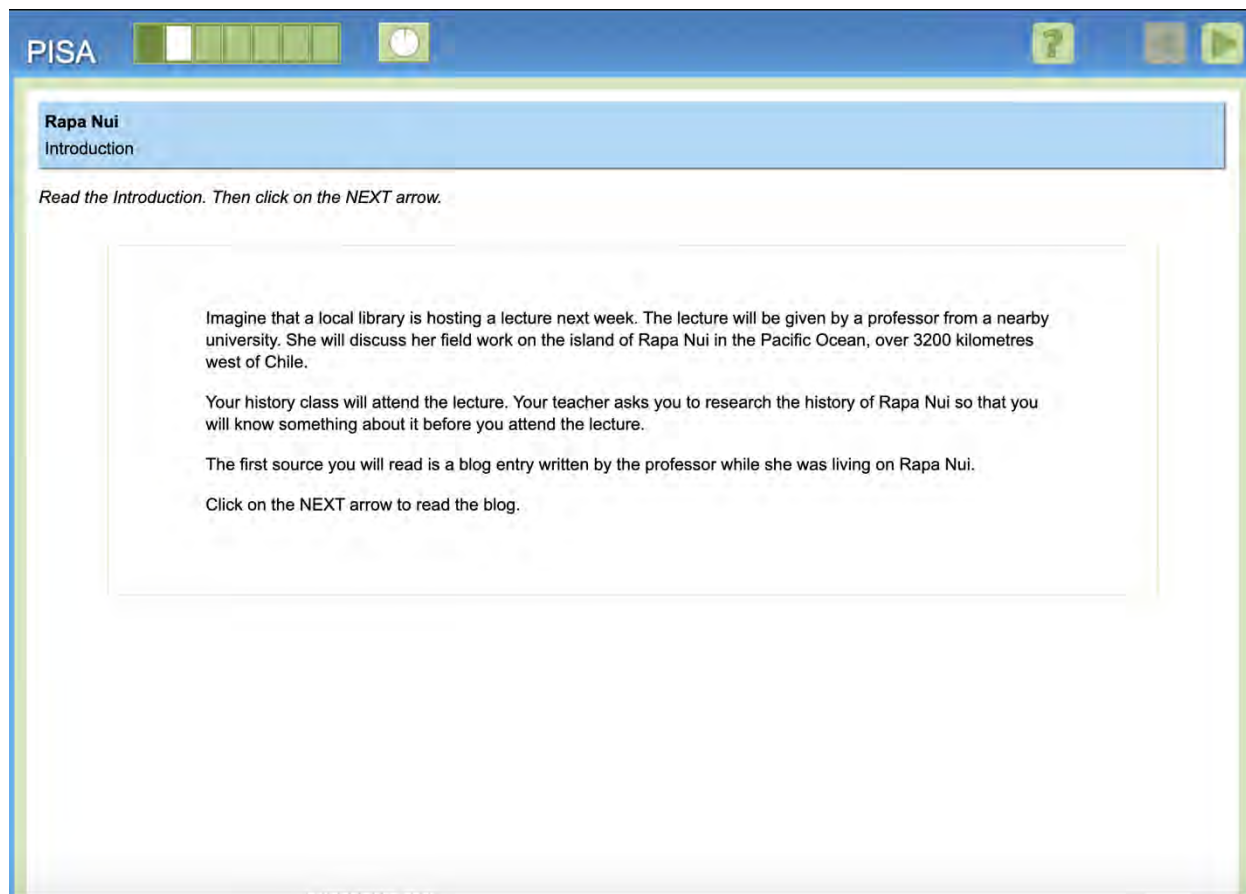
Exhibit 7, from PISA's Reading Literacy test for 15-year-olds, illustrates a short constructed response. Here, students are given a small text box and asked to write about a key difference they read about in the approach taken to two different conservation programs.

Exhibit 8. Example of a Fill in the Blank Item Response Format from ePIRLS' Mars Block

Exhibit 8, from ePIRLS for grade 4 students, illustrates the use of a fill in the blank item response format for a digital website text that is a graphic. Here, students are asked to use the graphic to identify the "names of the three planets between Mars and the Sun." To give their answers, students type each name ("Mercury," "Venus," and "Earth") into three separate text fields.

Exhibit 1. Example of a Specific Reading Purpose and a Knowledge-based UDE from PISA's Rapa Nui Block

Exhibit 1, from PISA's Reading Literacy test for 15-year-olds, illustrates how readers are situated, at the beginning of the block, within a specific reading purpose: To conduct research on the history of Rapa Nui in order to prepare for a lecture at a local library. This example also illustrates a knowledge-based UDE in which students are introduced to the first source they will read – a blog entry written by a professor while living in Rapa Nui.

Exhibit 2. Example of a Task-based UDE from the Smarter Balanced Items Published by The Regents of the University of California

Student Directions for Part 2

You will now review your sources, take notes, and plan, draft, revise, and edit your article. You may use your notes and refer to the sources. Now read your assignment and the information about how your article will be scored; then begin your work.

Your Assignment:

In your school, the Science Club is encouraging students to provide articles for its new website. For your contribution to the website, you will write an explanatory article about improving memory.

Using more than one source, develop a thesis/controlling idea to explain how to improve memory. Once you have a thesis/controlling idea, select the most relevant information to support your thesis/controlling idea. Then, write a multi-paragraph explanatory article explaining your thesis/controlling idea. Clearly organize your article and elaborate on your own ideas. Develop your ideas clearly and use your own words, except when quoting directly from the sources. Be sure to reference the source title or number when quoting or paraphrasing details or facts from the sources.

Explanatory Scoring

Your explanatory article will be scored using the following:

1. **Organization/Purpose:** How well did you state your thesis/controlling idea and maintain your thesis/controlling idea with a logical progression of ideas from beginning to end? How well did you narrow your thesis/controlling idea so you can develop and elaborate on the conclusion? How well did you consistently use a variety of transitions? How effective were your introduction and your conclusion?

2. **Elaboration/Evidence:** How well did you integrate relevant and specific information from the sources? How effective were your elaborative techniques? How well did you clearly state ideas using precise language that is appropriate for your audience and purpose?

3. **Conventions:** How well did you follow the rules of grammar usage, punctuation, capitalization, and spelling?

Now begin work on your explanatory article. Manage your time carefully so that you can:

- plan your multi-paragraph article,
- write your multi-paragraph article, and
- revise and edit the final draft of your multi-paragraph article.

Word-processing tools and spell-check are available to you.

For Part 2, you are being asked to write a multi-paragraph article, so please be as thorough as possible. Type your response in the space provided. The box will expand as you type.

Remember to check your notes and your prewriting/planning as you write, and then revise and edit your article.

Exhibit 2, from the Smarter Balanced test for grade 8 students, illustrates a task-based UDE in the form of scoring criteria and steps for writing an explanatory article. Additionally, the example illustrates the use of an extended constructed response item in the form of what would be a Use and Apply comprehension target in the 2026 NAEP Reading Assessment. See Appendix E for additional examples of different response formats.

Exhibit 3. Example of a Motivational UDE, from NAEP's "Tough as Daisy" Block

I'm the only girl at
the sign-up desk.

Tough as Daisy

by David M. Simon

The sign on the YMCA door says *Wrestling Tournament Today.*

I enter the gym and take a deep breath. It smells like old sweat socks and the stuff they use to wash wrestling mats.

I love that smell. Weird, huh? Not to me.

My dad always says, "Pound for pound, no one's as tough as Daisy."

I see my family in the stands. I wave to them and smile, but I'm nervous.

Lots of boys are already on the mats, loosening up. I'm the only girl at the sign-up desk. Some of

Exhibit 3, from a NAEP grade 4 block, illustrates a motivational UDE in the form of an illustration and caption. Together, the illustration and caption reading, "I'm the only girl at the sign-up desk." serve to pique readers' interest in the text. The illustration and caption also serve as a knowledge-based UDE because they introduce the text by offering key plot information (a girl standing in line, among only boys).

Exhibit 4. Example of Two Knowledge-based UDEs from NAEP's "Five Boiled Eggs" Block

Introduction: *Nasreddin Hodja, a character in this story, is familiar in many Turkish legends. "Hodja" means teacher.*

Long ago, a poor country boy left home to seek his fortune. Day and night he traveled, stopping to eat at inns along the way. Though he ate sparingly, his money quickly dwindled until, one day, no silver *akches* remained.

An *akche* is a unit of Turkish money.

Still, the boy kept walking, began to ache. Staggering up to the next inn he saw, he approached the innkeeper.

Exhibit 4, from a NAEP Grade 4 block, illustrates two knowledge-based UDEs. The first knowledge-based UDE appears in the form of an introduction to the story "Five Boiled Eggs," which introduces students to Nasreddin Hodja, a character in the story whose last name means "teacher" in Turkish. The second knowledge-based UDE appears in the form of a vocabulary pop-up box defining the Turkish word "akche."

Exhibit 5. Two Examples of Knowledge-based UDEs in the Form of Passage Introductions from a Released NAEP 2019 Block on E. B. White

Example 1

> **Meet the author:** *E. B. White, the author of children's classics* Charlotte's Web *and* Stuart Little, *was also a great essayist.*
>
> # Not Just for Kids Anymore
>
> "I have a lot of the cat in me," said author E. B. White, "and cats are not joiners."
>
> Perhaps that is why White, one of the country's greatest writers, is so hard to label. His essays for *The New Yorker* appealed to an urbane crowd, but he is best remembered for his

Example 2

> **E. B. White** *was not only a great author for children, he was also the preeminent essayist of his time. This essay, written as a "Talk of the Town" piece for* The New Yorker, *provides a hint of his powers.*
>
> # Twins
>
> by E. B. White
>
> On a warm, miserable morning last week we went up to the Bronx Zoo to see the moose calf and to break in a new pair of black shoes. We

Exhibit 5 illustrates two different written introductions, one for each of two texts. In Example 1, a knowledge-based UDE appears in the form of an introduction to an article about the writer E. B. White. In Example 2, a knowledge-based UDE appears in the form of an introduction to an essay by E. B. White, which explains that the author of the essay is also a children's author.

Exhibit 6. Example of Three Knowledge-based UDEs in the Form of Passage Introductions from the Michigan Student Test of Educational Progress

Source #1
You have found an article that describes how animals survive in different environments, the places where plants and animals live.

Source #2
You have found an article from *Appleseeds* magazine that describes how some animals build their homes.

Source #3
You have found an article that discusses plants and animals that live in the same place. The article describes how these plants and animals depend on each other to stay alive.

Exhibit 6, from Michigan's reading assessment for grade 4 students, illustrates three knowledge-based UDEs in the form of passage introductions for each of three different sources within a block. In this task, students are asked to learn from reading each source and to then write an informational article using what they have learned.

(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.

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.

Damasio, 1995

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.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review, 101*(3), 371-395.

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

Johnstone, C., Altman, J., & Thurlow, M. (2006). A state guide to the development of universally designed assessments. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

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*. London, UK: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., Enciso, 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

Markus and Kitayama 1991

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, C. (1979). The effect of background knowledge on young children's children's comprehension of explicit and implicit information. Journal of Reading Behavior, 11(3), 201-209210.

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. <u>Handbook of Multiple Source Use</u>, 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. (~~2010). Assessing the cultural validity of assessment practices: An introduction. Basterra, In M. R., Trumbull, E., & Solano-Flores, G. (Eds.). *Cultural validity in assessment: A guide for educators* (pp. 3-21). New York: Routledge.~~

~~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.

~~Taft, M. L., & Leslie, L. (1985). The effects of prior knowledge and oral reading accuracy on miscues and comprehension. *Journal of Reading Behavior, 17*(2), 163-179.~~

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/

~~Thompson, S. J., Johnstone, C. J., & Thurlow, M. L. (2002). Universal design applied to large scale assessments (Synthesis Report 44). Minneapolis, MN: National Center on~~

Educational Outcomes, University of Minnesota. Retrieved from
https://nceo.umn.edu/docs/OnlinePubs/Synth44.pdf

Thompson, S. J., Thurlow, M., & Malouf, D. (2004). Creating better tests for everyone through universally designed assessments. *Journal of Applied Testing Technology, 6*, 1–15.

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 den Broek, P., Risden, K., Fletcher, C. R., & Thurlow, R. (1996). A "landscape" view of reading. Fluctuating patterns of activation and the construction of a stable memory representation. In B. K. Britton, & A. C. Greasser (Eds.), *Models of understanding text* (pp. 165–187). Hillsdale, NJ: Lawrence Erlbaum Associates.

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.

~~Weiner 1985~~

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

Overview of the NAEP Reading Assessment and Projections

April 23, 2021

This document has been prepared in response to questions from the National Assessment Governing Board membership and staff regarding the current operational NAEP Reading Assessment in relation to the most recent draft of the 2026 updated framework. This document has three parts:

I. **Description of the Current Operational NAEP Reading Assessment in Relation to the Most Recent Draft of the 2026 Updated Reading Framework**
II. **Support Features, Relevant Research, and Development Processes in the Current Operational NAEP Reading Assessment**
III. **Implementing the Updated Framework and Maintaining Trend**

I. **DESCRIPTION OF THE CURRENT OPERATIONAL NAEP READING ASSESSMENT IN RELATION TO THE MOST RECENT DRAFT OF THE 2026 UPDATED READING FRAMEWORK**

Starting with the 1992 NAEP Reading Framework, a driving principle for the NAEP Reading assessment has been authenticity as a means of establishing face validity. Authenticity in the context of the NAEP reading assessment means that, to the extent possible, the assessment should reflect the reading experiences of students outside of the testing context. For example, the 1992 NAEP reading assessment was one of the first large-scale assessments to use only full-length, naturally occurring texts as its stimulus reading materials. The move to digital assessment under the current framework has allowed the NAEP reading assessment to reflect the digital reading experiences students encounter on a daily basis both inside and outside of school contexts. The draft 2026 updated framework continues to reflect the principle of authenticity.

Definition

The current NAEP Reading Framework lists the following definition: "The NAEP Reading Assessment is guided by a definition of reading that reflects scientific research, draws on multiple sources, and conceptualizes reading as a dynamic cognitive process." This definition

applies to the assessment of reading achievement on NAEP and states that reading is an active and complex process that involves:

- Understanding written text;
- Developing and interpreting meaning; and
- Using meaning as appropriate to type of text, purpose, and situation.

The draft 2026 updated framework maintains the current construct of reading comprehension while expanding the definition to include, "to explicitly acknowledge the sociocognitive processes of reading. Reading comprehension is defined as making meaning with text and four key features are highlighted—contexts, readers, texts, and activities." More specifically, the draft 2026 framework says the following:

"Reading comprehension is making meaning with text, a complex cognitive process shaped by students' social and cultural influences. To comprehend, readers:

- Engage with text in print and multimodal forms;
- Employ personal resources that include foundational reading skills, language, knowledge, and motivations;
- Extract, construct, integrate, critique, and apply meaning in activities across a range of contexts."

Testing Experience

The NAEP reading assessment transitioned from a paper-based assessment (PBA) to a digitally-based assessment (DBA) in 2017. In the most recent DBA in 2019, each student's assessment session began with a tutorial that included student interactions with the tools and interface and concluded with a 3-minute practice session. Following the tutorial and practice session, students worked through two 30-minute cognitive blocks. The second block was followed by a 15-minute survey questionnaire.

Texts

In accordance with the 2019 NAEP Reading Framework, which was first implemented in 2009, there are two broad categories of passages that make up the NAEP reading assessment: literary and informational. Literary texts include fiction, literary non-fiction, and poetry. Informational texts include exposition, argumentation or persuasive texts, and procedural texts.

The draft 2026 framework calls for three types of texts—literature, social studies, and science— and the texts in the 2019 operational pool fall easily into these three categories.

Items

After the passages are reviewed and approved by the Governing Board, items are written to assess three *cognitive* targets under the current framework. The current framework specifies the three cognitive targets as: Locate/Recall, Integrate/Interpret, and Critique/Evaluate.

The draft 2026 updated framework proposes four ***comprehension*** targets: Locate/Recall, Integrate/Interpret, Analyze/Evaluate, and Use and Apply. The addition of Use and Apply addresses the need to assess students' ability to apply the understandings they have gained from interacting with the stimulus materials for a given purpose (e.g., preparing a page of a website or writing a message to the school board).

The current NAEP Reading Framework calls for the following **item types**:

- **Selected response** – This item type encompasses traditional, single-answer, multiple-choice items as well as more complex items that require multiple selections to be answered correctly. NAEP's shift to digitally-based assessment allowed for the introduction of technology-enhanced items, which include matching (drag and drop), grid, and select-in-passage items. Most selected response items are scored dichotomously (correct or incorrect), but more complex selected response items may be scored for partial credit.
- **Constructed response, short and extended** – This item type requires students to generate a written response. Short constructed response items can be answered with a few words or sentences and extended constructed response items may elicit a short paragraph. These items are scored by humans, using a scoring rubric. Short constructed response items are scored with 2- or 3-point rubrics. Extended constructed response items use a 4-point rubric.

Percentages of each item type are specified in the framework for each grade. Typically, NAEP reading blocks include one extended constructed response item, three to five short constructed response items, and three to seven selected response items. The typical NAEP reading block includes a total of 9–11 items.

The draft 2026 updated framework recommends continuing with these item types and provides percentage ranges for selected response, short constructed response, and extended constructed response items. The draft framework also encourages the continued use and exploration of technology enhanced item types.

Reporting

Results of the NAEP reading assessment are reported on a 0–500-point scale. Three scores are reported at each grade level: a composite, or overall reading score, and two sub-scale scores, one for literary texts and one for informational texts. The draft 2026 updated framework maintains the 0–500-point scale and recommends reporting at each grade level: a composite score and three sub-scale scores—reading to engage in literature, reading to engage in science, and reading to engage in social studies contexts.

NAEP Contextual Questionnaire Items

Following the completion of two cognitive blocks, students respond to a 15-minute survey questionnaire. There are two sections to the Contextual Questionnaire: Core and Reading-specific. Core survey items collect information on students' demographic characteristics, opportunities to learn in and out of the classroom, and educational experiences.

Reading-specific survey items focus on reading-related activities and experiences in and out of school. These items are designed to inform interpretations of the results.

In addition to the student questionnaires, teachers and administrators in schools that participate in NAEP also complete their own NAEP Questionnaires.

The draft 2026 updated framework maintains the current approach to the survey questionnaires along with recommendations for changes to the specific items in the reading surveys.

Assembling the NAEP Assessment Via Assessment Blocks

Each NAEP reading assessment is administered in two 30-minute assessment blocks, followed by a 15-minute block of contextual items. Although each student sees only two blocks, there are multiple blocks in each operational assessment as shown in the chart below. Matrix sampling of students and blocks enables NAEP to cover a broad range of content, while also minimizing the burden for students and schools.

Table 1a summarizes the number of NAEP reading assessment blocks administered in the 2019 operational assessment for grades 4, 8, and 12. Typically, each block contains 9–11 items.

Table 1a. 2019 Operational NAEP Reading Assessment Blocks and Item Pool

	Grade 4	Grade 8	Grade 12
Total Number of Blocks	12	15	15
Total Number of Items	118	149	132

Types of Assessment Blocks

Currently, two types of blocks make up the NAEP operational reading assessment: discrete blocks and scenario-based task blocks.

Discrete item (DI) blocks provide general instructions for students to read the passage and provide answers to each assessment item relating to the passages that are presented. All texts and all items are always available for student access and use. The current operational pool of DI blocks is comprised of both transadapted and newly developed blocks as described below.

- **Transadapted blocks** – These blocks are digital renditions of the assessment blocks used in the paper and pencil era of NAEP. These DI blocks make up about two-thirds of the current operational assessment.

- **Newly developed blocks** – These blocks were developed specifically for a digital platform. To take full advantage of the digital format, some of these blocks use print texts and some use texts that are "digitally native" and multi-modal. Some passages

contain embedded hyperlinks and videos. (Note that videos are not used as introductions to texts.) Items addressing video content do so in relation to passage content.

Scenario-based Task (SBT) blocks use both print and digitally native, multi-modal texts. In contrast with DI blocks, students can only access texts and questions sequentially, as the SBTs control the order in which students read texts and items and respond to questions. In this way, students are presented with sources and stimulus materials as needed to respond to items. Videos appear both as part of the texts that students read and as additional content but are not used as introductions to texts. Items addressing video information always do so in relation to the written text.

