Quarterly Meeting of November 18-19, 2021

Via Zoom
and
Hilton Washington DC Capitol Hill
525 New Jersey Avenue, NW
Washington, DC 20001
(202) 628.2100

AGENDA

Thursday, November 18: 9:45 am – 5:15 pm (EST)

9:45 – 11:30 am	**Executive Committee Meeting** (see separate agenda) *Bev Perdue, Chair*
11:30 am – 12:45 pm	***Lunch Break***
12:45 – 1:00 pm	**Welcome** **Approval of November 2021 Agenda** **Approval of August 2021 Minutes** *Bev Perdue, Chair*
1:00 – 1:15 pm	**New Members' Oath of Office** *Bev Perdue*
1:15 – 1:45 pm	**New and Reappointed Members' Remarks** *Dana Boyd, Elementary School Principal* *Tyler Cramer, General Public Representative* *Viola Garcia, Local School Board Member* *Gary Herbert, Governor* *Scott Marion, Testing and Measurement Expert* *Bev Perdue, Governor*

1:45 – 2:00 pm	**Executive Director's Update** *Lesley Muldoon, Executive Director*
2:00 – 2:30 pm	**ACTION: NAEP 2026 Reading Assessment and Item Specifications** *Dana Boyd, Chair, Assessment Development Committee* *Suzanne Lane, Chair, Committee on Standards, Design and Methodology*
2:30 – 2:45 pm	***Break***
2:45 – 4:00 pm	**Discussion of Potential Changes to Framework Update Process** *Mark Miller, Vice Chair, Assessment Development Committee* *Patrick Kelly, Member, Assessment Development Committee*
4:00 – 5:15 pm	**Initial Public Comment on Current NAEP Science Framework** *Dana Boyd* *Christine Cunningham*

Friday, November 19: 9:00 am – 4:00 pm (EST)

9:00 – 11:00 am	**NAEP Budget and Assessment Schedule (CLOSED)** *Peggy Carr, Commissioner, National Center for Education Statistics* *Lesley Muldoon*
11:00 – 11:15 am	***Break***
11:15 am – 12:30 pm	**Strategic Vision: Looking Ahead to Year 2** *Bev Perdue*
12:30 – 1:45 pm	**WORKING LUNCH** **Mirrors or Windows: Briefing and Discussion** *Ray Hart, Executive Director, Council of the Great City Schools*
1:45 – 2:15 pm	**Member Discussion**

2:15 – 4:00 pm	**Planned and Potential Innovations for NAEP** *Tony Alpert, Executive Director, Smarter Balanced* *Alison Deigan, Director of Technology Systems, National Center for Education Statistics* *Marianne Perie, President, Measurement in Practice* *Holly Spurlock, Program Director, National Center for Education Statistics*

2021 - 2023 QUARTERLY BOARD MEETING DATES AND LOCATIONS

March 3 - 5, 2022	**TBD**
May 12 - 14, 2022	**TBD**
August 4 - 6, 2022	**TBD**
November 17 - 19 , 2022	**TBD**
March 3 - 4, 2023	**TBD**

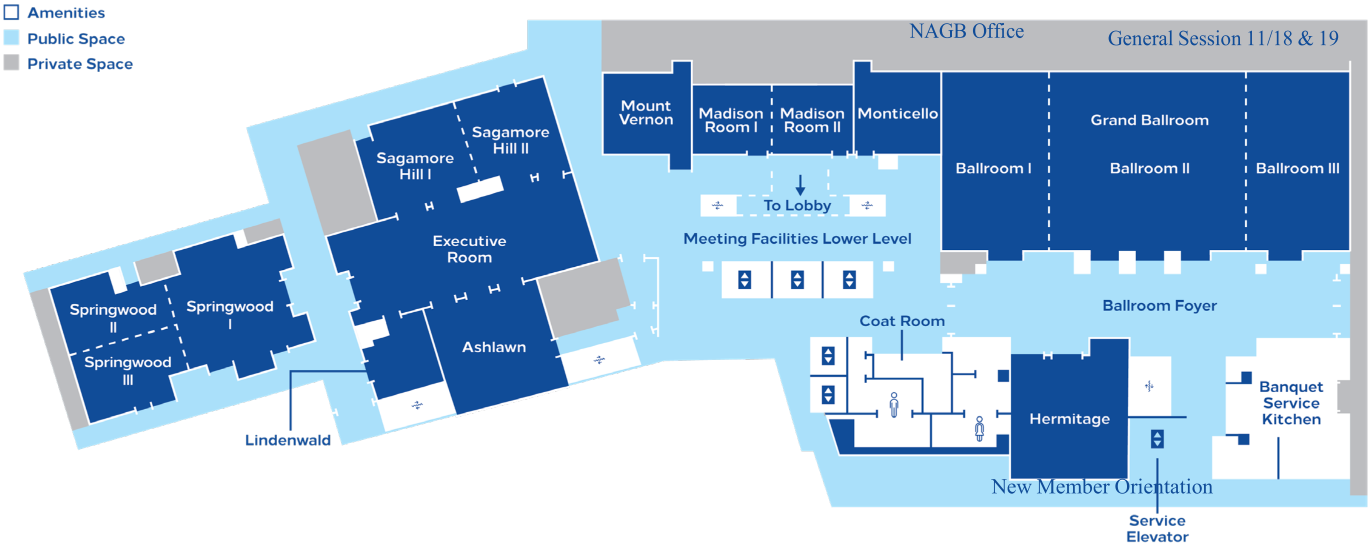
HILTON WASHINGTON DC CAPITOL HILL

FLOOR PLANS

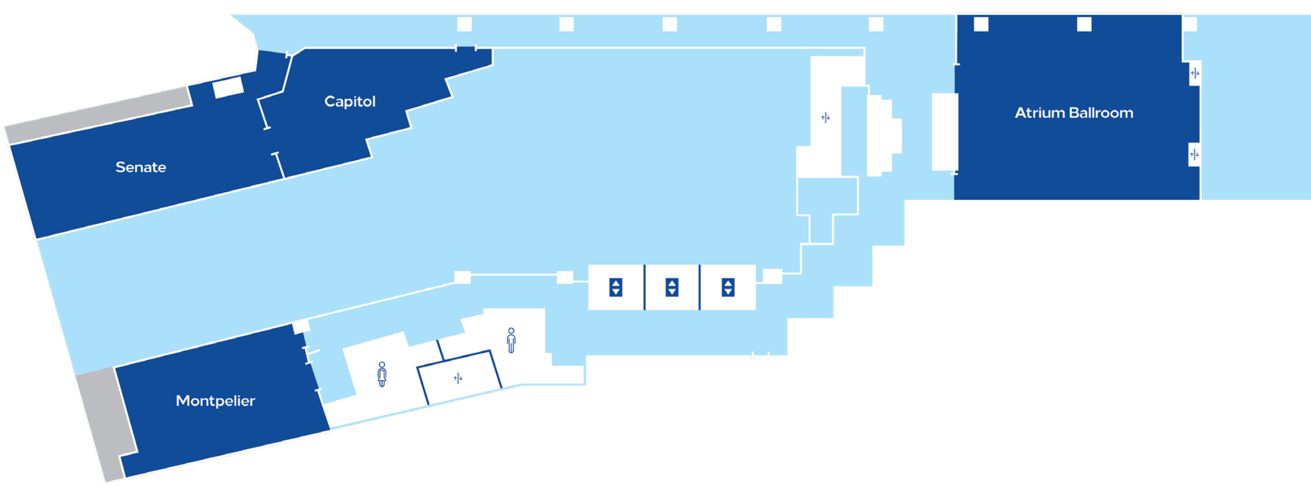
Lower Lobby Level

FLOOR MAP KEY

- Meeting/Conference Rooms
- Amenities
- Public Space
- Private Space



Mezzanine Level



Governing Board Members
2021 - 2022

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

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

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

Honorable Haley Barbour
BGR Group
Founding Partner
Yazoo City, Mississippi

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

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

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

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

National Assessment Governing Board

Committee Structure
(2021-2022)

Assessment Development Committee

Chair	Dana Boyd
Vice Chair	Mark Miller
	Christine Cunningham
	Frank Edelblut
	Viola Garcia
	Patrick Kelly
	Reginald McGregor
	Nardi Routten
	Sharyn Rosenberg (Staff)

Committee on Standards, Design and Methodology

Chair	Suzanne Lane
Vice Chair	Carey Wright
	Rick Hanushek
	Scott Marion
	Alice Peisch
	Julia Rafael-Baer
	Russ Whitehurst
	Lisa Stooksberry (Staff)

Executive Committee

Chair	Beverly Perdue
Vice Chair	Alice Peisch
	Haley Barbour
	Dana Boyd
	Paul Gasparini
	Suzanne Lane
	Tonya Matthews
	Mark Miller
	Marty West
	Carey Wright
	Matt Stern (Staff)

Reporting and Dissemination Committee

Chair	Tonya Matthews
Vice Chair	Marty West
	Alberto Carvalho
	Tyler Cramer
	Paul Gasparini
	Ron Reynolds
	Mark White
	Laura LoGerfo (Staff)

Nominations Committee

Chair	Paul Gasparini
	Dana Boyd
	Tyler Cramer
	Suzanne Lane
	Tonya Matthews
	Reginald McGregor
	Mark Miller
	Alice Peisch
	Munira Mwalimu (Staff)
	Lisa Stooksberry (Staff)

Updated November 1, 2021

National Assessment Governing Board

Meeting of August 5 - 6, 2021
McLean, VA and Virtual

OFFICIAL SUMMARY OF GOVERNING BOARD MEETING

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
Beverly Perdue
Julia Rafal-Baer
Ron Reynolds
Nardi Routten
Martin West
Mark White
Grover (Russ) Whitehurst
Carey Wright
Mark Schneider (ex-officio)

National Assessment Governing Board Staff

Lesley Muldoon, Executive Director
Lisa Stooksberry, Deputy Executive Director
Stephaan Harris
Donnetta Kennedy
Laura LoGerfo
Munira Mwalimu
Tessa Regis
Sharyn Rosenberg
Angela Scott
Matt Stern
Anthony White

National Center for Education Statistics (NCES)

Peggy Carr, Acting Commissioner
Tammie Adams
Samantha Burg
Jing Chen
Brian Cramer
James Deaton
Alison Deigan
Enis Dogan
Veda Edwards
Pat Etienne
Eunice Greer
Linda Hamilton
Shawn Kline
Tina Love
Holly Martin
Daniel McGrath
Nadia McLaughlin
Stephen Provasnik
Taslima Rahman
Eddie Rivers
Holly Spurlock
William Tirre
Ebony Walton
William Ward
Grady Wilburn

American Institutes for Research (AIR)
Brittany Boyd
Markus Broer
Kim Gattis
Courtney Gross
Cadelle Hemphill
Young Kim
Christy Talbot
Yan Wang

Council of Chief State School Officers, CCSSO
Fen Chou
Scott Norton

CRP, Inc.
Shamai Carter
Monica Duda
Subin Hona
Edward Wofford
Anthony Velez

Educational Testing Service (ETS)
Marc Berger
Jay Campbell
Carmen Dahlberg
Gloria Dion
Patricia Donahue
Amy Dresher
Kadriye Ercikan
Gary Feng
Robert Finnegan
Janel Gill
Yue Jia
Ranu Palta-Upreti
Rupal Patel
David Pelovitz
Hilary Persky
Emilie Pooler
Bobby Rampey
Shannon Richards

Sarah Rodgers
Simone Todd
Lisa Ward
Karen Wixson

Hager Sharp
James Elias
David Hoff
Cailin Jason
Joanne Lim
Debra Silimeo

The Hatcher Group
Jenny Beard
Sami Ghani
Robert Johnston
Zoey Lichtenheld
David Loewenberg
Michael Mershon
Melissa Rogers
Alexandra Sanfuentes
Devin Simpson
Nandini Singh
Jenna Tomasello

Optimal Solutions Group
Imer Arnautovic
Peter Simmons

Pearson
Joy Heitland
Eric Moyer
Pat Stearns
Llana Williams

Westat
Chris Averett
Greg Binzer
Lauren Byrne
Zully Hilton

Kavemuii Murangi
Jason Nicholas
Robert Perkins
Lisa Rodriguez
Rick Rogers
Keith Rust

WestEd
Georgia Garcia
Mira-Lisa Katz
Mark Loveland
Sonya Powers
Megan Schneider
Steve Schneider
Sarah Warner

Other Attendees
Diana Arya, University of California, Santa Barbara
Michelle Blair, Duke University
Nancy Brynelson, California State University, Chancellor's Office
Gina Cervetti, University of Michigan
Julie Coiro, University of Rhode Island
Susan Cramer, Remarc Group
Kelly Crowley, (Closed Captioner)
Brandon Dart, Management Strategies
Andrea Faulkner, North Carolina Department of Public Instruction
James Forester, U.S. Department of Education, Office of Legislation and Congressional Affairs
John Galisky, University of California, Santa Barbara
Joseph Garry, Neal Gross (Court Reporter)
Laura Goadrich, Arkansas Department of Education
Aleah Guthrie, SCORE
Ray Hart, Council of the Great City Schools
Andrew Ho, Harvard University
Christy Hovanetz, ExcelinEd
Linda Jacobson, The 74
Carol Jago, UCLA
Jon Kemp, DCI Group
Andrew Kolstad, P20 Strategies LLC
Beth LaDuca, Oregon Department of Education
Jason Lautenbacher, U.S. Department of Education, Office of the Chief Information Officer

Somer Levine, Pepperdine University
Cheryl Little, (Closed Captioner)
Brian Lloyd, Michigan Department of Education
Shelley Loving-Ryder, Virginia Department of Education
Raina Moulian, Alaska Department of Education & Early Development
Christian Myers
Alexa Patrick, Intern, U.S. Department of Education
P. David Pearson, University of California, Berkeley
Christine Pitts, Center on Reinventing Public Education
John Richard, Ohio Department of Education
Alexa Patrick Rodriguez, U.S. Department of Education, Office of Legislation and Congressional Affairs
Marco Sanchez, U.S. Department of Education, Office of Legislation and Congressional Affairs
Renee Savoie, Connecticut State Department of Education
Sarah Schwartz, Education Week
Sydney Smith, U.S. Department of Education, Office of Legislation and Congressional Affairs
Sheila Valencia, University of Washington
Valerie Walton, Public School
Natalie Wexler
Julie Williams, California Department of Education

Opening Remarks

Haley Barbour, Chair, called the session to order at 1:18 pm and welcomed attendees to the August 2021 National Assessment Governing Board (Governing Board) meeting held by webinar and in-person at the Ritz-Carlton in McLean, Virginia.

Approval of August 2021 Agenda

Barbour requested a motion for approval of the August 2021 agenda. A motion to accept the proposed agenda was made by Tyler Cramer and seconded by Martin West. No discussion ensued and the motion passed unanimously.

Approval of May 2021 Board Meeting Minutes

Barbour requested a motion for approval of the minutes of the May 2021 Governing Board meeting. Frank Edelblut made a motion to approve the May 2021 minutes and Mark White seconded the motion. There was no discussion and the motion passed unanimously.

Executive Director's Update

Governing Board Executive Director Lesley Muldoon provided a quarterly update. Muldoon welcomed Board members by expressing her gratitude to see them, as they convened for the August 2021 Board session. Prior to delivering her update, Muldoon explained best practices for hybrid meetings with attendees both online and in-person.

Muldoon began by stating that much had changed, among the Board and in the nation, since the Board had last met in person, 17 months ago. Muldoon pointed to two specific issues that had dominated the Board's agenda since March 2020. The first, Muldoon explained, was shifting the National Assessment of Educational Progress (NAEP) reading and math assessments from 2021 to 2022 due to the COVID-19 pandemic. She noted the ongoing need for NAEP to help inform the nation about the impact of the pandemic on student learning.

The second issue was the 2026 NAEP Reading Framework update, which has consumed significant Board attention over the past four Board meetings. Urgency around the Reading Framework grew, she added, due to noticeable declines in NAEP reading scores, particularly among eighth grade students, and due to the pandemic's impact. Muldoon stated that the Board strives to reach consensus on significant policy matters and applauded the Board's hard work on the Framework over the past few months. Muldoon congratulated the Board on crafting a Framework that is broadly representative of education stakeholders across the country.

Next, Muldoon outlined activities for the fall and highlighted two developments in Congress: First, the House Appropriations Committee recommended a $40 million increase for the NAEP program in the coming budget. This sum would fund general programmatic expenses to ensure that NAEP could administer the full assessment schedule. The Appropriations Committee also proposed an additional $25 million to conduct a NAEP civics assessment in 2024, since civics has played an important role in education this past year. Congress has expressed interest in understanding and measuring students' civics proficiency. The House passed this proposal, but it awaits approval from the Senate.

The second is the Senate Health Education Labor and Pensions (HELP) Committee's interest in reauthorizing the law that governs NAEP, known as the Education Sciences Reform Act, or ESRA. The Senate last took action to reauthorize ESRA in 2015, but the reauthorization never passed the House. The Senate HELP Committee is considering reintroducing an updated proposal in 2021. The Institute of Education Sciences (IES) and the National Center for Education Statistics (NCES) are slated to discuss this reauthorization with the Committee. Muldoon clarified that some legislation may be introduced after Congress resumes in August. If everything stays on schedule, action could be taken by the Senate in the fall.

Muldoon urged the Board to refocus their strategic priorities. Since the Reading Framework update consumed the Board's time over the past year, other efforts require attention. This coming November marks the first anniversary of the adoption of Strategic Vision 2025. Staff will deliver its first annual report on their progress toward the Vision's goals.

Muldoon elaborated on initiatives underway: The Committee on Standards, Design and Methodology (COSDAM) is overseeing work to review and revise the achievement level descriptors for the NAEP reading and math assessments. The Assessment Development Committee (ADC) is launching an update to the NAEP Science Framework. The ADC intends to seek public comment on the NAEP Science Framework before the Board begins the update process later this fall. The Reporting and Dissemination (R&D) Committee has discussed ways to improve socioeconomic status measures on NAEP, which will be shared with the full board in coming months.

Muldoon concluded by bidding farewell to Board members Gregory Cizek and Jim Geringer. Muldoon mentioned that new Board members will be appointed in the fall. Barbour thanked Muldoon then introduced Peggy Carr, Acting Commissioner for Assessment, NCES.

<u>National Center for Education Statistics Update</u>

Carr thanked Barbour and began by discussing the recently implemented NAEP Monthly School Survey and dashboard. This monthly survey allows NCES to be nimbler in its collection of relevant, high-quality data. Carr added that this method will be considered for future use since it allows NCES to connect with stakeholders more efficiently.

Carr provided an overview of the survey and initial observations. The monthly survey asks nine questions, including about the demographic composition of in-person and virtual students; instructional mode offered; the implementation of in-school learning, remote learning, or hybrid instruction; attendance; and teachers' vaccination status. Carr expressed surprise that some remote instruction did not involve live-teacher instruction at all, relying heavily on pre-recorded instruction. Carr also highlighted declines in attendance, as well as findings that private schools across the nation know more about their teachers' vaccination status than public schools. The results are available on an interactive online dashboard.

Carr then presented enrollment data, focusing specifically on grades 4 and 8. One primary finding was that Black, Asian, and Hispanic students were less likely to attend in-person instruction, a trend Carr said may continue during the 2021-22 school year.

Next, Carr reminded Board members about NCES' "School Pulse Panel" survey, which will collect and report data in a matter of weeks, as was done for the monthly survey. Carr stated that NCES intends to rotate through different indicators, in a module format, over the year. Survey administration will begin in September, with data released within six weeks on a dashboard. Currently, about 1,000 schools contribute to the survey. Involving more schools would help NCES report data at the regional and national levels.

Carr described the first-ever IES Reading Summit, held in June 2021. More than 1,000 attendees and 70 speakers participated in the virtual event. Participants learned about NAEP data as well as data collected by many of the grantees out of the IES research centers. Based on the success of the reading summit, the Council of Great City Schools and IES agreed to partner again to plan a similar summit on math.

Carr then shifted to the National Academies of Sciences (NAS) efforts to examine what NCES does and to consider how NCES should evolve in the future. This work will push NCES to reflect on its current practices and identify new practices. Carr stated that another effort led by NAS will investigate how the National Center for Education Research (another center within IES) develops and prioritizes research topics for grants.

Carr concluded by discussing the Evidence Act, which was implemented in 2018, changing how federal agencies manage, use, collect, and share data. Carr explained that the Act focuses on four pillars of evidence: (1) performance management, (2) policy analysis, (3) program evaluation, and (4) foundational fact finding (FFF). For NCES, the cornerstone is FFF, which is operationalized in three ways: surveys, assessments, and administrative data. Carr clarified the roles of each person supporting these efforts: Carr, herself, is currently acting as the Statistical Officer; Gregory Fortelny is the current Chief Data Officer; and Matthew Soldner is the current Evaluation Officer. Carr stated that the Act gives the Chief Data Officer activities and scope to be accomplished across the four pillars. Carr added that the Act also requires IES to convene a governance council with the statistical official, who serves as ex-officio member. Carr noted her interest in these upcoming opportunities and looked forward to the Board's participation.

Barbour thanked Carr and opened the floor for questions. Alberto Carvalho asked Carr about the close partnership between NCES and the Governing Board, and if there was any further consideration, associated with the future release of NAEP data, for analyses and/or reports of the environmental or socioeconomic conditions students face? Carr stated that she had just received a congressional request for information about schools' physical environments. She stated that more data would be required from NCES and other departments. Carr noted the many privacy and confidentiality laws governing this requirement, but NCES looks forward to adding this information to their portfolio.

Carvalho stated that this collaboration would help measure district effects, through regression analysis, which could help equalize conditions specific to levels of poverty, education level of families, and family income. Carvalho said such data could help the Board identify states and districts where best practices are being implemented. Carvalho stated that, at some point, with the Chair's agreement, NCES could present those data points to the Board.

Jim Geringer referenced a previous slide that highlighted the racial backgrounds of students who attended school in-person and remotely. He was surprised to find that minority groups had higher rates of remote participation, as the Board had heard that there was a lack of access to broadband and remote capabilities among those groups. Carr said that although some of these students were learning remotely, it did not mean they had access to a live teacher. She stated that when NCES evaluated those data points, and disaggregated the data by race, they could see that those with infrequent access to live teachers were primarily Hispanic. Carr added that, based on a recently published NWEA report, achievement gaps would likely widen. Carr clarified that NCES cannot explain this trend, but students and parents chose differently—and had access to different choices—by race.

IES Director Mark Schneider expressed agreement with Carr's statement about administrative data, but believed it was incomplete. He stated that there were greater concerns with measurement and suggested that more contextual data could address those concerns. The Evidence Act, Schneider clarified, increases the opportunity and the actual obligations of statistical agencies like NCES to gather data from multiple sources and merge the data. This would fall into Carr's purview as Statistical Officer and the Chief Data Officer's as well. Schneider added that someday, if the College Transparency Act is passed, the NCES

Commissioner would be in a higher position of authority to gather data from multiple sources and merge those data points together. In the meantime, IES is working with Census data and a geospatial program, which would help Carvalho understand the contextual measures he suggested.

Schneider pointed to another program: the Statewide Longitudinal Data Systems (SLDS) program. First administered in 2005, the SLDS serves as the backbone of states' student-level data. Schneider said SLDS elementary and secondary data can be merged with post-secondary data and labor market outcomes. Schneider stated that he wanted to see more administrative data merged into other data sets, from NCES, to help decision makers. Many states now contribute social justice and health care data to the SLDS program. Schneider underscored IES' commitment to translating complicated data into data that are more useful, usable, and used. Recent NAEP reports exemplify this aim.

Carr responded to Schneider's comments by adding that the Edge program to which Schneider referred includes geospatial work that helps NCES develop a better understanding of socioeconomic status (SES) and validate school-provided information on SES with data on neighborhood socioeconomic poverty. Carr stated that it would be beneficial if large school districts partnered with NCES to improve the validity and reliability of these estimates. Currently, 15 states work with the Edge program.

Christine Cunningham then asked, based on the chart that showed the racial and ethnic breakdown of students by learning mode, if the data had been disaggregated by rural and suburban students. Carr replied that NCES had found that a slight interaction with race and region of the country, specifically that students learning remotely were more likely in the Great Plains and in-person learning occurred more often in the South and the Midwest. Cunningham stated that rural students are 80% white and are often more likely to attend school in-person. She then asked if there was an interaction between school size and school location. Carr stated that NCES tried to disentangle race from different locale variables, but the pattern was still there, though not as strong.

<u>**Committee Updates**</u>

Barbour opened the session with an update on the Executive Committee. He shared that he and Alice Peisch sent a letter to U.S. Education Secretary Miguel Cardona in March advocating for increased funding for NAEP to maintain civics, U.S. History, and science on the assessment schedule. Barbour stated that President Biden's request included a $15 million increase. The House Appropriations Committee included this increase and an additional $25 million for the civics assessment in 2024. Barbour said the Executive Committee would continue to monitor the appropriations process.

Barbour summarized recent Executive Committee activities, two of which the Board will take action on shortly. The Committee approved a motion to change the assessment schedule, swapping out the Long-Term Trend (LTT) assessment for 17-year-olds with the LTT for nine-year-olds, who were tested immediately prior to the COVID-19 pandemic. The Committee also approved the nomination of Alice Peisch for the Board Vice Chair position. Barbour recounted that the Executive Committee talked with representatives from the National Academies who are studying NAEP, as Commissioner Carr mentioned. The NAS representatives asked Board members to recommend how NAEP processes could be modernized and more efficient.

Assessment Development Committee Chair Dana Boyd then provided an update on the ADC, noting that the Committee had remained focused on the Reading Framework and had reviewed the Chair's draft of the framework. She also mentioned that ADC recently reviewed the 2022 subject-specific NAEP student questionnaires in math, reading, U.S. History, and civics to approve the questions that will capture information about learning during the pandemic. Boyd also shared that ADC is developing recommendations for revising the framework process in preparation for a joint meeting with COSDAM in September and a full Board discussion in November. The recommendations will seek to improve the process to update frameworks in the future, beginning with the upcoming Science Framework update.

ADC also approved a motion to move the 2026 NAEP Reading Framework forward to full Board action. Boyd said the Committee also discussed upcoming activities, including a review of the Reading Assessment and Item Specifications. Boyd noted that Mark Miller suggested ADC prioritize the creation of abridged frameworks for math and reading. Boyd concluded by noting that Cunningham suggested that the Committee ask researchers to identify the contextual variables that would be most useful in their analyses of NAEP data.

Next COSDAM Chair Cizek discussed the Committee's work, including their evaluation of the need to develop a greater number of easier items on some NAEP assessments to measure students scoring below *NAEP Basic*. He added that in some subjects where there are greater numbers of items available, the challenge of administering an adequate number of items to students to align with their abilities is complex. Cizek proposed that this may involve a modification of content standards, test delivery mode, and other factors. Cizek said that COSDAM heard presentations on papers related to measurement and reporting for students performing below *NAEP Basic*. Cizek noted that additional research and Board discussion is needed to explore feasible, appropriate options to describe student knowledge and ability below the *NAEP Basic* level and urged further discussions with ADC and NCES.

Cizek reported that COSDAM also discussed the ongoing study to review and revise the achievement level descriptions for NAEP reading and math. Cizek explained that if the NAEP achievement level descriptions claim students can accomplish a task, then it is necessary to provide evidence that those students can perform those tasks effectively.

Nominations Committee Chair Jim Geringer reported that plans for new Board members whose appointments will begin in October were on schedule and that the Committee has already formulated an outreach strategy for the 2022 nominations campaign. Geringer stated that the outreach campaign will include a campaign website, social media videos, a webinar, and an outreach toolkit. Geringer listed the available positions the Board planned to fill in 2022: a fourth grade teacher, an eighth grade teacher, a secondary school principal, and a general public representative–Parent Leader. He encouraged Board members to share the campaign so the Nominations Committee could reach as many interested applicants as possible. Geringer praised Tessa Regis for her outstanding work on the nominations campaign.

R&D Chair Tonya Matthews then offered news from her committee. She reported that the virtual release of the 2019 NAEP Science results drew nearly 600 attendees. The release featured stakeholders from the National Science Teachers Association (NSTA), the National Association of Elementary School Principals (NAESP), and the National Association for Research in Science Teaching (NARST). In June, the Board partnered with NSTA and the Southern Regional Education Board (SREB) for a Twitter chat. She also stated that the R&D Committee discussed new approaches to improve the measure of SES within NAEP. Matthews stated that the R&D Committee also reviewed the communications and outreach plan, the Strategic Vision, as well as core contextual items. In July, the Committee reviewed the draft release plans for the Long-Term Trend Assessment and High School Transcript Study. Matthews requested a motion and second to approve the release plans.

Barbour accepted Matthews' motion and asked if anyone would second the motion. The motion was seconded by Beverly Perdue. The motion passed unanimously.

Returning to the Executive Committee's actions, Barbour acknowledged a recommendation by Jim Geringer, based on member feedback, to elect Alice Peisch as the Board's Vice Chair for the coming year. A motion was made by Wright and seconded by Cizek. The motion passed unanimously. Next, Barbour called on West, who made the motion to change the 2022 Long-Term Trend administration. Seconded by Matthews, the motion passed unanimously.

Recess

The meeting recessed at 2:41 p.m. and reconvened at 3:02 p.m.

ACTION: NAEP Reading Assessment Framework

Barbour opened this session by explaining that updating the 2026 Reading Framework fulfills one of the Board's primary legislative responsibilities. Barbour noted that members of the Chair's working group and ADC worked tirelessly over the past two months to reach a draft on which the Governing Board could reach consensus approval. Barbour clarified that the Board's Framework Development policy ensures a comprehensive process that accounts for professional standards, current research, and national needs for the NAEP Reading assessment. He added that the framework benefitted from the input of many experts and stakeholders, both internal to the Board and external. Barbour expressed his gratitude, on behalf of the Board, to the Visioning Panel and Development Panel who, under the leadership of David Pearson, devoted two years of expertise to update the Reading Framework. Barbour thanked the Technical Advisory Committee, stakeholders, and Board members. Barbour asked Board members Patrick Kelly and Carey Wright to provide context for the proposed framework.

Kelly explained that the Board is legislatively mandated to identify the content for each NAEP assessment. Congress does not indicate when frameworks must be updated, however, frameworks should maintain relevance in what is measured and reported. Kelly explained that in 2019 the Board determined an update was needed for the Reading Framework, which was last updated in 2004. This framework update was necessary due to changes made in the field of reading as well as NAEP's transition to a digital-based assessment.

Kelly referred to advances since the Board last adopted updates in 2004 in text comprehension research and societal changes that impact the ways students engage with text, especially digital text. Kelly stated that the updated version of the framework had a modestly updated definition of reading comprehension to include factors that influence student comprehension including social and cultural experiences. The updated version of the framework has also expanded the definition of text in response to the proliferation of digital media. It elevates the importance of disciplinary reading by creating sub-scales in science and social studies in addition to literature. The updated framework also employs principles of universal assessment design to support valid measurement of all students' reading comprehension, consistent with other large-scale assessments. Kelly stated that the updated framework aims to increase the relevance and usefulness of NAEP to the nation by prioritizing deeper levels of disaggregation in NAEP Report Cards, by disciplinary context, by looking at SES within race and ethnicity, and by increasing the amount of reporting around former and current English language learners who take the assessment.

Kelly added that the Development and the Visioning Panels spent two years developing a robust framework aligned to current practices and research. A public comment process invited interested parties to submit responses to the framework. Kelly said he felt confident the Board had reached consensus and that no member believed that the final product represented any one member's personal views. He asserted that the framework must be neutral on curriculum to fulfill NAEP's purpose as established by Congress.