Table 1b summarizes the number of NAEP reading assessment discrete and scenario-based blocks in the current operational assessment for grades 4, 8, and 12.

Table 1b. 2019 Operational NAEP Reading Assessment Discrete and Scenario-Based Blocks

Block Type	Grade 4	Grade 8	Grade 12
Scenario-based Task Blocks	2	2	2
Discrete Blocks (Transadapted)	7	10	11
Discrete Blocks (Newly developed for DBA)	3	3	2
Total	12	15	15

II. SUPPORT FEATURES, RELEVANT RESEARCH, AND DEVELOPMENT PROCESSES IN THE CURRENT NAEP OPERATIONAL READING ASSESSMENT

Consistent with the principle of authenticity, the current operational NAEP reading assessment uses *support features*, referred to as Universal Design Elements (UDEs) in the draft framework, that are intended to replicate the types of supports provided during reading instruction and practice in school and at home. One central principle is worth emphasizing: *all support features used in a particular block are available to all students who take that block.*

The types of support features available on the 2019 NAEP reading operational assessment include:

- Look-back buttons
- Pop-up notes
- Passage introductions
- Eliminate answer choice
- Highlighting and notetaking
- Text-to-speech on directions
- Zoom & selection of color themes
- Multi-part response frames

- Purpose statements*
- Avatars
- Graphic organizers
- Item foreshadowing
- Directions and transitions
- Item resetting

* Purpose statements are not considered UDEs in the draft 2026 updated framework.

Not all features are available in every block, but all of the current operational NAEP reading blocks include some support features. Some of these features are available for all reading blocks, and across other NAEP subjects, at the system level (e.g., highlighting, text-to-speech on directions, zoom, and color themes). Others are content-specific, including item look-back buttons, pop-up notes, passage introductions, and multi-part response frames (complex items with multiple components split into multiple response fields). Others appear only in SBTs, or a subset of SBTs, depending on the goals of the tasks, including block-specific purpose statements, avatars, graphic organizers and sequential directions and transitions.

The following subsection provides additional information about the use of pre-reading features, pop-up notes, and avatars and pop-up notes.

Pre-Reading Features

The current operational assessment includes two types of pre-reading features: ***block-specific purpose setting statements*** and ***introductions*** to specific texts, which have been developed in consultation with the Reading Standing Committee[1] and approved by the Assessment Development Committee (ADC) on behalf of the Governing Board. The current NAEP Reading Framework does not provide guidance on pre-reading features.

Purpose Setting

DI blocks include general directions to "read and answer the questions," but do not include block-specific purpose statements.

SBTs include both general directions and block-specific purpose statements. Block-specific purpose statements introduce a purpose for reading and describe the task students are to complete (e.g., gather information for a webpage or to compose an email message). Block-specific purpose statements focus on the tasks students will perform rather than on introducing specific texts. Block-specific purpose setting statements appear in six of the blocks (2 per grade) in the 2019 operational reading pool (17% of the pool).

Introductions

A small number of DI blocks include some information about the text students are about to read prior to reading. This prereading feature has appeared in NAEP Reading since before the

---

[1] The Reading Standing Committee is a diverse group of experts and state assessment staff in reading from across the nation. They advise as part of the assessment item development process, ensuring that NAEP assessment items align to the NAEP framework. There is a Standing Committee for each NAEP subject area assessment.

digitally-based assessment began in 2017. All introductions are written text; none are multi-media (video or audio) as was proposed in the draft 2026 framework.[2]

Passage-specific introductions appear in eight of the blocks across all three grades in the 2019 operational reading pool (23% of the pool). Five of these introductions were added by the test developers and three were part of the original source. In five of these instances, the introduction provides some information about the author. In two of these instances, the introduction provides context for passages that are excerpts.

Generally, there are no consensus assessment industry guidelines or standards for when/how to provide introductions, though there is an extensive research base on the role of prior knowledge in reading comprehension that provides some guidance. For example, seminal research on schema theory by John Bransford and his colleagues found that readers were only able to adequately demonstrate their reading comprehension skills with passages written in general terms when titles were provided that activated their schema/prior knowledge about the topics of the passages. This work, along with content analyses of instructional materials and cognitive labs with students, enabled NCES to implement passage introductions in the operational NAEP reading assessment.

In addition, introductions were deemed important by the Reading Standing Committee as a means of orienting the reader and as a response to the need for content and face validity evidence. In timed, on-demand assessments such as NAEP, brief framing can help to mitigate construct-irrelevant variance, and such introductions and framings are common in sources students encounter in their daily lives. Periodically, the NAEP program invites all states and participating Trial Urban District Assessment (TUDA) districts to review the entire pool of NAEP items. The most recent state/TUDA item review in 2015 included texts with introductions, and no concerns were raised regarding these features. Finally, text introductions appear in some state reading assessments, such as PARCC and Smarter Balanced.

The following examples of passage introductions from previous NAEP reading assessments come from released and publicly available blocks (i.e., not the current operational pool). The first is an introduction to a Turkish folktale called "Five Boiled Eggs." The second introduces an article about the writer, E. B. White, and the third introduces an essay by E. B. White, by explaining that the author of the essay they are about to read is also a children's author. The E.B. White passages appeared in the paper assessment and were released in 2011. The "Five Boiled Eggs" passage appeared in the paper assessment and was transadapted for the digital assessment in 2017 and released after that administration.

---

[2] Responding to the Governing Board's March 2021 Board meeting deliberations, the April 2021 draft of the 2026 framework update does not include multimedia introductions.

Example 1.

Introduction: *Nasreddin Hodja, a character in this story, is familiar in many Turkish legends. "Hodja" means teacher.*

Long ago, a poor country boy left home to seek his fortune. Day and night he traveled, stopping to eat at inns along the way. Though he ate sparingly, his money quickly dwindled until, one day, no silver *akches* ⊛ remained.

Still, the boy kept walking. Soon, however, his empty belly began to ache. Staggering up to the next inn he saw, he

Example 2.

Meet the author: *E. B. White, the author of children's classics* Charlotte's Web *and* Stuart Little*, was also a great essayist.*

Not Just for Kids Anymore

"I have a lot of the cat in me," said author E. B. White, "and cats are not joiners."

Perhaps that is why White, one of the country's greatest writers, is so hard to label. His essays for *The New Yorker* appealed to an urbane crowd, but he is best remembered for his

Example 3.

> **E. B. White** *was not only a great author for children, he was also the preeminent essayist of his time. This essay, written as a "Talk of the Town" piece for* The New Yorker, *provides a hint of his powers.*
>
> ## Twins
>
> by E. B. White
>
> On a warm, miserable morning last week we went up to the Bronx Zoo to see the moose calf and to break in a new pair of black shoes. We

Avatars

Avatars are task characters used to create a social context and facilitate purpose-setting and transitions in SBT blocks (no discrete blocks use them) but are in and of themselves neither purpose statements nor introductions. Two of the total pool of six SBT blocks, across grades, in the 2019 operational reading pool use avatars (6% of the blocks in the total pool).

Pop-up Notes

Pop-ups are indicated by buttons in the text that signal to students that they can read more about a word or phrase. These kinds of notes appeared on the paper-based assessment (PBA) as footnotes. Pop-up notes occur in three blocks in the 2019 operational pool (9% of the blocks in the total pool). Two of the three pop-up instances provide definitions of words/terms that may be unfamiliar to the reader and are important to overall understanding. The third instance presents information that was provided in the original text. There are no assessment items directly related to the information in the pop-up notes.

The following example shows a pop-up note from the passage "Five Boiled Eggs."

A substantial proportion, 63%, of the entire pool of reading blocks in the 2019 operational assessment *does not contain the pre-reading features* described above. These blocks could be characterized as providing opportunities for "cold reads" and will continue to be part of the operational assessment in 2026.

Relevant NAEP Research

As noted above, two types of blocks make up the NAEP operational reading assessment: Discrete Item (DI) blocks and Scenario-based Task (SBT) blocks. At each grade level in 2019 (as noted in Table 1b above), two of 12 grade 4 blocks are SBTs, two of 15 grade 8 blocks are SBTs, and two of 15 grade 12 blocks are SBTs. The remainder are DI blocks. A special study was conducted in 2018 to examine the SBT format, relative to the current framework. For this study, researchers created discrete versions of reading SBT blocks using the same texts and items for both versions. This special study compared student performance on the same set of items and passages in a DI block versus an SBT block.

Although this study was conducted before the framework update project began, it is relevant to conversations about the framework update because SBTs involve collections of support features, which are referred to in the framework update as Universal Design Elements (UDEs). Both SBT blocks and DI blocks include UDEs.

Three of the 15 UDEs in the draft 2026 updated framework only appear in SBT blocks (i.e., avatars, sequential directions and transitions, and item resetting). The remaining 12 UDEs,

including text introductions and pop-up notes, can appear in either SBT or DI blocks.[3] This study provides no information about specific UDEs. Instead, the study examines collections of UDEs in an SBT format.

This was a randomized control trial study with a total of 3,000 students, counterbalanced for version, genre (literary and informational), and block position at each grade. Both the SBT and DI versions of blocks were delivered on tablets. Consistent with the students' experience with DI and SBT blocks in the operational assessment, students were able to move among texts and items at will in the discrete version, whereas movement between texts and items was sequential in the SBT versions.

Key findings (The differences summarized below are statistically significant.):

- Students taking the SBT versions of blocks outperformed students taking the DI versions of block in four of the six blocks.
- The advantage for the SBT versions was consistent across all NAEP subgroups (gender, race, SES, disability, ELL). In other words, there is no differential effect for any subgroup.
- The advantage of the versions with support features was consistent for low- and high-performing students in four of the six blocks.
- For the four blocks for which performance on the SBT version was significantly higher, the differences in percent correct ranged between 2% and 8%, with an average of 5%.
- The SBT-DI special study provides some indication that SBT versions of items tend to be more engaging/motivating to students than DI versions of items. This tendency could contribute to students' higher performance on SBT versions of items, compared with DI versions of items.
- Generally, reading SBT blocks tend to be equally or more difficult than DI blocks, but when comparing SBT and DI versions of the same set of items, SBTs tend to be less difficult than their DI versions.[4]
- Speededness was more of an issue in SBT versions. Revisions were made to reduce speededness before these blocks became part of the operational assessment.

---

[3] Of the 15 UDEs listed in the February 26, 2021 draft of the reading framework update, 13 already appear in the reading assessment. The 2 additions would have been: student exemplars as mentor texts (a task-based UDE) and multimedia passage introductions (a knowledge-based UDE). However, multimedia passage introductions were removed from the latest draft of the framework update. Text introductions already appear on the assessment – see earlier sections of this document on (1) pre-reading features and (2) existence of "cold reads".

[4] Because NAEP uses an Item Response Theory (IRT) model to generate scores, adding more difficult items to the NAEP Reading Item Pool will improve measurement at the high end of the score scale, i.e., detect smaller differences in student achievement for higher performers. Conversely, adding less difficult items will improve measurement on the low end of the score scale, i.e., detect smaller differences in student achievement for lower performers. The IRT methodology for scoring ensures that adding harder items to the item pool will not artificially lower scores and that adding easier items to the item pool will not artificially inflate scores.

Other Standard Research and Reviews in NAEP Item Development

NCES implements a routine research and development cycle to develop every assessment block carefully before it is introduced to the operational NAEP assessment. Each new block undergoes systematic scrutiny, typically including these steps:

1. **Text Selection.** Texts and text sets are identified by the ETS reading item development team at a rate of four for every one text or text set expected to become part of the operational assessment. Proposed texts are reviewed by the ETS bias and sensitivity review team and the ETS editorial staff and are ultimately reviewed and approved by the NCES item development staff and contractors, and the Governing Board Assessment Development Committee (ADC).

2. **Initial Item Reviews.** After passages are approved, items are developed by the ETS reading item development team. Once draft items are completed, ETS reviewers conduct editorial, cold read, bias and sensitivity, and language accessibility reviews. They are then reviewed by NCES item development staff and contractors and the reading standing committee.

3. **Pretesting.** Following initial item review, items and support features are pretested, using:
   a. Cognitive interviews with individual students to determine how they respond to proposed new texts and comprehension test items. The purpose is to determine whether the tasks actually engage students in the intended comprehension processes.
   b. Tryouts under "live" testing conditions with 50–200 students from the target population to determine whether a wide range of students can complete the blocks within the allocated time and whether all of the parts of the block are working as intended.
   c. Usability studies, which test new item or passage interactions with small groups of students.

4. **Revised Item Reviews**. After items are pretested and revised by item developers, ETS reviewers conduct editorial, bias and sensitivity, and language accessibility reviews. They are then reviewed by NCES item development staff and contractors and the standing committee. Item revisions are adjudicated with NCES item development staff, and items are submitted to the Governing Board ADC for final review and clearance for piloting. Before piloting, state/TUDA reviews may occur.

5. **Piloting.** Proposed new blocks are folded into the administration of operational blocks of a live assessment. By comparing student and item performance across the new and the old blocks, NAEP developers can determine whether the new blocks effectively scale together with the old, measuring the same underlying comprehension construct.

6. **Post-pilot Reviews**. Following the collection of pilot data (n=2500–3000 students per form), the following groups review pilot data, item level analyses, texts, and items:

- ETS reading item development team
- ETS data analysis and reporting team
- ETS Differential Item Functioning (DIF) panel
- ETS bias and sensitivity review team
- NCES item development staff and contractors
- NCES data analysis and reporting staff and contractors
- ETS editorial staff
- ETS Reading Standing Committee
- Governing Board Assessment Development Committee (ADC)

III.   IMPLEMENTING THE UPDATED FRAMEWORK AND MAINTAINING TREND

This section provides information about the implementation of the updated framework and is based on the contents of the latest draft of that document.

Following Board adoption of an updated framework, it will take time to develop the assessment. As new content is piloted and approved, old content, in particular blocks transadapted from the paper-based assessment, can be phased out. Most importantly, this gradual item development for the updated framework allows for trend to be maintained.

The 2022 and 2024 assessments will be the last operational assessments that are fully aligned to the current framework. The 2026 assessment is projected to be the first operational assessment under the updated framework. The 2026 assessment would include both trend blocks from the 2022 and 2024 operational assessments and newly developed blocks piloted in 2024, being used for the first time in an operational assessment.

In the Governing Board's previous discussions of the updated framework, concerns were expressed that there would be insufficient carryover of content to maintain trend in 2026. However, the information below indicates that maintaining trend in 2026 is possible with careful planning. The projected contents of the next three operational assessments are as follows:

- 2022:
  - Grades 4 and 8 – trend content only (re-administration of 2019)
  - Grade 12 – no assessment
- 2024:
  - Grades 4 and 8 – trend content (all blocks carried over from 2022) plus new operational content (drawn from blocks piloted in 2017 and 2019)
  - Grade 12 – trend content only (re-administration of 2019)

- 2026:
  - Grades 4 and 8 – trend content (all blocks carried over from 2024) and new operational content (drawn from blocks piloted in 2024)
  - Grade 12 – no assessment

Projected Numbers of Blocks Available for the 2026 Operational Reading Assessment

The tables below include information about the numbers of blocks in each of the following two categories that will make up the 2026 operational assessment.

1. **Trend blocks**, which consist of discrete blocks from the 2022 and 2024 operational assessments, which do not include block-specific purpose statements, and SBT blocks from the 2024 operational assessments, which do include block-specific purpose statements.
2. **New operational blocks** developed to address new aspects of the updated framework, including block-specific purpose statements and the updated comprehension targets. These blocks are being used for the first time in the 2026 operational assessment and will not become trend blocks until they are administered operationally for the second time.

The proposed approach to a gradual implementation of the updated framework has been revised since the original Overview document was submitted to Governing Board staff just prior to the March 2021 Board meeting. The March 2021 version of this document suggested adding block-specific purpose statements to three existing discrete blocks at each grade and re-piloting them in 2024. However, the most recent plan retains the existing discrete blocks, as is, and redirects the funds that would have been used for modification and re-piloting of existing blocks to the development of new blocks under the aegis of the updated framework. The current plan provides for a carryover of blocks from the 2024 to the 2026 assessment of 80% at grade 4 and 83% at grade 8. **Although ultimately an empirical question, these percentages of carryover should allow for the maintenance of trend.[5]** (See Table 3a below.)

The current plan for the 2026 development proposes new pilot development of six blocks at grades 4 and 8[6] to yield four new operational blocks. It also assumes that blocks piloted in 2017, 2019, and 2024 will be approved for operational use and that there are no public releases prior to the 2026 assessment.

Tables 3a, 3b, and 3c provide information about the composition of the 2026 operational assessment based on the current plan.

---

[5] The current NAEP reading framework – adopted in 2004 and first implemented in 2009 – included no carryover from the previous framework (0 percent) and trend was maintained. To learn more about how trend was maintained for the 2009 NAEP Reading Assessment, see the Reading Trend Study description at https://nces.ed.gov/nationsreportcard/reading/trend_study.asp.
[6] Grade 12 will not be administered in 2026 and new grade 12 development is out of scope.

Table 3a. Projected Numbers of Blocks by Status available for the 2026 NAEP Operational Reading Assessment at Grades 4 and 8

Blocks	Grade 4	Grade 8
TREND	16 (80%)	20 (83%)
NEW OPERATIONAL	4	4
Total Blocks	**20**	**24**

As a result of needing to both maintain trend and introduce new content aligned with the updated framework, the 2026 operational assessment is projected to include more blocks at each grade than the 2022 operational assessment. The grade 4 assessment would contain 11 blocks in 2022[7] and as many as 20 in 2026, and the grade 8 assessment would contain 14 blocks in 2022 and as many as 24 blocks in 2026.[8] A larger item pool is also required to support reporting goals for the updated framework, including reporting for three subscales instead of the two subscales reported under the current framework.

All of the passages and items in the blocks that would be carried over from 2024 to 2026 are consistent with the updated framework. The block-specific purposes required by the updated framework will be present in 40% of the blocks at grade 4 and 33% of the blocks at grade 8.

Tables 3b and 3c describe the contents of the projected 2026 operational assessment at each grade broken down by subscale.

---

[7] Tables 1a and 1b showed that the 2019 assessment included 12 blocks at grade 4 and 15 blocks at grade 8. However, one cross-grade 4/8 block has been dropped for sensitivity reasons, resulting in 11 blocks at grade 4 and 14 blocks at grade 8 for the 2022 assessment. Blocks sometimes need to be dropped for sensitivity reasons if they address topics that might be disturbing because of recent or ongoing current events, e.g., a hurricane, a pandemic, etc.

[8] The actual number of 2026 blocks is contingent on the contents of possible public releases in 2022 and 2024.

Table 3b. Projected Number of Blocks available for the 2026 NAEP Operational Reading Assessment by Status and Subscale at Grade 4

Blocks	Reading in Literature	Reading in Social Studies	Reading in Science	Total Blocks
TREND	7	4	5	16 (80%)
NEW OPERATIONAL	*New development would include at least one block in each of the reading in social studies and science contexts.*			4
Total Blocks				**20**

Table 3c. Projected Number of Blocks available for the 2026 NAEP Operational Reading Assessment by Status and Subscale at Grade 8

Blocks	Reading in Literature	Reading in Social Studies	Reading in Science	Total Blocks
TREND	8	6	6	20 (83%)
NEW OPERATIONAL	*New development would include at least one reading in literature block.*			4
Total Blocks				**24**

Appendices 1 and 2 on the following pages depict the movement of blocks across the 2022, 2024, and 2026 assessments at grades 4 and 8, as well as the addition of newly developed blocks.

Appendix 1. Proposed Composition of the 2022, 2024, and 2026 Assessments at Grade 4 by Context and Status

2022 Assessment	2024 Assessment	2026 Assessment
Rdg in Lit Block 1	Rdg in Lit Block 1	Rdg in Lit Block 1
Rdg in Lit Block 2	Rdg in Lit Block 2	Rdg in Lit Block 2
Rdg in Lit Block 3	Rdg in Lit Block 3	Rdg in Lit Block 3
Rdg in Lit Block 4	Rdg in Lit Block 4	Rdg in Lit Block 4
Rdg in Lit Block 5	Rdg in Lit Block 5	Rdg in Lit Block 5
	Rdg in Lit Block 6	Rdg in Lit Block 6
	Rdg in Lit Block 7	Rdg in Lit Block 7
Rdg in Science Block 1	Rdg in Science Block 1	Rdg in Science Block 1
Rdg in Science Block 2	Rdg in Science Block 2	Rdg in Science Block 2
Rdg in Science Block 3	Rdg in Science Block 3	Rdg in Science Block 3
Rdg in Science Block 4	Rdg in Science Block 4	Rdg in Science Block 4
	Rdg in Science Block 5	Rdg in Science Block 5
Rdg in SocSt Block 1	Rdg in SocSt Block 1	Rdg in SocSt Block 1
Rdg in SocSt Block 2	Rdg in SocSt Block 2	Rdg in SocSt Block 2
	Rdg in SocSt Block 3	Rdg in SocSt Block 3
	Rdg in SocSt Block 4	Rdg in SocSt Block 4
	Pilot Block A	New Op Block
	Pilot Block B	New Op Block
	Pilot Block C	New Op Block
	Pilot Block D	New Op Block
	Pilot Block E	
	Pilot Block F	

KEY
Trend Block
New Operational Block
Pilot Block

Appendix 2. Proposed Composition of the 2022, 2024, and 2026 Assessments at Grade 8 by Context and Status

2022 Assessment	2024 Assessment	2026 Assessment
Rdg in Lit Block 1	Rdg in Lit Block 1	Rdg in Lit Block 1
Rdg in Lit Block 2	Rdg in Lit Block 2	Rdg in Lit Block 2
Rdg in Lit Block 3	Rdg in Lit Block 3	Rdg in Lit Block 3
Rdg in Lit Block 4	Rdg in Lit Block 4	Rdg in Lit Block 4
Rdg in Lit Block 5	Rdg in Lit Block 5	Rdg in Lit Block 5
Rdg in Lit Block 6	Rdg in Lit Block 6	Rdg in Lit Block 6
	Rdg in Lit Block 7	Rdg in Lit Block 7
	Rdg in Lit Block 8	Rdg in Lit Block 8
Rdg in Science Block 1	Rdg in Science Block 1	Rdg in Science Block 1
Rdg in Science Block 2	Rdg in Science Block 2	Rdg in Science Block 2
Rdg in Science Block 3	Rdg in Science Block 3	Rdg in Science Block 3
Rdg in Science Block 4	Rdg in Science Block 4	Rdg in Science Block 4
Rdg in Science Block 5	Rdg in Science Block 5	Rdg in Science Block 5
Rdg in SocSt Block 1	Rdg in Science Block 6	Rdg in Science Block 6
Rdg in SocSt Block 2	Rdg in SocSt Block 1	Rdg in SocSt Block 1
Rdg in SocSt Block 3	Rdg in SocSt Block 2	Rdg in SocSt Block 2
	Rdg in SocSt Block 3	Rdg in SocSt Block 3
	Rdg in SocSt Block 4	Rdg in SocSt Block 4
	Rdg in SocSt Block 5	Rdg in SocSt Block 5
	Rdg in SocSt Block 6	Rdg in SocSt Block 6
	Pilot Block A	New Op Block
	Pilot Block B	New Op Block
	Pilot Block C	New Op Block
	Pilot Block D	New Op Block
	Pilot Block E	
	Pilot Block F	

KEY
Trend Block
New Operational Block
Pilot Block

NCES Response to the Committee on Standards, Design and Methodology (COSDAM) Reading Framework Questions

April 22, 2021

This document has been prepared in response to questions from the National Assessment Governing Board's COSDAM regarding the current operational NAEP Reading Assessment in relation to the most recent draft of the 2026 updated framework. Three groups of questions are addressed in this document: 1) questions regarding Universal Design Elements (UDEs); 2) questions about the construct(s) being measured and the feasibility of maintaining trend; and 3) questions about implementation plans, projections, and budget considerations.

Universal Design Elements (UDEs) Questions

> What research evidence was used to implement the "support features" on the assessment, in particular the passage introductions?
>
> Is there existing evidence that knowledge-based UDEs are differentially effective based on students' prior knowledge?
>
> Is there any existing evidence regarding the "effect size" of UDEs on performance?

UDEs, such as introductions, have been part of the NAEP Reading Assessment since before NAEP became a digital assessment in 2017. In general, there are no assessment industry guidelines or standards for when/how to provide introductions, though there is an extensive research base on the role of prior topic knowledge in reading comprehension that provides some guidance. For example, seminal research on schema theory by John Bransford and his colleagues found that readers were only able to adequately demonstrate their reading comprehension skills with passages written in general terms when titles were provided that served to activate their schema/prior knowledge about the topics of the passages. This work, along with content analyses of instructional materials and cognitive interviews with students, provided justification for NCES to implement passage introductions in the operational NAEP Reading Assessment. It is also the case that text introductions appear in some state reading assessments, such as the Partnership for Assessment of Readiness for College and Careers (PARCC) and Smarter Balanced Assessment Consortium.

Examining the differential effectiveness of introductions and pop-up notes based on students' prior topic knowledge would require a study that includes measures of students' background knowledge. The NAEP program has not conducted any such study. Similarly, the NAEP program does not have evidence regarding the "effect size" of UDEs on performance because NAEP is not primarily a research program. NAEP relies on data from a variety of quantitative and qualitative sources to inform its development including cognitive interviews, small-scale tryouts, content reviews, and, occasionally, special studies. All new NAEP reading blocks are evaluated in a nationally representative pilot, followed by rigorous, block- and item-level analyses, and submitted for further review by the Governing Board's Assessment Development Committee (ADC). Weak or problematic blocks are not moved on for inclusion in the operational assessment.

Pretesting through cognitive interviews and small-scale tryouts was used to explore new UDEs introduced with Scenario-based Tasks (SBTs). The evidence from pretesting indicated that the majority of students reported that SBT UDEs were helpful and not distracting.

> How much time do existing UDEs add to testing? Would additional UDEs exacerbate this further?
>
> Is there any evidence about whether this additional time might hinder performance or be distracting?

The incorporation of UDEs in reading blocks is an integral part of the development of 30-minute blocks, as opposed to an "add on." Any potential time or cognitive burden they may pose is evaluated for each block as part of the development and pretesting processes via cognitive interviews and small-scale tryouts. Independent of the impact of UDEs, speededness is evaluated and addressed for all blocks as part of the development, pretesting, and piloting processes. The majority of the UDEs recommended in the draft framework are already included in the reading blocks in the NAEP operational assessment.

> Is NCES concerned about the framework's characterization of "support features" as UDEs given how the NAEP program already characterizes "Universal Design Elements?"

NCES describes "Universal Design Elements" as a form of accommodation that is available to all students. Some of the UDEs described by NCES, such as highlighting and zoom, are considered "task-based UDEs" in the updated framework. NCES is not concerned that the updated framework adds UDEs that are not specified in its current description of UDEs.

> What should be the main takeaways from the Scenario-based Task—Discrete Block (SBT-DI) study, relating to this framework update?

The main takeaway from the SBT-DI study is that regardless of ability (low vs. high), students performed better on the SBTs[1]. The main difference between the SBT and DI versions was the use of purpose-driven introductions and a broad range of UDEs in the SBTs. The support features in SBTs examined in the study are consistent with the updated framework.

Construct/Trend Questions

> In NCES's view, does the current framework and the framework update both allow for "cold reads?"

The assessment has not been operationalized in terms of "cold reads" as the term is not defined or discussed in either the current or draft updated framework.

> What evidence/ongoing studies/best guesses are there related to the likelihood of maintaining trend? Is this only a function of how many new blocks are needed?
>
> In NCES's view, do the newly proposed UDEs (mentor texts and multi-media introductions) represent a change in the construct of reading that can threaten trend (on a conceptual level)?
>
> In NCES's view, do other strictly digital UDEs from the framework represent a change in the construct of reading that should have already threatened trend (on a conceptual level)?

The likelihood of maintaining trend is a function of both how many new blocks are needed and whether these blocks differ qualitatively from the existing blocks in terms of what they measure. That said, there is a high likelihood of maintaining trend under the updated framework. Evidence of this comes from several sources. First, the construct of reading comprehension in the updated framework has changed very little from the construct in the current framework. This means that the passages and items developed under the updated framework will not differ significantly from those that were developed under the current framework. Second, the current plan of gradual implementation results in carryover of approximately 80% of blocks from the 2024 to the 2026 operational assessments. Finally, it should be noted that trend was maintained in the implementation of the current framework when there was no carryover from the previous assessment.

The possibility that the new UDEs recommended in the draft framework (mentor texts and multimodal introductions) could threaten the construct is also an empirical question that will be investigated through means such as pretesting and, possibly, special studies and evaluated

---

[1] The magnitude of the improvement, in terms of percent correct, ranged between 2% and 8% with an average of 5%. Although the SBT versions were less difficult than the DI versions of the same texts and items, evidence from the operational assessment indicates that SBTs are of equal or greater difficulty than the DI blocks in the operational pool.

through the various reviews described in the most recent memorandum—*Overview of the NAEP Reading Assessment and Projections*—to the ADC.

The "strictly digital" UDEs introduced with SBTs in 2019 are not a threat to trend either conceptually or empirically. Conceptually, they are consistent with the construct of reading comprehension in the current framework as a means of measuring students' ability to "use meaning as appropriate to type of text, purpose, and situation" (part of the definition from the current NAEP Reading Framework [2009]). Empirically, SBT blocks containing these UDEs scale with Discrete Blocks that do not include these UDEs.

Implementation and Budget Questions

> To what extent can the current reading item pool be used to implement the framework update? Specifically, how much re-field testing is needed and how much new item development is needed?
>
> What is the cost of implementing the framework update?

NCES's proposed approach to a gradual implementation of the updated framework has been revised since the March 2021 Board meeting (see also in this packet of materials—*Overview of the NAEP Reading Assessment and Projections*). The current plan increases the percentage of trend blocks carried over from the 2024 to the 2026 assessments, to 80% at grade 4 and 83% at grade 8. Although ultimately an empirical question, these percentages of carryover should allow for the maintenance of trend.

All of the passages and items in the blocks that would be carried over from 2024 to 2026 are consistent with the updated framework. The block-specific purposes required by the updated framework will be present in 40% of the blocks at grade 4 and 33% of the blocks at grade 8. The remainder of trend blocks include general purpose statements.

NCES expects to develop and pilot six new blocks at each of grades 4 and 8 to yield four new operational blocks. It also assumes that blocks piloted in 2017, 2019, and 2024 are approved for operational use, and there are no public releases prior to the 2026 assessment.

The cost of implementing the updated framework can only be determined when an updated framework has been approved.

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.

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

---

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.

> *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

Information from State Assessment Programs
April 28, 2021

21 of the 50 states responded to a recent poll conducted by the Council of Chief State School Officers (CCSSO).[1] Additionally, Board staff shared the same questions with representatives from Smarter Balanced and New Meridian (which manages the assessments formerly called PARCC).

- All 21 (plus New Meridian and Smarter Balanced) have definitions of reading comprehension that <u>do not explicitly</u> include topic or background knowledge as part of what is intended to be measured on their summative assessments.

- 19 of the 21 states (plus New Meridian and Smarter Balanced) attempt to mitigate the impact of background knowledge through selecting a wide range of passage topics.

- 17 of 20 states (plus New Meridian and Smarter Balanced) attempt to mitigate the impact of background knowledge by providing general information about the passage in the assessment, e.g., briefly introducing the topic, time period, author, or context of the passage. (1 state did not respond.)

- New Meridian, Smarter Balanced, and all 21 responding states attempt to mitigate the impact of background knowledge through defining terms in the passages that may be unfamiliar to the reader.

- New Meridian, Smarter Balanced, and 14 of 15 states attempt to mitigate the impact of background knowledge through another means. (6 states did not respond)

---

[1] The 21 states that responded include: Alabama, Arizona, Arkansas, Delaware, Georgia, Iowa, Kentucky, Maryland, Mississippi, New York, North Carolina, Ohio, Oklahoma, Oregon, South Carolina, South Dakota, Virginia, Washington, West Virginia, Wisconsin, and Wyoming

Draft Resolution in Honor of Michael Casserly
Executive Director of the Council of the Great City Schools

Approved XXX

Whereas, Michael Casserly has served as Executive Director of the Council of the Great City Schools for 29 years and is stepping down from his leadership position and assuming a role of consultant and advisor;

Whereas, through his insightful leadership of the Council, Michael Casserly inspired and initiated the Trial Urban District Assessment Program (TUDA) of the National Assessment of Educational Progress (NAEP), which assesses representative samples of students in urban districts across the United States through which invaluable data are provided for these districts to understand and to improve the educational experiences and outcomes of their cities' students;

Whereas, Michael Casserly strengthened and sustained the TUDA program for more than two decades, expanding the program from six districts in its first year to 27 districts that currently participate by eloquently explaining the unique value and immense power of TUDA data to urban school district policy and function;

Whereas, the TUDA data facilitate urban school districts' efforts to improve student performance and close achievement gaps by allowing districts to conduct comparative analyses with districts similar in demographic profile, to learn lessons from peers' experiences and successes, and to discern and implement effective education practices, with large cities making significant score gains and their performance on NAEP Reading and Mathematics improving faster than the nation;

Whereas, Michael Casserly's leadership of the TUDA Task Force with the Governing Board affords vitally important feedback to the Governing Board to inform and improve policy, research, and communications related to NAEP and to the TUDA program, leading to significant improvements;

Whereas, Michael Casserly's ardent advocacy for NAEP neither began nor concluded with the TUDA program in that he consistently and persuasively championed for support for the Nation's Report Card and worked tirelessly to advance the NAEP program through his savvy, sage, expert advice, his keen insights on strategic implementation of assessment programs, and his thoughtful, collegial collaboration with both the National Center for Education Statistics and the Governing Board members and staff;

Therefore, be it resolved that the National Assessment Governing Board expresses its profound appreciation and gratitude for Michael Casserly's unwavering support and monumental contributions to NAEP and student achievement in our nation;

Be it further resolved that a copy of this resolution be entered permanently into the minutes of the National Assessment Governing Board meeting of May 2021.

Signed on this Thirteenth Day of May, Two-Thousand and Twenty-One

Haley Barbour, Chair
National Assessment Governing Board

Executive Committee

May 5, 2021
3:00 pm – 5:00 pm ET
Zoom link to be sent

AGENDA

3:00 – 3:05 pm **Agenda Overview and Opening Remarks**
Haley Barbour, Chair

3:05 – 4:15 pm **NAEP Budget and Assessment Schedule (CLOSED)**
Lesley Muldoon, Executive Director
Peggy Carr, Associate Commissioner, National Center for Education Statistics

4:15 – 5:00 pm **Governing Board Priorities (CLOSED)**
Haley Barbour
Lesley Muldoon

5:00 pm **Adjourn**
Haley Barbour

Information Item   Strategic Vision 2025 Update   Attachment A

Strategic Vision 2025 Update

On September 29, 2020, the National Assessment Governing Board unanimously adopted Strategic Vision 2025. This approval marked the beginning of the development and implementation phase, which is managed by the staff and overseen by the Executive Committee. Staff will provide quarterly updates at the committee level and produce a comprehensive annual progress report every November. Because staff are in the process of developing work plans, this informational item is provided to the Executive Committee only in May 2021 and describes the current progress across committees. Subsequently, staff will provide updates through each of the standing committees.

Strategic Vision 2025 is organized by three pillars: *Inform, Innovate, and Engage*. Housed under the three pillars are eight strategic priorities. Since the March 2021 Quarterly Board Meeting, staff have initiated draft work plans that reflect the strategic priorities to be led by the Executive Committee; Assessment Development Committee (ADC); the Committee on Standards, Design and Methodology (COSDAM); and the Reporting and Dissemination Committee (R&D). The priorities led by each committee are:

Executive Committee	***INNOVATE:*** *Monitor and make decisions about the NAEP assessment schedule based on the Board's policy priorities of utility, frequency, and efficiency to ensure NAEP results are policy relevant. (SV 5)*[1]
ADC	***INNOVATE:*** *Optimize the utility, relevance, and timing of NAEP subject-area frameworks and assessment updates to measure expectations valued by the public. (SV 4)*
COSDAM	***INFORM:*** *Link NAEP resources with external data sources and disseminate what is learned from the sources so that NAEP can inform policy and practice in understandable and actionable ways. (SV 3)* ***INNOVATE:*** *Develop a body of evidence to improve the interpretation and communication of NAEP achievement levels to ensure that they are reasonable, valid, and informative to the public. (SV 6)*
R&D	***INFORM:*** *Identify the needs of stakeholders and refine resources to promote sustained use of NAEP data, enabling educators, researchers, advocates, and policymakers to understand and improve student achievement. (SV 1)* ***INFORM:*** *Elevate high-quality uses of NAEP resources to demonstrate NAEP's utility and to highlight the unique value of the Nation's Report Card to inform education policy and practice. (SV 2)* ***ENGAGE:*** *Develop, sustain, and deepen strategic partnerships to ensure that NAEP remains a trusted, relevant, and useful resource. (SV 7)* ***ENGAGE:*** *Help stakeholders understand how the Governing Board and NAEP can illuminate important skills for postsecondary education pathways. (SV 8)*

---

[1] To avoid the perception that the priorities are rank ordered, they are not numbered in Strategic Vision 2025. However, for the purposes of working documents, numbers are used for ease and clarity.

In developing work plans, staff focused first on the next 15 months; that is, identifying the work streams to be pursued through September 30, 2022. In drafting the work plans, staff are managing against the comprehensive resources of the team, including competing demands and timelines across priorities. Staff are also in the process of identifying the resources needed to reach targeted accomplishments, including whether contractual support is needed in any areas.

Below is a brief summary, organized by committee and priority, of the work planned through September 2022. Staff will continue to refine the work plans based on committee guidance and as resources are confirmed (i.e., budget and contractual vehicles).

Executive Committee

The Executive Committee is responsible for oversight of the Strategic Vision, on behalf of the Governing Board, and leads activities associated with the NAEP Assessment Schedule (SV 5). In service of SV 5, over the next 15 months the Board and staff will prioritize new and existing relationships with stakeholders in Congress, the Department of Education, and the government relations staff of partner organizations. As part of these partnership activities, the Executive Committee will collaborate with R&D on developing a strategy for communicating NAEP results to members of Congress and identifying productive ways to engage Congressional staff. These activities also contribute to SV 1 and 7, which are led by R&D and described later in this report.

To support the NAEP Assessment Schedule and the Board's policy priorities of utility, frequency, and efficiency, SV 5 will guide the Board's deliberations related to NAEP funding and the evolution of the next generation NAEP delivery platform.

Assessment Development Committee (ADC)

ADC is charged with leading the priority focused on NAEP subject-area frameworks and assessments (SV 4). Work plans for SV 4 involve completion of the Reading Framework update and initiation of the preliminary activities to inform a potential update to the Science Framework. To the latter, preliminary activities in Science include conducting an initial public comment period to collect input from the field on the state and relevance of the existing NAEP Science Framework; identifying experts to conduct a formal review of the existing framework; and hosting a panel discussion among those experts during a committee meeting to be held sometime in the fall. Turning the corner into Fiscal Year 2022, ADC will proceed, as determined by ADC and the Board, with revisions to the Science Framework.

The Board has commissioned two white papers to inform the design of future framework processes. In late spring or early summer, ADC will hold a joint meeting with COSDAM to discuss the current framework process and, as part of a continuous improvement effort, consider potential revisions to the process.

Associated with framework updates and a review of framework processes is ADC's plan to draft definitions of utility, relevance, and timing, which are central terms specified in SV 4. After drafting these definitions, ADC will engage the full Board in review and revision of those definitions, leading to development of a procedures manual to inform NAEP framework updates. As part of long-term planning, the ADC will map a timeline and schedule for upcoming framework reviews in accordance with the Governing Board's currently adopted assessment schedule, which extends through 2030. The current assessment schedule indicates that an update for NAEP Science will be implemented, if needed, in 2028 and updates for NAEP Writing, Civics, and U.S. History will be implemented, if needed, in 2030.

Finally, ADC plans to revise the Board's Item Development and Review Policy. This revision will include designing new item review procedures, trying out the new procedures, and creating an orientation video on those procedures so that all Board members -- and especially new Board members assigned to ADC -- are informed and well prepared.

Committee on Standards, Design, and Methodology (COSDAM)

COSDAM leads two priorities. The first focuses on linking NAEP to external data sources (SV 3) and the other on developing a body of evidence to improve interpretation and communication of NAEP achievement levels (SV 6).