Wright reminded Board members that, as a nonpartisan body, the Board strives for consensus on major decisions. The path to consensus for the 2026 NAEP Reading Framework involved many stakeholders whose key recommendations were focused on how to assess reading – not on reading instruction. The framework was crafted to clarify the Board's commitment to maintain NAEP's trend lines and a commitment to rigorous, inclusive, and unbiased assessment. Wright stated that the Board's discussions had led to a final version that reflects these expectations. She emphasized that, guided by the Board's Strategic Vision, the framework will support the Board's efforts to inform, innovate, and engage. Wright stated that recent NAEP results have shown that a crisis in early literacy and the COVID-19 pandemic will have only made it worse, leaving students further behind in their learning. Wright concluded by saying that the framework update reflects NAEP's emphasis on rigor, quality, and ability to chart trend; she expressed pride in the Board's work.

Barbour thanked Wright and paused for comments or questions. Upon hearing none, Boyd moved the 2026 NAEP Reading Framework for adoption by the full Board. Miller seconded the motion. Barbour thanked Miller and asked if anyone was inclined to engage in further discussion. Miller thanked everyone involved in the Reading Framework process, including Board staff, former staff member Michelle Blair, CCSSO, the Chair's working group, and others, without whom, he said, the Board would not have reached consensus. Kelly added to Miller's commendations, expressing his thanks to Blair as well as to Sharyn Rosenberg.

The motion passed unanimously. Barbour thanked and congratulated the Board. He then recognized Cunningham to provide an overview of the Technology and Engineering Literacy (TEL) Assessment.

Overview of the Technology and Engineering Literacy Assessment

Cunningham provided a brief history of the TEL assessment. She stated that the impetus for this assessment came from National Academy of Engineering and from National Research Council reports. In 2002, the Technically Speaking report outlined why all Americans needed to know more about technology. Cunningham explained that this was followed by the Tech Tally report in 2006, which provided steps to address this technological knowledge gap via assessment. The report recommended, specifically, that the Governing Board develop a framework for a NAEP assessment on technology and engineering literacy. Cunningham stated that TEL debuted in 2014 and was administered again in 2018. A third administration is currently planned for 2024.

Cunningham stated that TEL assesses technology and engineering literacy, which is defined as the capacity to use, understand, and evaluate technology, as well as to understand technological principles and strategies needed to develop solutions and achieve goals. She said that technology, according to the TEL Framework, is any modification of the natural world that is done to fulfill human needs or desires such as a toothbrush or bandage or a water or waste system. She stated that the one thing these technologies have in common is that they are created and refined through an engineering process. Cunningham clarified that the TEL Framework defines engineering as a systematic and often iterative approach to designing objects, processes, and systems that meet human needs and wants. She added that TEL is an assessment for all students, not just those pursuing a STEM education or occupation.

Cunningham stated that the current TEL Framework includes three assessment areas. The first assessment area, technology and society, encapsulates the effects that technology has on society and the natural world. It also allows students to explore ethical questions that can arise from the use of those technologies. The three subareas that exist under the technology and society section are the interaction of technology and humans; the effects of technology on the natural world; and effects of technology on the world of information and knowledge. The second major area of TEL focuses on design and systems. This section gauges students' understanding of how important the design process is in comprehending and accessing technologies. She said that four subareas in this section are outlined in detail in the assessment framework.

Cunningham said the third major area is information and communication (ICT) technologies. She clarified that this area covers computer and software learning tools, networking systems and protocols, handheld digital devices, and other technologies for accessing, creating and communicating information, and for facilitating creative expression. She said that the ICT domain was made up of five subareas: the construction and exchange of ideas and solutions, information research, investigation of problems, acknowledgment of ideas and information, and selection and use of digital tools.

Cunningham described the three cross-cutting practices in the TEL assessment: understanding technological principles; developing solutions and achieving goals; and communicating and collaborating. Cunningham turned the presentation over to Bill Ward of NCES, to provide Board members with a brief overview of the TEL operational assessment and to describe current challenges.

Ward clarified that the purpose of the TEL assessment is to measure students' knowledge and abilities in the areas of technology and engineering. Ward outlined the assessment's three content areas: technology and society; design and systems; and information and communication technology. Ward provided an overview of the item types on the TEL assessment. The TEL assessment has discrete items of various lengths, all of which are interactive and one to three minutes long. The assessment also features interactive scenario-based tasks, also of varying lengths, between 10 and 30 minutes. Ward stated that the framework calls for 80% of overall student testing time to be spent on scenario-based tasks and 20% to be spent on discrete items and shared examples of each item type. Scenario-based tasks proved to be labor-intensive in design and development. The 2014 assessment was device- and operating-system- dependent and was only able to run on Windows XP. Similar challenges remain in advance of 2024.

As 2024 nears, Ward said, TEL needs to transition to the NextGen eNAEP delivery platform. This next generation platform supports online delivery of multiple NAEP subjects, and can support the reading, math, civics, and U.S. history assessments. This platform is online and would be device- and operating system-independent. He added that platforms must be simple, sustainable, and easy to maintain so as to meet evolving technological changes. Right now, however, operations for TEL suffer from outdated laptops, older versions of web browsers, and development platforms.

Ward then outlined preliminary plans for 2024. He stated that NCES has preliminary plans to reprogram the TEL tasks within the NextGen eNAEP platform, since all other subjects, such as reading or math, would be delivered on this platform. Ward stated that they planned to reprogram only a portion of the assessment for 2024 due to budgetary constraints. This would allow for a special one-time reporting of scores for one subscale, but no composite scores.

Cunningham stressed that many educators view the TEL assessment as the gold standard for providing data about what students can do. She also stated that the Next Generation Science Standards (NGSS) were released in 2013, after much of the TEL work had been completed. She noted that these science standards were the first to include engineering explicitly. She added that many states now have technology and engineering as part of their state science standards, regardless of whether they have officially adopted the NGSS. Across the country, teachers are engaging students in engineering and technology concepts and practices. Cunningham said that the Board needs to consider whether to incorporate technology and engineering into the 2028 NAEP Science Framework update more directly. Ideally, the Board would delay decisions about TEL on the assessment schedule until the Board decides whether the TEL content will be incorporated into the Science Framework update. Cunningham opened the floor for questions and comments.

Kelly sought greater understanding of the type of knowledge the Board wanted high school graduates to know. Kelly applauded the TEL assessment for its creativity but asked how the assessment would test for soft skills.

Geringer stated that TEL initially had pushback and should remain a stand-alone assessment. Edelblut agreed with Geringer. He said it is necessary to separate knowledge from applied knowledge. TEL, from Edelblut's perspective, reflects only applied knowledge. Edelblut asked if it was possible to determine if a poor performance on an assessment item derives from a lack of core knowledge or logical thinking or from an inability to apply that knowledge.

Cunningham explained that the many fields within engineering means that the assessment challenges students to approach a problem from many different angles. Cunningham stated that there was not a substantial amount of underlying knowledge required for the TEL assessment. Most of the information needed for students to engage in deductive reasoning is provided on the assessment. Ward added there is no assumption that students have extensive math or science knowledge when they take the assessment. If anything, the assessment is used to understand the ways students wield knowledge and information to solve problems. Any information students may need to arrive at their individual conclusions is provided to them. Ward stated that, at present, they do not currently have a way of discerning logical thinking and applied knowledge.

West asked about whether any research has examined the predictive validity of students' performance on the TEL assessment. West also asked for clarification on the need for mastery of the underlying content knowledge that students have to draw on and apply. Cunningham deferred to Carr, who made two points. First, both item types on the assessment were developed with an evidence-centered design (ECD) approach. Part of that approach is to collect evidence of validity as the task is developed. NCES has some validity evidence about those tasks that may be informative. Carr's second point was that NCES participated in an international assessment that was similar to TEL and the distribution of scores lined up very well. NCES plans on participating in that assessment again.

Wright was particularly struck by the scenario-based task example that was provided. She stated that the metrics associated with student responses to this assessment item, especially the drop off of six percent, indicated that students had difficulty explaining their reasoning. Wright emphasized that this would be important to observe across assessments, not just in TEL. She urged the Board to think of the best ways to support students in this area. Cunningham stated that explanation and justification was just one of the eight practices in the science standards.

Nardi Routten asked if the assessment was administered only to eighth grade students. Routten then wondered if the Board merged the science and TEL assessments, would fourth- graders have to answer grade 8 questions? Cunningham stated that the Framework covers fourth, eighth, and twelfth grades but, thus far, an assessment exists only for eighth grade. Ward clarified that the eighth grade assessment would not be given to fourth graders.

Gasparini stated that the TEL assessment should not be distributed through old technology. He also stated that assessing twelfth graders would not provide much actionable information.

Matthews drew from her experiences leading a student focus group on the TEL assessment during a release event. Some of the highest performers on TEL were not the students who wanted to pursue careers in engineering. Matthews reminded the Board not to interpret TEL as only representing STEM-bound students. Instead, there are multiple ways to arrive at the correct answer and that this assessment is more representative of process data. Matthews said some students may have taken a science-based approach to arrive at an answer while other students may have leaned into civics.

Eric Hanushek asked if the TEL assessment was more expensive compared to other assessments. Ward divulged that the TEL assessment was relatively expensive to develop and administer. Hanushek responded saying this could lead to a budget issue.

McGregor stated that when talking about engineering design and principles, there are multiple ways of getting to the right answer. He also responded to Gasparini's comments, stating that it was more imperative for educators to bring fourth graders into the fold. McGregor also addressed comments made by Ward, stating that everyone is somewhat familiar with technology. If anything, it is more important to acknowledge this familiarity within the assessment so students are aware of their competence.

Cizek recommended that NCES research if measuring how students arrive at solutions through written responses, as TEL does, is valid. He questioned if, by only retrieving written responses, some students would be disadvantaged, because they were tasked with writing instead of expressing their solutions through an alternative medium.

Reynolds stated that he found the TEL-specific definitions of technology and engineering to be beneficial. He added that if the assessment were to ask grade 8 students about technologies, the participants would point to real-life examples. He then asked if the TEL assessment is designed to gauge students' knowledge of underlying concepts of engineering and technology. He believed this would strongly correlate with achievement.

Carr replied to several of these comments. First, NCES and the Board need to look at the twelfth-grade framework since TEL was administered only to eighth graders and has not been operationalized for other grades. Carr interpreted Cizek's question as one that addressed issues surrounding equity. Carr cautioned that although NCES now has process data, the TEL assessment may not capture as much process data as other assessments.

Cunningham closed with responses to other Board members' remarks. She said that Cizek's comment about TEL and writing reflects what she often hears from teachers – that doing engineering in the classroom gives them insight into how students think, especially English language learners and others who have difficulty expressing themselves in English. She said it is important for the Governing Board to think about how students' skills can be assessed through process data and other methods that are not reliant on students' verbal abilities. Cunningham then addressed Reynolds' question on how students define technology and engineering. She referenced multiple studies which found that students relate the terms to digital technologies, not simple technologies such as Band-Aids, bicycles, etc. She said that students are not asked to define technology and engineering on the TEL assessment, but the examples in the assessment inspire students to think beyond digital technologies.

The meeting adjourned for the day at 4:42 p.m.

NAEP Budget And Assessment Schedule (CLOSED)

Under the provisions of exemption 9(B) of §552b(c) of Title 5 U.S.C. and exemption 9(b) of §552b(c) of Title 5 U.S.C., the National Assessment Governing Board (Governing Board) met in closed session on Friday, August 6, 2021, from 10:00 a.m. to 11:33 a.m. to receive a briefing from Peggy Carr, Acting Commissioner, NCES, on the NAEP budget and Assessment Schedule. Chair Barbour announced that the session was closed to the public and that online participation would be monitored to assure only approved attendees participated in the session.

Lesley Muldoon, the Governing Board's Executive Director, referenced the prior day's Executive Committee meeting, where members received an update on NAEP appropriations. Muldoon stated that the House bill included an additional $40 million, and the Senate was working on the legislation. She noted that the legislation may include provision for additional annual NAEP appropriations to administer the NAEP Schedule of Assessments as adopted by the Board. Muldoon noted the appropriations update as the context for the Board's discussion of the NAEP budget.

Next, Carr provided a briefing on the NAEP budget and its impact on the NAEP Assessment Schedule. The briefing covered three areas: anticipated implications of COVID in the 2022 administration; budget implications through FY 2024, and costs of upcoming assessments. To the latter, Carr provided budget information for 2022 Long-Term Trend, 2024 Civics, and 2024 TEL.

Carr addressed the Board's questions throughout her presentation.

Recess

The meeting recessed at 11:33 a.m. and reconvened at 11:47 a.m.

Briefing on Upcoming NAEP Releases (CLOSED)

Under the provisions of exemption 9(B) of §552b(c) of Title 5 U.S.C., the Governing Board met in the second closed session on Friday, August 6, 2021, from 11:47 a.m. to 1:15 p.m. to receive briefings on upcoming NAEP Report Card releases of Long-Term Trend (LTT) Reading and Mathematics (9- and 13-year olds) and the High School Transcript Study (HSTS). The briefings were conducted by Grady Wilburn and Linda Hamilton, respectively, both of NCES.

After providing a brief background on the LTT assessments, Wilburn reported that the LTT assessment national sample was drawn for 9- and 13-year-old students. The assessment was administered on paper during the 2019-2020 school year with age 9 students in January-March 2020 and age 13 students in October-December 2019. The 2020 reading and mathematics performance metrics are reported as national average scores (0-500 scale), percentile scores, student group scores, and LTT performance levels scores (300, 250, 200 and 150) in both reading and math at different age groups.

Wilburn then shared highlights from the 2020 LTT results. Results were reported by race, ethnicity and gender, performance levels, achievement gaps, percentages of students reading for fun at ages 13 and 9, and course taking patterns in math compared to prior years.

Members asked questions during and after the presentation, which Wilburn addressed.

Next, Linda Hamilton previewed the 2019 NAEP HSTS results. She noted that the study is an administrative data collection of transcripts linked to the NAEP 12th grade mathematics and science assessments. Transcript data are collected from a nationally representative sample of graduating seniors in public and private high schools, about 47,000 high school graduates in 1,400 public and private schools. HSTS captures the types of courses that graduates take, covering grades 9 through 12, the number of credits they earn, and grade point averages earned along with the students' NAEP performance.

As in prior years, the inclusion criteria for the 2019 HSTS are that students must have graduated with a regular or honors diploma; completed at least three years of coursework that includes the 12th grade assessment year (i.e., 2018-2019 school year); earned at last 16 Carnegie credits; and earned a positive number of Carnegie credits in English courses.

Results were reported by gender, high school graduation rates, English learners, race/ethnicity, school locale, and student disability status as tracked by individualized education plans. Courses reviewed covered three major categories: Academic (English, mathematics, science, social studies, visual and performing arts, world languages); Career/Technical Education (computer-related studies, Other CTE courses); and Other (Physical and health education, religion, military science, and all other courses).

Hamilton shared preliminary results describing core measures—(1) average course credits earned, (2) Grade Point Average (GPA), and (3) curriculum levels (standard, midlevel and rigorous). GPA results were reported overall and by course type, student gender, student race and ethnicity as well as compared to GPAs of high school graduates in previous rounds of HSTS – 1990, 2000, 2009, and 2019.

Members asked questions after the presentation, which Hamilton addressed.

Records Management Briefing and Discussion (CLOSED)

Under the provisions of exemption 9(B) of §552b(c) of Title 5 U.S.C., the National Assessment Governing Board (Governing Board) met in closed session from 1:15 p.m. to 2:01 p.m. to receive an administrative briefing on federal records management requirements. Jason Lautenbacher, Chief, Information Branch at the Department of Education, presented the briefing for Governing Board members, who are considered Special Government Employees (SGEs).

Lautenbacher defined federal information, stating that it is any information that is created or received in conjunction with work related to the transaction of Department of Education business. Lautenbacher highlighted the responsibilities and obligations of members in preserving and protecting federal records. He recommended that SGEs avoid creating paper information as much as possible in accordance with OMB directive M-19-21; ensure all work-related information not publicly available is always encrypted or protected; and ensure all work-related information is forwarded to a department point of contact during member's tenure. Further, he indicated that members cannot retain any work-related information after their tenure expires. He provided members contact information for any questions.

Lesley Muldoon recommended a process for identifying and forwarding records, noting that members and staff have responsibilities in categorizing and preserving permanent records or temporary records. Muldoon explained that records created during the course of Governing Board business are already managed and preserved by staff. This includes records such as meeting minutes, Governing Board actions, and formal communications related to Controlled Unclassified Information (CUI).

Board members engaged in a question-and-answer session that referred to encrypting drives and files, forwarding documents, marking controlled unclassified information, and redacting personal information.

Recess

The meeting recessed at 2:01 p.m. and reconvened at 2:17 p.m.

Across the Board: Understanding Recent NAEP Results

The Board reconvened in open session, at which time Barbour introduced Ebony Walton of NCES. Walton was invited to share insights from an analysis of NAEP reading, mathematics, and science data.

Before summarizing the NAEP reporting team's findings, Walton stated that Carr wanted to bring the Board's attention to two studies: The first is the math curricula study. That study included an analysis that showed labels for algebra and geometry courses can be deceiving. In some cases, courses labeled honors do not cover advanced content. Walton stated that NCES's Daniel McGrath would send Board members a link to the study. The second study compared long-term trend to main NAEP. The study controlled for demographic changes since the 1990s to analyze score changes. Walton stated that McGrath would also share this study with the Board.

After these preliminaries, Walton began her presentation, "A Decade of Monitoring Study Progress (or Lack Thereof) Through the Lens of NAEP." To establish context, Walton listed the high-profile education topics from 10 years ago such as the adoption of the Common Core, demographic changes, and the educational impacts of the Great Recession. Walton gave an overview of NAEP data collected from 2009 to 2019 at grades 4, 8, and 12. Walton posed four key questions: (1) How have eighth graders performed across multiple subjects? (2) Looking at grades 4, 8, and 12, how has student performance changed? (3) Which states or TUDA districts stand out for having made gains or declines on NAEP mathematics and reading over the last decade? And, (4) Who are the lower-performing students?

Walton stated that, on average between 2014 and 2019, eighth graders' scores improved in TEL, declined in geography, reading, and U.S. History, and did not change significantly in math, science, and civics. Walton asked Board members to ponder what may explain this trend. Miller posited that engagement in the assessments may be a factor, since TEL goes beyond asking students basic questions and presents them with opportunities to elaborate. Matthews added that geography, reading, and U.S. history were seemingly related, which may have something to do with information recall or comprehension. Carr added that TEL is a literacy assessment while the others are not.

Walton pointed out that there are TEL components integrated in other subjects such as U.S. History, science, and geography. She then delved beyond the overall averages and showed that scores declined for lower-performers, while higher-performers either held steady or improved. A similar divergence appeared in comparisons between 2009 and 2019. In science, grade 8 students showed an improvement overall and at both ends of the score distribution. Walton noted, however, that while science results had improved over the decade, lower-performing students' scores have declined in more recent years.

Walton then addressed her second question: "Looking at grades 4, 8, and 12, how has student performance changed?" Walton summarized score changes and patterns across all three grades from 2009 to 2019 in math, reading, and science. Lower-performing students' scores declined across the board, except in grade 8 science. At the same time, higher-performing students' scores improved or stayed the same across grades and subjects, with the exception of grade 12 scores.

Next, Walton disaggregated the data by student race/ethnicity: white students in the 10th percentile saw scores declining, except in grade 8 science, but white students in the 90th percentile saw overall improvements in their scores. A similar trend is observed among Black students. Walton stated that Hispanic students in the 10th percentile have held steady, with some improvements in science. She also noted that Hispanic students at the 90th percentile scores have improved. For Asian students, there was no significant change at either the 10th or 90th percentiles in most grades and subject areas. And, there is little change in scores among Native Americans, Alaskan natives, students of two or more races.

Walton summarized the overall pattern of divergence between higher- and lower-performing students across races, grades, and subject areas. She added that reporting overall scores for student groups by race/ethnicity can obscure the changes happening *within* the groups. For example, over the course of the decade, average scores for white students did not change. However, scores within the group did change as higher- and lower-performing students' scores diverged. Walton also noted that lower-performing students' scores declined in all subjects and grades except grades 4 and 8 science. Higher-performing students' scores improved in grades 4 and 8 math, reading, and science. And lower- and middle-performing students' scores declined in grade 12 math and reading.

Walton moved on to the third question: "Which states or TUDA districts stood out for making gains or declining on NAEP mathematics and reading over the last decade?" Walton stated that four jurisdictions had overall score increases between 2009 and 2019 – California, the District of Columbia, the Department of Defense Schools, and Mississippi. Five jurisdictions had overall score decreases in the same time period– Arkansas, Kansas, Montana, North Dakota, and Vermont. Walton invited reactions from Board members.

Cunningham and Kelly observed that the jurisdictions with overall score decreases seemed highly rural, with the exception of Mississippi. Walton stated that scores have increased among students at the 90th percentiles in all jurisdictions with overall score increases. By contrast, the states that experienced a decline in overall scores all showed declines among students at the 10th percentile. Walton distilled these findings to support that jurisdictions' overall score increases are generally driven by increases in higher-performing students' scores and jurisdictions' score decreases are generally driven by declines in lower-performing students' scores.

Walton then showed how scores have changed at the 90th and 10th percentiles among TUDA districts. Overall, students in TUDA districts have made gains in grade 4 math and grade 8 math and reading. Grade 4 reading scores have not changed significantly. Gains are particularly strong among higher-performing students. No TUDA districts had overall score declines across grades and subjects. D.C. made gains across all subjects, and three TUDA districts made gains in three of four subject/grade combinations: Atlanta, Chicago, and Miami-Dade.

Walton closed by sharing insights with the Board on lower-performing students. In both math and reading in grades 4, 8, and 12, scores have dropped for students at the 10th percentile over the past decade. The percentage of students scoring below *NAEP Basic* has also increased in most grades and subject areas. Students scoring at or below the 25th percentile comprise about a third white students, a quarter Black students, a third Hispanic students, and 3% Asian students, 1% Native American and Alaskan Native students; and 3% students of two or more races. Slightly more than two-thirds of these lower-performing students are eligible for the National School Lunch Program; 41% had parents who did not graduate from college; 31% were classified as having a disability; and 19% are English learners.

Walton cited an article from the Organization of Economic Cooperation and Development (OECD) discussing lower-performing students and why they fall behind. She quoted the article, stating, "Poor performance is not the result of a single risk factor, but rather a combination and accumulation of barriers that affect students throughout their lives." Walton stated that she hoped, moving forward, the reporting team would have the opportunity to analyze the complexities of lower-performing students, especially in the wake of the COVID-19 pandemic.

Looking ahead, Walton outlined the assessment schedule for grades 4, 8, and 12 between 2022 and 2030. Walton left members of the Board with questions about the future: Will eighth graders make progress in subjects? Will lower-performing students' scores continue to decline? Will higher-performing students' scores continue improving? Will higher-performing students continue to drive overall score improvements? And, how will COVID-19 affect this?

Walton reported on additional NCES activities related to this topic. An expert panel provided recommendations on how to better measure and report on the skills of students who perform below *NAEP Basic*. Walton mentioned survey efforts like the NAEP 2021 School Survey, the upcoming monthly IES School Pulse Panel, and NAEP reporting efforts, which will examine skills of students across the score distribution. Walton thanked the Board and concluded her presentation, opening the floor for questions and comments.

Suzanne Lane asked, among states and cities with score declines, were demographics of students in 2009 similar to those in 2019? Walton stated that shifts in demographics would be picked up more at the state level than the national level. She continued, saying much of her analysis was done at the national level, but the reporting team would investigate Lane's question further. Walton stated that D.C. stood out, noting that its populations of white and Hispanic students had increased, and perhaps this shift could be observed in other cities as well.

Hanushek stated that it seemed as though the data were primarily focused on the two end points of the distribution and that there could be sampling errors since students below *NAEP Basic* may only be able to answer two or three questions on the test. He suggested a more reliable gauge of change over time would be to draw a regression line through all data points rather than relying so much on two end points of the distribution.

Cramer expressed concerns that educational inputs, at any level, take time to work. He stated that he was particularly concerned that these assessments did not seem to measure the length of a time a student was enrolled in a state or TUDA district. NAEP should measure this in order to evaluate student ability more accurately, which will provide more insights about the 10th percentile.

Rafal-Baer said she took issue with the bluntness of the NCES socioeconomic status (SES) data, saying that she felt greater insights into SES could be achieved with better data, especially regarding student access (or lack thereof) to necessary technology.

Carr asked to respond to Hanushek's comments. She restated that Hanushek's concern was a dearth of questions that students at the low end of the distribution are able to answer. She agreed that this was a legitimate concern. In terms of modeling the results, the reporting team is less concerned, because the sampling error is the predominant component of the standard error and just 20% of the error is measurement error.

West thanked Walton for her presentation, stating that it was of great value and this type of analysis needed to be shared more widely. West asked if the divergent score pattern was reflective of certain regions improving and others declining, or if the divergence was happening within regions. Walton said there is a mixed bag when it comes to distributions of scores within the states and stated that the range of average scores among states has narrowed. However, within states and districts, there are examples in which the divergent score pattern has not been observed.

Matthews reflected on a pertinent question: Who are we (members of the Board) assuming are on the bottom? Matthews stated that Walton's presentation challenges assumptions about the makeup of students at different ends of the performance distribution, and the potential impacts of the COVID-19 pandemic on student learning and achievement. Walton replied that a challenge for the reporting team is finding effective ways to describe lower-performing students.

Barbour asked about the degree to which a child being read to at home affected their achievement. Walton referred to contextual data about students reading for fun and the positive relationship that exists with reading scores.

Barbour then transitioned into the final segment of the meeting, farewell remarks from Geringer and Cizek.

Farewell Remarks

Barbour first expressed appreciation to Alice Peisch, whom he looks forward to seeing in-person at their next meeting. He then invited Geringer and Cizek to provide any parting remarks as they conclude their service on the Board.

Geringer expressed gratitude for Barbour's statement and thanked all those in attendance, particularly those facilitating the event. Geringer expressed appreciation for Carr, specifically, stating that he admired her ability to present data and answer any and every question asked by Board members. He concluded by thanking the Board staff. Barbour thanked Geringer and called upon Cizek for his farewell remarks.

Cizek said he wished to address three things: gratitude, admission of personal failures, and policy advice for the future. Cizek acknowledged his COSDAM colleagues: He expressed appreciation for Peisch for her leadership, especially in her masterful management of achieving framework consensus; Rafal-Baer for her engagement in framework development, her commitment to getting broader input for consensus, her friendship, and her encouragement; Hanushek for pushing the Committee members to widen their perspectives; Whitehurst and Wright for their service on the Chair's working group; and Lane for her insightful advice. Cizek reiterated his appreciation for every member of COSDAM, stating that they are a group of low multitudes, high aptitudes, and stellar attitudes (a reference to Geringer).

Cizek then outlined three main regrets. The first was that he should have listened more to Sharyn Rosenberg. He stated that her knowledge of psychometrics and of NAEP is extremely beneficial to the Board and that she will be an asset to ADC in her new role. His second regret was his lack of progress in pushing forward a new process for framework development. Cizek urged the Board to revisit the composition of framework panels. He asserted that membership on the panels should include greater representation by people who teach the subjects discussed and insisted that those doing the work – teachers – must be given the platform to amplify their voices. Cizek's final regret was that the labels of *NAEP Proficient*, *NAEP Basic*, and *NAEP Advanced* were still in trial status. He described the NAEP achievement levels as the signature reporting and interpretation mechanism for NAEP results, relied on by policymakers, and the standard by which states judge their own achievement levels. He added that it would be unwise to consider adding new levels such as below basic before the trial status is resolved. In conclusion, Cizek advised members of the Board to speak up immediately as Board terms pass quickly and time waiting to learn the fundamentals is wasted. Cizek stated it had been an honor to serve on the Board.

Barbour concluded the meeting by offering thanks to the staff for their work in organizing the Board meeting. Barbour concluded his remarks, stating the next meeting was scheduled for November.

The meeting adjourned at 3:29 p.m.

I certify the accuracy of these minutes.

______________________________ 11/4/2021
Chair Haley Barbour Date

National Assessment Governing Board
Executive Committee Meeting
Report of August 5, 2021

OPEN SESSION

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

National Assessment Governing Board Members: Alberto Carvalho, Tyler Cramer, Christine Cunningham, Frank Edelbut, Paul Gasparini, Eric Hanushek, Patrick Kelly, Suzanne Lane, Reginald McGregor, Julia Rafal-Baer, Ron Reynolds, Nardi Routten, Mark White, Russ Whitehurst.

National Assessment Governing Board Staff: 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: Tammi Adams, Samantha Burg, Jing Chen, Brian Cramer, James Deaton, Enis Dogan, Pat Etienne, Eunice Greer, Shawn Kline, Taslima Rahman, Holly Spurlock, Ebony Walton, Grady Wilburn, William Tirre.

U.S. Department of Education Staff: None.

Other attendees: Chris Averett, Vickie Baker, Greg Binzer, Brittany Boyd, Lauren Byrne, Jay Campbell, Randon Dart, Gloria Dion, Amy Dresher, Stuart Elliot, Gary Feng, Kim Gattis, Joy Heitland, Andrew Ho, Subin Hona, David Huff, Young Kim, Sami Kitmitto, Judith Koenig, Andrew Kolstad, Beth LaDuca, Joanne Lim, Richard Luecht,Nadia McLauglin, Jon Noble, Ranu Palta-Upreti, Emilie Pooler, Sonya Powers, Shannon Richards, Lisa Rodriguez, Rick Rogers, Keith Rust, Renee Savoie, Debra Silimeo, Peter Simmons, Anthony Velez, Llana Williams, Karen Wixson, Edward Wofford.

The Executive Committee met in open session from 10:30 a.m. to 11:10 a.m. to consider a change to the assessment schedule, to take action on the nomination for Vice Chair of the Governing Board, as well as to meet with representatives from the National Academies of Science, Engineering, and Medicine.

The session was called to order by Chair Haley Barbour at 10:30 a.m.

Barbour reminded everyone in attendance that the meeting is being conducted in a hybrid environment and set the rules and procedures for participating.

Barbour proceeded to two actions.

First, the Executive Committee considered a modification to the assessment schedule, replacing Long-Term Trend (LTT) Administration for age group 17-year-olds in 2022 with age group 9-year-olds. Having raised the idea at the May Board Meeting, Marty West expressed the need for the change in schedule because LTT 9-year-olds were the last age group assessed before the COVID pandemic. West noted this moment in time as a unique opportunity to better understand student progress and the impact of the COVID pandemic. Barbour called for a motion. Tonya Matthews made the motion to accept the proposed change to the assessment schedule, and it was seconded by West. The Executive Committee voted unanimously in favor of this assessment schedule change.

Second, the Executive Committee took up the nomination of Board Vice Chair to serve the next annual term in 2021-2022. Barbour asked Jim Geringer, who is completing his second and final term on the Board, to lead the discussion. Geringer had reached out individually to Board members to gauge interest in who should serve as Vice Chair. Geringer reported back overwhelming support for Alice Peisch to be renominated and elected as Vice Chair. Barbour asked for a motion. Marty West moved to nominate Alice Peisch as Board Vice Chair for 2021-2022, and it was seconded by Mark Miller. The Executive Committee voted unanimously in support of Peisch continuing to serve as Vice Chair for the next term. Barbour thanked Peisch for her incredible partnership and service over the last year.