With SV 3 over the next 15 months, COSDAM plans to collaborate with R&D to catalogue data already linked to NAEP and disseminate this information to stakeholders and partners. From there, COSDAM will lead the creation of a plan for increasing linkages, determine which additional linkages to take on, and begin conducting the necessary studies to do so.

With SV 6, COSDAM will use the existing National Assessment Governing Board Achievement Levels Work Plan, adopted in March 2020 and updated in July of that same year. Activities over the next year and half include:
- Hold panel meetings and conduct pilot and operational studies to review and revise achievement level descriptors (ALDs) in mathematics and reading, bringing forward from those studies recommendations to the full Board;
- Begin the studies to review ALDs in other subject areas;
- Convene ongoing advisory groups to discuss and provide feedback on materials for communicating NAEP achievement levels;
- Collect and evaluate information about current uses of NAEP achievement levels; and
- In collaboration with R&D, develop an interpretative guide for the NAEP achievement levels.

Reporting and Dissemination Committee (R&D)

R&D leads four cross-cutting priorities focused on NAEP--the resources needed (SV 2, 8) and used by NAEP stakeholders and partners (SV 1,7). Staff first identified how these four priorities intersect. They then considered the role and responsibilities of other committees in executing work on one or more of

these four priorities. A goal of the Board is that the Strategic Vision facilitates cross-committee collaboration. The priorities led by R&D are a natural starting point given the role, for example, of stakeholders, partners, and resources across the entirety of the Strategic Vision.

During its May 10 meeting, R&D will spend time discussing its Strategic Vision work plan. This work plan is highly informed by the Board's Outreach and Communications plan, also led by R&D and on its May agenda. Over the next 15 months, the Committee will:

- Prioritize partnerships and means of engagement;
- Meet with partners to determine their needs vis-a-vis NAEP and the Governing Board;
- Use insights from these partner meetings to develop events and resources, e.g., presentations, videos, graphics, and other materials which demonstrate NAEP's usefulness and relevance;
- Plan and execute 2019 NAEP Science initial release event and conduct post-release activities;
- Conduct activities that focus on the divergent trend line pattern seen in NAEP data across several subjects;
- Plan and host NAEP Transcript Study Release event and post-release activities that address partners' needs and connect to post-secondary preparedness;
- Design a strategy for ensuring appropriate interpretation and use of NAEP 2022 results; and
- Create a strategy to elicit deeper stakeholder engagement, building on FY2021 efforts.

Embedded within each of these high-level actions are numerous activities. For instance, prioritizing partner activities involves steps such as culling information gathered over the last two years from stakeholder input; conducting individual calls and meetings; convening focus groups; implementing a conference strategy, i.e., annual convenings hosted by partners; and establishing biannual meetings with partners to develop and sustain collaboration.

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

TABLE OF CONTENTS

iii

iv

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)

x

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

Paola Uccelli*
Professor, Education
Harvard University
Belmont, MA

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

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 century 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

---

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

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

3

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,

	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)	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	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 contextual variables of interest	Data collected from student, teacher, and administrator questionnaires on expanded set of contextual variables
	Some data collected from students' test taking behaviors (process data) in digital administrations	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*,

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.

9

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 cptures 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

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.

14

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

15

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

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

Assessment Components	Features of the NAEP Definition of Reading Comprehension			
	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

20

	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.

	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.

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.

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

23

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

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.

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.

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.

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

Welcome

You will read the story, *Hana Hashimoto, Sixth Violin*, by Chieri Uegaki and Qin Leng to prepare for a book discussion.

First, you will learn about **important events** in the story and **characters' thoughts, feelings, and actions.**

Then, you will **write about what the main character, Hana, is like as a person so that you are ready to discuss the book** with three peers.

Your teacher for this project will be Mr. Obas:

You will work with three classmates in your discussion group:

Gia   Gabe   Luisa

NEXT

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.

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).

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

27

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

28

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).

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).

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.

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

Deleted: accurate

Deleted: interpretations

Deleted: the validity of

Deleted: scores across

Deleted: valid

	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

34

	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."

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

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

(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.

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

36

Exhibit 3.8 Example of Grid Response Format from PISA

Chicken Forum Released Item #3

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.

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: 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).

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.

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¶

In this story, the main chara... Hana, decides to enter a tal... show to perform the violin, ... stringed instrument.

Before you read the story, s... the play button to watch a ... video of students playing st... instruments to hear the wa... sound.

After you watch th...

¶
As with task-based UDEs, these kinds of motivational UDEs align with UDA principles calling for "accessible, n... [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),

- 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.

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.

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.

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.

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.

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.

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.

49

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).

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

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 differences 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 context malleable 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

---

Margin annotations (tracked changes):

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).

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.

(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.

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/10.1177/0022057416196000303

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., Enciso, 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: [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:

o   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.

o   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.

- o 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.

- o 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 <u>does</u> test individual students but <u>does not</u> 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 – <u>Elements of Frameworks</u>: 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. (<u>Note</u>: Combines 2002 Principles 1 and 5.)

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

Principle 3 – <u>Framework Review</u>: 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. (<u>Note</u>: 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 – <u>Resources for the Process</u>: 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 – <u>Elements of Specifications</u>: shall be developed for use by NCES as the blueprint for constructing the NAEP assessment and items.

Principle 6 – <u>Role of the Governing Board</u>: 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 <u>NAEP subject area frameworks</u> 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 <u>NAEP subject-area frameworks and assessment updates</u> 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 <u>Education and the Workforce</u> of the House of Representatives, and the Committee on <u>Health, Education, Labor, and Pensions</u> 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 <u>and update</u> processes to ensure that the final Governing Board-adopted framework, specifications, contextual variables documents, <u>and their development processes comply</u> 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 <u>guidelines</u> 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. <u>Elements of Frameworks:</u> 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. <u>Define what will be tested and how, as well as</u>

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 <u>and updates</u>. 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 <u>contextual</u> 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.

Appendix E

TABLE OF CONTENTS

Appendix E

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

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.

Committee on Standards, Design and Methodology

May 3, 2021
12:00 – 2:00 pm ET (Virtual)

AGENDA

12:00 – 12:05 pm	**Welcome and Overview of Agenda** *Gregory Cizek, Chair*	
12:05 – 12:20 pm	**Update: Review and Revision of Mathematics and Reading Achievement Level Descriptions** *Eric Moyer, Pearson* *Sharyn Rosenberg, Governing Board staff*	Attachment A
12:20 – 12:35 pm	**Update: Framework Development Processes** *Gregory Cizek*	Attachment B
12:35 – 1:30 pm	**Discussion of NAEP Reading Assessment** *Gregory Cizek*	Attachment C
1:30 – 2:00 pm	**Discussion of 2022 NAEP Long-Term Trend** *Enis Dogan, National Center for Education Statistics* *Gregory Cizek*	Attachment D

Studies to Review and Revise NAEP Achievement Level Descriptions (ALDs) for Mathematics, Reading, and Other Subjects

Background

On September 24, 2020, the National Assessment Governing Board (Governing Board) awarded contract# 91995920C0004 to Pearson (as a result of a competitive bidding process) for conducting studies to review and revise NAEP achievement level descriptions (ALDs) in mathematics and reading using the 2019 NAEP assessments at grades 4, 8, and 12[1]. This work is intended to address the first recommendation of the evaluation of NAEP achievement levels that was conducted by the National Academies of Sciences, Engineering, and Medicine:

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

The Board committed to conducting studies to review and revise the NAEP ALDs in its initial response to the evaluation that was formally adopted and sent to the Secretary of Education and Congress in December 2016. The Board's Achievement Levels Work Plan, adopted in March 2020, further describes the intention for this work: "Addressing Recommendation #1 should focus on the current reporting ALDs for mathematics and reading at grades 4, 8, and 12. The methodology will be similar to what was done to evaluate the alignment and revise the 2009 NAEP Reading ALDs for grades 4, 8, and 12 (Donohue, Pitoniak, & Beaulieu, 2010) and the 2009 NAEP Mathematics ALDs for grade 12 (Pitoniak, Dion, & Garber, 2010). This process will generate new reporting ALDs that comply with the revised Board policy statement" (p. 3).

According to Principle 1a of the Board policy on Developing Student Achievement Levels for NAEP, "Content achievement level descriptions translate the policy definitions into specific

---

[1] The base period of this contract includes the review and revision of ALDs in mathematics and reading at grades 4, 8, and 12; in addition, an option may be exercised for a second phase of the contract focusing on review and revision of ALDs in U.S. history, civics, science, technology and engineering literacy (TEL) at grade 8 based on data from the most recent administrations of those assessments in 2018 and 2019.

expectations about student knowledge and skills in a particular content area, at each achievement level, for each subject and grade. Content ALDs provide descriptions of specific expected knowledge, skills, or abilities of students performing at each achievement level. They reflect the range of performance that items and tasks should measure. When setting achievement levels, the content ALDs provide consistency and specificity for panelist interpretations of policy definitions for a given assessment. During reporting, content ALDs communicate the specific knowledge and skills represented by *NAEP Basic*, *NAEP Proficient*, and *NAEP Advanced* for a given assessment" (p. 5).

Principles 3g and 4a of the Board policy apply specifically to this project of reviewing and revising the current ALDs and creating reporting ALDs (based on empirical data) that indicate what students at each achievement level *do* know and *can* do rather than what they *should* know and *should* be able to do[2]. Additional details for carrying out the work described by principles 3g and 4a are included in the Achievement Levels Procedures Manual.

The basis for the evaluation of NAEP achievement levels (and subsequently for this project) is the *existing* NAEP frameworks and item pools, not the new NAEP Mathematics Framework currently scheduled for implementation in 2025 or the NAEP Reading Framework that is currently under development and consideration by the Board. In accordance with principle 4b of the Board policy, the achievement levels and/or ALDs will need to be reviewed again once the new frameworks are implemented. Such work is beyond the scope of this project.

Project Overview

Dr. Eric Moyer is the project director at Pearson and Dr. Jennifer Galindo is the assistant project director at Pearson. Pearson will conduct a pilot study and an operational meeting using scale anchoring studies where panels of content experts judge the alignment of the current mathematics and reading ALDs and produce a set of recommended reporting ALDs for the Governing Board to consider in reporting the results from the next regular administration of the NAEP reading and mathematics assessments at grades 4, 8, and 12. The Governing Board is expected to take action on the reporting ALDs for mathematics and reading at grades 4, 8, and 12 in advance of the next release of these results.

Based on careful review of the history of ALD development, review, and revisions for NAEP mathematics and reading, a model-based anchored approach for reviewing the alignment of the ALDs for NAEP mathematics and reading will be used. The methodology for this alignment review study is based on that of previous studies, including the ALD development and review meeting held in 2009. The methodology was specified by the Board's Achievement Levels Work Plan and was selected to reduce the potential for possible inconsistencies from the use of different methods. The process of the model-based anchored approach will result in organizing

---

[2] According to the Board policy, ALDs will continue to describe what students *should* know and *should* be able to do for the purposes of item development and standard setting; only the reporting ALDs will be written in terms of what students *do* know and *can* do.

specific NAEP items by achievement level, which will serve as a key referent for panelists in reviewing and revising the current ALDs.

The model-based anchored approach includes three stages. The first stage will involve conducting statistical analyses to determine the items from the subject and grade that are anchored to a level corresponding to the score range within cut scores set to represent the achievement level descriptors (ALDs). The second stage relies on panels of content experts for each individual assessment. The panelists individually review the items that are anchored to each performance level and create summary descriptions of what students in each level are expected to know and be able to demonstrate based on the knowledge and skills measured by the items. In the final stage, the panelists compare the current ALDs for the respective assessment with their summary descriptions. The panelists note the similarities and differences, to make a recommendation regarding whether the current ALDs accurately describe what students in each level are expected to know and be able to demonstrate or if revisions to the current ALDs are needed to improve alignment. The final alignment judgment will be used to report whether the panels determined that there exists alignment between the current ALDs and student expectations. The final panel summary descriptions will be used to revise the current ALDs to create reporting ALDs that indicate what students at each achievement level do know and can do.

There is a technical advisory committee (TAC) consisting of the following experts in ALDs:

Dr. Karla Egan (Principal, EdMetric)

Dr. Ellen Forte (CEO and Chief Scientist, edCount)

Dr. Susan Loomis (Independent Consultant)

Dr. Marianne Perie (President, Measurement in Practice)

Dr. Mark Reckase (University Distinguished Professor Emeritus, Michigan State University)

Dr. Lauress Wise (Principal Scientist, Human Resources Research Organization)

The TAC is scheduled to meet for more than 100 hours (approximately 4 hours per month, with additional meeting time following the pilot and operational meetings) to provide technical advice on all aspects of the project to review and revise the mathematics and reading ALDs; this is intended to help ensure that all procedures, materials, and reports are carried out in accordance with current best practices, providing additional validity evidence for the process and results. In addition to frequent meetings and reviews of materials, two TAC members will attend the pilot and operational meetings to observe and provide feedback on the process.

Project Update (May 2021)

The COSDAM meeting on December 7, 2020 included a discussion of the proposed study design and plans for recruiting panelists and conducting the panel meetings virtually given the infeasibility of convening in-person meetings during early-to-mid 2021 in the midst of the COVID-19 pandemic. Committee members asked questions about and emphasized the

importance of protecting secure items in a virtual setting. Moyer explained that plans for maintaining item security were being documented and included the following safeguards: having panelists sign the NAEP non-disclosure agreement (which notes the severe penalties for violations) and repeatedly referring to it; providing Pearson laptops that are locked down and cannot be used for printing documents; using a secure Pearson server with high security protocols; setting up the standard setting platform with a single logon; ensuring that panelists cannot access secure materials outside of the scheduled meeting times; and visually monitoring panelists via Zoom video while they are working with secure materials.

The statement of work for this contract that the Governing Board issued on July 6, 2020 stated that the pilot study shall take place no later than February 2021 (to report results to COSDAM by March 2021) and that the operational study take place no later than early May 2021 (to hold a focused briefing session with COSDAM by the end of May 2021). This timeline was driven by the need for Board action in August 2021 in order to use the ALDs in reporting results for the NAEP 2021 Reading and Mathematics assessments for grades 4 and 8.

On December 27, 2020, Congress passed the Consolidated Appropriations Act of 2021, which rescheduled the mandated NAEP Reading and Mathematics assessments from 2021 to 2022. Consequently, the Board action to adopt reporting ALDs to be used for the release of these results is not needed until August 2022 rather than August 2021 as initially planned. Although there are many planned safeguards for protecting secure materials, there are more limitations inherent in a virtual environment. The recent change to when study results are needed for reporting the next administration means that Board action on the ALDs could be delayed from August 2021 to August 2022 to allow for the increased likelihood of conducting in-person meetings in late 2021 and early 2022.

In conjunction with the March 2021 COSDAM meeting, there was a brief project update indicating that there were plans to modify the project schedule to account for conducting the panel meetings in person in late 2021 and early 2022. A contract modification was executed on March 25 to implement these changes. With this revised schedule, the pilot meeting is being planned for October 25-28, 2021 and the operational meeting is being planned for February 22-25, 2022. Both of these meetings are intended to be in-person, with two TAC members attending to observe the process. To facilitate holding the meetings in-person, the project schedule for reviewing and revising the NAEP Reading and Mathematics ALDs has been extended, so the final ALDs will be presented for Board discussion at the May 2022 Board meeting and Board action at the August 2022 Board meeting. The intention is for the ALDs from this project to be used in the reporting of NAEP results in fall 2022.

The extended schedule also provided an opportunity for the project director, with the involvement of the TAC, to review multiple methods for anchoring assessment items to the achievement levels. The item anchoring to achievement levels is an important aspect of the ALD review process, given that the panelists' review of items associated with an achievement level will inform their statements about what students classified into each level actually know and can do. Several variations of methods for anchoring items to achievement levels were investigated and reviewed with the TAC, including the method used during the 2009 NAEP ALD review

studies, the anchoring method used for the Trends in International Math and Science Study (TIMSS), and anchoring methods using Item Response Theory (IRT) item mapping. The TAC recommended maintaining consistency with previous NAEP anchoring studies (as was also requested by the Governing Board in the Achievement Levels Work Plan) but developing a clear statement of what inferences can be made based on the methodology used.

The review of the anchoring method also included a reevaluation of the criteria for removing items based on item discrimination. In this context, item discrimination is used to differentiate between students classified into adjacent achievement levels. Various discrimination criteria were reviewed with the TAC to determine which process would be best for the study. The TAC recommended not using discrimination criteria even though it was used in previous NAEP anchoring studies because very few items would be removed anyway, and it is possible that these items may be needed for content reasons. That is, if items are dropped from reporting categories that already have a small number of items anchored to them, that will reduce the number of items even further and likely make the task more difficult for panelists.

The Design Document (attached) has been updated to reflect recent changes to the project. The document is very similar to the version reviewed by COSDAM in December 2020; primary changes are highlighted below:

- The panel meetings are intended to take place in-person rather than virtually per the March 2021 contract modification
- The discrimination criteria for determining anchor sets has been removed (page 22)
- At the recommendation of the TAC, the items for Reading will be reviewed by passage, with the items in the passage set being sorted by achievement level (pages 22 and 25)

Next Steps

Over the next few months, project staff will continue working to develop meeting materials and presentations; identify meeting space to hold the panel meetings[3]; and begin panelist recruitment.

---

[3] Project staff are working closely with Governing Board staff to plan the in-person meetings in compliance with current guidelines from the Centers for Disease Control and the Department of Education's Conference Approval group.

National Assessment Governing Board

Developing Achievement Level Descriptions for Mathematics and Reading

Submitted: April 2021

NAEP ALD Development Design Document

Table of Contents

The National Assessment of Educational Progress (NAEP), known as the "The Nation's Report Card," provides information on what students in the United States know and can do in various subject areas. As part of its legislative mandate for overseeing and setting policy for NAEP, the National Assessment Governing Board (Governing Board) develops achievement levels that further define expectations of what students should know and be able to do. Achievement on all NAEP assessments is reported using the following achievement levels, in accordance with the Board policy on Developing Student Achievement Levels for NAEP, which are defined as follows:

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

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

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

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

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

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

that they represent the knowledge and skills of students at each achievement level. Moreover, additional work to verify alignment for grade 4 reading and grade 12 mathematics is needed.

Recommendation #3: To maintain the validity and usefulness of achievement levels, there should be regular recurring reviews of the achievement-level descriptors, with updates as needed, to ensure they reflect both the frameworks and the incorporation of those frameworks in NAEP assessments.

In response to this evaluation of the NAEP achievement levels, the Governing Board developed and approved an Achievement Levels Work Plan in March 2020. To address the recommendations regarding the ALDs from the evaluation, the Governing Board issued a contract to Pearson in September 2020 to develop and conduct anchoring studies using NAEP 2019 data to review and revise the ALDs for grades 4, 8, and 12.

On behalf of the Governing Board, Pearson has developed this Design Document, which describes in detail the activities for the anchoring studies for grades 4, 8, and 12 NAEP mathematics and reading. This document is intended to provide the foundation for all ALD alignment review and revision activities. The Design Document will guide all aspects of the ALD review and revision process, including (1) a set of judgments about the alignment between the policy descriptions and current ALDs and expectations of what students should know and be able to do, based on the items that anchor to each achievement level for each assessment and (2) a set of revised ALDs that could be used as reporting ALDs, describing what students actually know and can do.

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

The model-based approach includes three stages. The first stage involves conducting statistical analysis to determine the items from the subject and grade that are anchored to each achievement level.

The second stage relies on panels of content experts for each individual assessment. Each panelist will review the items that are anchored to each achievement level to identify the knowledge and skills needed to respond to the items associated with a level. The panelists will then work together to develop common summary descriptions of the knowledge and skills that students in each level know and can do, based on the review of all items anchored to an achievement level. In the third stage, the panelists complete several alignment judgment rounds, comparing the current ALDs for the respective assessment with the summary descriptions. During the alignment judgment rounds, the panelists rate the degree of alignment and take notes regarding the similarities and differences between the ALDs and summary descriptions. The goal of this process is for the panel to make a recommendation regarding whether the current ALDs accurately describe what students in each level are expected to know and be able to demonstrate or if revisions to the current ALDs are needed not only to improve alignment but also to more accurately represent what students within each achievement level know and are able to do.

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

Below is a summary of what each section in the Design Document includes.

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

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

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

Section 4: *Achievement Level Descriptions Review Tasks and Procedures* describes the ALD review tasks, the nature of the tasks, and the procedures to be implemented prior to and as a part of the

operational ALD review meeting—including how panelists are trained and supported in implementing all activities.

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

Section 1: Achievement Level Descriptions Review Panels

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

Grade 4 Mathematics

Grade 8 Mathematics

Grade 12 Mathematics

Grade 4 Reading

Grade 8 Reading

Grade 12 Reading

A maximum of 48 panelists will be recruited for the operational study, with up to eight panelists recruited for each panel. Additionally, a maximum of 48 panelists with similar background distribution will be recruited for the pilot study, with up to eight panelists recruited for each panel. Panelists in each panel will be assigned to one of two replicate groups, allowing for a comparison of results across groups.

Panels for the ALD alignment review and revision operational and pilot studies will reflect an overall balance of gender, race/ethnicity, geographic location, and urbanicity, i.e., no more than 75% homogenous. Classroom teachers currently engaged in instruction in the respective grade and subject area will compose at least half of the panelists in each panel. At least two of the panelists recruited will be non-classroom educators with curriculum experience within the respective subject, such as state or local curriculum coordinators or higher-education faculty teaching education courses associated with the respective subject and level.

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

Overview of Panelist Recruiting Process

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

Phase 1: Identify nominators through allied organization and state departments of education; contact nominators and ask them to nominate outstanding classroom and non-classroom educators using an online nomination form (e.g., name, contact information, and basic qualifications). Nominators will be asked to briefly describe the rationale for that judgment when provided.

Phase 2: Notify nominees; request résumés and completed panelist forms (e.g., background in instruction with respective subject and grade, professional achievements, experience with students). Nominees will be asked to specify training and experience that makes them an outstanding candidate for panel selection.

Phase 3: Evaluate nominated candidates based on their background and experience; select the most qualified panelists and assign them to panel groups with respect to gender, race/ ethnicity, geographic location, instruction experience, type of institutional affiliation, and urbanicity.

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

To optimize recruitment, email and phone calls will be used to communicate with prospective panelists. An honorarium will be paid to panelists for the ALD alignment review and revision operational and pilot studies. Substitute teacher costs will be reimbursed directly to schools based on actual school costs for substitute teacher payments.

Identification of Panelist Nominators

Panelist nominators will be recruited using multiple sources. One source for nominators will be professional organizations that have a strong background in providing professional development in mathematics education or reading and literacy education. Focused will also be placed on professional organizations for minorities. Professional mathematics education organizations that should be used to recruit mathematics panelists should include:

- National Council of Teachers of Mathematics (NCTM)

- National Council of Supervisors of Mathematics (NCSM)

- Association of Mathematics Teacher Educators (AMTE)

Professional English language arts reading, and literacy education organizations that should be used to recruit reading panelists should include:

- National Council of Teachers of English (NCTE)

- International Literacy Association (ILA) (formerly the International Reading Association)

- American Literacy Corporation (ALC)

In addition to these organizations, state education organizations in mathematics or reading and literacy, state superintendents and departments of education, school board presidents, and district and school administrators of public and private education entities will be contacted in the four NAEP regions to propose qualified nominators across both panelist type (classroom teacher and non-classroom teacher educator).

Based on previous experience in recruiting NAEP panelists for achievement level setting meetings, Pearson estimates that 20 percent of the nominators will respond by submitting at least one nominee for consideration. Pearson further estimates that no more than 20 percent of the nominees would meet the qualifications, satisfy the requirements for representation, and agree to serve on the panel. For the pilot and operational studies, an estimated 2400 nominators must be identified to yield at least 480 active nominators, resulting in at least 480 nominees. Assuming that 20 percent of those

nominees will be eligible, meet the distribution requirements for representation on the panels, and be available and agree to serve as panelists, the yield would be the target of 96 panelists, with a target of 48 panelists for the pilot study and 48 panelists for the operational meeting. Pearson will supplement the number of nominators, as needed, to attain the panelists targets.

Prior to the beginning of identifying panelist nominators for the pilot meeting, the project director will meet with the COR and the Pearson meetings team to evaluate the current environment with regard to hosting in-person meetings and determine the feasibility of doing so. This will be reviewed prior to this step to ensure that communications to panelist nominators are accurate concerning in-person meetings. If it is determined, at this point, that in-person meetings will not be feasible, discussions will occur about other options prior to starting recruitment of panelist nominators.

Selection of Panelists

Nominees will be asked to complete an online questionnaire regarding their qualifications and experiences for serving on the panel. Candidates that present the credentials required will be contacted by phone to collect any missing information, verify the information provided, and confirm their willingness to serve on the panel, if selected. The goal is to select the most qualified panelists who are knowledgeable about the related subject at the appropriate level, while maintaining the goal of recruiting a mix of classroom teachers and non-classroom teacher educators for each panel.

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

Classroom Teacher Qualifications:

The nominee must meet <u>all</u> of the following qualifications:

- At least five years of overall teaching experience

- At least two years of experience teaching the related subject and the grade

- Judged to be "outstanding" in their professional performance by a nominator

Non-Classroom Teacher Educator Qualifications:

The nominee must meet <u>one</u> of the following qualifications:

- Non-teacher educational staff within school or district with education experience in the related subject and grade

- Curriculum director or content specialist serving schools at the related level or state department of education with education experience in the related subject and grade

- Postsecondary teacher education faculty teaching courses in the related subject and level

The credentials of panelists will be evaluated based on the number and importance of the credentials that are presented. Nominees having no distinguishing credentials will be scored low. Nominees having extensive credentials, including having been named outstanding teacher/teacher of the year and/or being actively engaged at the state or national level in professional activities within the specific subject and level, will score high. The scoring scheme differs for each panelist type (classroom teacher and non-classroom teacher educator). Nominees with the highest scores are given top priority by being considered the best-qualified candidates and being placed at the beginning of the candidate list. The selection process then selects panelists to reach the targets for representation listed above, with nominees having the highest qualifications being the first selected each time. All panels will be selected to have approximately equal proportions of males and females and equal proportions of representation from each of the four NAEP geographic regions. Every attempt will be made to create panels in which at least 25 percent of the panelists self-identify as a minority.

Prior to the beginning of both recruiting panelists and the final selection and notification of panelists, the project director will meet with the COR and the Pearson meetings team to evaluate the current environment with regard to hosting in-person meetings and determine the feasibility of doing so. This will be reviewed prior to this step to ensure that communications to panelists are accurate concerning in-person meetings. If it is determined, at either point, that in-person meetings will not be feasible, discussions will occur about other options prior to starting recruing panelist.

Each panelist for the pilot study and operational ALD review and revision meeting will be given an honorarium. Pearson acknowledges that the funds available to offer panelists are not commensurate with their contribution. They will emphasize that panelists' participation in the ALD review and revision project represents an exceptional contribution to education in the United States.

Section 2: Briefing Materials

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

- Confidentiality agreement

- Purpose and overview of the meeting

- Meeting roles and responsibilities

- Request for reimbursement form

- Meeting agenda

- NAEP framework for the relevant assessment

- Other materials identified as appropriate

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

In late October 2021, Pearson will conduct a pilot study of 4 days to implement the exact meeting procedure planned for the operational ALD alignment review and revision meeting. Conducting the pilot study at this time offers an opportunity to preview, revise, and resolve issues prior to the operational ALD alignment review and revision meeting in February 2022. To maintain uniformity of conditions, the pilot study and operational ALD alignment review and revision meetings will be held using the same process, including the same agenda of activities, and all steps in the pilot study will be the same as those planned for the operational ALD alignment review and revision meeting. By fully replicating the process of the operational study, the pilot will provide the information needed to determine whether any modifications are needed for the operational study.

The Technical Advisory Committee (TAC) will be consulted for advice and recommendations regarding details of the design prior to and after the pilot study. Up to two members of the TAC will be asked to observe the pilot study. All pilot and operational study materials will be based on data from the 2019 operational administration of the NAEP mathematics and reading assessments. The pilot study has the following goals:

- Determine whether modifications for training, instructions, materials, timing, and logistics will be needed for the operational ALD alignment review and revision study.

- Provide an opportunity for facilitators to practice the process before moving to the operational ALD alignment review and revision meeting.

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

A pilot study report will be prepared no later than December 2021, for presentation to COSDAM

during a webinar. The timing of the pilot study allows for preparation of the report and review of the

report by the COR, COSDAM and TAC so that improvements to the process can be made in advance of

the operational ALD review meeting.

Section 4: ALD Alignment Review Tasks and Procedures

Pearson will conduct an operational ALD alignment review meeting in February 2022. To maintain uniformity of conditions, the operational ALD alignment review meeting will be held with the same agenda of activities used for the pilot study, with the exception of adjustments made based on improvement to the pilot study process. This section describes the ALD alignment review procedures and tasks that Pearson will implement during both the pilot study and operational ALD alignment review meeting and includes information about the configuration of panels and materials, training of panelists, the collection of panelists' ratings, and the feedback given to panelists.

The current plan is to conduct both the pilot and operational meetings in-person. At this time, it is believed that an in-person meeting can be held safely and successfully, implementing sanitization and social distancing where possible. The CDC has prioritized educators in the vaccination schedule and it is expected that this will not negatively impact recruitment. However, the pandemic is being carefully monitored, and contingency plans are being developed.

The operational meeting will involve 48 panelists, with approximately eight per subject/grade. The panelists within each subject/grade will be assigned to two groups with approximately equal representation, which will be used as replicate panels during the meeting, since they will complete the same process receiving training and modeling from the same facilitator, however they group facilitation will be different. The TAC will be consulted prior to the operational meeting, and up to two members of the TAC will be asked to observe the meeting. All operational materials will be based on data from the most recent 2019 administration of the grade 4, grade 8, and grade 12 NAEP mathematics and reading assessments.

As previously indicated, Pearson will implement a model-based anchor approach and use the Pearson Standard Setting website platform to facilitate key aspects of the ALD alignment review process, including panelist training, review of ALDs, housing and reviewing of selected anchor item

sets, recording of panelists' feedback and ratings during each round of review, provision of feedback, and evaluation of the ALD alignment review process. The Pearson Standard Setting website will be set up to guide panelists through the steps of the ALD review process, with facilitators having the ability to restrict or provide panelists with access to sections of the site or activities, as needed. The night before the first day of the pilot study and the operational meeting, Pearson will reset the panelists' passwords for all panelists to a common meeting password. Additionally, access to the sites is controlled by Pearson administrators. The sites are deactivated overnight and during any extended breaks during the day.

The model-based anchored approach includes three stages. The first stage will involve conducting statistical analyses to determine the items that anchor to each achievement level (*NAEP Basic, NAEP Proficient,* and *NAEP Advanced*). Additionally, an anchor set will be developed for items that map to the higher end of the region below *NAEP Basic*. The second stage relies on panels of content experts for each individual assessment. The panelists in each committee individually review the items that are anchored to each performance level. While reviewing they items, they will write a description of the knowledge/skills required to answer the item correctly or receive a specific score point. It is from those individual item descriptions that the group level summary descriptions are developed for each achievement level. In replicate panels, they create summary descriptions of what students in each level know and can do based on the body of knowledge and skills measured by the items. In the final stage, the panelists compare the current ALDs for the respective assessment with the summary descriptions, noting the similarities and differences, to make a recommendation regarding whether the current ALDs describe what students in each level know and can do or if revisions to the current ALDs are needed to improve alignment. If it is determined that revisions are recommended, the panel provides recommended modifications to the ALDs.

Computer Use and Software

Pearson will use computers during the pilot study and the operational ALD review meeting. The use of computers and an online interface will reduce the time required for panelists to complete most steps in the ALD review activities. In addition, the use of computers will allow the panelists to interact with the items as students did, such that panelists are better able to understand what examinees would have to know or be able to do.

The computer will be a laptop computer (provided by Pearson) that is used by panelists to access the online Pearson standard setting website interface. With the review of materials required during the meeting, the panelists will also be provided an external monitor, which will provide them more screen area during the meeting to complete activities. As a part of supporting both efficiency and accuracy of the ALD review process, Pearson will use computers with access to the Pearson Standard Setting website, with a section specifically designed for the NAEP ALD review meeting. The NAEP website can be accessed only through a user identification code (ID) assigned by Pearson. Permissions will be set up for each user ID so each panelist can access only the materials he/she will be using.

The use of the website will reduce the time required for panelists to complete multiple steps in the process, since they will be able to access materials and complete activities within the website interface. As a website-based system, all materials and data will be stored, organized, and accessed through the website, which will ease the demands of development of materials and ensure the consistency of the materials with which the panelists interact. The use of the website will also ensure the security of the materials during the standard setting meeting, since the system will be used throughout the ALD review process and requires a secure login by all users. Pearson designed the interface for the NAEP ALD review process to have the following features:

- Simultaneous access by multiple users, with each individual user assigned a profile which defines their level of access to the site, including the materials the user can access and the specific functionalities available. Facilitators and site administrators can use conditional access

features based on date, activity completion, or custom settings to define access to materials and functionalities of the site.

- User access prior to the ALD review meeting, so users can interact with materials and experience training that will prepare them for the meeting and the ALD review process.

- Consistent user interface throughout the entire ALD review process, from pre-meeting work to post-meeting feedback.

- Management of storage and access to materials through the website, as a single access point for the meeting.

- Assignment of panelists to materials and items, based on panelists grouping.

- Embedded quality control features through conditional settings within activity that limit the range of responses to valid entries only and ensure no blank entries are accepted.

- Export of panelists' item review notes, as a group or individually. Facilitators will be able to view the panelists' notes within the website, to verify panelists' responses.

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

Preparing for the ALD Alignment Review Meeting

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

Development of Anchor Item Sets.

Prior to selecting the proposed anchoring approach outlined in this section, several variations were investigated to determine the effect of using different approaches, including the anchoring method used during the 2009 NAEP anchoring studies, item response theory item mapping approaches, and the anchoring method used for the Trends in International Math and Science Study (TIMSS). The investigated methods and the data were shared and discussed with the TAC. Each method resulted in different anchor item sets, which resulted from different conceptualizations of the set of students used to calculate the conditional probability utilized to anchor the item to an achievement level. It was

recommended by the TAC that the methodology from the previous alignment studies for mathematics and reading in 2009 (and specified by the Governing Board in the Achievement Levels Work Plan) would be the most defensible approach. This is because of the need to make consistent inferences about what the ALDs represent across the years by using the same anchoring methodology. That is, the anchor sets across years are created in the same manner and therefore the descriptions of the achievement levels represent the same range of students.

The development of the anchor item sets starts by grouping performances representing individual students from the most recent (2019) administration of the grade 4, grade 8, and grade 12 NAEP mathematics and reading assessments into achievement levels. The achievement level classification for each student is based on the average of their NAEP "plausible values" and the relationship to the boundaries of the achievement levels for the respective assessment. A student will be classified into either *NAEP Basic*, *NAEP Proficient*, or *NAEP Advanced* if their mean plausible value is at greater than or equal to the cut score for the respective achievement level. A student is classified into the region just below *NAEP Basic* when their average plausible value is below the cut score for *NAEP Basic* and the region between the mean plausible value and two standard errors above the mean includes the cut score for *NAEP Basic*. This approach will use all students in the NAEP sample from the most recent administration in 2019 to ensure that there are sufficient students associated with each achievement level for the analysis to determine each anchor item set. This is an approach has been utilized in previous NAEP anchor studies.

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

conditional *p*-values, one each for below *NAEP Basic, NAEP Basic, NAEP Proficient,* and *NAEP Advanced,* which represent the average performance on the item of the typical student within the four achievement levels. Items will be anchored to the first achievement level where the conditional *p*-values for the achievement level are greater than or equal to 0.67.

In 2009, in addition to the conditional *p*-values, item discrimination values were calculated for each item and achievement level and used as a determinant in the final anchor sets. Analysis of various discrimination criteria were reviewed with the TAC, including the initial criateria of the 40th percentile, a standard deviation criteria, a fixed value criteria, and using no criteria. Preliminary analysis across the different methods indicated that the discrimination criteria was removing very few items. In some grades of mathematics, however, the content areas had a limited number of items that anchored to them and dropping any item could risk adequate content coverage. The minimal number of items dropped due to the different discrimination criteria and the possible impact on the content coverage of the items resulted in the recommendation to not utilize a discrimination criteria for this study. After discussion with the TAC, it was decided that a discrimination criteria would not be used as part of the anchoring process.

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

working through the items that anchor to that achievement level. For the reading assessment, the items within a content area will be presented as sets with their associated passage. The passages will be ordered by average *p*-value. The items withing a passage set will also be ordered by achievement level and by average conditional *p*-value.

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

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

Training of Facilitators. The ALD alignment review and revision study will involve two content facilitators as well as a process facilitator per panel. There are six panels for each study, three

reading panels and three math panels. The content facilitators are selected for their expertise and experience in instruction in the related subject and grade as the framework they bring. The process facilitator is selected for the expertise and experience conducting meetings they have. Recognizing that facilitators may introduce individual difference that can result in slightly different instructions, content and process facilitators will be properly trained to implement the process uniformly and as intended. Pearson will prepare the PowerPoint presentations that facilitators will use during the meeting. In addition, facilitator handbooks will include the tables and graphs, a script for providing instructions, a description of the activities and an explanation of the feedback. Facilitators will attend a one-day virtual training prior to both the pilot study and the operational meeting. The project director overseeing the activities will lead the training. In addition, the facilitators and project director will do a walkthrough of the entire meeting the day before the pilot study and the day before the operational meeting.

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

Preparing Panelists for the ALD Alignment Review Process

Provide an Orientation. The operational meeting will have 48 panelists, with approximately eight panelists assigned to each panel. The six panels, along with the process and content facilitators, will begin the meeting with introductions and a description of the panelist recruitment process.

Provide an Overview of the Alignment Review and Revision Process. Facilitators will provide an overview of the purpose of the ALD alignment review and revision in general and description of

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

Individual Item Review and Summary Descriptions

Item Review. During the individual item review process, panelists will review all items within the anchor item set. Prior to panelists beginning the item review process, the facilitator will demonstrate how to use the website to review each item, view item information, and collect individual panelists' notes about each item using a practice anchor set. Additionally, the content facilitator will model to the panel how to approach the items in the anchor set and the process of developing item descriptions. The facilitator will also show how panelists how to convert the item description into a summary statement about what students know and can do for an achievement level. The panelists will work independently to review each item and create a description of the knowledge and skills demonstrated by students who answered the item correctly or who received a particular score for constructed-response items. During this step, panelists will review items within the same content area. For mathematics, items will be grouped by increasing achievement level (including the items anchored to just below the *NAEP Basic* level) and conditional *p*-value, from easiest to hardest. For reading, the passages will be ordered by increasing difficulty (from easiest to most difficult, based on average conditional *p*-value) and all items within a passage set will be presented to panelists before they review the next passage. Items within a passage will be ordered by increasing achievement level and conditional *p*-value. When

panelists have completed their review of the anchor items for an achievement level, they will be asked to write a single summary statement about what students within the three NAEP achievement levels (NAEP Basic, NAEP Proficient, and NAEP Advanced) know and can do, based on the item descriptions. Once panelists have completed their review of items across all achievement levels, they will be asked to consider what students in one achievement level can do compared to the adjacent achievement level(s). Panelists will complete the entire process for a content area before progressing to the next content area.

    Summary Descriptions. After individually writing descriptions for each item for a content area, the panelists will convene in their replicate panel to review the collected set of knowledge and skills associated with the items within an achievement level and construct a summary of what students performing in that content area and achievement level know and can do. (For reading, the panelist's item descriptions will be sorted by achievement level and conditional p-value so that they can reference the items during the discussions). Prior to beginning the replicate panel development of summary descriptions, the facilitator will model the process of using the item descriptions to develop summary descriptions by achievement level. This process will start with the *NAEP Basic* level, then the *NAEP Proficient* level, then the *NAEP Advanced* level. They will create their summary descriptions for each achievement level, within a content area, in the website using a shared document. The summary descriptions will be captured by the content facilitator assigned to that replicate panel. Once the replicate panel has completed the summary descriptions for a content area, they will then begin individual work for the next content area. After each content area of individual review, they will meet in the replicate panels to create the summary descriptions for that content area. When all content areas are completed, the panelists will progress to the alignment judgment rounds with replicate panel discussions, whole panel discussions, and ALDs revisions, if needed.