Barbour then invited several representatives from the National Academies of Science, Engineering, and Medicine to discuss a study they are conducting of the National Assessment of Educational Progress program titled “Opportunities For NAEP In An Age of AI And Pervasive Computation: A Pragmatic Vision.” The presentation was led by Stuart Elliot, Study Director, and Karen Mitchell, Study Panel Chair, who shared details about the study and asked for advice on how to make the recommendations clear and actionable. They also asked the Board for ideas on how to achieve cost-efficiencies for the NAEP program.

Several members offered suggestions. Barbour shared his support for maintaining two-year periodicity of reading and mathematics assessments, arguing that changing periodicity to every four years should not be the mechanism for reducing the program’s costs. Tonya Matthews expressed an interest in learning more about the study’s recommendations as educational assessment technology is developed and increases efficiency. Reminding everyone of the importance of motivating and engaging students taking NAEP, Mark Miller noted that students have gone from fill-in-the-bubbles to scenario-based tasks.

Greg Cizek commended the National Academies on the work they are conducting. Cizek continued that sometimes improvements cost more money but can lead to efficiencies longer term. Cizek gave the example of technological advancements like automated scoring which can be more efficient than human scoring of assessments. Cizek suggested that the National Academies keep in mind the scale of innovations or improvements and shared that sometimes it is difficult to know how much or how little these changes are going to cost.

Lesley Muldoon, Executive Director, shared that Governing Board staff met with the representatives from the National Academies several weeks ago to also discuss this study. During that meeting, Board staff discussed topics such as automated scoring and updating frameworks.

Reginald McGregor noted the importance of keeping in mind the needs of industry and the workforce when developing assessments. In addition, McGregor talked about the need to increase efficiency and that NAEP needs to be updated to keep up with technological advancements.

At 11:10 a.m. Chair Barbour ended the open session.

CLOSED SESSION

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

National Assessment Governing Board Members: Tyler Cramer, Christine Cunningham, Frank Edelbut, Paul Gasparini, Eric Hanushek, Patrick Kelly, Reginald McGregor, Ron Reynolds, Nardi Routten, Mark White.

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

National Center for Education Statistics Staff: Peggy Carr, Enis Dogan, Veda Edwards, Pat Etienne, Enuice Greer, Shawn Kline, Dan McGrath. Nadia McLaughlin, Holly Spurlock, William Tirre, Ebony Walton.

U.S. Department of Education Staff: None.

The Executive Committee met in closed session from 11:30 a.m. to 12:00 p.m. to discuss the NAEP budget and assessment schedule, in addition to other Governing Board priorities.

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.

Barbour introduced Lesley Muldoon, Executive Director, who provided an overview of the assessment schedule and an update on the Fiscal Year 2022 congressional appropriations process.

Barbour then introduced Peggy Carr, Acting Commissioner, National Center for Education Statistics (NCES). Carr led a presentation on the Budget and Assessment Schedule. Carr provided information about projected costs for the program, the impact of COVID and school closures on the costs, the projected budget implications for the assessment schedule, an update on the congressional appropriations process, and projected costs for research and development.

At 12:00 p.m. Chair Barbour adjourned the meeting.

I certify the accuracy of these minutes.

Haley Barbour, Chair

10/21/2021
Date

National Assessment Governing Board

Assessment Development Committee

Report of August 5, 2021

ADC Members: Dana Boyd (Chair), Mark Miller (Vice Chair), Christine Cunningham, Frank Edelblut, Patrick Kelly, Reginald McGregor and Nardi Routten.

Governing Board Staff: Executive Director Lesley Muldoon, Deputy Executive Director Lisa Stooksberry, Stephaan Harris, Sharyn Rosenberg and Angela Scott.

NCES Staff: Tammie Adams, Enis Dogan, Eunice Greer, Shawn Kline, Nadia McLaughlin, Taslima Rahman and Ebony Walton.

Other Attendees: American Institutes for Research: Brittany Boyd, Markus Broer, Kim Gattis, Cadelle Hemphill and Xiaying Zheng. CRP: Shamai Carter, Subin Hona, Anthony Velez and Edward Wooford. Educational Testing Service: Jay Campbell, Gloria Dion, Kadriye Ercikan, Hilary Persky, Emilie Pooler and Karen Wixson. Hager Sharp: David Hoff and Joanne Lim. The Hatcher Group: Jenny Beard and Jenna Tomasello. Management Strategies: Brandon Dart. Pearson: Joy Heitland and Eric Moyer. Westat: Greg Binzer, Lauren Bryne, Lisa Rodriguez and Rick Rogers. WestEd: Mark Loveland and Sonya Powers. Other: Vickie Baker (West Virginia Department of Education), Laura Goadrich (Arkansas Department of Education), Renee Savoie (Connecticut Department of Education) and Sarah Schwartz (Education Week).

Welcome and Review of Agenda

Chair Dana Boyd called the meeting to order at 9:05 a.m. ET and noted that this was the first hybrid Governing Board meeting; all ADC members were present in person but there were many audience members attending via zoom. Boyd welcomed Sharyn Rosenberg, Assistant Director for Assessment Development, to her new role supporting the ADC. Boyd asked each ADC member to share a recent highlight from their life.

ACTION: 2026 NAEP Reading Framework

Vice Chair Mark Miller noted that the Governing Board has the responsibility of determining what should be tested on NAEP, and the ADC leads and oversees NAEP framework development. Board action on the NAEP Reading Framework at this meeting allows NCES to implement the new assessment for the 2026 NAEP administration.

Miller described recent events that took place since the May Governing Board meeting, including the small group of Board members convened by Chair Haley Barbour and Vice Chair Alice Peisch to serve as the "Chair's Working Group" with a goal of making additional edits to the framework to achieve greater consensus. This was a cross-committee effort and ADC was represented by Patrick Kelly and Reginald McGregor. The full ADC also reviewed and provided feedback on the Chair's draft of the framework before it was finalized.

Miller asked whether there were any questions or comments; upon hearing none, Miller requested a motion from an ADC member to reflect the Committee's recommendation that the 2026 NAEP Reading Framework be approved and adopted by the Governing Board. The motion was made by McGregor and seconded by Nardi Routten; it was unanimously approved. Miller noted that this was an important milestone; he thanked everyone who contributed to this effort, including ADC members; Panel Chair David Pearson; the Visioning and Development Panels and the Technical Advisory Committee; WestEd staff; Governing Board staff; and NCES staff and contractors.

Upcoming ADC Activities and Priorities

Boyd stated that this was an opportunity to briefly discuss what is on the horizon for the ADC over the next year now that the NAEP Reading Framework is nearing completion. She encouraged ADC members to ask questions and provide feedback.

Rosenberg described the following upcoming activities: reviewing the NAEP Reading Assessment and Item Specifications; reviewing cognitive items and contextual variables; reviewing and revising framework processes; creating a framework development procedures manual; launching the Science Framework updates; implementing the Strategic Vision; and reviewing and revising the Board policy on item development and review.

Given the large number of upcoming activities, Frank Edelblut identified a need to establish priorities. Miller suggested prioritizing an additional activity to create abridged versions of the 2026 NAEP Mathematics and Reading Frameworks for use in dissemination. Christine Cunningham suggested asking researchers for feedback on what additional contextual variables would be most useful in secondary analyses of NAEP data.

Boyd adjourned the meeting at 9:35 a.m. ET.

I certify the accuracy of these minutes.

Dana K Boyd

Dana Boyd, Chair

September 23, 2021

Date

National Assessment Governing Board

Committee on Standards, Design, and Methodology

Report of August 3, 2021

Closed Session 1:00 – 1:40 p.m.

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

Governing Board Staff: Executive Director Lesley Muldoon, Deputy Executive Director Lisa Stooksberry, Sharyn Rosenberg, and Angela Scott.

NCES Staff: Tammie Adams, Jing Chen, Brian Cramer, Enis Dogan, Veda Edwards, Pat Etienne, Eunice Greer, Daniel McGrath, Nadia McLaughlin, Taslima Rahman, Holly Spurlock, Bill Tirre, and Ebony Walton.

Other Attendees: American Institutes for Research: Brittany Boyd, Markus Broer, Kim Gattis, Cadelle Hemphill, Saki Ikoma, Young Yee Kim, Sami Kitmitto, Ting Zhang, and James Zheng. CRP: Shamai Carter, Subin Hona, and Anthony Velez. Educational Testing Service: Jay Campbell, Amy Dresher, Robert Finnegan, Helena Jia, Hilary Persky, Luis Saldivia, Karen Wixson, and Meng Wu. Hager Sharp: David Hoff. Pearson: Scott Becker and Pat Stearns. Optimal Solutions: Imer Arnautovic. The Hatcher Group: Jenny Beard, Alex Sanfuentes, Jenna Tomasello.

Item Difficulty and Student Ability Distributions (Closed)

Under the provisions of exemption 9(B) of 552b of Title 5 U.S.C., COSDAM met in closed session from 1:00 p.m. to 1:40 p.m. to receive a briefing and discuss information related to secure NAEP item pools.

Chair Gregory Cizek called the meeting to order at 1:02 pm ET, noting the Committee would begin in closed session and then transition to open session. Cizek indicated the first agenda item would focus on concerns around measuring and reporting at the lower end of the NAEP scale. Cizek introduced Enis Dogan from the National Center for Education Statistics (NCES).

Dogan opened with the general observation that there are large groups of students performing at the lower end of the NAEP scale. However, because fewer test items exist in this low-performing range and because measurement error is larger, it is challenging to measure and report what students know and can do at this level. For instance, in 2019 NAEP Reading at Grade 4 the proportion of students performing below *NAEP Basic* reaches as high as 34 percent. Across the four mandated national assessments in some subgroups and urban districts those percentages are over 50 percent and, in a few cases, reach as high as 70 percent. Dogan described the difficulty level of the item pool as largely a function of NAEP frameworks, which reflect the rigor and cognitive complexity associated with particular objectives.

Dogan then turned to four examples across Grades 4 and 8 Reading and Mathematics to demonstrate the alignment between student score distributions and the difficulty level of test items. Although there was variation by grade and subject, in general the item-person maps showed that there were more students than items at the lower end of the scale and more items than students at the upper end of the scale.

Dogan described a special effort that NCES had undertaken to increase the number of NAEP Mathematics items targeting the lower end of the scale; these items are known as “KaSA items,” or Knowledge and Skills Appropriate items. This effort involved providing additional clarifications to the Mathematics Assessment and Item Specifications to describe how some of the framework objectives could be further constrained to produce easier items. Dogan ended by describing continued efforts to create items in the lower score range, and additional possibilities provided by adaptive testing.

Suzanne Lane asked about the number of multiple-choice items versus lower-level constructed response items, wondering about the extent to which constructed response items allow students to engage at the lowest levels. Dogan promised to follow up with a response after the meeting.

A discussion ensued regarding how items are developed and approved for the assessment. Committee members noted that the frameworks per se do not necessarily constrain the difficulty of items to a narrow range and expressed interest in better understanding and exploring additional efforts for producing more items targeted at the lower end of the scale.

Cizek thanked the Committee for taking the first step in better understanding what students in the lower-performing range might need from NAEP. He identified a need for further discussion on this topic, in conjunction with the Assessment Development Committee and NCES.

Cizek concluded the closed session at 1:40 p.m. and the Committee recessed for five minutes to transition to open session.

Open Session 1:45 – 3:00 p.m.

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

Governing Board Staff: Executive Director Lesley Muldoon, Deputy Executive Director Lisa Stooksberry, Sharyn Rosenberg, Angela Scott, and Matthew Stern.

NCES Staff: Tammie Adams, Jing Chen, Brian Cramer, Alison Deigan, Enis Dogan, Veda Edwards, Eunice Greer, Daniel McGrath, Nadia McLaughlin, Taslima Rahman, Holly Spurlock, Bill Tirre, and Ebony Walton.

Other Attendees: American Institutes for Research: Brittany Boyd, Markus Broer, Mary Ann Fox, Kim Gattis, Cadelle Hemphill, Saki Ikoma, Young Yee Kim, Sami Kitmitto, and Xiaying Zheng. CRP: Subin Hona, Anthony Velez, and Edward Wofford. Educational Testing Service: Jay Campbell, Amy Dresher, Kadriye Ercikan, Helena Jia, and Karen Wixson. Hager Sharp: David Hoff and Joanne Lim. Pearson: Joy Heitland and Eric Moyer. Optimal Solutions: Imer Arnautovic. The Hatcher Group: Jenny Beard, Alex Sanfuentes, Devin Simpson, Nandini Singh, and Jenna Tomasello. WestEd: Sonya Powers. Westat: Chris Averett, Greg Binzer, Lauren Byrne, Rick Rogers, and Leslie Wallace. Other: Karla Egan (EdMetric), Beth LaDuca (Oregon Department of Education), Andrew Kolstad (P20 Strategies LLC), and Jill Hendrickson Lohmeier (University of Massachusetts, Lowell).

Improving Information about Students Scoring Below the NAEP Basic Achievement Level

Chair Cizek opened the session at 1:45 p.m. ET. Cizek offered additional thanks to Enis Dogan of NCES for the presentation in closed session and noted the Committee will continue its discussion of below *NAEP Basic* performance with three background presentations to be followed by discussion. Cizek introduced Karla Egan of EdMetric.

Egan opened by characterizing the landscape of lowest-performing achievement levels, indicating there are a lot of opinions but little literature on the subject. Egan noted that of 46 states that have lowest-level achievement descriptors 43 of those could be located by searching publicly available information and following up with state departments of Education. TIMSS and PISA also use descriptors for lowest levels of performance. Egan noted that NAEP may not have sufficient items at the lowest end of the score range, which can result in a lack of measurement precision alongside the large population of students that fall below *NAEP Basic*. Egan raised the concern that of the 31 percent of students below *NAEP Basic* in 2019, high percentages are students of color and those who participate in the federal free or reduced-price lunch program.

Cizek asked about the qualitative differences in descriptors for the lowest category among the states that use them. For example, it seems the lowest level descriptors often shift from what students can do to what they cannot do or what they can do minimally. Egan observed that many states parse the language more in the lowest level but did not locate an extensive use of what students cannot do. Cizek noted from the earlier presentation, when we talk about a student at Basic, Proficient, or Advanced, we are saying they have a 67 percent probability after guessing that they can do these sorts of things, whereas at the lowest level we are saying some students may be able to do this.

Suzanne Lane expressed surprise at the detail in Grade 8 descriptors in Minnesota and Virginia, asking if Egan obtained any background on how those states wrote their descriptors and how their items banks provide information for the lowest levels. Egan does not know the strategies those or other states employed in writing their descriptors, reiterating the difficulty generally of locating states' item maps. Lane suggested it might be useful to follow up with Minnesota and Virginia to seek additional information about how they are able to provide this level of detail in reporting the lowest category of performance.

As Cizek thanked Egan and prepared to transition to the next presenter, he noted that Egan referenced Computer Adaptive Testing (CAT), which also came up in the closed session. Cizek mentioned that there is no guarantee the lower-performing students would receive lowest-level items unless NAEP moves to a different test administration paradigm–one that would preferentially administer items targeted to students' ability levels. Cizek suggested that CAT may be part of a solution that also includes more item development at the lower end of the range. Julia Rafal-Baer sought clarification about the purpose of adding more items at the lower end of the scale. Is the purpose of adding these items to motivate and encourage students who might otherwise give up because the items were too difficult from the start? Moreover, Rafal-Baer raised a concern that if items are added at the lower end of the scale but not the higher, is there a risk of overcorrecting? Cizek explained that the reason for including more items for lower performing students is to get a more accurate measurement of their level of performance; building students' confidence by introducing more difficulty is unlikely to have much of an effect on their performance. Further, moving to adaptive testing will provide as much information about higher performing students as lower performing students.

Next, Taslima Rahman presented results from the NCES-hosted Below NAEP Basic Panel Meeting in December 2020. The purpose of the panel was to share data and seek recommendations from experts about what NAEP can do to help the public and policymakers understand performance below the *NAEP Basic* achievement level. The panel expressed concern at the large proportion of students below the *NAEP Basic* level, particularly among some subgroups and districts. Rahman noted the increase in the population of students performing below *NAEP Basic* between 2013 and 2019. The panel gave four recommendations: to create a label and description for below *NAEP Basic*, increase the number of items at this level, gather additional data about how students below the *NAEP Basic* level approach items compared to students performing at other achievement levels, and increase reporting on the lowest performing students.

Hanushek asked why NAEP insists on having only three achievement level categories, noting that PISA has six categories and subdivides the lowest category into three parts. Cizek responded from the policy perspective rather than a measurement one, saying there is a lot of value in understanding what students below *NAEP Basic* know and can do. There are many ways to do that without creating new labels and categories including changing performance by improving policies. The Governing Board has a voice and a responsibility to report on the nation's educational progress, including pointing out that more needs to be done and illustrating what students at the lowest levels know and can do without requiring a new label.

Referencing the earlier presentation, Lane indicated that based on Grade 8 reading it might be possible to provide a descriptor based on existing items below *NAEP Basic*. For instance, a future study using an anchor-based method in mathematics would lend itself to looking at the items below *NAEP Basic* for Grades 4 and 8 to determine the extent to which some valuable achievement level descriptors could be obtained. For Grade 4 reading, additional efforts would have to be made to develop more items that are providing information for students below the NAEP Basic achievement level. Lane indicated that it might be useful to hear more about these methods to determine the extent to which achievement level descriptors can be obtained. Cizek asked about the need to bound the lowest level above zero in order to describe confidently what students know or can do. Lane agreed and noted the need to limit the range of students described in the lowest category, for example, by saying that the average student below *NAEP Basic* (which does not include everyone) may be able to do the things noted in the description.

Whitehurst agreed with Lane about the challenges in Grade 4 reading, noting that there are many fourth graders who cannot read single words fluently at a reasonable rate. Whitehurst argued the Board has to address the question of whether NAEP should be measuring those complex pre-requisite skills. Given the usefulness of such information, Whitehurst hopes the Board will take up this topic in the near future.

Cizek then introduced Jing Chen for the final presentation on the topic. Jing described NCES' 2018 study on oral reading fluency, which focused on students performing below *NAEP Basic*. Involving 180 public schools, 1800 grade four students were sampled. The study revealed that students performing at the lowest end of below *NAEP Basic*: read connected text with difficulty at half the words per minute of a fourth grader performing at the *NAEP Proficient* level; misread one out of every six words; focused on individual words, phrases, or clauses instead of meaning; read aloud in a manner than indicated lack of comprehension; recognized with difficulty words they were likely to know when listening or speaking; and showed limited knowledge of spelling-sound correspondence. Whitehurst cited frustration with these results, expressing the belief that with few exceptions a child could have been taught what they were being asked to do. Without the data and these kinds of examples, we could end up with an assessment that is not useful. Whitehurst urged the Committee and the Board to do a better job of assessing those students, inquiring if that might require NAEP to assess pre-reading skills in some cases (not just the ability to comprehend printed text) and noted the importance of getting this at the top of the Board's agenda in the year ahead.

Cizek cited the need for COSDAM to take a leadership role in this endeavor. While appreciating the outcomes from the NCES panel on below *NAEP Basic*, Cizek stated that it should have been the Governing Board that hosted such a panel and expressed hope for doing so in the future. Further, the Board needs to investigate and make policy recommendations around CAT, the pre-requisite skills NAEP might measure, and how to increase item coverage.

Update: Review and Revision of Mathematics and Reading Achievement Level Descriptions

Cizek introduced Eric Moyer of Pearson and described Moyer's update as an important part of the Board's work that focuses on understanding what students at various achievement levels know and can do and helps build public confidence in NAEP's claims about what students know and can do.

Moyer noted the study's goal to look at the NAEP achievement level descriptors based on NAEP framework definitions of what students should know. Using 2019 NAEP items, the aim is to classify students into achievement levels and create statements of what students can demonstrate within each achievement level. Part of the study involves alignment, comparing what the frameworks claim that students are able to do at each achievement level with what students actually demonstrate they can do based on their performance on the assessment.

Since the last report to COSDAM Pearson has taken steps to ensure a representative panel of participants, continued developing and reviewing materials for the meetings, and is reevaluating the possibility of holding meetings virtually due to ongoing health and safety concerns in the United States related to COVID-19.

Lane asked about the purpose of having only one panel per grade level rather than two. Moyer noted that Pearson is creating eight-person groups with replicate groups of four within the panels. Hanushek requested a description of the ideal panelist, to which Moyer noted the aim of identifying current or former educators with at least five years of classroom experience at the grade level and in the subject area. It is highly desirable to secure panelists who have experience in item review or standard setting as well as familiarity with NAEP and its sampling methodology. When Hanushek asked how teachers' effectiveness would be judged, Moyer noted the desire to identify panelists who hold National Board Certification or are leaders in their state or district. Jim Geringer noted the importance of teacher effectiveness and the value add a teacher can bring to students' learning and success.

In follow up from earlier, Lane asked about the rationale for eight panelists at each grade level, suggesting that increasing the size of the panel would make for better representation. Moyer noted the Technical Advisory Committee had spent a lot of time on this topic, having started with the plan of a six-person group that evolved into eight when it was decided that replicate groups would be needed.

Cizek thanked Moyer and referenced the expectation of another update at the next meeting.

Next Steps

In his final meeting as COSDAM Chair, Cizek offered thanks to Board staff, especially Sharyn Rosenberg, who is deeply valued by Cizek and all members of the Committee for her expertise. Cizek then turned to current and future examples where COSDAM and its members can and should play a leading role, from potential changes to board policy on achievement levels to the recent involvement of Whitehurst, Alice Peisch, and Carey Wright in the Reading Framework working group. Cizek acknowledged Peisch's leadership in the working group and the deep expertise in reading that Whitehurst and Wright brought to those discussions. Cizek recognized Rafal-Baer for leadership in calling for public comment to promote consensus on the Reading Framework and for her encouragement of his leadership of COSDAM and its role in the framework. Cizek concluded that COSDAM has been represented well in the reading endeavor and the framework will serve students well for many years to come.

Cizek noted the importance of forthcoming joint discussions with the Assessment Development Committee (ADC) to use lessons learned in reading to improve upon the framework update process. Cizek recommended improvements to the vetting of panelists panels and bringing more diverse perspectives to the table, including increasing the number of teachers who serve on panels. Cizek recognized the value of frameworks standing the test of time but recommended a more timely, incremental approach when revising frameworks in the future. Cizek expressed the need to have the full Board engaged earlier in the revision process and applauded the role that COSDAM can and should play based on the call in the by-laws that the committee oversees NAEP design and methodology. Cizek concluded with his appreciation of such great colleagues, citing the exceptional leadership, contributions, and attitudes among the COSDAM members.

Geringer thanked Cizek for his expertise and leadership, noting Cizek's ability to articulate and summarize complicated issues. Rafal-Baer praised Cizek's welcoming disposition and expectation from day one that each member of the Board and this Committee make their voices heard.

Cizek acknowledged the absence of COSDAM Vice Chair Carey Wright and Committee member and Board Vice Chair Alice Peisch, noting the opportunity to see them in-person or virtually at the full Board meeting later in the week.

Cizek adjourned the meeting at 2:57 p.m. ET.

I certify the accuracy of these minutes.

Gregory Cizek, Chair	09/27/2021 Date

National Assessment Governing Board

Reporting and Dissemination Committee

Report of July 22, 2021

3:00 - 5:00 pm

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

Governing Board Staff: Stephaan Harris, Laura LoGerfo, Lesley Muldoon, Munira Mwalimu, Angela Scott

National Center for Education Statistics Staff: Peggy Carr, Jing Chen, Enis Dogan, Veda Edwards, Patricia Etienne, Eunice Greer, Linda Hamilton, Daniel McGrath, Nadia McLaughlin, Taslima Rahman, Holly Spurlock, William Tirre, Ebony Walton, William Ward, Grady Wilburn

Department of Education: Tammie Adams

Contractors: AIR: George Bohrnstedt, Brittany Boyd, Markus Broer, Mary Ann Fox, Kim Gattis, Martin Hooper, Cadelle Hemphill, Young Kim, Sami Kitmitto, Yan Wang, Darrick Yee; CRP: Shamai Carter, Jasmine Fletcher, Anthony Velez, Edward Wofford; ETS: Jonas Bertling, Gloria Dion, Patricia Donahue, Amy Dresher, Gloria Dion, Robert Finnegan, Kate Faherty, Sami Kitmitto, Courtney Sibley, Karen Wixson; Hager Sharp: James Elias, David Hoff, Joanne Lim; HII-TSD: Michael Slattery. The Hatcher Group: Jenny Beard, Robert Johnston, Zoey Lichtenheld, David Loewenberg, Alex Sanfuentes, Nandini Singh, Jenna Tomasello; Management Strategies: Brandon Dart. P20 Strategies: Andrew Kolstad. Pearson: Scott Becker, Joy Heitland, Eric Moyer, Stanley Rabinowitz, Pat Stearns; Silimeo Group: Debra Silimeo; Westat: Lauren Byrne, Kavemuii Murangi, Jason Nicholas

Other: Rebecca Bennett (Massachusetts Department of Education), Vickie Baker (West Virginia Department of Education), Kathilia Delp, Donna Dubey (New Hampshire Department of Education), Jeremy Ellis (Missouri Department of Education), Jasmine Fletcher-For, Laura Goadrich (Arkansas Department of Education), Lynn Hardy (TBG), Beth LaDuca (Oregon Department of Education), Regina Lewis (Maine Department of Education), Rebecca Logan (Oklahoma Department of Education), Raina Moulian (Alaska Department of Education), Renee Savoie (Connecticut Department of Education)

Vice Chair Marty West called the Reporting and Dissemination Committee meeting to order at 3:01 pm on Thursday, July 22, 2021. West welcomed everyone and provided an overview of the agenda and the goals for the meeting.

Release Plan for 2020 NAEP Long-Term Trend

The national results of the 2020 National Assessment of Educational Progress (NAEP) Long-Term Trend (LTT) assessment for 9-year-olds and 13-year-olds will be released to the public in September 2021. The LTT assessment for which data will be reported this September occurred in 2020 and marked the last national assessment before schools were closed due to the COVID-19 crisis.

Typically, this assessment includes data for 17-year-olds, however, school closures to prevent the spread of COVID-19 in 2020 prevented the administration of the assessment to that age cohort. Originally, the Governing Board expected that NCES would administer the LTT to 17-year-olds in spring 2022, to resume assessing the 17-year-old cohort who could not participate when schools closed.

However, at the August 2021 quarterly meeting of the National Assessment Governing Board, the Board amended the assessment schedule to administer the 9-year-old LTT assessment in the 2021-2022 school year instead. By assessing 9-year-olds immediately prior to school closures in 2020 and again this upcoming school year when the vast majority of schools will reopen with full-time schedules, NAEP will capture student performance at two timepoints at the narrowest temporal boundaries of the COVID-19 potential impacts.

Laura LoGerfo, assistant director for reporting and analysis, described the release plan for the LTT results. The Board will introduce the LTT report and create one or two videos sharing and explaining the data. The video(s) will use graphics and simple animation to help introduce and explain the data. Excerpts from interviews with Lesley Muldoon, the Governing Board's Executive Director, and Dr. Peggy Carr (acting Commissioner of the National Center for Education Statistics) will provide context and highlight key findings. Committee members supported this plan.

In addition, Tyler Cramer urged his fellow committee members to read more about the NAEP Long-Term Trend program to learn its past and future. He promised to distribute various PowerPoint presentations and papers produced when the Board was deciding the fate of the Long-Term Trend assessment several years ago. Cramer asked about the intended audience for the LTT release and encouraged Board staff to consider who in Congress supports the LTT so strongly as to allocate funds for its administration. Marty West replied that the Board could host a special briefing for those LTT advocates in Congress.

Paul Gasparini suggested that the Board host a follow-up event to the LTT release featuring researchers who delve deep into LTT data, spotlighting the most ardent LTT fans and offering a preview of how the 2022 LTT administration to nine-year-olds will inform the nation about the impact of the pandemic.

Mark White posed an existential question, wondering what goals the Board has for this release. In response, Marty West connected these release events to the Strategic Vision which promotes the spotlighting of NAEP's value and utility to stakeholders and broader audiences. In general, release events encourage analysts to delve more deeply into NAEP resources and show the potential impact of NAEP data on informing education. In this specific case, West likened NAEP LTT to studying the climate (long-term), not weather (short-term), because LTT's periodicity is less frequent and its historical timeline longer than main NAEP's.

Bev Perdue summarized her disappointment in the fifty years of LTT results which show little improvement in performance despite billions of federal, state, and local funds allotted to schools. Despite all these investments, the nation ostensibly remains impotent in helping students learn what they should know. One caveat to this inference is the immense demographic shift in the population who took the LTT assessment in the 1970s and those who participated in 2020. The LTT sample now represents more minority students, more English learners, and more economically disadvantaged students, who often score lower on assessments. Thus a lack of apparent change overall belies relatively strong performances by subgroups.

Release Plan for High School Transcript Study

With agreement on the approach to releasing the LTT results, attention turned to the proposal for the release of findings from the NAEP High School Transcript Study (HSTS). In 2019, a nationally representative sample of grade 12 students took the NAEP Reading and Mathematics assessments in a nationally representative sample of America's high schools. In that year, the NAEP team requested that the sampled high schools provide transcripts for sampled students with complete transcripts, i.e., high school graduates.

The High School Transcript Study collects and reports data on the high school graduates' course-taking patterns and rigor, credit accumulation, and grade point averages. The transcript data include demographic information on sampled graduates and can be linked to NAEP scores from 2019. The release will focus on sharing results, stimulating conversation around high school coursework, and expanding the audience for these data.

This release will occur at approximately the same time when the Nation's Report Card is released biennially to cement the idea of NAEP Day in the last week of October. The release will combine a town hall approach with the feel of a moderated news talk show, e.g., C-SPAN, with an in-person component for speakers and a livestream for virtual attendees. The Board will tap its social media channels to crowdsource questions NAEP stakeholders have about high school graduates' schoolwork. Questions may cover high school course-taking trends, equitable access to rigorous courses, and concerns about academic preparations for postsecondary life.

The questions will be posed in a one-hour facilitated conversation that would 1) summarize HSTS results generally and 2) respond to specific questions from the field. The approach will be interactive and not static. A dynamic facilitator will foster a robust conversation based on the selected questions and provide an opportunity for the NCES Acting Commissioner Peggy Carr to share highlights from the data. Shining a spotlight on a few themes emerging from the complex data may help the audience grasp important findings more easily.

A Governing Board member or two will introduce the event; secondary school principal representative Paul Gasparini has graciously agreed to participate in the release. The Board may consider inviting a few questioners to submit their queries via video. In support of the release, the Board will produce and promote a video involving clips from interviews of high school seniors about their course-taking choices, to build interest in HSTS findings and connect the data to real life, not causally, but topically.

Marty West strongly supported this plan but cautioned that the release event and any promotional materials should emphasize that transcripts come from high school graduates only, not from seniors who did not graduate. The committee members appreciated the release of these data as an opportunity to reconsider what high school graduation means and how this meaning has evolved over time. Ron Reynolds commended the plan for making the transcript data more accessible and humanizing it through the video of high school students.

Tyler Cramer moved to approve both release plans for action by the full Governing Board at the upcoming August quarterly meeting, which Mark White seconded. The committee approved the plans unanimously.

<u>Review of Core Contextual Variables</u>

A primary responsibility of the Reporting and Dissemination Committee is to review and approve the core contextual variables on the NAEP student, teacher, and school administrator questionnaires. At this meeting, the committee reviewed items related to education during the COVID-19 pandemic added both to main NAEP and to LTT NAEP questionnaires. Holly Spurlock of the National Center for Education Statistics presented useful background information for the committee review. In response to a question from Tyler Cramer, Spurlock explained that contextual items emerge from research, prior surveys, R&D recommendations and reviews, suggestions from experts on the NAEP program's Questionnaire Standing Committee, and feedback from both the government (e.g., Office of Management and Budget) and stakeholders.