Alignment Judgment Rounds with Panel Discussions

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

Alignment Judgment Rounds. The replicate panels will be a significant part of the review and judgment process. After the panelists complete their individual item review the panelists will then meet in their assigned replicate panels to discuss their individual summary statements and to develop a replicate panel set of summary statements for that content area. They will toggle between individual review of items by content area and replicate panel development of summary descriptions until all sub-content areas have been addressed. To identify areas of alignment and lack of alignment, panelists will then work independently to make comparisons between the replicate panel summary descriptions with the policy descriptions and current ALDs. Panelists will be asked to assess the degree of alignment of the replicate panel summary descriptions by content area and achievement level to the policy descriptions and to the ALDs. If they indicate weak to moderate alignment, they will be prompted to provide suggestions/comments for why the alignment is weak and what modifications could be made. Additionally, they will make a judgment on the overall alignment of the summary descriptions as a whole to the policy descriptions and current ALDs.

After the first round of individual alignment judgments, panelists will meet back in their replicate panels and discuss why they made their specific alignment judgments. If necessary, they can make adjustments to the summary descriptions to clarify or modify the language to better indicate the original intended meaning. The panelists will then proceed with another round of individual alignment

judgments. They will then meet back as a whole panel and review summary statistics of the alignment judgments made and have a whole panel discussion similar to the replicate panel discussion. Where necessary, they will visit the summary descriptions to evaluate if the difference in the judgments relates to differences in the summary descriptions between the replicate panels. Panelists will be shown the summary descriptions from the two replicate panels side-by-side to evaluate the degree to which they are similar or different and develop a set of whole panel summary descriptions. The panelists will have one more opportunity to make individual alignment judgments based on the whole panel summary descriptions.

After the third round of individual alignment judgments, a whole panel discussion will take place. For the final whole panel discussion, panelists will be asked to focus on any content areas that still indicate greater than 50% panelist agreement of weak alignment or greater than 66% panelist agreement of moderate or weak agreement with the current ALDs. The panelist will review the ALDs for these subcategories and have the opportunity to suggest edits to the ALDs to improve alignment.

Cross-grade Review. Before the end of the meeting, the panelists will be brought together to review the reporting ALDs developed by each group, to review the terminology used across the achievement levels and grades to ensure that they clearly delineate progression of skills across grades and levels. This discussion will be led by one of the process and content facilitator teams. Any suggested revisions to the reporting ALDs will be determined by the panelists that constructed them, since they are most familiar with the content and discussions, but recommended changes may be provided by any panelists. The recommended ALDs will be the result of this final meeting.

Process Evaluations. Procedural evidence refers to the appropriateness of the procedures and how well those procedures were implemented. Evidence for procedural validity may come from a number of sources, including criteria for selecting panelists, the justification for the method, the quality of the implementation of the procedure, and the completeness of the documentation of the process. As

another source of evidence of procedural validity, panelists will be asked to complete evaluation forms after each major activity of the process. Evaluations will include both selected-response and open-ended questions that address the panelists understanding of the process and confidence in the results. No key-entry by staff is required because panelists will use the secure website interface to complete their evaluations. Panelist entries will be available for viewing using the facilitator login to the website. Facilitators will scan written responses for possible problems as they are collected during each day. Summary statistics will be computed for all ratings items and written responses. These analyses will be reviewed in real time throughout each day, and any sources of confusion will be identified for clarification with individual panelists or the panel as a whole. The operational meeting must be completed in time for the recommended ALDs to be used for reporting the next NAEP mathematics and reading assessments in grades 4, 8, and 12 during fall 2022.

Section 5: External Feedback

Pearson understands the potential contribution of external feedback to the review and revision of ALDs for the NAEP mathematics and reading assessments. Pearson has designed a strategy for collecting and using feedback from content experts and NAEP stakeholders for the panel recommendations for revising the ALDs.

Pearson plans to implement a vigorous and targeted approach to soliciting feedback regarding the panel recommendations through personal contact with key leaders and members of stakeholder groups, and through the use of a simplified and directed format for reviewer response. Pearson will create a website to obtain external feedback on the panel recommendations for revising the ALDs resulting from the operational ALD review meeting. The website will provide a means for stakeholders and the public to find information about the study design and the panel recommendations and to provide feedback. Prior to opening the site to the public, Pearson will submit the site to Governing Board staff for review and approval.

Pearson will solicit comments from content persons, technical persons, and persons in education-related organizations that are known to use NAEP. When organizations are engaged, a personal call from project staff will establish the initial contact and explain the importance of the request. For collection of comments from members of organizations, Pearson will ask the organization leadership to communicate the request for feedback to the membership and to encourage members' cooperation and response. In all cases, a personalized email message will be sent to explain the purpose of the request for comment and to provide a link to the website having background about the project, instructions, and questions for respondents to consider. Reviewer comments can be recorded through the link, downloaded by Pearson staff, and saved for analysis. Follow-up emails will be sent to encourage responses and to contact individuals for clarification and additional information, should that be necessary. The feedback will be summarized and presented to members of the panel during a virtual meeting, to allow the panelists to make any revisions to their recommendations before the final ALDs are presented to the Governing Board.

References

Bourque, M. L., (2009). *A history of NAEP achievement levels: Issues, implementation, and impact, 1989-2009*. Washington, D.C.: National Assessment Governing Board.

Donahue, P., Beaulieu, N., Freund, D., & Pitoniak, M. (2009). *Report on the 2009 Reading Achievement Level and Scale-Anchoring Study (Draft)*. Princeton, NJ: Educational Testing Service.

Egan, K. (2021). *Reviewing achievement levels and cut scores.* Washington D.C.: National Assessment Governing Board.

Loomis, S. (2016). *Anchor Studies for Analysis of NAEP Achievement Levels*. Prepared for presentation to an Expert Panel convened by HumRRO for the National Assessment Governing Board, Arlington, Virginia, on July 12–13, 2018.

Mullis, I.V.S, Cotter, K.E, Centurino, V.A.S, Fisbein, B.G & Liu, J. (2016). Using Scale Anchoring to Interpret the TIMSS 2015 Achievement Scales. In M.O. Martin, I.V.S. Mullis, & M. Hooper (Eds.), *Methods and Procedures in TIMSS 2015* (pp. 14.1-14.47). Retrieved from Boston College, TIMSS & PIRLS International Study Center website:

http://timss.bc.edu/publications/timss/2015-methods/chapter-14.html

Pearson (2016). *Developing achievement levels on the National Assessment of Educational Progress in grade 8 technology and engineering literacy: Technical report*. Iowa City, IA: Author.

Perie, M. (2018). *Validating achievement level descriptors.* Paper prepared for presentation to an Expert Panel convened by HumRRO for the National Assessment Governing Board, Arlington, Virginia, on July 12–13, 2018.

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 is in the process of being scheduled to occur in June.

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 <u>does</u> test individual students but <u>does not</u> 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.
> <u>Framework panels</u> 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.

> <u>Public comment</u> 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 <u>NAEP subject area frameworks</u> 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 <u>NAEP subject-area frameworks and assessment updates</u> 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;

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.

Page B-5

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 <u>and update</u> processes to ensure that the final Governing Board-adopted framework, specifications, contextual variables documents, <u>and their development processes comply</u> 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 <u>guidelines</u> 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. <u>Elements of Frameworks:</u> 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. <u>Define what will be tested and how, as well as</u>

Page C-2

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

Page C-3

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.

Page C-4

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
Visioning Panel Deliberations (includes Development Panel members) Purpose: 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

Page D-2

95

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) <u>Item specifications</u> – 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) <u>Contextual variables</u> – recommendations on related contextual variables to be collected from students, teachers, and school administrators	**- Review - Approve or modify**	- Monitor* - Approve	- Facilitate Process - Produce Reports	
<u>Implement Assessment in collaboration with NCES.</u>		- 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.

Appendix E

TABLE OF CONTENTS

Appendix E

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

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

---

Memorandum on National Educational Assessments Used by Foreign Countries   3

- 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.

<div align="center">

Key Aspects of Framework Processes

</div>

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.

Discussion of NAEP Reading Assessment and Implications of Proposed Framework

The Governing Board has been discussing proposed updates to the NAEP Reading Framework since last summer. Some of the more recent discussions and questions have focused on the need to better understand the nature and rationale for features that are already included in the current reading assessment.

COSDAM Chair Gregory Cizek and Vice Chair Carey Wright determined that it would be useful for Committee members to have an opportunity to engage in discussion on this topic at the upcoming COSDAM meeting. To support this discussion, they generated a list of questions and requested that the National Center for Education Statistics (NCES) provide written responses in advance of the COSDAM meeting.

The attached materials include: 1) an Overview document produced by NCES as background for discussions during the May quarterly meeting on the proposed NAEP Reading Framework update, and 2) the specific questions and answers produced to inform this COSDAM discussion. The materials for the May quarterly meeting include additional documents to support the full Board discussions on the proposed NAEP Reading Framework update.

IES Institute of Education Sciences

Overview of the NAEP Reading Assessment and Projections

April 23, 2021

This document has been prepared in response to questions from the National Assessment Governing Board membership and staff regarding the current operational NAEP Reading Assessment in relation to the most recent draft of the 2026 updated framework. This document has three parts:

    I.   **Description of the Current Operational NAEP Reading Assessment in Relation to the Most Recent Draft of the 2026 Updated Reading Framework**

    II.   **Support Features, Relevant Research, and Development Processes in the Current Operational NAEP Reading Assessment**

    III.   **Implementing the Updated Framework and Maintaining Trend**

    I.   DESCRIPTION OF THE CURRENT OPERATIONAL NAEP READING ASSESSMENT IN RELATION TO THE MOST RECENT DRAFT OF THE 2026 UPDATED READING FRAMEWORK

Starting with the 1992 NAEP Reading Framework, a driving principle for the NAEP Reading assessment has been authenticity as a means of establishing face validity. Authenticity in the context of the NAEP reading assessment means that, to the extent possible, the assessment should reflect the reading experiences of students outside of the testing context. For example, the 1992 NAEP reading assessment was one of the first large-scale assessments to use only full-length, naturally occurring texts as its stimulus reading materials. The move to digital assessment under the current framework has allowed the NAEP reading assessment to reflect the digital reading experiences students encounter on a daily basis both inside and outside of school contexts. The draft 2026 updated framework continues to reflect the principle of authenticity.

Definition

The current NAEP Reading Framework lists the following definition: "The NAEP Reading Assessment is guided by a definition of reading that reflects scientific research, draws on multiple sources, and conceptualizes reading as a dynamic cognitive process." This definition

applies to the assessment of reading achievement on NAEP and states that reading is an active and complex process that involves:

- Understanding written text;
- Developing and interpreting meaning; and
- Using meaning as appropriate to type of text, purpose, and situation.

The draft 2026 updated framework maintains the current construct of reading comprehension while expanding the definition to include, "to explicitly acknowledge the sociocognitive processes of reading. Reading comprehension is defined as making meaning with text and four key features are highlighted—contexts, readers, texts, and activities." More specifically, the draft 2026 framework says the following:

"Reading comprehension is making meaning with text, a complex cognitive process shaped by students' social and cultural influences. To comprehend, readers:

- Engage with text in print and multimodal forms;
- Employ personal resources that include foundational reading skills, language, knowledge, and motivations;
- Extract, construct, integrate, critique, and apply meaning in activities across a range of contexts."

Testing Experience

The NAEP reading assessment transitioned from a paper-based assessment (PBA) to a digitally-based assessment (DBA) in 2017. In the most recent DBA in 2019, each student's assessment session began with a tutorial that included student interactions with the tools and interface and concluded with a 3-minute practice session. Following the tutorial and practice session, students worked through two 30-minute cognitive blocks. The second block was followed by a 15-minute survey questionnaire.

Texts

In accordance with the 2019 NAEP Reading Framework, which was first implemented in 2009, there are two broad categories of passages that make up the NAEP reading assessment: literary and informational. Literary texts include fiction, literary non-fiction, and poetry. Informational texts include exposition, argumentation or persuasive texts, and procedural texts.

The draft 2026 framework calls for three types of texts—literature, social studies, and science—and the texts in the 2019 operational pool fall easily into these three categories.

Items

After the passages are reviewed and approved by the Governing Board, items are written to assess three *cognitive* targets under the current framework. The current framework specifies the three cognitive targets as: Locate/Recall, Integrate/Interpret, and Critique/Evaluate.

The draft 2026 updated framework proposes four ***comprehension*** targets: Locate/Recall, Integrate/Interpret, Analyze/Evaluate, and Use and Apply. The addition of Use and Apply addresses the need to assess students' ability to apply the understandings they have gained from interacting with the stimulus materials for a given purpose (e.g., preparing a page of a website or writing a message to the school board).

The current NAEP Reading Framework calls for the following **item types**:

- **Selected response** – This item type encompasses traditional, single-answer, multiple-choice items as well as more complex items that require multiple selections to be answered correctly. NAEP's shift to digitally-based assessment allowed for the introduction of technology-enhanced items, which include matching (drag and drop), grid, and select-in-passage items. Most selected response items are scored dichotomously (correct or incorrect), but more complex selected response items may be scored for partial credit.
- **Constructed response, short and extended** – This item type requires students to generate a written response. Short constructed response items can be answered with a few words or sentences and extended constructed response items may elicit a short paragraph. These items are scored by humans, using a scoring rubric. Short constructed response items are scored with 2- or 3-point rubrics. Extended constructed response items use a 4-point rubric.

Percentages of each item type are specified in the framework for each grade. Typically, NAEP reading blocks include one extended constructed response item, three to five short constructed response items, and three to seven selected response items. The typical NAEP reading block includes a total of 9–11 items.

The draft 2026 updated framework recommends continuing with these item types and provides percentage ranges for selected response, short constructed response, and extended constructed response items. The draft framework also encourages the continued use and exploration of technology enhanced item types.

Reporting

Results of the NAEP reading assessment are reported on a 0–500-point scale. Three scores are reported at each grade level: a composite, or overall reading score, and two sub-scale scores, one for literary texts and one for informational texts. The draft 2026 updated framework maintains the 0–500-point scale and recommends reporting at each grade level: a composite score and three sub-scale scores—reading to engage in literature, reading to engage in science, and reading to engage in social studies contexts.

NAEP Contextual Questionnaire Items

Following the completion of two cognitive blocks, students respond to a 15-minute survey questionnaire. There are two sections to the Contextual Questionnaire: Core and Reading-specific. Core survey items collect information on students' demographic characteristics, opportunities to learn in and out of the classroom, and educational experiences.

Reading-specific survey items focus on reading-related activities and experiences in and out of school. These items are designed to inform interpretations of the results.

In addition to the student questionnaires, teachers and administrators in schools that participate in NAEP also complete their own NAEP Questionnaires.

The draft 2026 updated framework maintains the current approach to the survey questionnaires along with recommendations for changes to the specific items in the reading surveys.

Assembling the NAEP Assessment Via Assessment Blocks

Each NAEP reading assessment is administered in two 30-minute assessment blocks, followed by a 15-minute block of contextual items. Although each student sees only two blocks, there are multiple blocks in each operational assessment as shown in the chart below. Matrix sampling of students and blocks enables NAEP to cover a broad range of content, while also minimizing the burden for students and schools.

Table 1a summarizes the number of NAEP reading assessment blocks administered in the 2019 operational assessment for grades 4, 8, and 12. Typically, each block contains 9–11 items.

Table 1a. 2019 Operational NAEP Reading Assessment Blocks and Item Pool

	Grade 4	Grade 8	Grade 12
Total Number of Blocks	12	15	15
Total Number of Items	118	149	132

Types of Assessment Blocks

Currently, two types of blocks make up the NAEP operational reading assessment: discrete blocks and scenario-based task blocks.

Discrete item (DI) blocks provide general instructions for students to read the passage and provide answers to each assessment item relating to the passages that are presented. All texts and all items are always available for student access and use. The current operational pool of DI blocks is comprised of both transadapted and newly developed blocks as described below.

- **Transadapted blocks** – These blocks are digital renditions of the assessment blocks used in the paper and pencil era of NAEP. These DI blocks make up about two-thirds of the current operational assessment.

- **Newly developed blocks** – These blocks were developed specifically for a digital platform. To take full advantage of the digital format, some of these blocks use print texts and some use texts that are "digitally native" and multi-modal. Some passages

contain embedded hyperlinks and videos. (Note that videos are not used as introductions to texts.) Items addressing video content do so in relation to passage content.

Scenario-based Task (SBT) blocks use both print and digitally native, multi-modal texts. In contrast with DI blocks, students can only access texts and questions sequentially, as the SBTs control the order in which students read texts and items and respond to questions. In this way, students are presented with sources and stimulus materials as needed to respond to items. Videos appear both as part of the texts that students read and as additional content but are not used as introductions to texts. Items addressing video information always do so in relation to the written text.

Table 1b summarizes the number of NAEP reading assessment discrete and scenario-based blocks in the current operational assessment for grades 4, 8, and 12.

Table 1b. 2019 Operational NAEP Reading Assessment Discrete and Scenario-Based Blocks

Block Type	Grade 4	Grade 8	Grade 12
Scenario-based Task Blocks	2	2	2
Discrete Blocks (Transadapted)	7	10	11
Discrete Blocks (Newly developed for DBA)	3	3	2
Total	**12**	**15**	**15**

II. SUPPORT FEATURES, RELEVANT RESEARCH, AND DEVELOPMENT PROCESSES IN THE CURRENT NAEP OPERATIONAL READING ASSESSMENT

Consistent with the principle of authenticity, the current operational NAEP reading assessment uses *support features*, referred to as Universal Design Elements (UDEs) in the draft framework, that are intended to replicate the types of supports provided during reading instruction and practice in school and at home. One central principle is worth emphasizing: *all support features used in a particular block are available to **all** students who take that block.*

The types of support features available on the 2019 NAEP reading operational assessment include:

- Look-back buttons
- Pop-up notes
- Passage introductions
- Eliminate answer choice
- Highlighting and notetaking
- Text-to-speech on directions
- Zoom & selection of color themes
- Multi-part response frames

- Purpose statements*
- Avatars
- Graphic organizers
- Item foreshadowing
- Directions and transitions
- Item resetting

* Purpose statements are not considered UDEs in the draft 2026 updated framework.

Not all features are available in every block, but all of the current operational NAEP reading blocks include some support features. Some of these features are available for all reading blocks, and across other NAEP subjects, at the system level (e.g., highlighting, text-to-speech on directions, zoom, and color themes). Others are content-specific, including item look-back buttons, pop-up notes, passage introductions, and multi-part response frames (complex items with multiple components split into multiple response fields). Others appear only in SBTs, or a subset of SBTs, depending on the goals of the tasks, including block-specific purpose statements, avatars, graphic organizers and sequential directions and transitions.

The following subsection provides additional information about the use of pre-reading features, pop-up notes, and avatars and pop-up notes.

Pre-Reading Features

The current operational assessment includes two types of pre-reading features: **block-specific purpose setting statements** and **introductions** to specific texts, which have been developed in consultation with the Reading Standing Committee[1] and approved by the Assessment Development Committee (ADC) on behalf of the Governing Board. The current NAEP Reading Framework does not provide guidance on pre-reading features.

Purpose Setting

DI blocks include general directions to "read and answer the questions," but do not include block-specific purpose statements.

SBTs include both general directions and block-specific purpose statements. Block-specific purpose statements introduce a purpose for reading and describe the task students are to complete (e.g., gather information for a webpage or to compose an email message). Block-specific purpose statements focus on the tasks students will perform rather than on introducing specific texts. Block-specific purpose setting statements appear in six of the blocks (2 per grade) in the 2019 operational reading pool (17% of the pool).

Introductions

A small number of DI blocks include some information about the text students are about to read prior to reading. This prereading feature has appeared in NAEP Reading since before the

---

[1] The Reading Standing Committee is a diverse group of experts and state assessment staff in reading from across the nation. They advise as part of the assessment item development process, ensuring that NAEP assessment items align to the NAEP framework. There is a Standing Committee for each NAEP subject area assessment.

digitally-based assessment began in 2017. All introductions are written text; none are multi-media (video or audio) as was proposed in the draft 2026 framework.[2]

Passage-specific introductions appear in eight of the blocks across all three grades in the 2019 operational reading pool (23% of the pool). Five of these introductions were added by the test developers and three were part of the original source. In five of these instances, the introduction provides some information about the author. In two of these instances, the introduction provides context for passages that are excerpts.

Generally, there are no consensus assessment industry guidelines or standards for when/how to provide introductions, though there is an extensive research base on the role of prior knowledge in reading comprehension that provides some guidance. For example, seminal research on schema theory by John Bransford and his colleagues found that readers were only able to adequately demonstrate their reading comprehension skills with passages written in general terms when titles were provided that activated their schema/prior knowledge about the topics of the passages. This work, along with content analyses of instructional materials and cognitive labs with students, enabled NCES to implement passage introductions in the operational NAEP reading assessment.

In addition, introductions were deemed important by the Reading Standing Committee as a means of orienting the reader and as a response to the need for content and face validity evidence. In timed, on-demand assessments such as NAEP, brief framing can help to mitigate construct-irrelevant variance, and such introductions and framings are common in sources students encounter in their daily lives. Periodically, the NAEP program invites all states and participating Trial Urban District Assessment (TUDA) districts to review the entire pool of NAEP items. The most recent state/TUDA item review in 2015 included texts with introductions, and no concerns were raised regarding these features. Finally, text introductions appear in some state reading assessments, such as PARCC and Smarter Balanced.

The following examples of passage introductions from previous NAEP reading assessments come from released and publicly available blocks (i.e., not the current operational pool). The first is an introduction to a Turkish folktale called "Five Boiled Eggs." The second introduces an article about the writer, E. B. White, and the third introduces an essay by E. B. White, by explaining that the author of the essay they are about to read is also a children's author. The E.B. White passages appeared in the paper assessment and were released in 2011. The "Five Boiled Eggs" passage appeared in the paper assessment and was transadapted for the digital assessment in 2017 and released after that administration.

---

[2] Responding to the Governing Board's March 2021 Board meeting deliberations, the April 2021 draft of the 2026 framework update does not include multimedia introductions.

Example 1.

Introduction: *Nasreddin Hodja, a character in this story, is familiar in many Turkish legends. "Hodja" means teacher.*

Long ago, a poor country boy left home to seek his fortune. Day and night he traveled, stopping to eat at inns along the way. Though he ate sparingly, his money quickly dwindled until, one day, no silver *akches* ✸ remained.

Still, the boy kept walking. Soon, however, his empty belly began to ache. Staggering up to the next inn he saw, he

Example 2.

Meet the author: *E. B. White, the author of children's classics* Charlotte's Web *and* Stuart Little*, was also a great essayist.*

Not Just for Kids Anymore

"I have a lot of the cat in me," said author E. B. White, "and cats are not joiners."

Perhaps that is why White, one of the country's greatest writers, is so hard to label. His essays for *The New Yorker* appealed to an urbane crowd, but he is best remembered for his

Example 3.

E. B. White was not only a great author for children, he was also the preeminent essayist of his time. This essay, written as a *"Talk of the Town"* piece for The New Yorker, *provides a hint of his powers.*

Twins

by E. B. White

On a warm, miserable morning last week we went up to the Bronx Zoo to see the moose calf and to break in a new pair of black shoes. We

Avatars

Avatars are task characters used to create a social context and facilitate purpose-setting and transitions in SBT blocks (no discrete blocks use them) but are in and of themselves neither purpose statements nor introductions. Two of the total pool of six SBT blocks, across grades, in the 2019 operational reading pool use avatars (6% of the blocks in the total pool).

Pop-up Notes

Pop-ups are indicated by buttons in the text that signal to students that they can read more about a word or phrase. These kinds of notes appeared on the paper-based assessment (PBA) as footnotes. Pop-up notes occur in three blocks in the 2019 operational pool (9% of the blocks in the total pool). Two of the three pop-up instances provide definitions of words/terms that may be unfamiliar to the reader and are important to overall understanding. The third instance presents information that was provided in the original text. There are no assessment items directly related to the information in the pop-up notes.

The following example shows a pop-up note from the passage "Five Boiled Eggs."

Example 1.

A substantial proportion, 63%, of the entire pool of reading blocks in the 2019 operational assessment *does not contain the pre-reading features* described above. These blocks could be characterized as providing opportunities for "cold reads" and will continue to be part of the operational assessment in 2026.

Relevant NAEP Research

As noted above, two types of blocks make up the NAEP operational reading assessment: Discrete Item (DI) blocks and Scenario-based Task (SBT) blocks. At each grade level in 2019 (as noted in Table 1b above), two of 12 grade 4 blocks are SBTs, two of 15 grade 8 blocks are SBTs, and two of 15 grade 12 blocks are SBTs. The remainder are DI blocks. A special study was conducted in 2018 to examine the SBT format, relative to the current framework. For this study, researchers created discrete versions of reading SBT blocks using the same texts and items for both versions. This special study compared student performance on the same set of items and passages in a DI block versus an SBT block.

Although this study was conducted before the framework update project began, it is relevant to conversations about the framework update because SBTs involve collections of support features, which are referred to in the framework update as Universal Design Elements (UDEs). Both SBT blocks and DI blocks include UDEs.

Three of the 15 UDEs in the draft 2026 updated framework only appear in SBT blocks (i.e., avatars, sequential directions and transitions, and item resetting). The remaining 12 UDEs,

including text introductions and pop-up notes, can appear in either SBT or DI blocks.[3] This study provides no information about specific UDEs. Instead, the study examines collections of UDEs in an SBT format.

This was a randomized control trial study with a total of 3,000 students, counterbalanced for version, genre (literary and informational), and block position at each grade. Both the SBT and DI versions of blocks were delivered on tablets. Consistent with the students' experience with DI and SBT blocks in the operational assessment, students were able to move among texts and items at will in the discrete version, whereas movement between texts and items was sequential in the SBT versions.

Key findings (The differences summarized below are statistically significant.):

- Students taking the SBT versions of blocks outperformed students taking the DI versions of block in four of the six blocks.
- The advantage for the SBT versions was consistent across all NAEP subgroups (gender, race, SES, disability, ELL). In other words, there is no differential effect for any subgroup.
- The advantage of the versions with support features was consistent for low- and high-performing students in four of the six blocks.
- For the four blocks for which performance on the SBT version was significantly higher, the differences in percent correct ranged between 2% and 8%, with an average of 5%.
- The SBT-DI special study provides some indication that SBT versions of items tend to be more engaging/motivating to students than DI versions of items. This tendency could contribute to students' higher performance on SBT versions of items, compared with DI versions of items.
- Generally, reading SBT blocks tend to be equally or more difficult than DI blocks, but when comparing SBT and DI versions of the same set of items, SBTs tend to be less difficult than their DI versions.[4]
- Speededness was more of an issue in SBT versions. Revisions were made to reduce speededness before these blocks became part of the operational assessment.

---

[3] Of the 15 UDEs listed in the February 26, 2021 draft of the reading framework update, 13 already appear in the reading assessment. The 2 additions would have been: student exemplars as mentor texts (a task-based UDE) and multimedia passage introductions (a knowledge-based UDE). However, multimedia passage introductions were removed from the latest draft of the framework update. Text introductions already appear on the assessment – see earlier sections of this document on (1) pre-reading features and (2) existence of "cold reads".

[4] Because NAEP uses an Item Response Theory (IRT) model to generate scores, adding more difficult items to the NAEP Reading Item Pool will improve measurement at the high end of the score scale, i.e., detect smaller differences in student achievement for higher performers. Conversely, adding less difficult items will improve measurement on the low end of the score scale, i.e., detect smaller differences in student achievement for lower performers. The IRT methodology for scoring ensures that adding harder items to the item pool will not artificially lower scores and that adding easier items to the item pool will not artificially inflate scores.

Other Standard Research and Reviews in NAEP Item Development

NCES implements a routine research and development cycle to develop every assessment block carefully before it is introduced to the operational NAEP assessment. Each new block undergoes systematic scrutiny, typically including these steps:

1. **Text Selection.** Texts and text sets are identified by the ETS reading item development team at a rate of four for every one text or text set expected to become part of the operational assessment. Proposed texts are reviewed by the ETS bias and sensitivity review team and the ETS editorial staff and are ultimately reviewed and approved by the NCES item development staff and contractors, and the Governing Board Assessment Development Committee (ADC).

2. **Initial Item Reviews.** After passages are approved, items are developed by the ETS reading item development team. Once draft items are completed, ETS reviewers conduct editorial, cold read, bias and sensitivity, and language accessibility reviews. They are then reviewed by NCES item development staff and contractors and the reading standing committee.

3. **Pretesting.** Following initial item review, items and support features are pretested, using:
   a. Cognitive interviews with individual students to determine how they respond to proposed new texts and comprehension test items. The purpose is to determine whether the tasks actually engage students in the intended comprehension processes.
   b. Tryouts under "live" testing conditions with 50–200 students from the target population to determine whether a wide range of students can complete the blocks within the allocated time and whether all of the parts of the block are working as intended.
   c. Usability studies, which test new item or passage interactions with small groups of students.

4. **Revised Item Reviews**. After items are pretested and revised by item developers, ETS reviewers conduct editorial, bias and sensitivity, and language accessibility reviews. They are then reviewed by NCES item development staff and contractors and the standing committee. Item revisions are adjudicated with NCES item development staff, and items are submitted to the Governing Board ADC for final review and clearance for piloting. Before piloting, state/TUDA reviews may occur.

5. **Piloting.** Proposed new blocks are folded into the administration of operational blocks of a live assessment. By comparing student and item performance across the new and the old blocks, NAEP developers can determine whether the new blocks effectively scale together with the old, measuring the same underlying comprehension construct.

6. **Post-pilot Reviews**. Following the collection of pilot data (n=2500–3000 students per form), the following groups review pilot data, item level analyses, texts, and items:
- ETS reading item development team
- ETS data analysis and reporting team
- ETS Differential Item Functioning (DIF) panel
- ETS bias and sensitivity review team
- NCES item development staff and contractors
- NCES data analysis and reporting staff and contractors
- ETS editorial staff
- ETS Reading Standing Committee
- Governing Board Assessment Development Committee (ADC)

III. IMPLEMENTING THE UPDATED FRAMEWORK AND MAINTAINING TREND

This section provides information about the implementation of the updated framework and is based on the contents of the latest draft of that document.

Following Board adoption of an updated framework, it will take time to develop the assessment. As new content is piloted and approved, old content, in particular blocks transadapted from the paper-based assessment, can be phased out. Most importantly, this gradual item development for the updated framework allows for trend to be maintained.

The 2022 and 2024 assessments will be the last operational assessments that are fully aligned to the current framework. The 2026 assessment is projected to be the first operational assessment under the updated framework. The 2026 assessment would include both trend blocks from the 2022 and 2024 operational assessments and newly developed blocks piloted in 2024, being used for the first time in an operational assessment.

In the Governing Board's previous discussions of the updated framework, concerns were expressed that there would be insufficient carryover of content to maintain trend in 2026. However, the information below indicates that maintaining trend in 2026 is possible with careful planning. The projected contents of the next three operational assessments are as follows:

- 2022:
  - Grades 4 and 8 – trend content only (re-administration of 2019)
  - Grade 12 – no assessment
- 2024:
  - Grades 4 and 8 – trend content (all blocks carried over from 2022) plus new operational content (drawn from blocks piloted in 2017 and 2019)
  - Grade 12 – trend content only (re-administration of 2019)

- 2026:
    - Grades 4 and 8 – trend content (all blocks carried over from 2024) and new operational content (drawn from blocks piloted in 2024)
    - Grade 12 – no assessment

Projected Numbers of Blocks Available for the 2026 Operational Reading Assessment

The tables below include information about the numbers of blocks in each of the following two categories that will make up the 2026 operational assessment.

1. **Trend blocks**, which consist of discrete blocks from the 2022 and 2024 operational assessments, which do not include block-specific purpose statements, and SBT blocks from the 2024 operational assessments, which do include block-specific purpose statements.
2. **New operational blocks** developed to address new aspects of the updated framework, including block-specific purpose statements and the updated comprehension targets. These blocks are being used for the first time in the 2026 operational assessment and will not become trend blocks until they are administered operationally for the second time.

The proposed approach to a gradual implementation of the updated framework has been revised since the original Overview document was submitted to Governing Board staff just prior to the March 2021 Board meeting. The March 2021 version of this document suggested adding block-specific purpose statements to three existing discrete blocks at each grade and re-piloting them in 2024. However, the most recent plan retains the existing discrete blocks, as is, and redirects the funds that would have been used for modification and re-piloting of existing blocks to the development of new blocks under the aegis of the updated framework. The current plan provides for a carryover of blocks from the 2024 to the 2026 assessment of 80% at grade 4 and 83% at grade 8. **Although ultimately an empirical question, these percentages of carryover should allow for the maintenance of trend.[5]** (See Table 3a below.)

The current plan for the 2026 development proposes new pilot development of six blocks at grades 4 and 8[6] to yield four new operational blocks. It also assumes that blocks piloted in 2017, 2019, and 2024 will be approved for operational use and that there are no public releases prior to the 2026 assessment.

Tables 3a, 3b, and 3c provide information about the composition of the 2026 operational assessment based on the current plan.

---

[5] The current NAEP reading framework – adopted in 2004 and first implemented in 2009 – included no carryover from the previous framework (0 percent) and trend was maintained. To learn more about how trend was maintained for the 2009 NAEP Reading Assessment, see the Reading Trend Study description at https://nces.ed.gov/nationsreportcard/reading/trend_study.asp.
[6] Grade 12 will not be administered in 2026 and new grade 12 development is out of scope.

Table 3a. Projected Numbers of Blocks by Status available for the 2026 NAEP Operational Reading Assessment at Grades 4 and 8

Blocks	Grade 4	Grade 8
TREND	16 (80%)	20 (83%)
NEW OPERATIONAL	4	4
Total Blocks	**20**	**24**

As a result of needing to both maintain trend and introduce new content aligned with the updated framework, the 2026 operational assessment is projected to include more blocks at each grade than the 2022 operational assessment. The grade 4 assessment would contain 11 blocks in 2022[7] and as many as 20 in 2026, and the grade 8 assessment would contain 14 blocks in 2022 and as many as 24 blocks in 2026.[8] A larger item pool is also required to support reporting goals for the updated framework, including reporting for three subscales instead of the two subscales reported under the current framework.

All of the passages and items in the blocks that would be carried over from 2024 to 2026 are consistent with the updated framework. The block-specific purposes required by the updated framework will be present in 40% of the blocks at grade 4 and 33% of the blocks at grade 8.

Tables 3b and 3c describe the contents of the projected 2026 operational assessment at each grade broken down by subscale.

---

[7] Tables 1a and 1b showed that the 2019 assessment included 12 blocks at grade 4 and 15 blocks at grade 8. However, one cross-grade 4/8 block has been dropped for sensitivity reasons, resulting in 11 blocks at grade 4 and 14 blocks at grade 8 for the 2022 assessment. Blocks sometimes need to be dropped for sensitivity reasons if they address topics that might be disturbing because of recent or ongoing current events, e.g., a hurricane, a pandemic, etc.

[8] The actual number of 2026 blocks is contingent on the contents of possible public releases in 2022 and 2024.

Table 3b. Projected Number of Blocks available for the 2026 NAEP Operational Reading Assessment by Status and Subscale at Grade 4

Blocks	Reading in Literature	Reading in Social Studies	Reading in Science	Total Blocks
TREND	7	4	5	16 (80%)
NEW OPERATIONAL	*New development would include at least one block in each of the reading in social studies and science contexts.*			4
Total Blocks				**20**

Table 3c. Projected Number of Blocks available for the 2026 NAEP Operational Reading Assessment by Status and Subscale at Grade 8

Blocks	Reading in Literature	Reading in Social Studies	Reading in Science	Total Blocks
TREND	8	6	6	20 (83%)
NEW OPERATIONAL	*New development would include at least one reading in literature block.*			4
Total Blocks				**24**

Appendices 1 and 2 on the following pages depict the movement of blocks across the 2022, 2024, and 2026 assessments at grades 4 and 8, as well as the addition of newly developed blocks.

Appendix 1. Proposed Composition of the 2022, 2024, and 2026 Assessments at Grade 4 by Context and Status

2022 Assessment	2024 Assessment	2026 Assessment
Rdg in Lit Block 1	Rdg in Lit Block 1	Rdg in Lit Block 1
Rdg in Lit Block 2	Rdg in Lit Block 2	Rdg in Lit Block 2
Rdg in Lit Block 3	Rdg in Lit Block 3	Rdg in Lit Block 3
Rdg in Lit Block 4	Rdg in Lit Block 4	Rdg in Lit Block 4
Rdg in Lit Block 5	Rdg in Lit Block 5	Rdg in Lit Block 5
	Rdg in Lit Block 6	Rdg in Lit Block 6
	Rdg in Lit Block 7	Rdg in Lit Block 7
Rdg in Science Block 1	Rdg in Science Block 1	Rdg in Science Block 1
Rdg in Science Block 2	Rdg in Science Block 2	Rdg in Science Block 2
Rdg in Science Block 3	Rdg in Science Block 3	Rdg in Science Block 3
Rdg in Science Block 4	Rdg in Science Block 4	Rdg in Science Block 4
	Rdg in Science Block 5	Rdg in Science Block 5
Rdg in SocSt Block 1	Rdg in SocSt Block 1	Rdg in SocSt Block 1
Rdg in SocSt Block 2	Rdg in SocSt Block 2	Rdg in SocSt Block 2
	Rdg in SocSt Block 3	Rdg in SocSt Block 3
	Rdg in SocSt Block 4	Rdg in SocSt Block 4
	Pilot Block A	New Op Block
	Pilot Block B	New Op Block
	Pilot Block C	New Op Block
	Pilot Block D	New Op Block
	Pilot Block E	
	Pilot Block F	

KEY
Trend Block
New Operational Block
Pilot Block

Appendix 2. Proposed Composition of the 2022, 2024, and 2026 Assessments at Grade 8 by Context and Status

2022 Assessment	2024 Assessment	2026 Assessment
Rdg in Lit Block 1	Rdg in Lit Block 1	Rdg in Lit Block 1
Rdg in Lit Block 2	Rdg in Lit Block 2	Rdg in Lit Block 2
Rdg in Lit Block 3	Rdg in Lit Block 3	Rdg in Lit Block 3
Rdg in Lit Block 4	Rdg in Lit Block 4	Rdg in Lit Block 4
Rdg in Lit Block 5	Rdg in Lit Block 5	Rdg in Lit Block 5
Rdg in Lit Block 6	Rdg in Lit Block 6	Rdg in Lit Block 6
	Rdg in Lit Block 7	Rdg in Lit Block 7
	Rdg in Lit Block 8	Rdg in Lit Block 8
Rdg in Science Block 1	Rdg in Science Block 1	Rdg in Science Block 1
Rdg in Science Block 2	Rdg in Science Block 2	Rdg in Science Block 2
Rdg in Science Block 3	Rdg in Science Block 3	Rdg in Science Block 3
Rdg in Science Block 4	Rdg in Science Block 4	Rdg in Science Block 4
Rdg in Science Block 5	Rdg in Science Block 5	Rdg in Science Block 5
Rdg in SocSt Block 1	Rdg in Science Block 6	Rdg in Science Block 6
Rdg in SocSt Block 2	Rdg in SocSt Block 1	Rdg in SocSt Block 1
Rdg in SocSt Block 3	Rdg in SocSt Block 2	Rdg in SocSt Block 2
	Rdg in SocSt Block 3	Rdg in SocSt Block 3
	Rdg in SocSt Block 4	Rdg in SocSt Block 4
	Rdg in SocSt Block 5	Rdg in SocSt Block 5
	Rdg in SocSt Block 6	Rdg in SocSt Block 6
	Pilot Block A	New Op Block
	Pilot Block B	New Op Block
	Pilot Block C	New Op Block
	Pilot Block D	New Op Block
	Pilot Block E	
	Pilot Block F	

KEY
Trend Block
New Operational Block
Pilot Block

NCES Response to the Committee on Standards, Design and Methodology (COSDAM) Reading Framework Questions

April 22, 2021

This document has been prepared in response to questions from the National Assessment Governing Board's COSDAM regarding the current operational NAEP Reading Assessment in relation to the most recent draft of the 2026 updated framework. Three groups of questions are addressed in this document: 1) questions regarding Universal Design Elements (UDEs); 2) questions about the construct(s) being measured and the feasibility of maintaining trend; and 3) questions about implementation plans, projections, and budget considerations.