Tonya Matthews asked if different modifiers of the word computer, i.e., desktop, laptop, tablet, may produce inconsistent responses. She wondered if modifications could help clarify what information each of these technology items seek or if these items appear sufficiently general to cover most instances of technology. Spurlock responded that in pilot tests, the items seemed general enough and sufficiently inclusive to facilitate interpretation and reporting. The word computer did not evoke thoughts about tablets, and students do not equate the various hardware.

Tyler Cramer wondered why some of the items present responses in negative to positive order (e.g., never to always left to right) but other items are presented as positive to negative, such as yes before no. Jonas Bertling, survey lead for the NAEP contractor, ETS, shared that survey researchers construct responses in terms of yes/no, then offer different nuances of no (i.e., never, not once, not all the time), so the yes must appear first. However, items that range in frequency, such as never to always or 0 for not very likely to 5 for very likely, the reply options show the intuitive left-to-right increase as participants may expect to see in a number line. Outliers to this behavior are long-standing items, preserved in the exact same way over decades for trend analyses.

Next, committee members cautioned that if schools closed again in the fall due to the next wave of COVID-19, the items would require additional revision. Tyler Cramer noted that item #7 on page 16 of the review package should be clarified to determine if that does or does not include the teacher. Cramer also wanted to know why there are no LTT questions for school administrators to report on the percentage of students who are new to the school.

Ron Reynolds conveyed disappointment in the dropping of the teacher sex variable, which means that the NAEP surveys cannot detect any gender or sex discrimination. Currently, the teacher sex variable includes only binary responses (male or female), but federal statistical agencies are examining how future surveys address questions of sex and gender. New instructions for surveys may indicate that additional response options must be included, which NCES awaits. Elaborating on Reynolds' query, Matthews asked if the approval process for items such as teacher sex, which is optional, compels as rigorous a review process as the rest of the items. Spurlock noted that all questionnaire items are optional, thus the same review process applies. Considering some states' reactions to particular questionnaire items, ensuring that the items reflect purpose and federal policy becomes especially critical.

Paul Gasparini enthusiastically endorsed the items about grading policies and practices. He also noted that items about instruction mode during the pandemic (e.g., hybrid, in-person, distance) omit several important options. He expressed concern that the survey will miss vitally informative data. Bertling replied that the survey team developed these items last year, when the presented options captured the most frequent modes of instruction. Last year's development process left no paths to revise the survey. Thus, the NAEP team cannot add new sub-items. Bertling acknowledged this substantial limitation, but hopes the options may be meaningful, pointing out the option for respondents to check 'not applicable (N/A)' and noting the connection between these items and those included on the special COVID-19 school study. Gasparini recommended that the Board flag these items as needing supplemental guidance for interpretation, with which Matthews agreed and added that NAEP should find items where N/A is selected often, implying that these questions deserve revision.

LoGerfo thanked everyone for their diligent and thoughtful reviews of the items and encouraged committee members to send additional comments by July 25th at midnight. She will compile and send all feedback to the NAEP team by their deadline on July 26th.

State Mapping Study Briefing

Marty West introduced the session on the State Mapping Study by proclaiming this report's remarkable impact on policy, showing the "honesty gap" between what states purport is their educational standards' rigor and what NAEP shows is their rigor. Taslima Rahman, the lead author for this report at the National Center for Education Statistics, thanked West and proceeded to provide an in-depth, detailed, comprehensive look at the report's results.

The entire report merits reading, however headlines shared by Rahman include news that state math standards may be interpreted as more rigorous than their reading standards, i.e., more state standards mapped at the *NAEP Proficient* level in mathematics than in reading. Across both grades and both subjects, most state standards aligned at the *NAEP Basic* achievement level. And, compared to the previous decade, more state standards mapped at the *NAEP Proficient* level in 2019 than in 2009.

At the conclusion of Rahman's presentation, Matthews noted that this report always elicits universal acclaim at its release. Marty West echoed Matthews' praise, commending the study for its profound impact on education policy. The report allows for comparisons across states, which in earlier iterations, showed enormous variation in what states call grade-level work. West underscored a finding from the report that this variation has narrowed over time, calling attention to Tennessee's evolution. Mark White, one of the two state legislators on the Governing Board and a representative from Tennessee, remarked how findings from previous editions of the state mapping study motivated Tennessee to revise their state standards, which has resulted in improved NAEP scores.

General Updates

In the remaining minutes of the committee meeting, Marty West reminded committee members of their prior deliberations on improving the measure of socioeconomic status (SES) on NAEP. The NAEP team is making progress on providing income estimates that can be added to state administrative data systems which could in theory be reported back to NAEP. This progress is exciting, but only happening in states with active Statewide Longitudinal Data System (SLDS) grants and the timeline for the project's fruition is unclear. Realistically, a comprehensive solution to the issue will take a substantial amount of time, and even then, income represents only one component of socioeconomic status.

Thus the committee should encourage and monitor this work but also contemplate potential next steps or options:

1. Continue to gather relevant information in contextual questionnaires that researchers can use to construct SES proxies;
2. Revise contextual questionnaires to improve the quality of relevant information; and
3. Develop a recommended index of SES to be included in NAEP.

There was no time to discuss these possible options, but the conversation shall continue at the next committee meeting in November.

In conclusion, Perdue expressed kudos for a rich and productive meeting, and Tonya Matthews adjourned the meeting at 5:01 pm.

I certify the accuracy of these minutes.

____________________________ 10/27/2021

Tonya Matthews, Chair Date

National Assessment Governing Board

Nominations Committee

Closed Session

July 28, 2021

Nominations Committee Members: Governor Jim Geringer (Chair), Dana Boyd, Tyler Cramer, Tonya Matthews, Mark Miller, Reginald McGregor.

Board Member: Suzanne Lane

Members Absent:
Alice Peisch and Paul Gasparini

Board Staff: Stephaan Harris, Donnetta Kennedy, Lesley Muldoon, Munira Mwalimu, Tessa Regis, and Lisa Stooksberry.

Under the provisions of exemptions 2 and 6 of § 552b (c) of Title 5 U.S.C., the Nominations Committee met in a closed session on Tuesday, July 28, 2021 from 5:30 p.m. to 6:30 p.m. Eastern time to discuss the following agenda topics:

- An update on the nominations for board terms that begin on October 1, 2021
- Board vacancies for terms beginning October 1, 2022
- 2022 campaign plans and a proposed timeline
- Next steps in work plans

Governor Geringer called the meeting to order at 5:30 p.m. ET. After welcoming members, Geringer previewed the agenda topics for discussion.

Nominations for Board Terms Beginning October 1, 2021

Lisa Stooksberry, Deputy Executive Director, updated the committee on the status of 2021 appointments and recent communications with the Secretary's office. It is anticipated that appointments are on schedule for terms beginning October 1, 2021.

Board Vacancies for Terms Beginning October 1, 2022

Governor Geringer reported that the following four vacancies will need to be filled for terms that would begin on October 1, 2022:

1. Eighth-Grade Teacher
2. Fourth-Grade Teacher
3. General Public Representative – Parent Leader
4. Secondary School Principal

Outreach Strategy for 2022 Nominations Campaign

Stephan Harris, Assistant Director of Communications, briefed the committee on outreach strategies for the 2022 nominations campaign. Harris noted the dual purposes of the outreach campaign—to promote Board vacancies and building partnerships with targeted groups. The outreach efforts reflect priorities of Strategic Vision 2025.

The 2022 campaign will be launched via a website splash page this summer. A tool kit will be developed, and a webinar will be convened to attract candidates for all open categories with a focus on the General Public Representative–Parent Leader category.

Members discussed the need to clarify the General Public Representative slot as shown on the Board membership chart to clarify the two generalists and two parent leaders.

2022 Nominations Timeline

Tessa Regis briefed the committee on the 2022 nominations campaign timeline. She reported that the campaign will be launched on September 8, 2021 and will conclude with action on the slate of finalists at the March 2022 Quarterly Board meeting. Members discussed and concurred with the proposed timeline.

Looking ahead, Tonya Matthews noted that a large number of vacancies would occur in the 2023 cycle. Matthews suggested initiating recruitment strategies earlier than usual next year to allow adequate time to solicit nominations. She noted that the workload for the Nominations Committee would be very heavy and should take into account the needs of new members to familiarize themselves with the work plans. She suggested that a possible need for an additional Board member on the Nominations Committee to support rating work for the 2023 cycle.

Farewell Remarks

Matthews noted that this meeting of the Nominations Committee would be Chair Geiringer's last meeting as his term of office would conclude on September 30, 2021. Matthews read a poem she wrote commending Geringer for his contributions to the Board's work and thanking him for hosting a Board meeting in Wyoming. She then called on each member to provide remarks, following which Dana Boyd, Tyler Cramer, Mark Miller, and Reginald McGregor thanked Geringer for his leadership, insights, collegiality, and mentoring.

Geringer provided remarks on his service and thanked members for their contributions to the Board's work. He then turned to next steps in the work of the Nominations Committee and provided closing remarks. Geringer thanked Suzanne Lane for attending the meeting and noted that she would bring much-needed testing and measurement expertise to the Nominations Committee in the future.

The meeting adjourned at 6:30 p.m.

I certify the accuracy of these minutes.

Jim Geringer

Jim Geringer, Chair

October 14, 2021

Date

Executive Committee

November 18, 2021
9:45 am – 11:30 am ET
Grand Ballroom

AGENDA

9:45 – 9:50 am	**Agenda Overview and Opening Remarks** *Beverly Perdue, Chair*
9:50 – 10:15 am	**Vision for the Next Year** *Beverly Perdue* *Lesley Muldoon, Executive Director*
10:20 – 11:20 am	**Budget and Assessment Schedule (CLOSED)** *Peggy Carr, Commissioner, National Center for Education Statistics* *Lesley Muldoon*
11:20 – 11:25 am	**Transition to Open Session**
11:25 – 11:30 am	**ACTION: NAEP Assessment Schedule** *Beverly Perdue*
11:30 am	**Adjourn** *Beverly Perdue*

National Assessment of Educational Progress
Schedule of Assessments

Approved August 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 9-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 2025. 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.

Assessment Development Committee

November 10, 2021

3:15 – 4:30 pm ET

Zoom:

https://us02web.zoom.us/webinar/register/WN_CeZuwAwcQJmTqmecgSHCqg

AGENDA

3:15 – 3:20 pm	**Welcome and Review of Agenda** *Dana Boyd, Chair* *Mark Miller, Vice Chair*	
3:20 – 3:40 pm	**ACTION: 2026 NAEP Reading Assessment and Item Specifications** *Dana Boyd* *Mark Miller*	See plenary tab
3:40 – 4:30 pm	**Discussion of Initial Public Comment on Current NAEP Science Framework** *Dana Boyd* *Christine Cunningham*	Attachment A
Information Item	**Item Review Schedule**	Attachment B

Discussion of Initial Public Comment on Current NAEP Science Framework

The NAEP Assessment Schedule indicates that the Board will consider whether updates to the NAEP Science Framework are needed for the administration of the 2028 assessment and beyond.

Current NAEP Science Framework

The current framework was adopted by the Board in 2005 and implemented beginning with the 2009 NAEP science assessment at grades 4, 8, and 12. The framework includes two dimensions: content and practices.

The science content for NAEP is defined by a series of statements that describe key facts, concepts, principles, laws, and theories in three broad areas:

- Physical Science
- Life Science
- Earth and Space Sciences

Physical Science deals with matter, energy, and motion; Life Science deals with structures and functions of living systems and changes in living systems; and Earth and Space Sciences deal with Earth in space and time, Earth structures, and Earth systems.

The second dimension of the framework is defined by four science practices:

- Identifying Science Principles
- Using Science Principles
- Using Scientific Inquiry
- Using Technological Design

These practices can be combined with any science content statement to generate student performance expectations, and assessment items can then be developed based on these performance expectations.

The framework specifies that 50 percent of the assessment time should be devoted to multiple choice items and the remaining 50 percent should be constructed response items. For each grade level, the constructed response items are intended to include at least one hands-on performance task and at least one interactive computer task.

Trends in State Science Standards

The Board's Framework Development policy calls for using information about trends in state standards as one resource in the decision-making process of whether and how a framework should be updated. In 2016, the American Institutes for Research (under contract to the National Center for Education Statistics) conducted a comparison study of the Next Generation Science Standards (NGSS) and the NAEP Science, Technology and Engineering Literacy (TEL), and Mathematics frameworks. The degree of overlap between the NGSS and NAEP varied across grades and depending on whether the NGSS were compared to the NAEP Science Framework only or whether the TEL and/or Mathematics frameworks were also included. The summary and conclusions are detailed on PDF pages 103-108 of the technical report.

Earlier this year, Board staff commissioned an additional study under a previous contract with the Human Resources Research Organization (HumRRO) to better understand how the NAEP Science Framework overlaps with state standards for the states that did not fully adopt the NGSS – including states that partially adopted the NGSS and states that did not adopt the NGSS. As with the study of NAEP and NGSS, there was some overlap and some important differences between NAEP and state science standards, with variation across grades and content areas. The discussion and conclusions appear on PDF pages 35-36 of the report.

Public Comment

Under the leadership of the Assessment Development Committee (ADC), the Board has been discussing how to strengthen existing processes and procedures for updating NAEP frameworks. One proposed improvement is to conduct a public comment period on the current assessment framework to seek broad input upfront on whether and how the current framework should be updated. Consequently, the Board conducted an initial public comment on the current NAEP Science Framework from August 20 – October 15, 2021. Commenters were asked to address three questions:

- Whether the NAEP Science Assessment Framework needs to be updated
- If the framework needs to be updated, why a revision is needed
- What a revision to the framework should include

The purpose of seeking public comment on the current framework is to surface a broad range of views related to a given subject at the outset of the framework development process. This initial comment then can inform initial Board direction and the selection of panelists to represent diverse perspectives on the issues that are of most importance to the Board.

Thirty submissions were received from a variety of individuals, groups of individuals, and organizations. In addition, Board staff sought input from the National Center for Education Statistics (NCES) on operational issues and challenges associated with the current framework and assessment; a memo was submitted by NCES to summarize their feedback. The raw comments are attached, along with a summary of specific points raised by major theme.

Discussion

The purpose of the November plenary discussion is:

- To identify what information is needed for the Board to make a determination of whether and how the NAEP Science Framework should be updated;
- To identify the key issues/topics for which the Board may want to provide policy guidance to the framework panels; and

- To identify what additional input and expertise (e.g., commissioned white papers, expert panels) is needed to inform the policy guidance to be set forth in a Board charge to the framework panels

During the November 10th ADC meeting, ADC members will discuss initial recommendations to the full Board on the points listed above.

ADC Chair Dana Boyd and ADC member Christine Cunningham will facilitate the ADC and plenary discussions on this topic. Following the November Board meeting, Board staff will commission targeted expert input on the key issues identified to inform future Board decisions during spring 2022 on whether and how to update the NAEP Science Framework.

Summary of Public Comments Received on the Current NAEP Science Framework[1]

November 4, 2021

Contributors

Spurlock, Holly	National Center for Education Statistics
Pellegrino, James	University of Illinois Chicago (NAEP Validity Studies Panel white paper)
Petersen, Anne	Virginia Department of Education
Moulding, Brett	Retired Utah State Office of Education Curriculum Director and Instruction Former NAEP Science Advisory Committee Member
Sneider, Cary	Former NAGB Member
Gordon, David	CAST (originally Center for Applied Special Technology)
Finn Jr., Chester E.	Thomas B. Fordham Institute
Murphy, Stephen	Cognia
Heinz, Michael	Council of State Science Supervisors
Murphy, Danielle	*Affiliation not provided*
Reid, Ann	National Center for Science Education
Foster, Jacob	STEM learning Design, LLC
Huntoon, Jacqueline	Michigan Technological University
Barber-Lester, Kelly	University of North Carolina Pembroke
Wray, Kraig	Pennsylvania State University
Looy, Mark	Answers in Genesis
Lowry, Michael	The McCallie School
Wysession, Michael	NSF's Earth Science Literacy Initiative Earth and Space Science for the NRC's Framework for K-12 Science Education Earth and Space Science for the Next Generation Science Standards Washington University St. Louis
McCarthy, Michelle	Montana Office of Public Instruction
Multiple Authors	Georgia State University
Haverly, Christa Marie Marshall, Stephanie Kayumova, Shakhnoza Cheuk, Tina Basile, Vincent McDonald, Scott Taylor, Jonte' C.	Northwestern University University of Minnesota Twin Cities University of Massachusetts Dartmouth California Polytechnic State University San Luis Opispo Colorado State University Pennsylvania State University Pennsylvania State University
	National Science Teaching Association—Statement endorsed by the Council of State Science Supervisors and the National Science Education Leadership Association

[1] This summary was produced by Dr. Arthur Thacker of the Human Resources Research Organization under subcontract to the Manhattan Strategies Group as part of contract 919995921F0002, Technical and Logistical Services.

	National Science Education Leadership Association
Settlage, John	University of Connecticut
Schwartz, Renee	National Association for Research in Science Teaching (NARST) Georgia State University
Badrinarayan, Aneesha	State Performance Assessment Learning Community (SPA-LC) coordinated by the Learning Policy Institute
Sterling Burnett	Heartland
Codere, Susan	Multiple Literacies in Project Based Learning
Keller, Tom	STEM Education Strategies, LLC
Thomas Tretter	*Affiliation not provided*
Bryan, Carl	Wisconsin Department of Public Instruction

Overall Summary

Twenty eight of the 31 submitted comments recommended some level of revision for the NAEP Science Framework. Most of those comments focused on bringing the framework into alignment with state standards (including but not limited to the Next Generation Science Standards (NGSS)) and improving equity and fairness for all tested students. There were also several comments regarding assessment design and accessibility for all students. Suggested revisions ranged from minor editorial comments to significant overhaul of the framework. *(Note that not all submitters responded directly to the question of "Whether the 2019 NAEP Science Framework needs to be updated." The count is based on the content of the submissions and whether the submitters recommended changes to the current framework.)*

Alignment to NGSS/National Academies Framework (three dimensional standards)

Fifteen of the 31 submitted comments focused, either fully or in part, on updating the NAEP Science Framework to better align with the National Academies Framework and NGSS. Most comments centered around current changes in state standards and teaching and learning and concerns that NAEP assessments would not accurately reflect student performance due to a misalignment between what NAEP tests and what is happening in classrooms. Several of these comments suggest including content from the NAEP Technology and Engineering Literacy (TEL) Framework in the science assessment. A couple of comments suggest merging science and TEL, but there are cautions provided in the full text for that suggestion as well. Conversely, there were three comments cautioning the Governing Board not to make substantive changes in the framework (one specifically indicating that the Board's mission is not to follow NGSS). Summary comments follow in bullet form.

Specific comments received:

- The NAEP Science Framework does not approach science as three dimensions, Science and Engineering Practices (SEP), Disciplinary Core Ideas (DCI), and Crosscutting Concepts (CCC). Revisions should include a clear alignment to the National Academies Framework for K-12 Science Education.

- Merging content from the TEL would improve alignment to NGSS. The TEL might be eliminated, and engineering practices (and technology) incorporated into what is considered science. NGSS includes much of the first two TEL components—designs and systems and technology and society. The third, communications technology, is more closely related to English Language Arts (ELA) than science.
- Attend to shifts in grade levels for content learning progressions. This is especially relevant if NAEP adopts a three-dimensional framework, where the interactions among DCI, SEP, and CCC could potentially cross grade levels for a given phenomenon. It is vital that the assessment items measure constructs that are appropriate for the intended grade level.
- Consider changing the assessed science grade from 4 to 5. The NGSS organized elementary standards for grades K-5, middle school standards for grades 6-8. Many states administer their assessments in grade 5. This might make NAEP science results more comparable and relevant for states.
- Tease out research since the science framework was updated. States have largely changed their standards.
- Frameworks must redefine content, practices, and crosscutting concepts to align to the way they are operationalized in the NGSS. Framework practices overlap NGSS practices, but are too broad to focus on specific expectations of current science instruction.
- Crosscutting concepts in the current NAEP Framework are anchored in the content statements themselves. NGSS and more recent literature refer to crosscutting concepts in a more theme-based way, like the NAEP Science Framework did from 1996-2005. The NAEP framework should adopt the seven crosscutting concepts included in the NGSS, or relabel the current crosscutting content if more substantial revisions are not made.
- Two consensus studies of the National Academy of Science include *Taking Science to School: Learning and Teaching Science in Grades K-8* (2007) and a *Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas* (2013). Forty-four states (representing 71% of U.S. students) have science standards influenced by the Framework for K-12 Science Education.
- Assessment can drive instruction forward or backward. Coherence between federal and state assessment will provide state leaders with another tool to improve science instruction for all students.
- The current NAEP Science Framework has two separate components, science content and science practices. Framework for K-12 Science Education also defines distinct practices, core ideas, and crosscutting concepts—the difference is the expectation that they are integrated in instruction and assessment.

- Integration of science practices and content is vital and may require attention to the measure of each construct independently, plus a measure of the integrated abilities of students.
- The current framework is too differentiated by discipline. Interesting problems in science are less and less likely to be confined to one particular discipline.
- Frameworks for NAEP Science and NAEP TEL were developed before the NRC Framework and NGSS. All drew upon bodies of theory, research, and practice regarding the knowing, learning, and teaching of science and technology available at the time of their development. There are significant similarities, and substantial differences between the two NAEP frameworks and the NGSS.
- Alignment differences between NAEP and NGSS are magnified as grades increase from 4 to 8 to 12. NGSS is more interdisciplinary across grade levels, while NAEP shifts toward physical science in grades 8 and 12, especially grade 12.
- NGSS science practices are more demanding than NAEP practices and focus more on “doing science” rather than knowing science.
- NGSS performance expectations are viewed to demand more than NAEP performance expectations in terms of application of disciplinary content. This leads to misalignment even if the science content covered by both frameworks is similar.
- Combining NAEP Science and TEL might improve alignment to state standards, but the two NAEP frameworks are quite different. If content from the TEL is to be included in science, the high variability of overlapping content by grade must be accounted for. Items/tasks would also need to be redesigned as TEL tasks intentionally omit relevant science content. An assessment aligned to NGSS would look substantially different from assessments aligned to either NAEP Science or NAEP TEL.
- Given state science standards adoptions, the current NAEP Science Framework and assessment may be substantially at variance with a relatively pervasive national perspective on what is desired for students to know and be able to do in science at grades 4, 8, and 12 and how they could be expected to show proficiency via large-scale assessment.
- Evidence shows that adoption of the new science standards has been staggered across time since 2013, as has been the design and implementation of state large-scale assessments aligned to those new standards. The latter invariably lag two or more years behind adoption of new state standards. The most recent national survey of science education (2018) suggests that little changed between 2012 and 2018 in science instructional practice. Results from the NAEP science assessment from 2009 to 2019 also show little in the way of change in student performance across time.
- If substantive review of the frameworks is completed to better align with NRC and NGSS, then the meaning of science proficiency should also be considered. The ability to

integrate content and practice knowledge consistent with the separate but related considerations of science and engineering content is key.

- Consider inclusion of technology and engineering content similar to the TEL and whether it would be appropriate to merge the science and TEL frameworks.

- Integrating the NAEP Science and TEL assessments would have benefits in terms of cost savings and alignment, so the Governing Board may wish to consider merging the two frameworks.

- Remove the silos represented by traditional course disciplines in life, physical, and earth science and address the cross-fertilization that is currently happening in STEM (as found in NGSS).

- Emphasize the scientific practices modeling and argumentation. New assessment items should be heavily connected to the modeling process. Argumentation can foster students' abilities to evaluate claims using evidence and consider concepts like confirmation bias and other fallacies.

- Current standards are based on research that originated before 2005. It should be updated to reflect the more current understanding of science education described by the NGSS.

- The NAEP framework is broad but needs to more accurately reflect the depth of learning and application that is now expected of students.

- Given the likely scope of a revision to the NAEP Science Framework and the implications for the 2028 assessment, as well as the possibility of incorporating aspects of TEL in the new framework and assessment, it seems highly likely that preserving the science or TEL trend through 2028 will not be feasible or advisable. Priority should go to insuring the validity of the revised science framework and assessment for 2028 and beyond. Doing so should not be compromised in a possibly misguided effort to preserve trend at all costs.

Equity/Diversity

The second most prevalent comment topic regarding potential framework revisions had to do with ensuring equity among diverse populations of students. Fourteen of the 31 submissions included equity/diversity as a major theme in their comments. The comments ranged from general concerns about the ways that NAEP reports data on student subgroups, to very specific concerns regarding students' opportunities to learn and the representation of the majority group (higher socioeconomic white students) in the content of the test items. Several comments focused on ensuring that the represented science was not taken out of context, but that context be included to make the phenomena and problems more genuine for students.

Specific comments received:

- Lack of physics courses/teachers, especially during year one of high school, and especially for minority and high-poverty student populations, may conflate performance and opportunity to learn first year physics concepts.

- The COVID-19 pandemic shined a spotlight on inequities and unjust public education practices. As an organization that is not constrained by limitations created by statewide policies, the Governing Board should position itself to take up that work and to exemplify how large-scale assessments can provide equitable opportunities for all students to make their thinking visible.

- The following words and phrases are completely absent from the NAEP Science Frameworks—equity, equality, inequality, racism, bias, scientific racism, prejudice, sexism, or ethics. The term race is only used for tracking subgroup performance, and culture is limited to the role of science in influencing cultures. There is no discussion of bias or the mitigation of bias—a well-established and ongoing concern in education.

- The framework presents a vision and version of science as objective, neutral, and divorced from context, despite its unquestionably troubled history (and present) as it pertains to issues of inequity broadly, and specifically racism and sexism.

- Update references and acknowledge advances in understanding of student diversity and cultural relevance.

- Expand the meaning of diversity (beyond students with disabilities and English learners) consistent with more recent NAEP resources (e.g., NAEP TEL Framework).

- Emphasize diversity, equity, accessibility, and inclusion to support learning, increase engagement, and provide visible representation in content with a goal to improve diversity in representation of underrepresented groups in science fields of study and the workplace.

- Make students the focal point of the assessment and include meaningful feedback loops with the community as reflected by the students' contexts and communities.

- Create a practice for understanding diverse learners and connecting them to science activities, including outreach and engagement with family community members. This would inform assessment development, curriculum integration, and solving real problems.

- Adopt a "growth mindset" strategy for revisions that promotes self-efficacy and motivation to learn from mistakes, then expand scientific skills centered on real world/life problem solving and knowledge.

- Connect the performance expectations to students' lived experiences (e.g., relevant phenomena). Equitable and inclusive performance expectations guide the development of assessment items and tasks.

- Develop assessments that reflect the mindsets and habits of professionals in the field and that "this shift from students as consumers of information to practitioners of field knowledge is especially significant for Black, brown and Indigenous students, signaling that they belong to a larger intellectual community" (Safir and Dugan, 2021). The assessments that students encounter should include tasks that elicit authentic student performance to the extent practicable.

- Expand the definition of "assessment of design" to include other considerations beyond scientific principles (e.g., economic, social) to better engage students with more relevant problems based on their lived experiences and social justice.

- Incorporate cross-sectional views of item DIF (e.g., low SES Black females). Real differences may be being washed out by the ways student subgroups are currently defined.

- Include representatives from traditionally underrepresented subgroups in all development processes—from developing the frameworks to developing test blueprints, selecting phenomena for testing, item writing, and development of scoring rubrics/criteria.

- New research outlined in research like How People Learn II: Learners, Contexts, and Cultures (2018) provides further input regarding integration of content and practice for improved and more equitable outcomes. Students do not use their knowledge of content, practice, and cross-cutting concepts in isolation of one another. The knowledge interacts in ways that provide scaffolding for recall, integration and problem solving in the context of a novel or repeat phenomenon(a). As noted by the Achieve Framework for evaluating cognitive complexity, artificially separating these cognitive processes in assessment does not provide us with an accurate or equitable measure of student proficiency in science. It is in our best interest to align our measures with instructional practice.

- The new framework should endeavor to focus on interpretations within communities and populations based on opportunity to learn (OTL) metrics while also maintaining an 'asset' orientation in all interpretations, rather than traditional 'deficit' views that have been associated with large-scale assessments, such as NAEP, and the reporting of outcomes.

- OTL metrics must consider how students are given experiences to connect their science learning experiences through "forms of knowledge and ways of using language from their everyday experiences in families and communities." This means broadening the collection of OTL data from districts, communities, and schools.

- Interrogate the assumptions about science knowledge embedded in the standards (i.e., whose histories and narratives are and are not included in this body of knowledge and practices).

- Update the technical aspects of the assessments themselves to be more inclusive of historically marginalized student populations.

- Invite people to participate in this review process, including on the expert panel, who are multilingual, of color, differently abled, and so on; leverage their expertise and lived experiences; and provide them with authority and agency to make substantive changes to the program.

- NAEP should stop fostering deficit explanations about achievement gaps via NAEP science results. NAEP should proactively develop reporting approaches that redirect media, political, and layperson discussions in ways that disrupt widespread beliefs that demographics dictate destinies. Requires more disaggregation and should point toward discussion toward remedies rather than promote ideas about gap inevitability.

- Support secondary research on equity and diversity in science education by allowing access to data and promoting relevant studies on the intersections of student gender, race, and social class.

- When NAEP does include cases where concepts are embedded in context, the contexts (e.g., hares in state park) feature the lived experiences of the dominant groups in U. S. society (e.g., upper middle class).

Accessibility

In addition to comments about equity and diversity generally, there were several comments specifically about accessibility. These comments were mostly about ensuring access to the NAEP assessments for all students. There is concern that NAEP does not assess students with the most severe cognitive disabilities. There were also comments requesting that accessibility be built into all aspects of NAEP test development, from adoption of frameworks through reporting of results.

Specific comments received:

- Incorporate principles of Universal Design throughout the framework. Adopt an inclusive validity framework that considers construct irrelevant factors that learners bring to testing. Include additional accessibility features for all students (including Els, SWDs, and non-identified students).

- Find a way to include students with the most significant cognitive disabilities (reference on the frameworks and include in testing).

- Young students may have insufficient access to and training in computer use for fair inclusion in digital assessments.

- Communities in digital deserts may have insufficient access to broadband services to support digital assessment.

- A major tenet of fairness, as conceptualized in the testing standards, is that assessment administrators must provide access for all examinees in various populations, particularly in allowing for accommodations and modification for learners with different cognitive, linguistic, and physical abilities (AERA/APA/NCME, 2014).

- Sample NAEP science items are laden with dense language and vocabulary, particularly in context-driven items. More consideration for English learners, beyond the current statements, must be put into practice in the development of NAEP science.

- It would be very useful for NAEP to develop equity indicators with respect to achievement and school and community factors, like those used in international assessments. Intentional attention to equity and social justice within the science curriculum and instruction are essential for developing scientific literacy.

- There are interactions between item difficulty and a student access to demonstrate knowledge of science practices. A large proportion of students score in the "Below Basic" performance category, and the large amount of contextual information may limit their ability to demonstrate what they can do. More items in the lower range of difficulty are needed to assess lower ability students.

Cautions Regarding Wholesale Revisions

While most of the received comments requested revisions to the Science Frameworks, there were a few (3) that promoted maintaining the framework as is. These comments posited that the current frameworks were of high quality and that NAEP functions as it is intended currently. There were concerns about maintaining trend and about tracking subgroup performance. Others commented that changes should be made in moderation to maintain the parts of the frameworks that are functioning well (e.g., the inclusion of sample items, focusing on scientific phenomena).

Specific comments received:

- 2012 comparisons between the NAEP Science Framework and state standards conducted by the Fordham Institute determined that the NAEP framework was of very high quality compared to most state standards. Minor updates may be required, but more substantive changes should only be made if absolutely necessary.

- NAEP should continue to include sample test items and complete explanations regarding what those items measure, how they are scored, and how they fit into the larger measurement construct in any revisions.

- The NGSS are already nine years old. Any revisions to NAEP frameworks should include a current literature review to ensure that a new NAEP framework is not outdated before it comes into use.

- Continue to ground assessment items in science phenomena and engineering design problems. A focus on sense making is what we now aspire to for our students.

- The NAEP Science Framework faces a precarious challenge: standardizing the instrument across time to identify longitudinal patterns, while accommodating changes in science education.
- The stated purpose of NAEP science assessment is to evaluate trends in scientific literacy overall and by demographic group. The current content, practices, and test design accomplish this goal. NAEP's purpose is not to mirror NGSS.

Editorial Updates

Editorial updates were included in many of the submitted comments, including a "marked up" version of the current framework. The bullets in this section are examples, but do not constitute the full range of edits, corrections, and clarifications submitted.

Example comments received:

- Eliminate references to NCLB and update to reflect current legislation (e.g., ESSA).
- Eliminate the term "special needs" and replace with "students with disabilities."

Addressing Controversial Subject Matter

Comments about controversial subject matter were inconsistent. They included: a call for NAEP to lead states in teaching socially, but not scientifically, controversial subjects; a request to omit controversial topics from the framework; and a request to ensure that minority views (e.g., creation science) are allowable in science teaching. Specific comments received:

- Special attention should be given to socially but not scientifically controversial topics. These specifically include evolution, climate change, and vaccination, as well as to the nature of science. It is counterproductive to make allowances for states that have chosen to under-educate or miseducate their students.
- A general framework should avoid discussion of scientifically disputed or politically charged issues such as anthropogenic climate change or embryonic stem cell research. If climate change is included, address the controversy regarding the quality of scientific evidence available to support the widely held conclusions.
- Inclusion of controversial ideas in the teaching of science is both legal and beneficial, particularly criticisms of evolution, the earth's age, and the reliability of dating methods. Teachers should not be required to teach creation science of ideas that support a younger age of the earth, but they should have the academic freedom to teacher alternative ideas—even if they happen to be in the minority.

Assessment Design

This section includes comments made regarding the assessment design. The interactions among framework objectives, tested content, and score reports are reinforced by the comments provided here.

Specific comments received:

- NAEP developers must be extremely transparent and explicit about the interpretations—and non-interpretations—of the assessment results based on the methodology in comparison to each particular state's standards and approach.

- Pay close attention to cognitive complexity—as a revision of the frameworks will require more complex items to effectively address the intended measurement construct.

- Increase emphasis on innovative item types, especially constructed response items and "predict, observe, and explain (POE)" items. Items may need to be clustered to address science concepts.

- Include and expand hands-on performance tasks, as these are fundamental to doing science and necessary to demonstrate the application of science.

- Include and expand the use of interactive computer tasks (ICT).

- Illustrative NAEP questions are too narrow in scope and tend toward acquisition of principles and facts. Broader test items should mirror our expectations for science teaching and learning in classrooms, assessing students broader understanding, integration, and use of scientific knowledge.

- NAEP should lead the way in designing science assessments that go beyond traditional large-scale multiple-choice tests. New approaches to science instruction allow many opportunities for informal assessment as student engage in investigations, create representations, and discuss evidence. Meaningful formal assessments will require careful articulation of the desired learning goals and how students can demonstrate that they have achieved them.

- The revision should include:

    1. Modeling as a practice. Students should be asked to create, evaluate, and/or revise models, and use them to predict the result of changes to system components. The development of explanatory models can help students make their thinking visible and can be an equalizer for English Language Learners.
    2. Planning investigations. Students should be able to identify independent and dependent variables and to design scientifically valid investigations.
    3. Analyzing data. Students should be able to analyze complex, real-world data using graphing and graphing analysis tools.
    4. Engaging in argument from evidence. Students should be assessed on their ability to use evidence to construct and justify a scientific claim.

- Measuring of two dimensions (content and practice) are ambiguous. In many cases, the experiences of the student dictate whether they access learned content knowledge or engage in science practice when interpreting an item's content (familiarity with the content/context dictates how the student approaches the problem). Items must have greater specificity regarding the nature of exactly what they are measuring.

- Hands-on Performance Tasks (HOTs) may need to be changed to hybrid models and included as interactive computer tasks due to practical and logistical considerations. Further research is required to determine if they can replace HOTs in terms of psychometrics and content validity.

- Prioritize students' active engagement in phenomena and sense making (figuring out) as the mechanism for science teaching, learning, and assessment.

- Allow for deeper exploration of phenomena by having sets of multiple items digging into a particular phenomenon.

NCES Comments Summary

NCES submitted comments relating to challenges and considerations presented by the current NAEP Science Framework for operationalizing the science assessments. Their issues are categorized into:

1. Ambiguous Content
2. Ambitious Content
3. Standardized Assessment Constraints
4. Implementation Considerations

1. Ambiguous guidance

   Learning progressions (LPs) are referenced heavily in the Science Framework. LPs are not clearly explicated, and their development has not been sufficient to cover the intended science content. Currently, cognitive demands and science practices proved the mental model and structure for measuring student progression in understanding science.

2. Ambitious Content

   Measuring two dimensions (content and practice) is a requirement for science items. There is not enough specificity around expectations for measuring two dimensions. The example items in the current framework show varying approaches, but do not provide guidance on what is acceptable or preferable. In fact, whether a student approaches an item from a content or practice perspective may depend on that student's lived experiences and science background. Several examples are provided.

   There is also concern that the NAEP items are too difficult for many of the test takers. Given how large the proportion of Below Basic students there are, the number of items in that range of the score scale is low. This issue is complicated by the inclusion of language-heavy context provided with items. The context may be needed by lower ability students, but may also contribute to issues with cognitive load and fatigue.

Quantitative reasoning in science. The Science Framework indicates that students' mathematics knowledge should be 1-2 grade levels below their current grade in science. However, the quantitative reasoning may require much higher math skills than even their current grade. As an example, fourth graders must interpret multiple distributions of data on a graph. Further examples from the released items are provided.

3. Standardized Assessment Constraints

   Concept maps require more time than is reasonable given a 30-minute cognitive block. Many students do not reach the end of the task. This is true for partial concept maps as well (on 8th and 12th grade).

   There are design limitations with hands-on performance tasks (HOTs). The 30-minute block, space allotted to the student, and limitations on the materials provided mean that students cannot truly freely design an experiment. Experimental hybrid hands-on performance tasks (HHOTs), administered digitally and completed virtually show promise (especially in terms of speededness). These items will need to be researched to ensure content validity and psychometric soundness.

4. Implementation Considerations

   Hybrid hands-on performance tasks (HHOTs) are resource intensive. Task development is intense, plus these items require kit materials. They also require additional training for administrators.

   Alignment with future NAEP Innovations (like multi-stage testing, online, device agnostic, and reduced contact administration) may require substantial changes. These may include a designated staff administrator to monitor HHOTs. Scenario-based tasks like ICTs and HHOTs may require additional bandwidth. There are currently few easy items in the item pool and item development constraints make them challenging to create, which may limit how lower-difficulty stage adaptive item blocks can be developed.

   Increasing the number of HHOTs and ICTs may require increasing the number of printed booklets and, because they are often paired, may require increasing the required sample size. Increasing the number of these items may create challenges for monitoring trend. An increase in these items types should be implemented over several cycles.

   Further guidance on grade or skill progressions for scientific inquire would be helpful. There is no guidance in the framework for how scientific inquiry skills, like design, conduct, analyze, or draw conclusions from investigations may differ across grades.

Institute of Education Sciences

550 12th Street SW
Washington, DC 20202
Office: 202 123 4567

To: Sharyn Rosenberg, Ph.D.
Assistant Director for Assessment Development
National Assessment Governing Board

From: Holly Spurlock, Ph.D.
Branch Chief for National Assessment Operations
National Center for Education Statistics

Date: October 15, 2021

Subject: Implementation Challenges with the Current Science Framework

This memo summarizes implementation challenges and considerations presented by the current Science Framework for operationalizing a science assessment. The issues can be divided into several categories: ambiguous guidance, ambitious content, standardized assessment constraints, and additional implementation considerations. In addition, attached is a NAEP Validity Studies (NVS) Panel white paper titled "Revision of the NAEP Science Framework and Assessment".

Ambiguous guidance

Learning progressions. Learning progressions (LPs) are referenced heavily in the Science Framework as part of the cognitive and mental models that should be used to measure students successive understanding of complex science principles. While there are no rigid requirements of the framework to assess science content and knowledge using Learning progressions, NCES has not implemented LPs to the extent expected by the framework. This is an area where the field of science assessment development has not caught up with the forward-thinking nature of the science framework. In the field of science, LP development in science assessment development has been uneven and insufficient to fully cover framework content, and existing LPs are still being developed and validated by the science assessment field. Further, there are differing approaches to measuring LPs in a standardized assessment. The science framework views LPs as a mental model for how knowledge matures over time regardless of grade, while other assessment standards focus on grade-level progressions. Instead, NCES relies heavily on the cognitive demands and science practices outlined by the framework to provide the mental model and structure for measuring student progression in understanding science principles.

Ambitious content

Measuring two-dimensions (content and practice). The Science Framework requires that each item generate performance expectations for the integration of science content and practice knowledge. That is, each item must measure two-dimensions; "knowing" science and "doing"

Institute of Education Sciences

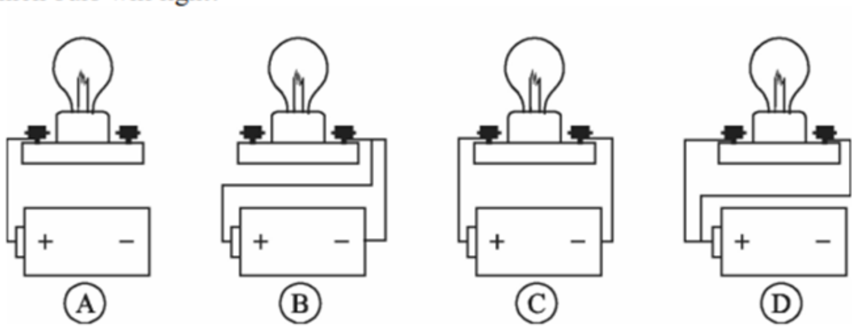
550 12th Street SW
Washington, DC 20202
Office: 202 123 4567

science. However, the framework does not provide enough specificity around the performance expectations for measuring two-dimensions (i.e., content and practice) for assessment developers and various stakeholders. The example items in the current framework show varying approaches that reflect debates among stakeholders, but it does not provide guidance on which approaches are acceptable or preferable. The example shown below from pages 65-66 of the Science Framework illustrates the challenge with measuring domain knowledge (i.e., content) and application of science skills (i.e., practice), as the latter can depend heavily on the former.

Figure 1. Illustrative item for measuring Using Science Principles (pages 65-66, Science Framework).

Illustrative Item

The pictures show a light bulb connected to a battery.
Which bulb will light?



Key: C

P4.11, Using Science Principles
Source: TIMSS 2003, Grade 4.

Student responses to this item are open to two interpretations. If students have had a great deal of exposure to these types of circuit representations, their responses would fall under Identifying Science Principles. However if, these circuit representations are relatively new for students, then they would need to apply more reasoning and their responses would fall under Using Science Principles.

The distinction between how students apply their content knowledge (e.g., “Identifying Science Principles” science practice or “Using Science Principles” science practice in Figure 1) depends heavily on the prior content knowledge students bring to the item. Further, there is not sufficient guidance for how much content knowledge should be measured in scientific-inquiry focused discrete items, hands-on performance tasks and interactive computer tasks – a topic that is heavily debated among the scientific assessment development community. The example shown

IES Institute of Education Sciences

550 12th Street SW
Washington, DC 20202
Office: 202 123 4567

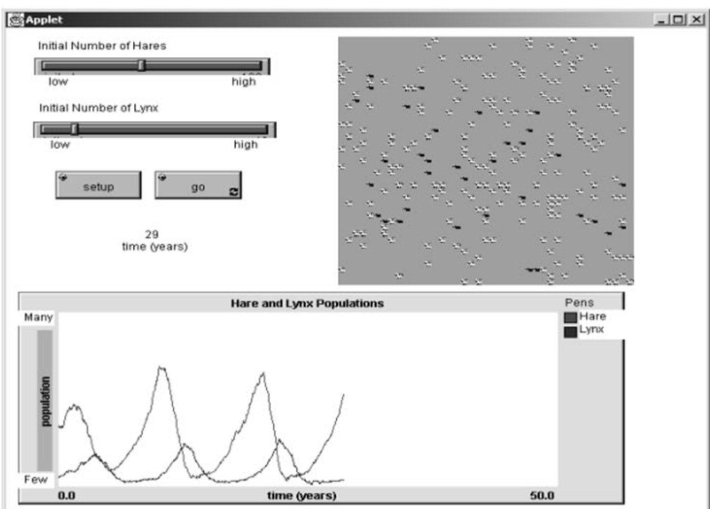
below from pages 69-71 of the Science Framework, was heavily debated among NCES's science standing committee[1] on whether this illustrative item assessed any content knowledge.

Figure 2. Illustrative Item for measuring Scientific Inquiry (pages 69-71, Science Framework).

Illustrative Item

This interactive computer task is one module in an extended assessment of students' abilities to use a range of technologies to investigate a complex problem, "Should lynx be reintroduced into a national park?" Students accessed, organized, and analyzed data on the number of hares in the park over a 25-year period, researched factors that would impact the population, and created a graph to analyze the trend. (See appendix C for a description of the full task.)

This module allows students to interact with a simulated predator/prey (lynx/hare) population model. Students use the modeling tool to observe population trends that result from different parameter values for the lynx and hare populations. The screen shot below is an example of what students see after they have selected parameters and run the simulation. Note that it is a single screen shot and represents only a small subset of the many screens that students actually see when engaged in this interactive computer task. After students have run the modeling software, they are asked a series of questions (e.g., size of the hare population over time).



L8.6, Using Scientific Inquiry
Source: Quellmalz et al. 2004.

Content experts could (and did during the Science item development process) argue that the illustrative item in Figure 2 measured how well students can manipulate variables to collect data **without** expectations for understanding content knowledge related to the interdependence of species.

[1] NCES's item development contractor utilizes subject-area standing committees composed of teachers and other content experts, state and local education agency representatives, and content area researchers, to review new item development.

IES Institute of Education Sciences

550 12th Street SW
Washington, DC 20202
Office: 202 123 4567

Greater specificity in future frameworks about approaches and examples demonstrating a consistent approach (or expected and clearly indicated range of approaches) for how to assess content and practice would be helpful. The framework does include a section on the Summary of Practices (page 76) with two examples of clarifications on sample performance expectation for two content statements. For brevity, only the Life Science example is included here.

Figure 3. Clarification: Sample Performance Expectations for a Life Science Content Statement (pages 77-78, Science Framework).

Clarification: Sample Performance Expectations for a Life Science Content Statement

The examples below are all related to the following grade 8 Life Science content statement:

L8.4: Plants are producers—they use the energy from light to make sugar molecules from the atoms of carbon dioxide and water. Plants use these sugars along with minerals from the soil to form fats, proteins, and carbohydrates. These products can be used immediately, incorporated into the plant's cells as the plant grows, or stored for later use.

All examples are also related to a specific situation:

Two different varieties of grass—one better adapted to full sunlight and one better adapted to shade—are each grown in sunlight and in shade.

The results of a controlled experiment along these lines might resemble the following:

Condition	Grass Type A	Grass Type B
Sunlight	Better growth*	Less good growth*
Shade	Less good growth*	Better growth*

* Several variables could be used to indicate growth: mass or dry mass of plants, thickness of stems, number of new sprouts, etc.

<u>Identifying Science Principles</u>

1. State from where a plant's food originates.
2. Classify the grass plants as producers or consumers.

The first performance calls for students to repeat information found in the content statement with little or no modification. The second performance asks students to use the definition of producers given in the content statement to classify or identify the plants.

<u>Using Science Principles</u>

1. Predict whether sugar will move up or down the stems of the grass plants and explain your prediction.
2. Explain where the mass of the growing grass originates.

These performances require students to use principles in the content statement to predict or explain specific observations (growing grass in this case). The content statement itself does not provide the answers to the questions.

Institute of Education Sciences
550 12th Street SW
Washington, DC 20202
Office: 202 123 4567

Using Scientific Inquiry

1. Given a data table showing the mass of grass plants of each type grown in the sunlight and shade, draw conclusions about which variety of grass is better adapted to each condition.
2. List other variables that should be controlled in order to feel confident about your conclusions.

The first performance is related to the content statement in that the importance of light for plant growth is useful background information for students. However, the performance requires interpretation of new information (the data table) that has to do with differences among types of plants, while the content statement contains generalizations about all plants. Thus, the performance requires students to use the data to develop new knowledge that they did not have previously. The second performance is in part an assessment of the students' understanding of experimental design. However, good answers would also require knowledge of this and related content statements to identify variables that are relevant to plant growth.

Using Technological Design

1. Given experimental results on the growth of different varieties of grass plants under sunlight and shade conditions, develop a plan for using different types of grass seed in different parts of a partially shaded park.

This performance requires students to use knowledge of the content statement and the experimental results to accomplish a practical goal (in this case, a park with grass growing well in areas that receive varying amounts of sunlight).

While these examples in the Science Framework and Specifications documents are not actual items, they provide considerations for how items can target different science practices. This would make it easier for assessment developers to know what expectations are, for example, for how much content knowledge should be measured in tasks, or whether content as context is sufficient. This would also be helpful in determining how a collection of two-dimension items across item types (DIs, ICTs, and HOTs) can cover the breadth and depth required by the framework.

Item difficulty. The Science Framework includes grade-level achievement level descriptors for each science content area and general statements about the science practices for *NAEP Basic*, *NAEP Proficient* and *NAEP Advanced*, and suggests that these descriptions can be used to develop a broad range of items for each achievement level. However, the framework also expects students to be exposed to challenging subject matter, e.g., "[In designing hands-on performance tasks] the NAEP assessment should provide students with a challenging problem… Hands-on performance tasks should be "content rich" in that they require knowledge of science principles to carry them out (Science Framework, pages 106-107)." Given the framework performance expectations for breadth and depth of content knowledge and its integration with practices, it is a challenge to develop items in the easier range while maintaining item rigor and measuring authentic knowing and doing science. If expectations for content knowledge are too high,

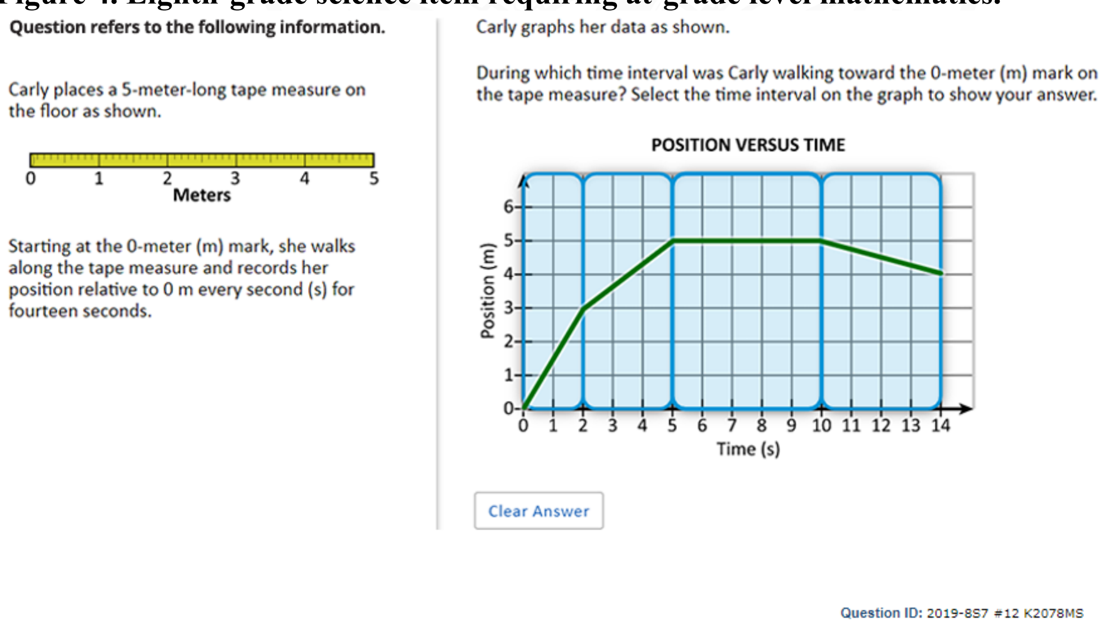
IES Institute of Education Sciences

550 12th Street SW
Washington, DC 20202
Office: 202 123 4567

students may not be able to demonstrate what they can do (i.e., science practice). However, if they are too low, the measurement of a practice may not be considered valid. Results from the 2019 Science assessments illustrate this point further: 27% of 4th graders, 33% of 8th graders, and 41% of 12th graders fall below *NAEP Basic*, however we have fewer items that measure these students compared to *NAEP Basic*, *NAEP Proficient* and *NAEP Advanced*. Further, the amount of contextual information that students must be given within an item in order to meaningfully engage with the content and practices can lead to higher cognitive load and burden, particularly for lower ability students who may need that context more so than higher ability students. While recent attempts have been made to identify and measure more basic scientific content and skills to develop easier items, the Science item pools continue to be difficult and may reflect a rigorous Science Framework.

Quantitative reasoning in science. The Science Framework Specifications state that the mathematics content required for quantitative reasoning in science content and practice knowledge should be 1-2 years below grade level (Science Framework Assessment and Item Specifications, page 21). However, NCES has had to use at- or above-grade level mathematics content knowledge in some science items to validly measure students' quantitative reasoning in science. For example, the NAEP Mathematics Framework does not expect fourth graders to read or interpret multiple distributions of data. However, displaying multiple distributions of data on a graph may be needed to assess fourth graders scientific inquiry skills of interpreting data and drawing conclusions from an experiment with two or more conditions, e.g., a graph with two or more lines. Figure 4 provides another example from a released eighth-grade science item.

Figure 4. Eighth-grade science item requiring at-grade level mathematics.



Institute of Education Sciences
550 12th Street SW
Washington, DC 20202
Office: 202 123 4567

The eighth-grade science item in Figure 4 asks students to interpret a line graph that describes Carly's position relative to a 5-meter-long tape measure for 14 seconds. This aligns with the science content objective, P8.14 "An object's position can be measured and graphed as a function of time" (Science Framework, page 34). However, students are not typically introduced to line graphs of this nature until eighth grade according to the common core state standard 8F.B.5 "Describe qualitatively the functional relationship between two quantities by analyzing a graph (e.g., where the function is increasing or decreasing, linear or nonlinear). Sketch a graph that exhibits the qualitative features of a function that has been described verbally." Further, the updated NAEP Mathematics Framework permits this type of graph at eighth grade, but it is not permitted at fourth grade. The examples in this section demonstrate the need to use at-and-above grade level mathematics content knowledge to validly measure students quantitative reasoning in science. This challenge is similarly expressed in the NVS white paper on "Revision of the NAEP Science Framework and Assessment", which states "NGSS performance expectations in science and engineering would likely require students to use some mathematics that is beyond the corresponding grade level".

Standardized assessment constraints

Timing constraints with concept maps. The framework recommends that each assessment include at least one concept-mapping interactive computer task (ICT) at eighth grade and twelfth grade. However, it is not feasible to develop authentic concept-mapping items that allow students to show the process of transferring their mental models into conceptual models as concept maps within a 30-minute cognitive block. NCES developed an ICT that included a partial concept-mapping task for the 2009 science assessment where students were asked to read and synthesize information from animal cards (i.e., habit and diet) to finish a partially constructed food web. However, 51% of students were not able to reach the final item of the task during pilot testing. Edits were made to the task to remove most of the concept-mapping portion so that students were only asked to fill in two missing organisms and their connecting arrows in the food web, but still 22% of students did not reach the end of the task. Given that prior attempts to develop a concept-mapping task within 30-minutes were not successful, NCES has not implemented concept-mapping in the Science assessments.

Design limitations with hands-on performance tasks (HOTs). The framework states that students should be able to freely design the experiment for HOTs, particularly given past criticism that the previous science framework allowed for prescriptive or "recipe"-like HOTs. However, the structure of a HOT and the materials a student can use are limited by assessment timing (i.e., 30-minute cognitive block), space allocated to the student on assessment day, safety, and what is provided in the kit materials. With the migration to hybrid hands-on tasks (HHOTs) for the 2015 pilot, where students were given digital instructions and could record their answers digitally, NCES developed tasks that allowed students flexibility in designing hands-on

Institute of Education Sciences
550 12th Street SW
Washington, DC 20202
Office: 202 123 4567

experiments and running multiple experimental trials. However, the 2015 pilot showed that hybrid hands-on tasks were speeded, and that speededness varied by grade and task. All three grade 4 and all three grade 12 tasks were speeded, from 23% to 72% of students not reaching the final item. Two out of three grade 8 tasks were speeded, from 75% to 81% of students not reaching the final item. After making considerable edits to constrain the experimental design of the hands-on tasks, the 2019 operational data shows that the HHOTs were much less speeded, ranging from 10% to 28% of students not reaching the final item. Development of hands-on tasks requires careful balance of the amount or depth of directions provided so that all students can engage in the task while designing and carrying out an experiment that can fit within the 30-minute assessment time and materials provided. There is the potential for hands-on tasks to become entirely virtual simulations as part of interactive computer tasks (ICTs). Further research is needed to investigate psychometric and content validity considerations to determine if ICTs can fully replace HOTs to measure scientifically inquiry.

Additional Considerations

Hybrid hands-on performance tasks (HHOTs) are resource intensive. HHOTs incur more expenses, additional resources and level of effort compared to any other item type found in NAEP. Extra resources are required prior to, during and post-data collection to develop and administer HHOTs alongside other science content. Below are some examples of the extra work required:

- In addition to rigorous task development that can cost more than discrete item development, item developers must also perform parallel processes to design and develop the associated kits (e.g., prototyping and testing). Once the kits are finalized, approved, and manufactured, additional quality assurance efforts are required to ensure that the digital tasks and the kits are in sync for a cohesive student experience and smooth administration.
- HHOTs require kit materials, which creates additional resources and costs for the Materials, Distribution, Processing and Scoring contractor to purchase, package and ship the kit materials to field staff. Further, some kit materials can be difficult or expensive to modify after piloting if changes are required.
- The Sampling and Data Collection contractor must hire an additional field staff member to the sample that includes HHOTs so they can monitor the students use of the kits and support the HHOT administration. This requires specialized administrator training and additional staffing to:
    - Receive and inventory kits
    - Distribute kits at appropriate time
    - Monitor kit use
    - Respond to questions in a standardized manner
    - Clean up after the kits

Institute of Education Sciences
550 12th Street SW
Washington, DC 20202
Office: 202 123 4567

- After the administration, administrators are asked to sort kit materials into goodie bags and waste to offer reusable materials for school use.
- Extra effort is required to develop scoring rubrics and training materials to support scoring of HHOTs. Scoring guides can be intensive given the open-ended nature of student responses to items assessing scientific inquiry.

However, providing students with opportunities to demonstrate their understanding of scientific inquiry and experimentation through designing, implementing, and drawing conclusions is an important part of the Science Framework. NCES continues to investigate ways of replacing hands-on activities with alternative, less-costly designs.

Alignment with future NAEP innovations. In recent years, the NAEP program has expressed an interest in moving towards more innovative and less costly administration models, like multi-stage testing and online, device agnostic and reduced contact administration. There are several aspects of the current Science that should be considered as NAEP moves towards these future innovations. Below are some examples.

- Having a designated field staff administrator to monitor HHOTs must be accounted for as NAEP program goals shift to a reduced contact and contactless administration model. In the reduced contact and contactless models, school staff will serve as administrators and may need further staff and training to accommodate administration of HHOTs.
- Scenario-based tasks, like ICTs and HHOTs, may require additional bandwidth to run resource-heavy science inquiry simulations. This may be challenging for online and device agnostic delivery models that require assessments to run on school internet with limited bandwidth and school devices with reduced processing speeds (e.g., RAM).
- As previously mentioned, the difficulty of the science item pools prohibits implementing adaptive design for the Science assessment as there are insufficient items to support development of easy, or even moderately easy targeted blocks. If there is a desire to implement adaptive design, there are also challenges associated with how to handle HHOTs and ICTs in an adaptive design (e.g., most HHOTs and ICTs target one science subscale).

Design constraints with increasing the number of hybrid hands-on tasks (HHOTs) and interactive computer tasks (ICTs). While it is difficult to predict what impact increasing the number of HHOTs and ICTs will have on measurement validity and reliability in the future, NCES anticipates several operational challenges that should be considered. Analyses from the 2019 science results indicate that a higher proportion of HHOTs and ICTs could have had a larger impact on group scores and consequently an impact on trend reporting. Further, increasing the number of HHOTs and ICTs would add more blocks to the assessment and consequently more booklets since HHOTs and ICTs should be paired, or linked, with each other and with discrete blocks according to balanced incomplete block (BIB) design. Increasing the number of

Institute of Education Sciences

550 12th Street SW
Washington, DC 20202
Office: 202 123 4567

booklets might increase the sample size requirement for some analyses and potentially increase the level of effort and resources needed to manage a larger item pool. Ultimately, if there is a desire to increase the number of HHOTs or ICTs in the science assessment, then NCES recommends that this increase be implemented gradually over several assessment cycles.

Grade or skill progressions for scientific inquiry. The Science framework does not provide any information as it relates to the application of science inquiry across grade levels and skill progressions. There is no guidance from the framework for how scientific inquiry skills, e.g., design, conduct, analyze or draw conclusions from investigations, may differ for fourth-graders, eighth-graders and twelfth-graders. NCES created evidence centered design (ECD) models to guide grade-level development of items and tasks that assessed scientific inquiry, but further guidance on this area would be helpful.

Enclosure: NAEP Validity Studies White Paper: Revision of the NAEP Science Framework and Assessment

NAEP Validity Studies White Paper: Revision of the NAEP Science Framework and Assessment

James W. Pellegrino
University of Illinois Chicago

October 2021
Commissioned by the NAEP Validity Studies (NVS) Panel

The NAEP Validity Studies (NVS) Panel was formed in 1995 to provide a technical review of NAEP plans and products and to identify technical concerns and promising techniques worthy of further study and research. The members of the panel have been charged with writing focused studies and issue papers on the most salient of the identified issues.