Universal Design Elements (UDEs) Questions

> What research evidence was used to implement the "support features" on the assessment, in particular the passage introductions?
>
> Is there existing evidence that knowledge-based UDEs are differentially effective based on students' prior knowledge?
>
> Is there any existing evidence regarding the "effect size" of UDEs on performance?

UDEs, such as introductions, have been part of the NAEP Reading Assessment since before NAEP became a digital assessment in 2017. In general, there are no assessment industry guidelines or standards for when/how to provide introductions, though there is an extensive research base on the role of prior topic knowledge in reading comprehension that provides some guidance. For example, seminal research on schema theory by John Bransford and his colleagues found that readers were only able to adequately demonstrate their reading comprehension skills with passages written in general terms when titles were provided that served to activate their schema/prior knowledge about the topics of the passages. This work, along with content analyses of instructional materials and cognitive interviews with students, provided justification for NCES to implement passage introductions in the operational NAEP Reading Assessment. It is also the case that text introductions appear in some state reading assessments, such as the Partnership for Assessment of Readiness for College and Careers (PARCC) and Smarter Balanced Assessment Consortium.

Examining the differential effectiveness of introductions and pop-up notes based on students' prior topic knowledge would require a study that includes measures of students' background knowledge. The NAEP program has not conducted any such study. Similarly, the NAEP program does not have evidence regarding the "effect size" of UDEs on performance because NAEP is not primarily a research program. NAEP relies on data from a variety of quantitative and qualitative sources to inform its development including cognitive interviews, small-scale tryouts, content reviews, and, occasionally, special studies. All new NAEP reading blocks are evaluated in a nationally representative pilot, followed by rigorous, block- and item-level analyses, and submitted for further review by the Governing Board's Assessment Development Committee (ADC). Weak or problematic blocks are not moved on for inclusion in the operational assessment.

Pretesting through cognitive interviews and small-scale tryouts was used to explore new UDEs introduced with Scenario-based Tasks (SBTs). The evidence from pretesting indicated that the majority of students reported that SBT UDEs were helpful and not distracting.

> How much time do existing UDEs add to testing? Would additional UDEs exacerbate this further?
>
> Is there any evidence about whether this additional time might hinder performance or be distracting?

The incorporation of UDEs in reading blocks is an integral part of the development of 30-minute blocks, as opposed to an "add on." Any potential time or cognitive burden they may pose is evaluated for each block as part of the development and pretesting processes via cognitive interviews and small-scale tryouts. Independent of the impact of UDEs, speededness is evaluated and addressed for all blocks as part of the development, pretesting, and piloting processes. The majority of the UDEs recommended in the draft framework are already included in the reading blocks in the NAEP operational assessment.

> Is NCES concerned about the framework's characterization of "support features" as UDEs given how the NAEP program already characterizes "Universal Design Elements?"

NCES describes "Universal Design Elements" as a form of accommodation that is available to all students. Some of the UDEs described by NCES, such as highlighting and zoom, are considered "task-based UDEs" in the updated framework. NCES is not concerned that the updated framework adds UDEs that are not specified in its current description of UDEs.

> What should be the main takeaways from the Scenario-based Task—Discrete Block (SBT-DI) study, relating to this framework update?

The main takeaway from the SBT-DI study is that regardless of ability (low vs. high), students performed better on the SBTs[1]. The main difference between the SBT and DI versions was the use of purpose-driven introductions and a broad range of UDEs in the SBTs. The support features in SBTs examined in the study are consistent with the updated framework.

Construct/Trend Questions

> In NCES's view, does the current framework and the framework update both allow for "cold reads?"

The assessment has not been operationalized in terms of "cold reads" as the term is not defined or discussed in either the current or draft updated framework.

> What evidence/ongoing studies/best guesses are there related to the likelihood of maintaining trend? Is this only a function of how many new blocks are needed?
>
> In NCES's view, do the newly proposed UDEs (mentor texts and multi-media introductions) represent a change in the construct of reading that can threaten trend (on a conceptual level)?
>
> In NCES's view, do other strictly digital UDEs from the framework represent a change in the construct of reading that should have already threatened trend (on a conceptual level)?

The likelihood of maintaining trend is a function of both how many new blocks are needed and whether these blocks differ qualitatively from the existing blocks in terms of what they measure. That said, there is a high likelihood of maintaining trend under the updated framework. Evidence of this comes from several sources. First, the construct of reading comprehension in the updated framework has changed very little from the construct in the current framework. This means that the passages and items developed under the updated framework will not differ significantly from those that were developed under the current framework. Second, the current plan of gradual implementation results in carryover of approximately 80% of blocks from the 2024 to the 2026 operational assessments. Finally, it should be noted that trend was maintained in the implementation of the current framework when there was no carryover from the previous assessment.

The possibility that the new UDEs recommended in the draft framework (mentor texts and multimodal introductions) could threaten the construct is also an empirical question that will be investigated through means such as pretesting and, possibly, special studies and evaluated

---

[1] The magnitude of the improvement, in terms of percent correct, ranged between 2% and 8% with an average of 5%. Although the SBT versions were less difficult than the DI versions of the same texts and items, evidence from the operational assessment indicates that SBTs are of equal or greater difficulty than the DI blocks in the operational pool.

through the various reviews described in the most recent memorandum—*Overview of the NAEP Reading Assessment and Projections*—to the ADC.

The "strictly digital" UDEs introduced with SBTs in 2019 are not a threat to trend either conceptually or empirically. Conceptually, they are consistent with the construct of reading comprehension in the current framework as a means of measuring students' ability to "use meaning as appropriate to type of text, purpose, and situation" (part of the definition from the current NAEP Reading Framework [2009]). Empirically, SBT blocks containing these UDEs scale with Discrete Blocks that do not include these UDEs.

Implementation and Budget Questions

> To what extent can the current reading item pool be used to implement the framework update? Specifically, how much re-field testing is needed and how much new item development is needed?
>
> What is the cost of implementing the framework update?

NCES's proposed approach to a gradual implementation of the updated framework has been revised since the March 2021 Board meeting (see also in this packet of materials—*Overview of the NAEP Reading Assessment and Projections*). The current plan increases the percentage of trend blocks carried over from the 2024 to the 2026 assessments, to 80% at grade 4 and 83% at grade 8. Although ultimately an empirical question, these percentages of carryover should allow for the maintenance of trend.

All of the passages and items in the blocks that would be carried over from 2024 to 2026 are consistent with the updated framework. The block-specific purposes required by the updated framework will be present in 40% of the blocks at grade 4 and 33% of the blocks at grade 8. The remainder of trend blocks include general purpose statements.

NCES expects to develop and pilot six new blocks at each of grades 4 and 8 to yield four new operational blocks. It also assumes that blocks piloted in 2017, 2019, and 2024 are approved for operational use, and there are no public releases prior to the 2026 assessment.

The cost of implementing the updated framework can only be determined when an updated framework has been approved.

Age 17 NAEP Long-Term Trend (LTT) Assessments in 2022

Due to the school closures in response to COVID-19, the age 17 NAEP LTT mathematics and reading assessments, which were originally scheduled to be administered in 2020 along with the corresponding LTT assessments for ages 9 and 13, have been postponed to 2022 and will be administered between end of March and May 2022. All items to be administered in 2022 come from the 2012 operational assessments.

The reporting scales for both Mathematics and Reading LTT are univariate with no subscales reported. In the base year (i.e., 1973 for Mathematics and 1971 for Reading), a cross-age scale was established, which placed the results of all three age groups on the same scale. After the base year, however, the IRT scaling of three age groups has been conducted separately. In other words, the data analyses from one age group have no impact on those of the other age groups. Therefore, the two-year gap between the age 17 assessment and the assessment of the other two age groups does not pose an issue in terms of scaling. On the other hand, there are a few potential issues that are worth considering.

One potential psychometric concern of NCES is that the age 17 LTT assessments, if administered in 2022, will have a ten-year gap when trending back to 2012. That is two years longer than age 9 and age 13 LTTs. Because of this larger gap and possible learning loss due to the pandemic, NCES anticipates more age 17 trend items to function differently between 2022 and 2012 and hence being split[1] in scaling, as compared to LTTs of ages 9 and 13. As a reference, the table below summarizes the number of items treated[2] for Mathematics LTT and Reading LTT in 2012 and 2020. The 2012 LTTs were 4 years apart from the previous LTT assessments (2008) and very few trend items were split in scaling for both subjects of all three age groups. In contrast, both age 9 and age 13 Mathematics LTTs in 2020 had about 14 percent of the item pool (19 and 22 items respectively) split. At what point the number of items treated items become problematic is an empirical question that likely depends on the specific assessment and the nature of the remaining trend items in terms of how representative they are of the construct being measured.

---

[1] When a trend item is "split" between the two consecutive years being linked, it is treated as two different items, and hence, it no longer serves as a trend item.
[2] There are three kinds of treatments: an item might be split, or dropped from the analysis all together, or one or more score categories might be collapsed.

Table 1. 2012 & 2020 Item Treatment Summary for the NAEP Long-Term Trend Assessments

Subject	Age	Year	Number of Items	Number of Items Treated	Number of Items with Categories Collapsed	Number of Items Split	Number of Items Dropped
Mathematics	Age 9	2012	136	3	1	0	2
		2020	135	22	1	19	2
	Age 13	2012	157	6	0	2	4
		2020	152	25	0	22	3
	Age 17	2012	155	5	0	2	3
Reading	Age 9	2012	88	3	3	1	0
		2020	78	2	2	0	0
	Age 13	2012	106	4	1	3	1
		2020	95	5	1	5	0
	Age 17	2012	103	2	0	2	0

Note: Details may not add up to total because some items had more than one type of treatment.

Another important factor is the degree to which the pandemic continues in 2022. If students take the LTT assessments under COVID-19 mitigation protocols in 2022, changes to the normal, pre-pandemic administration procedures may adversely affect the students' performance. This would introduce a potential confounding variable that will make the interpretation of the trend results for this age group challenging. Needless to say, this issue is not unique to age 17 LTT, but relevant to all assessments in 2022. In addition, given the likely learning losses due the ongoing pandemic, the results of this assessment might show scores for age 17 in 2022 that in the score range of age 13 cohort in 2020. This would render the interpretation of the results for this group very challenging. Note that, regardless of the actual results, the report card for the 2022 age 17 would not include the performance of the other two ages assessed in 2020.

A final issue is related to the utility of the age 17 LTT PBA data point in 2022. LTT assessments, originally scheduled for 2020 for all age groups, were supposed to provide an additional data point under the paper-based assessment (PBA) condition before the assessments are transitioned to DBA in 2024. Given the potential psychometric and validity issues described above and the fact that the next LTT for age 17 is in three years (2025), the value of an age 17 LTT PBA in 2022 might be questionable.

As discussed above, although not insurmountable, there are a few potential issues with the age 17 LTT in 2022. The budgetary aspects of this assessment will be discussed in closed session at the upcoming Governing Board meeting on May 14.

Reporting and Dissemination Committee

May 10, 2021
10:00 am – 12:00 pm

Zoom:
https://us02web.zoom.us/webinar/register/WN_uUHhJhKfQWiJ9j255DKnCA

AGENDA

10:00 – 10:15 am	**Strategic Vision 2025**	
	Tonya Matthews, Chair	
	Laura LoGerfo, Assistant Director	
10:15 – 11:00 am	**Board Outreach: Accomplishments and Plans**	*Attachment A*
	Stephaan Harris, Assistant Director for Communications	
	Robert Johnson, The Hatcher Group	
11:00 – 11:55 am	**Discussing Socioeconomic Status and NAEP**	
	Martin West, Vice Chair	
	Daniel McGrath, National Center for Education Statistics	
	Ebony Walton, National Center for Education Statistics	
	William Ward, National Center for Education Statistics	
11:55 am – 12:00 pm	**Updates / Queries**	*Attachment B*

NATIONAL ASSESSMENT GOVERNING BOARD
COMMUNICATIONS PLAN SUMMARY

APRIL 2021

PRIORITY AUDIENCES

Education Administrators	Researchers	Advocates for...	Policymakers
• School district superintendents • Assessment directors at the district and state levels	• Who study student achievement data • Who work at think tanks	• Improving student achievement • Education and opportunity equity	• Chief state school officers and their policy staff • State legislators and their staffs

COLLECTING STAKEHOLDER INPUT

- Conduct a focus group with superintendents to better understand how the Governing Board can build relationships with them and convey the relevance of NAEP to their work.

- Build interest and understanding of NAEP among policymakers and better understand how NAEP can be most useful to them through hosting a NAEP 101 and two-way Q&A sessions with new education staffers on Capitol Hill, state lawmakers who serve on education committees, and new state chiefs and their key staff members.

- Conduct interviews or have one-on-one meetings with national education and opportunity equity advocates and researchers who study K-12 education and student achievement at think tanks.

CORE MESSAGING THEMES

These core messages will continue to ground Governing Board communications:

- The Governing Board will continue to fulfill its Congressional mandate to answer: "How are our nation's students doing?"

- NAEP is the gold standard for student assessment.

- NAEP results are a catalyst for action to improve student achievement.

- Understanding what students know and can do is critical to effective and equitable education policy and practice.

Overarching themes and messages that will be addressed and/or used in 2021 communications:

- Guided by a new strategic vision, the Governing Board will disseminate NAEP resources to inform the work of education stakeholders and promote high-quality uses of NAEP that support improvements in policy and practice; ensure NAEP remains at the forefront of assessment design and technology; and strengthen partnerships and communications with stakeholder organizations.

- At a time of significant disruption to our nation's schools and students, NAEP provides trusted and objective information about the status and progress of American education.

- The Governing Board will continue to monitor and use NAEP to inform growing concerns about widening performance gaps.

- NAEP is informing efforts to address equity in state and urban school systems.

- NAEP data, including the 2019 NAEP Reading and Mathematics grade 12 release and the upcoming 2019 NAEP Science release, are one of few ways for the nation to understand student progress over time.

- NAEP will advance the nation's understanding of performance over time of student subgroups such as students with disabilities, English learners, and by family economic background, race, and ethnicity.

STRATEGIES AND TACTICS

Social Media

- Use social media to amplify 2019 NAEP Science results and contextual data and highlight upcoming Governing Board NAEP and TUDA use cases.

- Deepen stakeholder targeting and engagement by coordinating social posts with organizations directly.

- More frequently engage with other social media accounts and followers.

- Diversify Twitter content to include polls and questions.

- Expand the Governing Board's LinkedIn presence by increasing content about individual Board members and increasing Board member and staff engagement with the content.

- Host more Twitter chats.

- Expand use of social media advertising.

Email Outreach

- Complete an analysis of newsletter subscribers to better understand the current newsletter audience and segment the subscriber list to further reach subscribers with other relevant emails.

- Reactivate disengaged newsletter subscribers through a targeted email campaign.

- Grow the newsletter audience through a Facebook advertising campaign that links to the newsletter sign-up form.

- Survey highly engaged Governing Board audiences, such as newsletter subscribers and NAEP release attendees, to understand what they want to know more about.

- Improve the subscriber experience with a welcome email.

- Expand the Governing Board's email strategy to get resources such as videos, narratives, and frameworks in front of target audiences.

- Conduct a Hatcher-led training session for Governing Board staff on email outreach best practices.

Media Outreach and Placement

- Board members are underutilized spokespersons for the Governing Board. One of the most effective ways to build visibility is to start local. Raise the profile of Board members in their local media markets by:

    - Developing a short questionnaire for Board members exploring their interest in being sources for education stories in their local markets.

    - Drafting a media tip sheet about each Board member and distributing it to media outlets in local markets proposing a story on the respective member's participation on the Governing Board.

    - Providing a virtual media training for Board members.

    - Conducting outreach to local media on Board members' behalf.

- Identify podcast and blog opportunities for Board members and staff.

- Offer and set up annual big-picture check-in meetings with select education trade outlets.

- Offer background briefings to select reporters on Governing Board news developments and priority issues.

- Identify opportunities to present at journalism conferences or training sessions.

- Expand outreach in non-English language media.

Conference Strategy

We recommend these themes for conference presentations:

- NAEP data and materials are used to inform and develop state learning standards and assessments, and to understand student achievement trends—both recent and long-term.

- Expand and deepen audience understanding of NAEP, related tools and resources, and how NAEP data can be used to meet a target audience's specific needs.

- TUDA school districts are leaders in the use of NAEP data, with many using NAEP data to advance equity initiatives.

- NAEP data that show growing gaps in student achievement are just one example of how NAEP data and materials can help understand and track inequity in learning across student groups.

- A deep dive into 2019 NAEP Science results.

Events

Release Events

- Use video to elevate and illustrate NAEP assessments.

- Explore alternate themes for NAEP Day 2021, shifting the focus to achievement trends and equity or other themes that use NAEP data to generate timely and meaningful conversations.

- Leverage an upcoming event to elevate TUDA districts' role in supporting NAEP. Content could feature how TUDA districts are using NAEP to improve equity or feature TUDA districts that are making NAEP gains and how they are doing it.

- Pre-record questions asked by high-profile figures in the field for a Q&A segment or other form of outside participation for virtual events.

- Use Webex for release events because it provides the right capacity and options for a large, national event.

- Limit virtual events to 75 minutes to avoid audience fatigue and drop-off.

- Integrate surveys into release events to learn from attendees about what they enjoyed, learned, and would want to see done differently in future events.

- Continue to include virtual components to NAEP release events, even when the Governing Board is able to hold events in-person again.

Post-Release Activities

- Partner with the Southern Regional Education Board on a Twitter chat on science education and equity.

- Create a NAEP Day "post-release event" that continues a conversation about equity with an audience that is not a typical Governing Board primary target but could help expand the conversation and awareness about NAEP.

- Host a discussion about how NAEP data can help inform the broader national conversation on equity, race, and how achievement gaps are reported.

- Begin post-release activity planning during report card release event planning.

- Create post-release activities that go beyond NAEP results.

Creating Tools and Resources

Data Graphics

- Create simple data graphics with content relevant to priority audiences.

- Continue to experiment with animation in data graphics.

- Produce dissemination plans for each set of data graphics.

- Develop graphics that connect NAEP results and contextual data with equity themes.

Video

- Create at least one video in 2021 that communicates NAEP data.

- Use Facebook advertising more frequently to promote videos.

- Leverage the interest of state chiefs to share their stories.

Narratives

- Expand narrative content to include pieces that go beyond NAEP results.

- Create a dissemination plan and package with graphics and social toolkits for partners.

- Experiment with pairing narrative and video products to cover topics with greater depth and breadth and to allow for more dissemination opportunities.

Institute of Education Sciences

Upcoming NCES Reports as of April 2021

Report Cards / Initial NAEP Releases

Expected Release Date

2019 NAEP Science Report Cards at Grades 4, 8, 12	May 25, 2021
2019 High School Transcript Study	Fall 2021
2020 NAEP Long-term Trend Mathematics and Reading, Ages 9 and 13	Fall 2021
2021 NAEP School and Teacher Questionnaires	Fall 2021

Other IES/NAEP Reports

Expected Release Date

The 2018 NAEP Oral Reading Fluency Study	April 27, 2021
Updates to the NAEP 2021 Monthly School Survey Dashboard	May 5, June 10, July 8, 2021
National Indian Education Study 2019: American Indian and Alaska Native Students at Grades 4 and 8	May 18, 2021
Mapping State Proficiency Standards onto NAEP Scales: Results from the 2019 NAEP Mathematics and Reading Assessments	June 2021
Summary Report on the NAEP 2021 Monthly School Survey	Fall 2021

Nominations Committee

Tuesday, May 11, 2021
5:30 – 6:30 pm (ET)

Zoom Meeting

Meeting ID: 868 3814 1736
Passcode: 508812

AGENDA

5:30 – 5:35 pm	**Welcome and Agenda Overview**
	Governor Geringer, Chair
5:35 – 5:40 pm	**Update on Nominations for Board Terms Beginning October 1, 2021**
	Lisa Stooksberry
5:40 – 6:00 pm	**Discussion: Looking Ahead to 2022 Nominations Campaign**
	Stephaan Harris
6:00 – 6:25 pm	**Discussion: Updating Procedures Manual**
	Tessa Regis
6:25 – 6:30 pm	**Next steps**
	Governor Jim Geringer
6:30 pm	**Adjourn**