Panel Members:

Keena Arbuthnot
Louisiana State University

Peter Behuniak
Criterion Consulting, LLC

Jack Buckley
American Institutes for Research

James R. Chromy
Research Triangle Institute (retired)

Phil Daro
Strategic Education Research Partnership (SERP) Institute

Richard P. Durán
University of California, Santa Barbara

David Grissmer
University of Virginia

Larry Hedges
Northwestern University

Gerunda Hughes
Howard University

Ina V.S. Mullis
Boston College

Scott Norton
Council of Chief State School Officers

James Pellegrino
University of Illinois at Chicago

Gary Phillips
American Institutes for Research

Lorrie Shepard
University of Colorado Boulder

David Thissen
University of North Carolina, Chapel Hill

Gerald Tindal
University of Oregon

Sheila Valencia
University of Washington

Denny Way
College Board

Project Director:

Sami Kitmitto
American Institutes for Research

Project Officer:

Grady Wilburn
National Center for Education Statistics

For Information:

NAEP Validity Studies (NVS) Panel
American Institutes for Research
1333 Broadway Ave, Suite 300
Oakland, CA 94612
Email: skitmitto@air.org

CONTENTS

TABLES

FIGURES

OVERALL PURPOSE AND ORGANIZATION

The purpose of this white paper is to consider issues related to the scope and focus of a possible new framework for National Assessment of Educational Progress (NAEP) Science (hereafter, NAEP science), including its possible expansion to include aspects of what is represented in NAEP Technology and Engineering Literacy (TEL) (hereafter, NAEP TEL). The goal is to provide the NAEP Validity Studies (NVS) Panel and the NAEP program with input about possible directions for the future and the rationale for choosing among them. Five major sections comprise this paper.

Section I sets the stage for the sections that follow by providing brief background information about the history and projected future uses of the NAEP Science Framework and Assessment as well as the NAEP TEL Framework and Assessment. It also summarizes the National Center for Education Statistics (NCES) and the National Assessment Governing Board (NAGB) timeline for consideration of possible revisions to the NAEP science framework in anticipation of its use to guide the NAEP Science Assessment scheduled for 2028.

Section II contains information on analyses comparing the current NAEP science framework and the NAEP TEL framework to the overall science and technology framework and related set of standards that emerged in the United States in the early part of the last decade. The section begins with a brief synopsis of the content and focus of the NAEP Science and TEL frameworks followed by a brief synopsis of the National Research Council (NRC) *Framework for K–12 Science Education* (NRC, 2012) (hereafter, NRC framework) and the derivative *Next Generation Science Standards* (NGSS) (NRC, 2013). Following that, results are presented from an extensive study comparing the alignment between NAEP Science and NAEP TEL and NGSS (Neidorf et al., 2016). In doing so, the section also considers some of the implications regarding assessments aligned with each reference source.

Section III focuses on the status of science standards and assessments in individual states since the publication of the NRC framework and the NGSS. It reviews the current status regarding state adoptions of science standards that are either identical to NGSS or that are partially aligned with the NGSS (i.e., NRC framework and NGSS "alike"), as well as states with science standards that have no claimed alignment with either the NGSS or NRC framework. For those states with science standards that are NRC framework/NGSS alike, results are summarized from a study examining content alignment between those state standards and the NAEP science framework (Dickinson et al., 2021). The section also includes a summary of the status of the design and implementation of state science assessments relative to their currently adopted standards. This consideration is limited to states that have adopted the NGSS and those whose adopted standards are NRC framework/NGSS alike. The section includes a brief review of the status of the implementation of curricular and instructional practices in states relative to the NRC framework and NGSS. Results are based on the most recent (2018) National Survey of Science and Mathematics Education. The section concludes with a consideration of trends in NAEP science performance for the last 12 years and some possible implications for future NAEP science assessments.

Section IV provides a brief discussion of advances in technology as related to the assessment of science and engineering knowledge and skills. It considers how various developments in digital technologies should be considered in reviewing the existing NAEP Science framework and assessment and envisioning possibilities for their updating. Discussion focuses on the affordances of technology with respect to the constructs that could be included in a revised framework and the associated task design, data capture, and data analytic issues involved in an assessment aligned to an updated framework. The section concludes with a brief discussion of practical and equity concerns related to digitally based assessment of science and technology proficiency.

Section V contains a set of conclusions and recommendations as input to the NCES and NAGB process of reviewing the NAEP science framework and considering possible revision. Conclusions and recommendations are based on the major findings presented in the prior sections.

SECTION I: BACKGROUND, TIMELINE, AND INPUTS

Relevant History: NAEP Science and NAEP TEL

NAEP Science

NAEP science is based on a framework that was adopted in 2005 for the 2009 assessment (NCES, 2009, 2014). That framework was used for the 2015 and 2019 administration of science at grades 4, 8, and 12. It will be used once more for the 2024 (originally 2023) administration of science at eighth grade only. The 2028 (originally 2027) operational administration of the science assessment at grades 4 and 8 at the national, state, and large urban district levels is supposed to be based on an updated science framework.

NAEP TEL

The NAEP TEL assessment is based on a framework developed for grades 4, 8, and 12 in the 2011–2012 period for the 2014 assessment at grade 8. That framework was used for the 2018 TEL administration for grade 8. It will be used twice more for the 2024 (originally 2023) and 2028 (originally 2027) TEL administrations for grade 8. Both planned TEL administrations overlap with NAEP science administrations: 2024 overlaps with the current science framework and assessment, and 2028 overlaps with the new science framework and assessment.

NAEP Science and TEL—Possible Merger

Discussions have been held within NAGB about possibilities for combining NAEP science and TEL, especially because both are now digitally based assessments. Doing so may make logical sense given overlaps in conceptual coverage with contemporary U.S. science and technology frameworks. Another benefit could be cost savings realized by having a single assessment representing key aspects of knowledge and skill for science and technology. Such a merger clearly would be most beneficial for the planned 2028 administration of both science and TEL. NAGB therefore may wish to consider developing a single 2028 assessment based on a new integrated science and technology framework.

Status and Plans for Review, Update, and/or Revision of the NAEP Science Framework

NAGB has started the process needed to consider updating the science framework for application in the design of the 2028 grades 4 and 8 science assessment. Given the current timeline, it appears that a decision about the need for and the scope of a science framework revision will be completed during 2022. Work toward making such a decision includes:

- Detailed information available in an NCES report issued in 2016 titled *A Comparison Between the Next Generation Science Standards (NGSS) and the National Assessment of Educational Progress (NAEP) Frameworks in Science, Technology and Engineering Literacy, and Mathematics* (Neidorf et al., 2016). Information about the results of this study is presented in Section II.
- A recently completed study by HumRRO titled *Comparative Analysis of the NAEP Science Framework and State Science Standards* (Dickinson et al., 2021) in which content overlap was examined between the NAEP science framework and the science

standards of individual states. Classification of state standards was based on information from the National Science Teachers Association (NSTA) specifying which states have current standards that are identical to NGSS, partially NGSS, or non-NGSS. The focus for the analysis was on alignment between the NAEP science framework and the standards of the partial NGSS and non-NGSS states. Information about the results of this study is presented in Section III.

- Input from a group of five or more experts, each of whom would consider the information derived from the two studies mentioned above—the 2016 AIR comparison of NAEP to NGSS (Neidorf et al., 2016) and the more recent HumRRO analysis of state standards relative to NAEP (Dickinson et al., 2021)—as well as other factors given the expert's experience in the field of science education, to present their thoughts on whether the framework needs to be changed and why.
- NAGB recently issued a public call for input on the NAEP science framework regarding its revision. NAGB requested responses from interested parties by October 15, 2021.

NAGB is scheduled at its March 2022 meeting to consider whether to move ahead with a revision of the science framework for application in the design of the 2028 science assessment. The board also will consider the input received from the various sources mentioned above. The timing of these activities should NAGB choose to recommend a science framework revision would easily extend into 2023 if not beyond. Given existing statutes, NAGB will convene two panels based on their policy (NAGB, 2018a, p. 5):

- The **Framework Visioning Panel** shall formulate high-level guidance about the state of the field to inform the process, providing these in the form of guidelines. The major part of the Visioning Panel work will be at the beginning to provide initial guidance for developing a recommended framework. The Visioning Panel shall be composed of the stakeholders referenced in the introduction above. At least 20 percent of this panel shall have classroom teaching experience in the subject areas under consideration. This panel may include up to 30 members with additional members as needed.
- The **Framework Development Panel** shall develop drafts of the three project documents and engage in deliberations about how issues outlined in the Visioning Panel discussion should be reflected in a recommended framework. As a subset of the Visioning Panel, the Development Panel shall have a proportionally higher representation of content experts and educators, whose expertise collectively addresses all grade levels designated for the assessment under development. Educators shall be drawn from schools across the nation, who work with students from high-poverty and low-performing schools, as well as public and private schools. This panel may include up to 15 members, with additional members as needed.

The timeline for initiating and completing the work of the panels remains to be specified, and because the work of the development panel follows from the work of the visioning panel, its work would end sometime in 2023 or later, pending public review of a draft framework and commentary with subsequent revision and then final adoption by NAGB. A revised framework would be used to develop the design and tasks for the 2028 NAEP science assessment.

SECTION II. ANALYSIS OF THE NAEP SCIENCE FRAMEWORK RELATIVE TO OTHER CONTEMPORARY SCIENCE AND TECHNOLOGY FRAMEWORKS

This section examines how the NAEP science framework and assessment and NAEP TEL framework compare with the NRC *Framework for K–12 Science Education* (hereafter, NRC framework) and the derivative *Next Generation Science Standards* (NGSS). It begins with a brief description of key elements of each of the four reference sources and is followed by a summary of results from a detailed study of the correspondences between the two NAEP frameworks and the NGSS. Highlighted in the summary are important areas of similarity and dissimilarity and some of the implications relative to assessment.

Overview of the NAEP Science Framework and Assessment

As noted earlier, the current NAEP science assessment is based on a framework originally developed for the 2009 assessment administration at grades 4, 8, and 12. That framework also was used for the 2011 administration at grade 8 and the 2015 and 2019 administrations at grades 4, 8, and 12. The framework is scheduled to be used once more for the 2024 administration for eighth grade only. The scheduled 2028 operational administration of science for grades 4 and 8 is supposed to be based on an updated science framework.

The current NAEP science framework (NAGB, 2008, 2014) was developed approximately 4 years before the 2009 administration and incorporated ideas from contemporary theory and research on science learning and assessment including synthesis volumes from the NRC: *How People Learn: Brain, Mind, Experience and School* (Bransford et al., 2000); *Knowing What Students Know: The Science and Design of Educational Assessment* (Pellegrino et al., 2001); Systems of State Science Assessment (Wilson & Bertenthal, 2005) and *Taking Science to School* (National Research Council, 2007). The framework included important ideas about the learning and knowing of both science content and science practices with a particular emphasis on their integration as discussed below.

Science Content. The science content for NAEP is defined by a series of statements that describe key facts, concepts, principles, laws, and theories in three broad areas: physical sciences, life sciences, and Earth and space sciences. Table 1 shows the major topics and subtopics within each of the three major science domains. The nature of the specific content knowledge changes in both scope and sophistication across the three grade levels.

Table 1. NAEP science content areas and topics

Physical sciences	Life sciences	Earth and space sciences
Matter • Properties of matter • Changes in matter	Structures and functions of living systems • Organization and development • Matter and energy transformations • Interdependence	Earth in space and time • Objects in the universe • History of Earth
Energy • Forms of energy • Energy transfer and conservation		Earth structures • Properties of Earth materials • Tectonics
Motion	Changes in living systems	Earth systems

• Motion at the macroscopic level • Forces affecting motion	• Heredity and reproduction • Evolution and diversity	• Energy in Earth systems • Climate and weather • Biogeochemical cycles

SOURCE: National Assessment Governing Board, 2014, Exhibit 4, p. 19. Reprinted with permission.

Science Practices. The second dimension of the framework is defined by four science practices: Identifying Science Principles, Using Science Principles, Using Scientific Inquiry and Using Technological Design. In the NAEP science framework, the first two practices (Identifying Science Principles and Using Science Principles) generally are considered as "knowing science," and the last two practices (Using Scientific Inquiry and Using Technological Design) are considered as the application of that knowledge to "doing science" and "using science to solve real-world problems."

Table 2 provides a high-level description of the nature of each specific practice in terms of the types of cognitive demands placed on students as they engage in a practice as applied to a topic from a specific science content area.

Table 2. NAEP science practices: General labels and specific applications

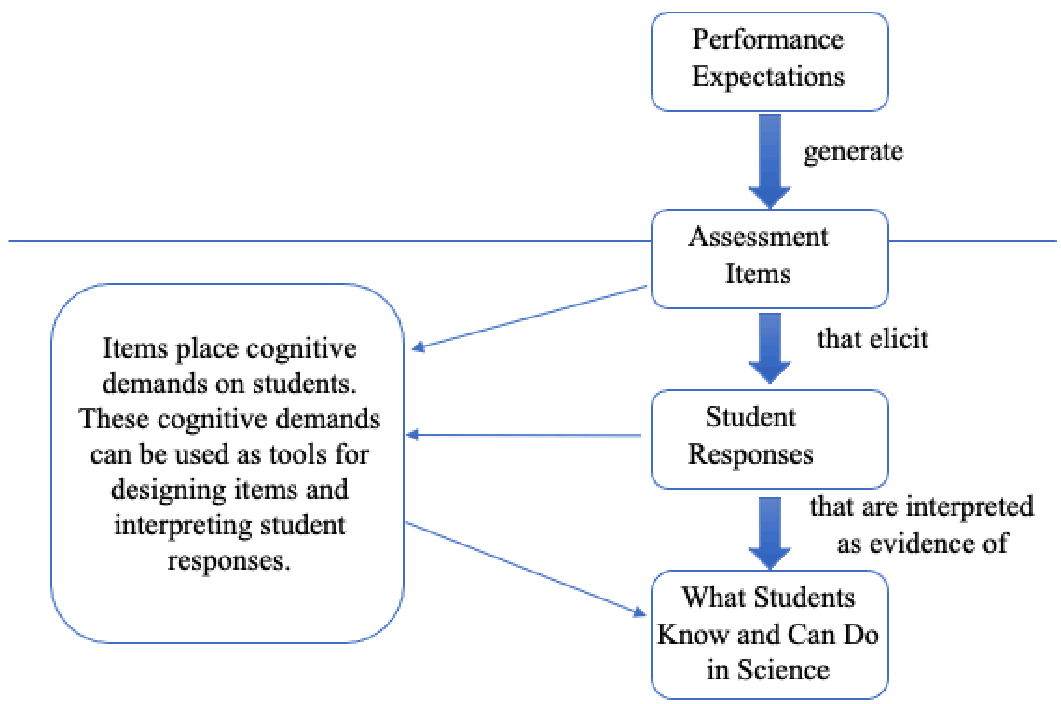
	Practice Label	Practice Applications			
←Communicate accurately and effectively→	Identifying Science Principles	Describe, measure, or classify observations.	State or recognize correct science principles.	Demonstrate relationships among closely related science principles.	Demonstrate relationships among different representations of principles.
	Using Science Principles	Explain observations of phenomena.	Predict observations of phenomena.	Suggest examples of observations that illustrate a science principle.	Propose, analyze, and/or evaluate alternative explanations or predictions.
	Using Scientific Inquiry	Design or critique aspects of scientific investigations.	Conduct scientific investigations using appropriate tools and techniques.	Identify patterns in data and/or relate patterns in data to theoretical models.	Use empirical evidence to validate or criticize conclusions about explanations and predictions.
	Using Technological Design	Propose or critique solutions to problems given criteria and scientific constraints.	Identify scientific tradeoffs in design decisions and choose among alternative solutions.	Apply science principles or data to anticipate effects of technological design decisions.	

SOURCE: National Assessment Governing Board, 2014, Exhibit 13, p. 76.

Performance Expectations—Combining Content and Practices. The design of the NAEP science assessment is guided by the framework's descriptions of both the science content and science practices to be assessed but with the key assumption that the practices are to be combined with a science content statement to generate specific student performance expectations that serve as the target for assessment. Assessment items are then developed based on the description of each specific performance expectation.

Using the logic of specific performance expectations as a guide for item development processes, items are then designed to vary the cognitive demands of tasks, a process that then influences the conclusions to be made about student performance. Such a process of item development can be represented schematically as shown in Figure 1.

Figure 1. NAEP assessment item development model



SOURCE: National Assessment Governing Board, 2014, Exhibit 2, p. 12.

In 2009, 2011, and 2015, NAEP science was administered as primarily a paper-and-pencil test. In 2019 a major shift occurred when NAEP science was administered for the first time as an entirely digitally based assessment. The Nation's Report Card (2019) provides a description of the new digital assessment:

> The NAEP digitally based science assessment consisted of standalone, discrete questions, and scenario-based tasks comprising a connected sequence of questions. Scenario-based tasks were designed to engage students in scientific inquiry through hands-on activities and computer simulations set in real-world contexts. The tasks provided students opportunities to demonstrate their knowledge and skills in each of three science content areas and four science practices. The science assessment included two types of scenario-based tasks:
>
> - Interactive computer tasks (ICTs). ICTs use real-world simulations to engage students in scientific investigations that require the use of science inquiry skills and application of scientific knowledge to solve problems.

- Hybrid hands-on tasks (HHOTs). Students perform hands-on scientific investigations using materials in kits provided by NCES. The “hybrid” in HHOTs denotes that these tasks combine hands-on investigations with digital activities. Students use NCES-supplied tablets to view kit instructions, record results and data, and answer assessment questions.

Overview of the NAEP TEL Framework and Assessment

As noted earlier, a TEL framework was developed for the first TEL assessment in 2014 at grade 8 and was used again for the 2018 TEL at grade 8. It is scheduled to be used twice more for the 2024 and 2028 TEL administrations at grade 8.

The development of this framework and assessment was motivated by several factors. In the science education community, a call for preparing students with technology and engineering literacy has been long awaited. The *Science for All Americans* report (American Association for the Advancement of Science, 1990) explicitly suggested that science education should incorporate technology and engineering as a form of scientific inquiry. Bybee (2010) proposed an advance to STEM education by integrating technology and engineering with science and mathematics education. He argued that “there are very few other things that influence our everyday existence more [than technology] and about which citizens know less” (Bybee, 2010, p. 30). Bybee suggested extending traditional information communication technology education by integrating ICTs with other subjects. He further pointed out that involving students in engineering activities could promote their abilities for both problem solving and innovation. He also acknowledged that engineering as typically presented in schools was inconsistent with its careers and contributions to society, and thus authentic scenarios needed to be developed for both learning and assessment (Bybee et al., 2009).

The NRC report, *Education for Life and Work: Developing Transferable Knowledge and Skills in the 21st Century*, identified information literacy and ICT literacy as two of the most frequently mentioned critical competencies for students to succeed in the 21st century (Pellegrino & Hilton, 2012). That report discussed various foundations for education, and STEM education in particular, including preparing future entrants to the labor market with the ability to adapt to technological changes in society rather than simply acquiring static bits of knowledge. Similarly, another 2012 NRC report, the *Framework for K–12 Science Education* (NRC, 2012), framed one of the overarching goals of science education as the development of students who “are careful consumers of scientific and technological information related to their everyday lives” (p. 1). The framework explicitly includes “Engineering, Technology, and Applications of Science” as one of four disciplinary core ideas and describes “defining problems, design solutions, and using computational thinking” as critical components of science and engineering practices. Further discussion of the NRC framework follows this section on TEL.

These and other trends related to technology and engineering literacy spurred the development of a TEL framework and inclusion of the TEL assessment as part of the NAEP program. The goal of TEL has been to obtain information about students’ understanding of technology and its effect on our society and environments, as well as students’ ability to design solutions to solve real-world problems. The TEL framework describes TEL as the “capability to use, understand, and evaluate technology as well as to understand technological principles and strategies needed to develop solutions and achieve goals" (NAGB, 2013, p. xi). Specifically, the framework identified three interconnected areas to be assessed (NAGB, 2018b, p. xii) as follows:

- *Technology and Society* deals with the effects that technology has on society and the natural world and with the sorts of ethical questions that arise from those effects. Knowledge and capabilities in this area are crucial for understanding the issues surrounding the development and use of various technologies and for participating in decisions regarding their use.
- *Design and Systems* covers the nature of technology, the engineering design process by which technologies are developed, and basic principles of dealing with everyday technologies, including maintenance and troubleshooting. An understanding of the design process is particularly valuable in assessing technologies, and it can also be applied in areas outside technology, since design is a broadly applicable skill.
- *Information and Communication Technology* includes computers and software learning tools, networking systems and protocols, hand-held digital devices, and other technologies for accessing, creating, and communicating information and for facilitating creative expression. Although it is just one among several types of technologies, it has achieved a special prominence in technology and engineering literacy because familiarity and facility with it is essential in virtually every profession in modern society.

Students taking the TEL assessment are expected to succeed in the following three types of thinking and reasoning practices:

- *Understanding technological principles* focuses on students' knowledge and understanding of technology and their capability to think and reason with that knowledge;
- *Developing solutions and achieving goals* refers to students' systematic application of technological knowledge, tools, and skills to address problems and achieve goals presented in societal, design, curriculum, and realistic contexts; and
- *Communicating and collaborating* centers on students' capabilities to use contemporary technologies to communicate for a variety of purposes and in a variety of ways, working individually or in teams. (NAGB, 2018b, pp. 3-2–3-3)

The TEL assessment has developed scenario-based tasks designed to engage students in multimedia environments to gauge students' understanding of technological and engineering principles and their ability to apply such principles to determine design solutions. Most of TEL's assessment tasks are computer simulation problems involving technology and engineering scenarios.

Overview of the NRC Science Education Framework and Next Generation Science Standards

Based on multiple sources of evidence and discussions about the knowing and learning of science, the nature of science education as it had been practiced in the United States, and evidence of relatively poor student achievement in science across K–16+, agreement emerged during the early part of this century about the need for substantial change in science standards, instruction, and assessment, including what we expect students to know and be able to do in science, how science should be taught, and how it should be assessed.

> Recognition of this science education problem can be found in reports spanning elementary, secondary, and postsecondary education (K–16+). These reports present a consistent description of the nature of competence in science and include NRC reports on K–8 science education in formal and informal learning environments (NRC, 2007, 2009); curriculum and assessment frameworks for Advanced Placement (AP) science courses (e.g., College Board, 2011a, 2011b); and even revisions in the nature of the science knowledge required for entry to medical school and assessed on the Medical College Admissions Test (e.g., American Association of Medical Colleges, 2012). (Pellegrino, 2016, p. 5)

Reconceptualization of the nature of science competence emergent from these many and diverse sources was captured to some extent in the College Board's standards for success in high school science (College Board, 2009). Their most complete expression for all K–12 science education was presented in the 2012 NRC report, *A Framework for K–12 Science Education. Practices, Crosscutting Concepts and Core Ideas.* The NRC framework report contains many important key ideas, including articulation of three interconnected dimensions of science competence as denoted in the report's title. The three dimensions are Disciplinary Core Ideas (DCIs), Crosscutting Concepts (CCCs), and Science and Engineering Practices (SEPs). The NRC framework provides detailed descriptions of each dimension, the concepts that each dimension encompasses, and the rationale for their inclusion. Figure 2 provides a list of the dimensions and their associated high-level concepts.

DCIs are the big ideas associated with a discipline, like life science, and which are essential to explaining phenomena. CCCs are ideas like systems thinking that are important across many science disciplines and provide a unique lens to examine phenomena. SEPs are the multiple ways of knowing and doing science and engineering, like developing models and constructing explanations that scientists and engineers use to study the natural and designed world. The framework focuses on the need for the integration of these three dimensions in science and engineering education. The knowledge associated with each of the three dimensions must be integrated in the teaching, learning, and doing of science and engineering, and in assessing what students know and can do. The framework emphasizes research indicating that learning about science and engineering "involves integration of the knowledge of scientific explanations (i.e., content knowledge) and the practices needed to engage in scientific inquiry and engineering design" (NRC, 2012, p. 11). The disciplinary core ideas, crosscutting concepts, and science and engineering practices serve as thinking tools that work together to enable scientists, engineers, and learners to design solutions to problems, reason with evidence, and make sense of phenomena. When learners engage in science and engineering practices integrated with DCIs and CCCs to make sense of compelling phenomena or design solutions to complex problems, they build new knowledge about all three dimensions and come to understand the nature of how scientific knowledge and engineering solutions develop.

Figure 2. The three dimensions of the NRC framework

Scientific and Engineering Practices	Crosscutting Concepts
▪ Asking questions (for science) and defining problems for engineering ▪ Developing and using models ▪ Planning and carrying out investigations ▪ Analyzing and interpreting data ▪ Using mathematics and computational thinking ▪ Constructing explanations (for science) and designing solutions (for engineering) ▪ Engaging in argument from evidence ▪ Obtaining, evaluating, and communicating information	▪ Patterns ▪ Cause and effect: Mechanism and explanation ▪ Scale, proportion, and quantity ▪ Systems and system models ▪ Energy and matter: Flows, cycles, and conservation ▪ Structure and function ▪ Stability and change
Disciplinary and Core Ideas *Physical Sciences* PS 1: Matter and its interactions PS 2: Motion and stability: Forces and interactions PS 3: Energy PS 4: Waves and their applications in technologies for information transfer *Earth and Space Sciences* ESS 1: Earth's place in the universe ESS 2: Earth's ecosystems ESS 3: Earth and human activity	 *Life Sciences* LS 1: From molecules to organisms: Structures and processes LS 2: Ecosystems: Interactions, energy, and dynamics LS 3: Heredity: Inheritance and variation of traits LS 4: Biological evolution: Unity and diversity *Engineering, Technology, and the Applications of Science* ETS 1: Engineering design ETS 2: Links among engineering, technology, science, and society

SOURCE: NRC 2012, Box S-1, p. 3.

The rationale for the choice of the specific DCIs is important to note here relative to other previous standards and frameworks. One criticism of U.S. K–12 science curricula relative to those of other countries was that they were "a mile wide and an inch deep" (Schmidt et al., 1997, p. 62). The same concerns about breadth versus depth were made in an NRC Report on advanced study of science in U.S. high schools (NRC, 2002). In reaction, the framework focused on core ideas in each of the four content domains with the directive that students should continue to be exposed to these core ideas with increased levels of complexity and explanatory power relative to a range of phenomena and problem contexts throughout their schooling.

While each of the three dimensions matters, a central argument of the framework is that proficiency is demonstrated through *performances* that require the integration of all three dimensions. Such demonstrations are labeled *Performance Expectations (PEs)* because they specify what students at various levels of educational experience should know and be able to do. The Next Generation Science Standards (NRC, 2013) are an expression of the integrated knowledge vision contained in the framework, and provide a set of standards expressed as performances expectations for students from Kindergarten to 12th grade. The NGSS appear as clusters of performance expectations related to particular aspects of a core disciplinary idea (see Figure 3 for an example at grade 4). Each performance expectation requires students to draw upon knowledge of a specific practice and a crosscutting concept in the context of specific elements of disciplinary core knowledge. Across the set of performance expectations at a given grade level or grade band, each practice and crosscutting concept appears in multiple standards. A student demonstrates grade-level proficiency by completing performances that demonstrate that they can make use of their knowledge. To truly know and understand science is to be able to use the three dimensions of scientific knowledge together to explain compelling phenomena and/or provide solutions to complex problems.

Figure 3. NGSS Performance Expectations for Grade 4 Life Science 1: From molecules to organisms: Structures and processes

4-LS1 **From Molecules to Organisms: Structures and Processes**

PERFORMANCE EXPECTATIONS

Students who demonstrate understanding can:

4-LS1-1. Construct an argument that plants and animals have internal and external structures that function to support survival, growth, behavior, and reproduction. [*Clarification Statement*: Examples of structures could include thorns, stems, roots, colored petals, heart, stomach, lung, brain, and skin.] [*Assessment Boundary*: Assessment is limited to macroscopic structures within plant and animal systems.]

4-LS1-2. Use a model to describe that animals receive different types of information through their senses, process the information in their brain, and respond to the information in different ways. [*Clarification Statement*: Emphasis is on systems of information transfer.] [*Assessment Boundary*: Assessment does not include the mechanisms by which the brain stores and recalls information or the mechanisms of how sensory receptors function.]

Science and Engineering Practices	Disciplinary Core Ideas	Crosscutting Concepts
Engaging in Argument from Evidence Engaging in argument from evidence in 3–5 builds on K–2 experiences and progresses to critiquing the scientific explanations or solutions proposed by peers by citing relevant evidence about the natural and designed world(s). • Construct an argument with evidence, data, and/or a model. (4-LS1-1) • Use a model to test interactions concerning the functioning of a natural system. (4-LS1-2)	**LS1.A: Structure and Function** • Plants and animals have both internal and external structures that serve various functions in growth, survival, behavior, and reproduction. (4-LS1-1) **LS1.D: Information Processing** • Different sense receptors are specialized for particular kinds of information, which may then be processed by an animal's brain. Animals are able to use their perceptions and memories to guide their actions. (4-LS1-2)	**Systems and System Models** • A system can be described in terms of its components and their interactions. (4-LS1-1), (4-LS1-2)

SOURCE: NRC, 2013, p. 38. Reprinted with permission.

An important issue relative to the present paper's discussion of NAEP Science and NAEP TEL is the NRC framework's emphasis on the connections among science, engineering, and technology. While these connections are somewhat separate across NAEP Science and TEL, key practices and ideas from engineering are included in the NRC framework because of important interconnections between science and engineering and because evidence shows that engaging in engineering design can help leverage student motivation and increase learning in science. One goal of including ideas related to engineering, technology, and the applications of science in the framework for science education is to help students understand the similarities and differences between science (the natural world) and engineering (the designed world) by making the connections between the two fields explicit and by providing all students with an introduction to engineering.

The NGSS expanded upon the framework's adoption of the logic of learning progressions to describe students' developing proficiency in the three intertwined domains across grades K–12, noting that "If mastery of a core idea in a science discipline is the ultimate educational destination, then well-designed learning progressions provide a map of the routes that can be taken to reach that destination" (NRC, 2012, p. 26). The stress on learning progressions is supported by research on science knowing and learning described in the 2005 NRC report *Systems of State Science Assessment,* the 2007 NRC report *Taking Science to School* and in other documents describing research on the progression of student learning and understanding in science (e.g., Alonzo & Gotwals, 2012; Corcoran et al., 2009). The framework built in the idea of a developmental progression of student understanding across the grades by specifying grade band end point targets at grades 2, 5, 8, and 12 for each component of each disciplinary core idea. For the practices and crosscutting concepts, the framework also provided sketches of possible progressions for learning each practice or concept but did not indicate the expectations at any particular grade level. The NGSS built on these suggestions and developed tables that define what each practice might encompass at each grade level. The NGSS also defined the expected uses of each crosscutting concept for students at each grade level.

The NRC framework and NGSS stand in sharp contrast to prior generations of U.S. science standards (e.g., American Association for the Advancement of Science, 1992; NRC, 1996, 2000) that treated content and inquiry as separate strands of science learning. Unfortunately, both instruction and assessment followed suit. The form the standards took contributed to this separation: Content standards stated what students should know, largely in the form of declarative knowledge, and inquiry standards stated what they should be able to do, largely in the form of procedural knowledge. Consequently, instruction often separated content learning from inquiry and vice versa. Science education often was often criticized as "lots of hands on but not much minds on." In a similar fashion, assessments separately measured content knowledge in the absence of application or inquiry practice components in the absence of content concerns. Thus, the NGSS idea of an integrated, *multidimensional science performance* represents a different way of thinking about science proficiency. Disciplinary core ideas and crosscutting concepts serve as thinking tools that work together with scientific and engineering practices to enable learners to solve problems, reason with evidence, and make sense of phenomena. Such a view of competence signifies that measuring proficiency solely as the acquisition of core content knowledge or as the ability to engage in general inquiry processes is neither appropriate nor sufficient.

In the context of assessment, the importance of this integrated perspective of what it means to know science is that one should be attempting to assess where a student can be placed along a sequence of progressively more "scientific" understandings of a given core idea and successively more sophisticated applications of practices and crosscutting concepts. This idea is relatively unfamiliar in the realm of science assessments, which more often have been viewed as simply measuring whether students know or do not know particular grade-level content (Pellegrino, 2013). To support an integrated and developmental approach to science learning, the framework explains that assessment tasks "must be designed to gather evidence of students' ability to apply the practices and their understanding of the crosscutting concepts in the contexts of specific applications in multiple disciplinary areas" (NRC, 2012, p. 218). Assessments must strive to be sensitive both to grade-level-appropriate understanding and to those understandings that may be appropriate at somewhat lower or higher grades. This is particularly important for assessment materials and resources to support ongoing classroom instruction. The challenges of designing such multidimensional assessments for classroom and large-scale assessment use are substantial. Potential approaches and solutions were discussed in detail in another NRC report, *Developing Assessments for the Next Generation Science Standards* (Pellegrino et al., 2014).

Comparing the NAEP Science and TEL Frameworks and NGSS

Given the brief descriptions provided above, it should be clear that there are multiple similarities and overlaps as well as differences between the NAEP science framework and the NGSS and between NAEP TEL and NGSS. Even though the NAEP science framework predates the 2012 NRC framework and the derivative 2013 NGSS, overlapping content exists, each has a description of science practices, and both make use of the idea of performance expectations that involve the intersection of content and practice. The NAEP TEL framework was developed about the same time as the NRC framework and overlaps with the latter's highlighting of engineering practices alongside science practices, and its inclusion of Engineering, Technology, and the Application of Science as one of the four disciplinary areas.

Although some of the ideas that are part of the NRC framework and NGSS have found their way over time into the NAEP Science assessment and NAEP TEL assessment, including the design of scenario-based tasks in both NAEP assessments and enacted through technology, neither NAEP framework is reflective of the more dramatic shifts found in the NRC framework and NGSS. NAEP TEL focuses on various aspects of technology and engineering literacy and shares certain things in common with the NRC framework and NGSS. In addition, when it was developed and implemented as a technology-based assessment, TEL included more innovative scenario-based item types than the paper-and-pencil NAEP science assessment. The 2019 digitally based NAEP science assessment has moved in a similar direction. Interestingly, when the NRC framework and NGSS were published, NCES leadership often used TEL items as illustrations of performance tasks in NAEP of the type implied by the NGSS, in part because the paper-and-pencil NAEP science assessment did not include such items at the time.

The most significant difference between NAEP science and NAEP TEL and the NRC framework and NGSS is the singular focus of the latter two on the idea of *knowledge in use*—that competence is demonstrated by being able to use DCI and CCC conceptual knowledge in the context of one or more SEPs to solve problems, explain phenomena, and/or design solutions to challenging problems (Harris et al., 2019). Thus, a major concern regarding the future of the NAEP science and TEL assessments is the nature and degree of the alignment between current NAEP frameworks and the NGSS, especially if most states have adopted NGSS or NRC framework/NGSS alike standards and have implemented state assessments aligned with those standards. A related question is whether states, districts, and schools have accordingly modified curricular choices and instructional practices in ways consistent with their own standards (NRC framework or NGSS) and assessments. If a serious misalignment between NAEP science and the science and technology instruction and assessment practiced in schools exists, the validity and value of the NAEP science assessment results for the 2024 or 2028 administrations could be seriously questioned.

The remainder of this section includes the results from a detailed examination of the alignment between each of NAEP science and TEL frameworks with NGSS.[1] These data are critical in thinking about whether changes are needed in NAEP to better align with contemporary U.S. frameworks and standards as well as the extent to which a single assessment framework more like the NGSS would suffice to create a NAEP science and technology assessment rather than two NAEP science and technology assessments as is currently the case. Section III examines the situation with respect to (a) state science standards relative to the NGSS, (b) state science assessments relative to their current standards, and (c) implementation of new science standards in terms of curricular choices and instructional practices in the field.

Comparative Study of the NAEP Science and TEL Frameworks and NGSS

The main purpose of *A Comparison Between the Next Generation Science Standards (NGSS) and the National Assessment of Educational Progress (NAEP) Frameworks in Science, Technology and Engineering Literacy, and Mathematics* (Neidorf et al., 2016) was "to determine the extent to which the NGSS performance expectations are aligned with the content objectives and definitions of practices in the NAEP science and TEL frameworks. An additional purpose was to determine the extent to which the NGSS performance expectations involving mathematics-related practices are aligned with the content objectives in the NAEP mathematics framework." (Neidorf et al., 2016, p. 2).[2]

A comparison of the NGSS with the NAEP STEM frameworks can yield multiple important outcomes with potential implications for a revision of NAEP science and a possible merger of NAEP science and TEL. Neidorf et al. (2016) listed the following (p. 2):

- For the science comparisons, similarities suggest areas where NAEP may provide useful science assessment examples and national achievement data on the student understandings in the natural sciences described in the NGSS. Differences suggest areas where NAEP and NGSS-based science assessments may each provide unique contributions.
- The TEL comparisons augment these findings by identifying additional areas of overlap with the engineering and technology content and practices in the NGSS. Together, these comparisons explore how completely the full range of content and practices in the NGSS are covered by the NAEP science and TEL frameworks as well as the unique aspects of each.
- The mathematics comparisons, while more limited, explore the degree of alignment between the mathematics-related performance expectations in the NGSS and the NAEP mathematics framework. The NGSS are not intended to guide mathematics

[1] The NAEP Science framework and assessment also can be compared to international large-scale science assessment programs in terms of content focus, assessment practices, and future directions. Doing so is beyond the scope of this paper, but for those interested in the PISA and TIMSS science assessment programs, such information is available in a forthcoming chapter on large-scale science assessment (Zhai & Pellegrino, in press).
[2] The Neidorf et al. (2016) study was conducted prior to the adoption of the 2019 math framework for administration in 2026.

assessments, and the performance expectations in science and engineering do not specify explicit mathematics requirements. However, the mathematics students may need to use in responding to items developed to assess these performance expectations can be inferred and compared to the mathematics included in NAEP across grades. Thus, such comparisons can provide information on how assessments based on the NGSS might compare with NAEP in terms of the level of mathematics and quantitative skills that would be required of students.

Three research questions guided this comparison study (Neidorf et al., 2016, p. 3):

1. *Related to the NAEP science framework*: How similar (or different) are the NGSS performance expectations in physical sciences, life sciences, and Earth and space sciences to the content and practices in the NAEP science framework at the corresponding grade levels?
2. *Related to the NAEP TEL framework*: How similar (or different) are the NGSS performance expectations in engineering, technology, and applications of science to the content and practices in the NAEP technology and engineering literacy framework at the corresponding grade levels?
3. *Related to the NAEP mathematics framework*: To what extent are the mathematics-related NGSS performance expectations and practices aligned with the content and skills specified in the NAEP mathematics framework, and at which grade(s)?

Major Findings

The report discusses multiple ways in which the NAEP science and TEL frameworks and the NGSS were compared and contrasted, including different directions and forms of comparison. A plethora of findings are reported and what follows is excerpted from a summary of the major results of those comparisons. It is taken directly from the AIR report.

> There was a moderate to substantial degree of ***content overlap*** between the NGSS and the NAEP science and TEL frameworks. About half of the NGSS performance expectations in the upper elementary grade band (grades 3–5) covered content that overlaps with NAEP science or TEL at grade 4. In contrast, there was much less content in NAEP science that overlapped with the NGSS at grade 4 (and in TEL that overlapped at any grade).
>
> Ninety percent or more of the NGSS performance expectations at the middle school and high school levels covered content that overlaps with NAEP science or TEL at grades 8 and 12, respectively. A somewhat lower, but still substantial, percentage of content in NAEP science at grades 8, and 12 (from 74 to 88 percent) overlapped with the NGSS.
>
> Because of differences in the depth, breadth, detail, or focus of the overlapping content, *content alignment* was lower than *content overlap* when the NGSS was compared to the NAEP science and TEL frameworks together. Moreover, when relevant performance expectations in the natural sciences (physical sciences, life sciences, and Earth and space sciences) and in engineering, technology, and applications of science (ETS) were compared to the NAEP science and TEL

frameworks individually, content alignment differed by grade and by content domain.

Across frameworks, content alignment of the NGSS with the NAEP science and TEL frameworks was moderate. Roughly half of the NGSS performance expectations aligned to NAEP (science or TEL) at each grade level. At grades 3–5, 38 percent of performance expectations were aligned with the science framework and 13 percent with the TEL framework, with 2 percent in the sciences aligned with both NAEP and TEL. At the middle school level, 44 percent of performance expectations were aligned with the science framework and 13 percent with the TEL framework, with 3 percent in the sciences aligned with both. At the high school level, 44 percent of performance expectations were aligned with the science framework and 13 percent with the TEL framework (with no performance expectations aligned with both).

When looking only at the performance expectations in science, the content alignment of the NGSS with the NAEP science framework was low at grade 4 (36 percent) and moderate at the middle school and high school levels (about 50 percent at each grade level). Comparing NAEP science to the NGSS, alignment at grades 4 and 8 was similarly low (23 percent) and moderate (56 percent), respectively; at grade 12, the alignment of NAEP to the NGSS was substantial (71 percent).

Across grades, the greatest degree of alignment between the NGSS and the NAEP science framework was in life sciences and the lowest was in physical sciences, based on the content similarity ratings at both the objective level and at the content area level as a whole. From 48 to 54 percent of NGSS performance expectations in life sciences were aligned with NAEP objectives compared to from 29 to 42 percent of NGSS performance expectations in physical sciences. Looking at the content areas as a whole, life sciences was the only content area rated as similar at two grades (grades 8 and 12) whereas physical sciences was rated as similar only at grade 12, and Earth and space sciences only at grade 8. None of the content areas as a whole were rated as similar at grade 4.

When looking only at the performance expectations in engineering, technology, and applications of science (ETS), content alignment to the NAEP TEL framework was strong for NGSS performance expectations in engineering design (at least 75 percent at each grade level), but weaker for those in the sciences with connections to ETS, especially at the upper grades (as low as 38 percent). The alignment of NAEP TEL with the NGSS, in contrast, was weak at all grade levels, because many more assessment targets are in NAEP TEL as well as assessment areas or subareas that do not have corresponding disciplinary core or component ideas in the NGSS. In addition to engineering design at all three grade levels, both the NGSS and NAEP TEL include the effects of technology on society and the natural world at the middle and high school levels.

The NGSS and NAEP science framework emphasize some content at different grades. That is, some content that was not similar at the corresponding grade level was aligned at a higher or lower grade level in the other framework. In general, the percentage of objectives aligned at a different grade was low—representing no more than one fifth of the objectives. The one exception was for NAEP science at grade 4, where 59 percent of content statements were aligned at a lower of higher grade in the NGSS. The percentage aligned at a different grade decreased over the grade levels for both the NGSS and the NAEP science framework.

Notably, the NGSS and NAEP objectives at middle school/grade 8 that were aligned to other grades were only aligned at the higher grade level in the other framework (high school/grade 12)—i.e., none of the middle school performance expectations were aligned with NAEP grade 4 content statements in science, and none of the NAEP grade 8 content statements in science were aligned with NGSS performance expectations in grades K–5. In addition, some objectives at high school/grade 12 in both the NGSS and NAEP were aligned at the middle school/grade 8 level in the other framework. Thus, the difference between the NGSS and NAEP science framework at grade 8 was more in terms of what content is emphasized in middle school versus high school.

Both the NGSS and the NAEP science and TEL frameworks include objectives at each grade level that cover *unique content*. This reflects nongrouped objectives covering content that is in one framework but not in its counterpart at any grade. (Examples are given in exhibits 10–12 for science and exhibit 13 for TEL). The unique content, together with content that overlapped but was not aligned at any grade in the counterpart framework, represented between 43 and 48 percent of NGSS performance expectations in science and between 18 and 28 percent of NAEP science content statements. Unique content also represented between 14 and 55 percent of NGSS performance expectations in ETS and between 72 and 87 percent of NAEP TEL assessment targets. Unique content reflects areas where each program can contribute different information about student outcomes.

Practices alignment was uniformly strong, but the emphasis of NGSS performance expectations across the NAEP science and TEL practices differed from the emphases specified in the NAEP frameworks.

Ninety-nine percent of NGSS performance expectations in science were aligned with NAEP science practices and 81 percent of performance expectations in ETS were aligned with NAEP TEL practices.

The NGSS performance expectations in science were more strongly concentrated in the NAEP science practice of *using science principles* (60 percent across grades) than was specified in the NAEP science framework (30 to 40 percent across grades). In contrast, very few of the NGSS performance expectations aligned with *identifying science principles* (4 percent across grades) compared to the 20 to 30 percent specified for NAEP across grades. The emphasis on *using scientific inquiry* (22 percent) and *using technological design* (13 percent) was more comparable to NAEP science (30 and 10 percent, respectively, across grades).

> The NGSS performance expectations in ETS were strongly concentrated in the NAEP TEL practice of *developing solutions and achieving goals* (62 percent across grades), which was greater than what is specified in the NAEP TEL frameworks (40 percent across grades). Only small percentages of NGSS performance expectations aligned with NAEP's *understanding technological principles* (12 percent) and *communicating and collaborating* (7 percent) (compared to 30 percent in each practice across grades in NAEP TEL).
>
> **However, despite some strong indications of alignment between the NGSS and NAEP content and practices dimensions separately, when both content and practices were considered together, the NGSS and NAEP science framework were found to be not aligned at the *overall framework level*. That is, at each grade level, the two frameworks were rated as not similar. This was generally because panelists thought that the individual NGSS performance expectations often went beyond what would be expected based on the descriptions of the practices in the NAEP framework when they are applied to specific content statements, even if the science content covered was similar to that in the NGSS. (Neidorf et al., 2016, pp. 94–97, emphasis added)**

Major Conclusions and Implications

The AIR report (Neidorf et al, 2016) also included a set of major conclusions about the relationships among the NAEP science and TEL frameworks and NGSS based on all the various comparisons executed in the study and the judgments made by experts. It focused on implications regarding possible similarities and differences in the demands of assessments aligned to each of the three reference sources. The following is taken directly from the AIR report.

> Together, the results from the various components of the comparison study suggest that NGSS-based assessments and NAEP science and TEL assessments would be aligned to some degree, but each would also have unique content and different emphases in terms of science and TEL practices. This is because some of the grouped NGSS and NAEP objectives with overlapping content—those that were aligned—would likely lead to similar assessment items, but some were different enough that they would likely lead to assessment items with a different content focus. Additionally, those objectives that were not grouped (and either aligned at a lower or higher grade or not aligned at all) would represent unique content at the given grade.
>
> For example, content alignment of an NGSS-based assessment with the NAEP science assessment would likely be low at grade 4—moderate if the entire upper elementary grade band was considered—and moderate at the middle and high school levels. The lower alignment at grade 4 relates to the greater breadth of content in NAEP (evidenced by the greater number of nongrouped objectives) and the fact that some of the content in NAEP at grade 4 may be covered at a different grade in the NGSS's upper elementary grade band.

> An NGSS-based assessment also would likely have a much greater emphasis—over half the assessment—on *using science principles* and a much lesser emphasis on *identifying science principles* than a NAEP science assessment—only 4 percent. This is not surprising given that NAEP explicitly includes declarative knowledge in this latter practice, where the NGSS emphasize the application of science knowledge.
>
> Another implication looking across the study is that the content and practices embodied in NGSS performance expectations that involve engineering design are not fully covered by either the NAEP science or NAEP TEL framework, despite strong alignment with the engineering design assessment targets in NAEP TEL. This includes both performance expectations in engineering design and those in the sciences that involve design applications. Thus, assessment tasks involving engineering design could look quite different in the two programs despite these areas of overlap.
>
> The NAEP science framework—which specifies the practice of *using technological design* (with which many of the NGSS performance expectations in science that involve design applications aligned)—is restricted to the consideration of scientific criteria, constraints, and trade-offs in making design decisions. This is in contrast to the NGSS (and NAEP TEL), which more fully reflect the engineering design process and include a broader range of considerations such as social and economic factors (excluded in NAEP science). Additionally, the NAEP TEL framework and assessments do not expect prior science content knowledge, in contrast to the NGSS, which require the application of science concepts. NAEP TEL, rather, provides the background on the science concepts needed to be successful on the items and tasks measuring the engineering design process.
>
> A final implication is that the tasks that could be developed to assess the NGSS performance expectations in science and engineering would likely require students to use some mathematics that is beyond the corresponding grade level in the NAEP mathematics framework; in contrast, the NAEP science and TEL assessments require mathematics at or below the corresponding grade. In other words, some of the mathematics that could be required in an NGSS-based assessment would be at a higher level than what is required in NAEP science and TEL assessments. (Neidorf et al., 2016, pp. 98–99)

SECTION III. ANALYSIS OF THE NAEP SCIENCE FRAMEWORK AND ASSESSMENT RELATIVE TO STATE SCIENCE POLICY AND PRACTICES: STANDARDS, ASSESSMENTS, AND CLASSROOM INSTRUCTION

This section examines how the NAEP science framework aligns with science standards and assessments that have been adopted and implemented in the states. Three main questions are of interest: (1) Since publication of the NRC framework and the NGSS, how many states have adopted the NGSS or standards that are similar in nature? (2) How do the standards of those states that have not completely adopted the NGSS align with NAEP? and (3) For those states that have adopted the NGSS or similar standards, what is the status of the design and implementation of their state assessments relative to their standards? The section then seeks to establish what the states are doing in the way of instruction as related to the NRC framework and NGSS. It closes with an examination of trends in NAEP science assessment performance between 2009 and 2019 and what those results might imply about the current state-of-science education. Overall, the information provided in this section has substantial implications for considering where states are likely to be in science instruction and assessment by the time the current NAEP science assessment is administered in grade 8 in 2024 and when the updated science assessment is administered in grades 4 and 8 in 2028.

NAEP, NGSS, and State Science Standards Comparisons

Since the publication of the NRC framework and NGSS states, 21 states have explicitly adopted the NGSS as their state science standards and 24 other states have adopted standards that NSTA has designated as partial NGSS in that they are multidimensional standards like the NGSS. In such cases they have based their standards development on the NRC framework and have typically adhered to the central idea of integrated performance expectations based on two or more dimensions as in the NGSS.

In February 2021, HumRRO published a report for NAGB entitled *Comparative Analysis of the NAEP Science Framework and State Science Standards* (Dickinson et al., 2021).

> The method used to conduct this comparative study relied heavily on obtaining experts' judgments regarding the overlap of subject matter between the NAEP science framework and states' science standards…. The comparative analysis included only the standards from states that did not fully adopt the NGSS (i.e., 6 states) and those that partially adopted the NGSS (i.e., 24 states, including the Department of Defense schools). The science standards from the partial NGSS adopting states, which are based on the NRC framework, were included in the study. However, NGSS performance expectations were excluded from the analysis, given the previous study comparing NAEP and NGSS. (Dickinson et al., 2021, p. 1.)

Table 3 below shows which state's standards were included in the analysis.

To execute this analysis. the HumRRO team started by pulling out all content statements, objectives, and performance expectations outside NGSS. The focus was on the content overlap and not the practice overlap. They did some preliminary distillation by matching state and NAEP content statements to look at state and NAEP content side by side to rate the overlap. Also, they identified content-related practices in state statements. They then developed a consensus statement to give the overall impression of where states are doing things differently. They tried to include only statements in the science domains and cut out technology and engineering statements if easy to do so. They did not look explicitly at the TEL framework. An important point to note is that in conducting this work, the comparison of NAEP to state standards is based on an aggregation of all the states' standards rather than a state-by-state individual comparison. Thus, the comparison paints a very broad picture of overlap between the NAEP framework and the partial NGSS and non-NGSS states as a whole. Further details about the methodology and specific sets of outcomes can be found in the complete report.

Table 3. Non-NGSS, partial NGSS, and full NGSS adopting states

Non-NGSS Adopting States	Partial NGSS Adopting States	Full NGSS Adopting States
Florida	Alaska	Arkansas
North Carolina	Alabama	California
Ohio	Arizona	Connecticut
Pennsylvania	Colorado	Delaware
Texas	Department of Defense Education Activity	District of Columbia
Virginia	Georgia	Hawaii
West Virginia	Idaho	Illinois
	Indiana	Iowa
	Louisiana	Kansas
	Massachusetts	Kentucky
	Minnesota	Maine
	Missouri	Maryland
	Mississippi	Michigan
	Montana	Nevada
	North Dakota	New Hampshire
	Nebraska	New Jersey
	New York	New Mexico
	Oklahoma	Oregon
	South Carolina	Rhode Island
	South Dakota	Vermont
	Tennessee	Washington
	Utah	
	Wisconsin	
	Wyoming	

SOURCE: Dickinson et al., 2021, p.12.

The following conclusions, based on the analyses completed by both the HumRRO staff and the outside experts, were offered in the report. They are reprinted here verbatim from that document (Dickinson et al., 2021, pp. 6–7).

1. When examining the content covered by the full set of states' science standards (with any NGSS performance expectations removed), there are many state statements that do not overlap in content with any NAEP statement.
   - At grade 4, 31 percent of all state content statements reviewed by HumRRO experts and external science experts were rated as not overlapping a NAEP content statement.

   - At grade 8, 32 percent of all state content statements reviewed by HumRRO experts and external science experts were rated as not overlapping a NAEP content statement.
   - At grade 12, 55 percent of all state content statements reviewed by HumRRO experts and external science experts were rated as not overlapping a NAEP content statement.
2. Considering only the state content statements that the experts reviewed, all NAEP statements at least partially overlap in content with at least one state statement. In most cases, NAEP statements overlap in content with multiple state statements. Finally, in some cases, NAEP content statements are fully reflected in a combination of multiple state content statements.
   - For each NAEP content statement HumRRO identified multiple state content statements with overlapping content. Review by external experts verified content overlap with at least one of these pairings for each NAEP content statement.
   - Experts noted that there were instances where a combination of state content statements would fully cover the content in a NAEP content statement.
3. Experts rated the least amount of content overlap between NAEP and states' standards at grade 12.
   - Overall, at grade 12, 19 percent of state content statements reviewed by expert panelists were rated as having no content overlap with a NAEP content statement.
4. As with the NAEP-to-NGSS comparison, experts rated the least amount of overlap in content between NAEP and states' standards for the Physical Science domain, especially at grades 8 and 12.
   - At grade 8, 9 percent of state Physical Science content statements reviewed by expert panelists were rated as not overlapping a NAEP content statement.
   - At grade 12, 25 percent of state Physical Science content statements reviewed by expert panelists were rated as not overlapping a NAEP content statement.
5. Science experts identified the grades 4 and 8 state content statements to most frequently reflect NAEP's Identifying Science Practices and the grade 12 state content statements to most frequently reflect NAEP's Using Science Practices. The experts least frequently identified the states' content statements to reflect NAEP's Using Technological Design.
   - At grades 4 and 8, 54 percent of all state content statements reviewed by expert panelists were rated as reflecting NAEP's *Identifying Science Practices*.
   - At grade 12, 51 percent of all state content statements reviewed by expert panelists were rated as reflecting NAEP's *Using Science Practices*.
   - Across the grade levels, between 1 percent and 5 percent of all state content statements reviewed by expert panelists were rated as reflecting NAEP's *Using Technological Design*.
6. Science experts noted that states whose standards are based on the NRC K–12 Science Framework have more in common with NAEP than states whose standards are not based on that framework.
   - Consensus statements developed by both the grade 8 and grade 12 expert panels included assertions that they observed more content overlap between NAEP and the science standards of states who based their standards on the NRC K–12 Science Framework.

State Science Policy and Practices: Standards, Assessments, and Classroom Instruction

Thus far we have established three important findings that bear on a judgment about the validity of results from the NAEP science assessment at the time of its next implementation in 2024 and subsequently in 2028 if substantial revision is not made to both the framework and the derivative assessment before the 2028 administration. First, as described in Section II, major differences exist between the NAEP framework and the NRC *Framework for K–12 Science Education* and the derivative *Next Generation Science Standards* in science content, science and engineering practices, and in their juxtaposition in the form of performance expectations. Second, currently, 45 states (including Department of Defense Education Activity) have either fully adopted the NGSS as their state standards (21) or adopted NGSS-like state science standards (24). Third, when the latter states' standards and those of non-NGSS adopting states (6) are compared with NAEP content, several substantive differences arise. Thus, it seems reasonable to conclude that the current NAEP science framework may be substantially at variance with and lagging a contemporary view of what we want students to know and be able to do in science at grades 4, 8, and 12 and how we would expect them to show proficiency. That view of proficiency has become policy for the preponderance of states and is realized via their state science standards.

How far out of synch the NAEP framework and assessment may be with what instruction and science assessment look like in most states in 2024 and 2028 and with what students know and can do in science depends very much on the following timelines: (a) state adoption of new standards following publication of the NRC framework and NGSS, (b) implementation of new state assessments aligned with those standards, (c) availability of curricular and instructional resources reflecting the new vision of science learning and instruction, and (d) implementation of teacher professional learning programs relative to each of a–c. We provide information relevant to these concerns in the following material.

Time Course for Adoption of New State Standards and Assessments

An article that includes information about adoption of new science standards by Smith (2020) discusses results from the two most recent National Survey of Science and Mathematics Education (NSSME) completed in 2012 and 2018 (see also Banilower et al, 2018). Table 4 shows the pattern of adoption of the NGSS or NGSS-like standards by the states as of 2018. The 16 early adopters did so between 2013 and 2015 while the 24 late adopters did so between 2015 and 2017, and non-adopters had not adopted by spring 2018 when NSSME collected data. Note that there are some differences between Table 4 and the Table 3 shown earlier regarding NGSS adoptions. For example, Florida, North Carolina, Ohio, Pennsylvania, Virginia, and Texas remain nonadopters as of 2021 and they have been joined by West Virginia, which was previously designated as a late adopter. In contrast, Arizona, Alaska, Maine, and Minnesota have moved from the nonadopter group into the late adopter group.

Table 4. Adoption of NGSS or NGSS-like standards – August 2018

Early Adopters	Late Adopters	Non-Adopters
California*	Alabama	Alaska
Delaware*	Arkansas*	Arizona*
District of Columbia	Colorado	Florida
Illinois*	Connecticut	Maine
Kansas*	Georgia*	Minnesota*
Kentucky*	Hawaii	North Carolina*
Maryland*	Idaho	North Dakota
Nevada	Indiana	Ohio*
New Hampshire	Iowa*	Pennsylvania
New Jersey*	Louisiana	Texas
Oklahoma	Massachusetts*	Virginia
Oregon*	Michigan*	
Rhode Island*	Missouri	
South Carolina	Mississippi	
Vermont*	Montana*	
Washington*	Nebraska	
	New Mexico	
	New York*	
	South Dakota*	
	Tennessee*	
	Utah	
	West Virginia*	
	Wisconsin	
	Wyoming	

* Lead state

SOURCE: Data are from Smith, 2020.

One of the many factors driving instructional practice relative to the vision of science teaching, learning, and assessment contained in the NRC framework and state science standards aligned with that vision is the status of each state's large-scale science assessment relative to its adopted standards. Consistent with federal requirements, states that have adopted new science standards are obligated to implement new assessments aligned with those standards having the minimum requirement for at least one assessment in each of the elementary school grade bands (grades 3–5), the middle school grade band (grades 6–8), and the high school grade band (grades 9–12). An analysis for this paper by AIR staff of the 21 states that have fully adopted the NGSS (14 of which are shown as lead adopters in the table above) reveals that all but one of those 21 states, Arkansas, has already developed and in most cases implemented a large-scale science assessment that they claim is aligned with the NGSS. The timeline of assessment implementation varies from 2014 to 2019, with some implementations planned for 2020 but delayed until 2021, given suspension of all large-scale assessments in spring 2020 due to the COVID-19 pandemic. The timelines for implementation of new science assessments for the states classified as partial NGSS are less clear although for the majority of those states their websites indicate that their standards and assessments require integration of the disciplinary core content and practices described in the NRC Framework and many include mention of the third dimension of crosscutting concepts. Some have adopted many if not all the performance expectations from the NGSS. For some states, the timeline for full implementation of new assessments extends to 2025.

Survey Information on Science Instructional Practices: 2018 vs. 2012

NSSME has provided periodic snapshots of K–12 science instruction in the United States for more than 40 years. Study topics include teacher backgrounds and beliefs, professional learning opportunities, course offerings, instructional objectives and activities, resources for instruction, and policies affecting instruction. The two most recent studies were conducted in 2012 and 2018. The 2012 study provides baseline data on multiple indicators prior to publication of the NGSS. From 2013 to 2018, 39 states and the District of Columbia adopted the NGSS or NGSS-like standards. By the time the 2018 survey was conducted, NGSS states accounted for more than two thirds of the nation's K–12 students. The 2018 study provides a snapshot of the state-of-science instruction in 2018 relative to the vision of the NRC framework and the NGSS, including the opportunity to observe any impact on instructional beliefs and practices relative to 2012 in light of the publication of the NRC framework in 2012 and the NGSS in 2013.

Smith's 2020 analysis and discussion of results from the 2018 NSSME (Banilower et al., 2018) shows that states have been slow in the full implementation of their new science standards in terms of making a difference in instructional practice. As discussed by Smith, one reason for the slowness is the lack of good curriculum materials aligned with the new standards. Another reason for the slowness is the need for substantial teacher professional development related to understanding the science and engineering practices as well as the meaning and manifestation of integration of the multiple dimensions expressed by the performance expectations. Related to the latter, valid, high-quality assessments reflecting the kinds of performances expected from students also have been lacking. In general, during the period in question there was a paucity of such examples for classroom use as well as at the large-scale state assessment level given the timeline for implementation of new NGSS-aligned assessments as described above from the analysis of state websites by AIR staff.

Regarding professional development, Smith (2020) reports that roughly four of five secondary science teachers (i.e., middle school and high school) participated in science-focused professional development in the preceding 3 years, in contrast to three of five elementary science teachers. Only about half of schools or districts offered any science-focused professional development in the preceding 3 years, and participation data were largely unchanged since 2012. About a third of secondary teachers participated in more than 35 hours of professional development in the 3 years preceding 2018, and more than 4 in 10 elementary teachers had none. As Smith notes, even 35 hours, spread over 3 years, is not much considering prominent instructional practices and the shifts that the framework and NGSS entail.

Among the other results summarized by Smith were results regarding data on instructional practices and emphases in elementary, middle school, and high school classrooms (see Smith 2020, Table 1). Most importantly, in 2018 the most frequent "heavy emphasis" instructional objective reported by Science teachers was "understanding science concepts," particularly in middle and high schools (47 percent of Science teachers in elementary schools, 77 percent in middle schools, and 76 percent in High schools). In contrast, the second most frequent objective with a heavy emphasis reported by teachers was "learning how to do science" but only in 26 percent of Science classes in elementary schools, 46 percent in middle schools, and 41 percent in High schools. Smith concluded that:

> Despite widespread adoption of the NGSS and NGSS-like standards, data from the NSSME+ point to few differences in science instruction compared to 2012. Further, the data from teachers in adopting states vary little from those in non-adopting states. Among the few differences, we do see encouraging signs. Among them, classes in adopting states were more likely to emphasize learning how to do engineering, and they were less likely to emphasize learning vocabulary and facts. In terms of instructional activities, classes in early-adopting states were less likely to rely on lecture and more likely to have students do hands-on activities. However, the data overall suggest that much work lies ahead to achieve the vision laid out in the framework and the standards themselves (Smith, 2020 p. 608).

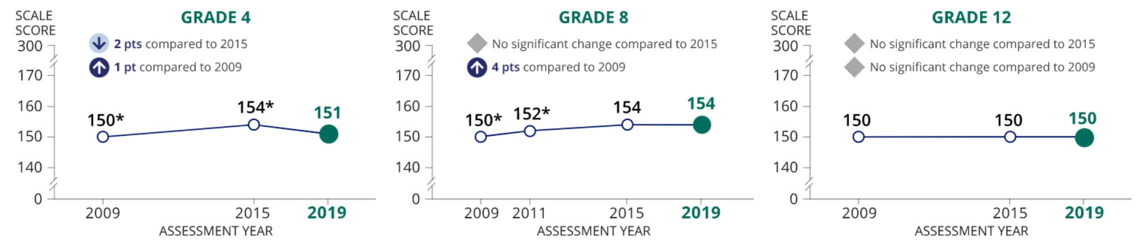
Perhaps not surprising is that substantial changes in science instructional practices were not observed in the 2018 NSSME survey relative to 2012 and that aspects of the vision for science teaching and learning embodied in the NRC framework and NGSS were less well represented in teacher beliefs and instructional practices. As noted by Smith (2020), 5 years may not be enough time. Many of the critical factors needed to spur change are only now becoming more prominent with further changes on the horizon during the next 2 years when NAEP science is set to be administered again for grade 8 only. Among the drivers of change are new state science assessments reflecting the NGSS or similar science standards. In addition, growth in both commercially available and open education resources (OER) aligned with the NGSS has been significant. One of the largest of the OER curricular initiatives is the foundation-funded OpenSciEd project (https://www.openscied.org/about/), which has generated instructional units covering all the middle school NGSS performance expectations and is working on similar materials for other grade levels. At the classroom level, assessment resources have been developed to support formative and summative assessment practices in ways aligned with the multidimensional assessment vision described in the 2014 NRC report, *Developing Assessments for the Next Generation Science Standards* (Pellegrino et al., 2014). See for example the materials available from the *Next Generation Science Assessment Project* (http://nextgenscienceassessment.org) and from the Stanford *NGSS Assessment Project* (https://scienceeducation.stanford.edu/assessment).

NAEP Science Performance Changes Over Time

One final source of information about possible changes in science education in the United States over time might be gleaned from an examination of performance on the NAEP science assessment for the period from 2009 when the new science framework and assessment were first implemented to 2019 when NAEP science was delivered as a digitally based assessment, in contrast to prior years. These data track student performance both before and after the NRC framework and NGSS.

The 2019 NAEP science scale score results are shown in Figure 4 for each of the grade levels in comparison to prior administrations back to 2009. As can be seen in Figure 4, the average science score for the nation at grade 4 was lower by 2 points compared to 2015, whereas average scale scores at grades 8 and 12 did not significantly differ from 2015. At grades 4 and 8, average scale scores were higher when compared to 2009, while the average scale score at grade 12 was not significantly different across years.

Figure 4. Average scores in NAEP science, by grade: 2009–2019

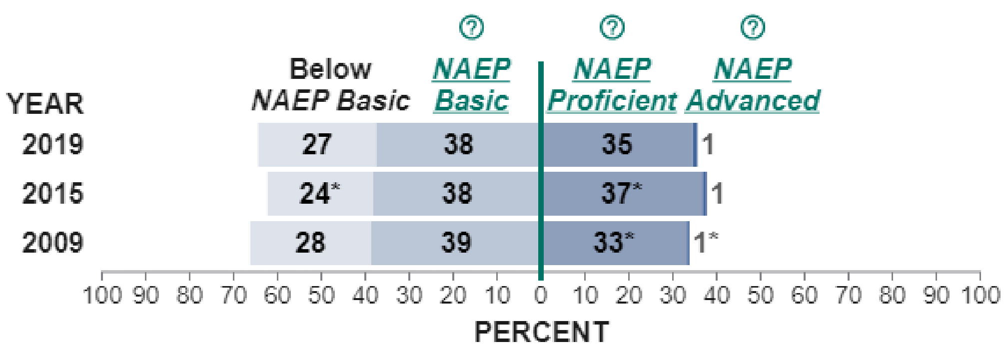


*Significant different ($p < .05$) from 2019.

SOURCE: The Nation's Report Card, 2019. Reprinted with permission.

Although the absolute levels of the scale scores and the trends in those scores are important indicators of student performance, of particular significance is the reporting of results in terms of achievement levels. As shown below in Figures 5, 6, and 7, the rates by which students were classified into the achievement levels varied across the grades with the highest rate of *Proficient* classifications occurring in grade 4, slightly lower levels of proficiency at grade 8 and substantially lower student proficiency classifications at grade 12. Note that at all three grade levels, there is a very low level of classification of student performance at the *Advanced* level. This finding holds across years.

Figure 5. NAEP achievement-level results in NAEP science for fourth-grade students: 2009, 2015, and 2019

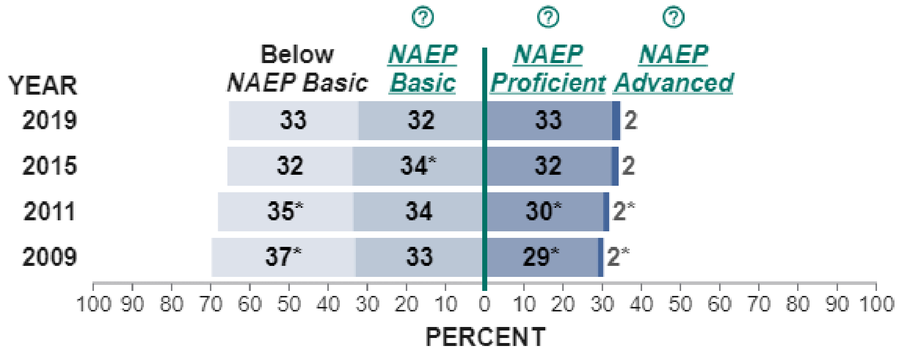


* Significantly different ($p < .05$) from 2019

Note: NAEP achievement levels are to be used on a trial basis and should be interpreted and used with caution.

SOURCE: The Nation's Report Card, 2019. Reprinted with permission.

Figure 6. NAEP achievement-level results in NAEP science for eighth-grade students: Various years, 2009–2019

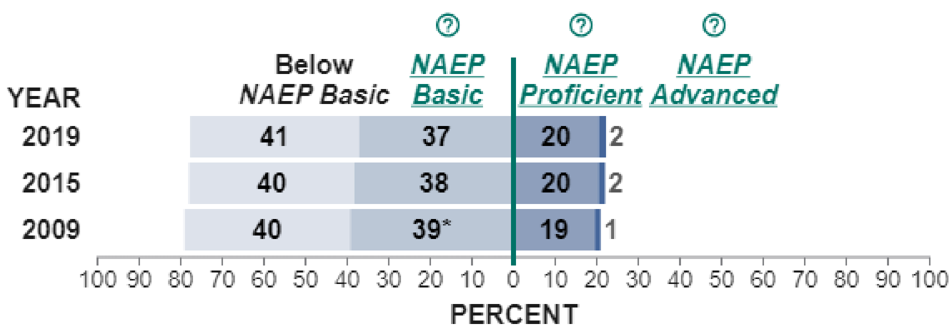


* Significantly different (*p* < .05) from 2019.

Note: NAEP achievement levels are to be used on a trial basis and should be interpreted and used with caution.

SOURCE: The Nation's Report Card, 2019. Reprinted with permission.

Figure 7. NAEP achievement-level results in NAEP science for twelfth-grade students: 2009, 2015, and 2019



* Significantly different (*p* < .05) from 2019.

Note: NAEP achievement levels are to be used on a trial basis and should be interpreted and used with caution.

SOURCE: The Nation's Report Card, 2019. Reprinted with permission.

Perhaps there are two major takeaways from this examination of the NAEP science assessment results. First, not much has changed over time implying that science instruction also has not changed substantially despite the existence and adoption of new standards with higher expectations about what students are supposed to know and be able to do. Despite their differences in content and format of science assessment, the most recent trend results from the PISA science assessment and the TIMSS science assessment largely corroborate the lack of change in U.S. science performance during the last decade. Second, those new standards are much needed because science performance across the grade bands is relatively poor and only declines across grades. The vast majority of students are below *Proficient* as defined by the NAEP achievement levels.

The real concerns then are threefold: (1) whether instruction aligned with the new standards will take hold in ways envisioned by the NRC framework and NGSS and change performance, (2) whether the NAEP science assessment can track the impact of those changes given the differences between the NAEP framework, the NGSS and the majority of state science standards, and (3) whether NAEP science and/or TEL will have sufficient instructional sensitivity to reveal what has and has not happened over time when next administered in 2024 or 2028.

SECTION IV. TECHNOLOGY IMPLICATIONS FOR NAEP SCIENCE

This section briefly considers how various developments in digital technologies need to be considered in reviewing the existing NAEP science framework and assessment and envisioning possibilities for their updating. The discussion that follows focuses on the affordances of technology regarding the constructs that could be included in a revised framework and the associated task design, data capture, and data analytic issues involved in an assessment aligned to an updated framework. The section concludes with a brief discussion of practical and equity concerns related to digitally based assessment of science and technology proficiency.

Technology and NAEP Assessment

During the last two decades, much has been written and speculation made about the power of technology to both improve and transform assessment across a range of assessment contexts and purposes (e.g., Behrens et al., 2019; Bennett, 2008; Drasgow, 2016; Gane et al., 2018, Pellegrino & Quellmalz, 2010; Pellegrino et al., 2001). Although technology's potential for improving and transforming assessment has yet to be fully realized, the vast majority of national-, international-, and state-level assessments of science and technology have moved almost entirely to digital presentations of materials accompanied by technology-based data capture for purposes of scoring, analysis, and reporting. Within the past decade, PISA (2015, 2018), eTIMSS (2019), NAEP Science (2019), and NAEP TEL (2014, 2018) have been delivered via technology using various types of devices including laptops, tablets, and desktops.

Not only has technology changed assessment delivery, response capture, and scoring, it also has had a significant effect on assessment design. This includes the types of tasks and situations that can be presented to students with the goal of tapping into various forms of scientific thinking and reasoning aligned with the practices of science and engineering as found in the NAEP science and TEL frameworks and NGSS. For the NAEP program, some of the newer task types that take advantage of some of technology's affordances were briefly described in Section II, including the scenario-based tasks added to the NAEP science assessment in 2019. The latter were modeled to a great extent after the digitally based tasks were first introduced in NAEP TEL in 2014. The literature on NAEP has considered a number of the affordances of technology for the assessment program, including implementation and analysis of the types of scenario-based tasks in science piloted by NAEP in 2015 and included as part of NAEP 2019, including analyses of student response data (e.g., Bennett, 2008; Bergner & von Davier, 2019; Duran et al., 2020; Lee at al., 2019; Mullis, 2019). The purpose of the discussion that follows is to briefly highlight some of the possibilities for the future of NAEP science as related to both the framework and the assessment.

Opportunities and Possibilities for NAEP Science

As discussed in prior sections of this paper, conceptions of scientific and technological competence have evolved during the last 10–15 years, some of which align with the current NAEP framework and assessment while others go beyond both. Thus, in considering possible changes for the design of the 2028 administration of the science assessment, it will be important to consider how some of the affordances of technology discussed below may influence the nature of the competencies included in the framework, the design of the assessment tasks needed to provide evidence of those competencies, and the associated measurement and interpretive challenges, especially in light of goals for reporting the results. The assessment as evidentiary reasoning argument presented in the NRC report *Knowing What Students Know: The Science and Design of Educational Assessment* (Pellegrino et al., 2001) frames the discussion. In Chapter 7 of that report many of the affordances of technology for advancement of assessment design and practice are discussed in terms of the three interconnected components of the assessment triangle: *Cognition*, *Observation,* and *Interpretation*. As argued in that report:

> The role of any given technology advance or tool can often be differentiated by its primary locus of effect within the assessment triangle. For linking *cognition* and *observation*, technology makes it possible to design tasks with more principled connections to cognitive theories of task demands and solution processes. Technology also makes it possible to design and present tasks that tap complex forms of knowledge and reasoning. These aspects of cognition would be difficult if not impossible to engage and assess through traditional methods. Related to the link between *observation* and *interpretation*, technology makes it possible to score and interpret multiple aspects of student performance on a wide range of tasks carefully chosen for their cognitive features, and to compare the resulting performance data against profiles that have interpretive value. (Pellegrino et al., 2001, p. 252)

The discussion that follows elaborates on these general ideas regarding NAEP science. It focuses is on the constructs that could be represented in an updated framework, the ways in which those constructs could be realized in the assessment environment, and some of the interpretive challenges and solutions associated with doing so for purposes of measurement and reporting.

The *Cognition* vertex of the assessment triangle. What matters in assessment is what we are trying to reason about – the contemporary conception of student *Cognition* in a domain like science that matters to scientists, educators, and society. A contemporary view of multidimensional proficiency in science includes the expectation that learners should be able to use their disciplinary core knowledge to engage in a variety of science practices in the service of explaining phenomena and designing solutions while answering challenging questions (NRC, 2012). As the conception of student cognition changes and expands in terms of what students are supposed to know and be able to do, as has been the case for science, technology affords opportunities for substantially changing and extending the *Observation* and *Interpretation* components of the assessment triangle in order to more adequately represent and provide evidence about the constructs of interest. Doing so enhances the entire evidentiary reasoning process and the validity of the NAEP science assessment given its intended interpretive use as an index of trends in U.S. science achievement.

The *Observation* vertex of the assessment triangle. Technology provides opportunities for presentation of dynamic stimuli (e.g., videos, graphics, 2- and 3-D simulations) that can be interacted with in the service of eliciting relevant sets of responses from students. Simultaneously, technology enables the generation and capture of a variety of response products, including situations in which students generate responses using multiple modalities (e.g., drawing and writing). In general, *technology-enhanced assessments* are defined by their capacity to provide novel stimuli and/or responses that would not be possible with traditional, paper-and-pencil assessment formats. Technology-enhanced assessments such as those included in NAEP science 2019 and NAEP TEL enable engagement with a variety of science and engineering practices (e.g., generating models, planning and carrying out investigations, engaging in computational thinking) by opening the door to interactive stimulus environments and response formats that better match the intended reasoning and response processes that form the basis for desired claims about student proficiency (Gorin & Mislevy, 2013).

Students' interactions with these technology-enhanced assessments can be logged to provide data on how they engage in particular processes. In certain applications such as engineering or experimental design, the process by which one completes the activity can be as important a piece of information about knowledge and skill as the final product. In these cases, understanding the operations that students performed in the process of creating the final product may be critical to evaluating students' proficiency. Log data offer the opportunity to reveal these actions, including where and how students spend their time, and what choices they make in situations like using a simulation. Such applications offer the potential to provide large volumes of "click-stream" and other forms of response process data that might be useful for inferences about student thinking as discussed by Ercikan and Pellegrino (2017). Such data can be complex, however, and must be segmented and analyzed in construct-relevant ways if they are to be reliable and valid for a given interpretive use. An ongoing challenge is identifying how to take massive volumes of log data and distill it into actionable information to make judgements about students' knowledge, skills, and abilities (e.g., Bergner & von Davier, 2019).

The *Interpretation* vertex of the assessment triangle. Technology offers significant opportunities for enhancement of the reasoning-from-evidence process given the types of observations described above. Collecting the types of data just mentioned in the discussion of observations makes little sense unless there are ways to reliably and meaningfully interpret them. This can evolve through mechanisms such as automated scoring of responses and application of complex parsing, statistical and inferential models for response process data. Much has been written recently about the opportunities of student-response-process data for capturing what students are doing when they solve problems and answer questions related to science and technology (see Ercikan & Pellegrino, 2017). Such data include the time taken to perform various actions, the actual activities chosen, and their sequence and organization. The potential exists for examining the global and local strategies students use while solving assessment problems and the implications, including how such strategies relate to the accuracy or appropriateness of final responses. Although capturing such data in a digital environment is "easy," making sense of the data is far more complicated. The same can be said for capturing data to constructed response questions where students may be expressing in written and/or graphical form an argument or explanation about some scientific problem or phenomenon, describing the design of a scientific investigation, or representing a model of some structure or process.

The data capture contexts described above are challenging regarding scoring and interpretation. It is here that AI and machine learning may play a significant role in future science assessments. Machine learning mimics human scoring processes by first "learning" from scoring by human experts to develop algorithmic models and then applying those models to automated scoring of new student responses (Zhai, Yin, et al., 2020). Advances have been made in the automated scoring of short, written, constructed responses for various topics and content in science and other subjects (see Beggrow et al., 2014; Nehm et al., 2012; Williamson et al., 2012). However, automated scoring of other types of constructed response products, such as the features that might be included in drawings and other forms of graphical representation associated with a practice like modeling, has not yet been explored in-depth (see Gerard et al. [2016] for one promising attempt). For both written and graphical responses, well-designed task models that define the features of responses that matter for scoring are needed. This likely will have a considerable impact on the development of automated scoring systems that are both reliable and practical for implementation across a variety of assessment contexts.

Developments in machine learning also may allow researchers to analyze complex response process data of the type described above (Zhai, 2021). Traditional statistical methods are often difficult or inappropriate to apply to such data. Machine learning, however, might assist in analyzing these types of data to reveal patterns that provide important insights into students' cognitive processes in problem solving (Zhai, Haudek, et al., 2020; Zhai, Yin, et al., 2020). Such data may prove to be especially informative about student thinking and reasoning and thus add greatly to the knowledge gained about student competence from large-scale assessments like NAEP that go beyond the performance accuracy data they now provide. An interesting example was provided in a recent study by Pohl et al. (2021). The authors showed that differences in student response processes, of the type described above, when combined with scoring methods, can significantly change the interpretation of a country's performance on a large-scale assessment such as PISA. Their study findings showed that current reporting practices in PISA confound differences in test-taking behavior with differences in competencies and can do so in a different way for different examinees, threatening the validity and fairness of comparisons. Thus, their argument is that test-taking behavior is not a confounding factor introducing construct-irrelevant variance, but that it is something that provides important information on how examinees approach tasks, which can be meaningful outside the testing situation. Disentangling and reporting all these factors as part of a performance portfolio could result in fairer comparisons across groups and enables a better understanding of student competencies and important possible causes of variations in performance. Explorations of the analysis and interpretation of response process data have been initiated for some of the NAEP science tasks (Bergner & von Davier, 2019; Lee at al., 2019) and the results suggest that this is a fertile area for future exploration, albeit taking into consideration some of the cautions mentioned below.

Areas of Concern for NAEP Science

Assessments that can tap into and measure multidimensional knowledge take the form of *knowledge-in-use* tasks (Harris et al., 2019). Technology can make practical the design, administration, and scoring of such tasks. An area of concern is that technology by itself is not enough: Technology cannot fix assessments that are poorly designed or misaligned with the desired learning targets. Instead, technology considerations need to be integrated with assessments through a transparent and principled design process. As the targets of assessment become more conceptually complicated, with demands such as jointly measuring science practices and conceptual knowledge, a principled design process is essential for developing relevant and valid assessment tasks (Gorin & Mislevy, 2013; Pellegrino et al., 2014). A principled design process like *Evidence Centered Design* (Mislevy, 2018; Mislevy & Haertel, 2006; Mislevy & Riconscente, 2006) that identifies task and response features that matter can also move the scoring process from a black box statistical approach to one that is more transparent and defensible. Explicit task and response models with defined response features can lead to improved human scoring as well. A caveat, in a general sense, for NAEP science is that if NAEP wants to capture more complex forms of scientific thinking and reasoning using digital environments, this cannot be done by simply applying technology to the sense-making process "after the fact," which seldom is well done or efficient. Thus, a very deliberate design process needs to be used for task design and data capture that takes into consideration the relevant forms of evidence and the means for interpretation of that evidence throughout the task design, task refinement, and task validation processes.

Although technology can enhance many aspects of large-scale assessment, concerns have arisen about the equity and fairness of digitally based assessment. An area of concern is comparability of results and validity of inferences derived from performance obtained across different modes of assessment, especially for varying groups of students (see Berman et al., 2020). As NAEP science has moved from paper-and-pencil assessment to digitally based assessment, the general focus has been on mode comparability and concerns about student familiarity and differential access to the hardware and software used (see Way & Strain-Seymour, 2021). As the digital assessment world advances, a significant issue for future large-scale science and technology assessments is determining how student background characteristics including language, culture, and educational experience influence performance on different types of tasks and innovative assessment designs that leverage the power of technology. As the tasks become more innovative, equity and fairness concerns may become even more important than general mode comparability effects.

Another area of concern relates to cost, efficiency, and feasibility. Complex, scenario-based tasks such as those found in NAEP science and TEL are challenging to design well and costly to create relative to more conventional tasks. They typically also take significant amounts of time for students to complete. Given the nature of the scenarios, they also tend to be memorable because they depict interesting, engaging, and often realistic problem-solving situations. They exemplify and perhaps magnify many of the challenges that have long been noted about the inclusion of performance tasks in large-scale testing programs such as NAEP. Davey et al. (2015) provided an excellent discussion of the many challenges associated with development and deployment of performance assessments for constructs represented in science standards such as the NGSS. Their report included a discussion of many of the measurement and statistical challenges associated with the interpretation and reporting of performance data. Thus, NAEP science will have to consider tradeoffs associated with inclusion of technology-based assessment tasks relative to adequate representation and sampling of the constructs of interest. The fact that NAEP science uses a matrix-sampled block design for selection and administration of tasks may mitigate some of the many concerns noted by Davey et al. (2015). NAEP can offer leadership to the large-scale science assessment field in providing a vision and examples of how science and technology competence can and should be assessed and reported.

SECTION V. CONCLUSIONS AND RECOMMENDATIONS

The purpose of this white paper is to consider the need for a revised NAEP science framework and its possible scope and focus including expansion to aspects of what is represented in NAEP TEL. The goal is to provide the NAEP NVS Panel and the NAEP program input about possible futures for NAEP science. As such, the paper can also serve as input to NAGB's deliberations in 2022 about the need and possible directions for a revision of the science framework that would in turn serve as the basis for development of the NAEP science assessment scheduled for 2028.

Topics Covered Across Sections I–IV

- A brief history of the current NAEP science and TEL frameworks and assessments and their projected use over the next seven years through 2028
- Brief descriptions of the content and focus of the NAEP science and NAEP TEL frameworks and assessments as well as the National Research Council's *Framework for K–12 Science Education* (NRC, 2012) and the derivative *Next Generation Science Standards* (NRC, 2013)
- Results from an extensive comparison of the content and focus of both NAEP frameworks with the NGSS
- Information on the timeline and status of state adoptions of the NGSS or similar science standards derived from the NRC framework
- Results from a study comparing the content of state science standards with the NAEP science framework for states with science standards similar but not identical to the NGSS together with states with standards unrelated to the NGSS or NRC framework
- Information about the status of development and implementation of standards-aligned, large-scale state science assessments for those states that have either adopted the NGSS or similar standards
- Information about the conditions of science instruction based on the 2012 and 2018 National Survey of Science and Mathematics Education
- Trends in NAEP science assessment performance for 2009–2019 for students at grades 4, 8, and 12
- A discussion of the affordances of technology for consideration in refinements and revisions to the NAEP science framework and assessment

Conclusions and Implications

Alignment of NAEP Science and NAEP TEL With Other Frameworks and Standards

The frameworks for NAEP science and NAEP TEL were developed before the NRC framework and NGSS and all within a window of approximately 6–7 years. All four drew upon bodies of theory, research, and practice regarding the knowing, learning, and teaching of science and technology available at the time of their development. Given time lags among them, it should come as no surprise that there are both significant similarities between the two NAEP frameworks and the NGSS and substantial differences as determined by a 2016 AIR comparison study (Neidorf et al., 2016).

Conclusion 1. Overlap exists between NAEP science and NGSS in terms of the focal science content areas—physical science, life science, and Earth and space science—and subtopic areas within each domain, but substantial differences exist in specific content. The differences are magnified in the movement from grade 4 to grade 8 to grade 12. One reason for the pattern of differences across grade levels is that the NGSS is based on a set of four disciplinary core ideas (DCIs) in each domain of science, and each DCI is elaborated across grades in terms of knowledge expectations. This was a deliberate design decision in the NRC framework that is replicated in the NGSS. In contrast, the NAEP framework changes content emphasis and focus across grades 4, 8, and 12 with an increasing emphasis on physical science content at grades 8 and 12, especially at grade 12.

Conclusion 2. Overlap exists between the NAEP framework and NGSS regarding the concept of science practices that describe ways of thinking about and reasoning with science content. The NAEP science practices and the NGSS science practices are different in at least two ways, however. Two of the four NAEP practices are considered to be more focused on "knowing science" in contrast to the other two that are more focused on "doing science." In contrast, the NGSS includes eight specific science and engineering practices, each of which fall under the category of science inquiry ("doing science") and/or engineering design. In general, the NGSS science and engineering practices are more demanding than at least two of the NAEP practices, and this is especially apparent when the practices are combined with content to form performance expectations as noted below.

Conclusion 3. Although both NAEP and NGSS express the targeted knowledge and skills for students in the form of performance expectations, the NGSS performance expectations are considered to demand much more in the way of application of disciplinary content knowledge to answer a question involving a science practice to demonstrate proficiency. Regarding the latter point, the 2016 AIR comparison study concluded: "… despite some strong indications of alignment between the NGSS and NAEP content and practice dimensions separately, when both content and practices were considered together, the NGSS and NAEP science framework were found to be not aligned at the *overall framework level*. That is, at each grade level, the two frameworks were rated as not similar. This was generally because panelists thought that the individual NGSS performance expectations often went beyond what would be expected based on the descriptions of the practices in the NAEP framework when they are applied to specific content statements, even if the science content covered was similar to that in the NGSS" (Neidorf et al., 2016, p. 97).

Conclusion 4. The NGSS includes a fourth dimension in its content framework—engineering, technology, and the applications of science as well as two engineering practices—defining problems and designing solutions. The AIR comparison study (Neidorf et al., 2016) showed that the NGSS has overlap with both NAEP science and NAEP TEL with respect to certain aspects of engineering, technology, and design. The overlap is highly variable, however, depending on grade level and direction of comparison. A significant difference between NGSS and TEL is that NGSS performance expectations related to technology and design require science content knowledge, which is not true of the TEL assessment that provides relevant science content in the task situation.

Conclusion 5. Given differences between NAEP science, NAEP TEL, and the NGSS in terms of content, practices, and performance expectations, the AIR study (Neidorf et al., 2016) concluded that an assessment aligned to the NGSS could look substantially different from assessments aligned with either NAEP science or NAEP TEL. Much of this difference is associated with the demands of the NGSS performance expectations for science DCIs, as noted above. The same concern applies to performance expectations for the DCI designated as engineering, technology, and applications of science as well as performance expectations involving the engineering practices when combined with science disciplinary content. For the most part, the NGSS performance expectations likely would lead to more challenging assessment tasks than those found in either NAEP science or NAEP TEL.

Status of State Science Standards, Assessments, and Instruction

Given substantial differences between the NAEP science and NAEP TEL frameworks and the NGSS, an obvious question is the degree to which states have adopted the NGSS or similar standards and the status of implementation of policies and practices associated with those standards. Included among the latter is implementation of state large-scale assessments aligned to their current standards. A related concern is penetration of the NRC framework's vision for science learning, teaching, and assessment at the level of classroom practice. Such information has implications for the validity of results from the NAEP science assessment when it is re-administered in grade 8 in 2024 and when an updated science assessment is administered in grades 4 and 8 in 2028.

Conclusion 6. Currently, 45 states (including the Department of Defense Education Activity) have either fully adopted the NGSS as their state standards (21) or adopted NGSS-like state science standards (24; Dickinson et al., 2021). These states represent a substantial proportion of the total U.S. student population across grades K–12. When the standards of states that have adopted NGSS-like standards (24) and those of non-NGSS-adopting states (6) are compared to the NAEP framework based solely on content, several differences arise. Such differences are not surprising given that standards based on the NRC framework are likely to show results that are highly similar to those obtained directly from comparison of content from the NAEP science framework with the NGSS. As mentioned above, the NRC framework and NGSS include a specific set of disciplinary core ideas that remain constant across grade levels while growing in depth and sophistication. State standards based on the NRC framework are likely to show the same pattern of content similarities and dissimilarities with NAEP within and across grades that were revealed in the AIR study (Neidorf et al., 2016) comparing NAEP and NGSS. Results reported in the HumRRO 2021 study of state content standards vis-à-vis NAEP are very similar in that regard (Dickinson et al., 2021). The implication is that at least at the policy level, significant differences exist between NAEP's view of science proficiency and its assessment and the view that has become policy for the preponderance of states and realized via their officially adopted state science standards. Given state science standards adoptions, the current NAEP science framework and assessment may be substantially at variance with a relatively pervasive national perspective on what is desired for students to know and be able to do in science at grades 4, 8, and 12 and how they could be expected to show proficiency via large-scale assessment.

Conclusion 7. The pace at which standards reflecting the NGSS or the NRC framework affects classroom teaching, learning, and assessment has been slow, perhaps not unexpectedly. Evidence shows that adoption of the new standards has been staggered across time since 2013, as has been the design and implementation of state large-scale assessments aligned to those new standards. The latter invariably lag two or more years behind adoption of new state standards. The most recent national survey of science education conducted suggests that little has changed between 2012 and 2018 in science instructional practice (Smith, 2020). Results from the NAEP science assessment from 2009 to 2019 also show little in the way of change in student performance across time (The Nation's Report Card, 2019). One major factor in the slow penetration at the classroom level appears to be limited availability and implementation of professional learning programs for teachers. Although state implementation of large-scale assessments aligned with the NGSS or NRC framework has progressed, and classroom instructional and assessment resources aligned with the NRC framework's vision of teaching, learning, and assessment have become more readily available, the current and future state of classroom practice remains to be determined. Regarding the latter, the National Academies of Science, Engineering, and Medicine (NASEM) is convening a two-day summit in October 2021 at which time the status of implementation of science standards with a focus on areas where additional work may be needed will be discussed. In summary, how far out of alignment the NAEP science framework and assessment may be with science instruction and assessment in most states in 2024 when the current assessment is to be used remains to be seen. It seems reasonable to conclude, however, that significant differences likely will exist in 2028 if the NAEP science framework and assessment are not updated and revised.

Technology and NAEP Science

Conclusion 8. Technology already has had a substantial impact on the NAEP program—and particularly on NAEP science. Both NAEP science and NAEP TEL currently are delivered as digitally based assessments and include new types of tasks that take advantage of some of the affordances of technology for task design, presentation and interaction, data capture, scoring, and analysis. Possibilities exist for capitalizing on the multiple affordances of technology in updating and revising the NAEP science framework and assessment. These include consideration of additional science and technology proficiencies that should be included in the framework, the capacity for their realization in the assessment in the form of tasks and situations that require particular forms of scientific and engineering reasoning, and opportunities for analysis and reporting of those proficiencies in ways that go well beyond overall accuracy. In general, innovative uses of technology offer NAEP science the possibility of leadership in the large-scale science assessment field by providing a vision and examples of how science and technology competence can and should be assessed and reported. Further movement in this direction must take into consideration design and analytic challenges together with equity, cost, and feasibility concerns.

Recommendations

Given the findings described, serious concerns exist about the capacity of the NAEP science assessment to fulfill its mission to provide valid and reliable information about the status of science achievement in the United States in 2028 and beyond unless a detailed review and revision of the NAEP science framework is recommended by NAGB in 2022 and then pursued by an appropriate framework visioning panel followed by a framework development panel.

The major threat to the validity of NAEP science involves adoption by a preponderance of states of science and technology education standards that differ substantially from the NAEP science framework. Assuming continued implementation of assessments, curriculum materials, instructional practices, and professional learning opportunities aligned with those standards, whether the NAEP science assessment can track the impact of those changes on science achievement and whether NAEP science and/or NAEP TEL will have sufficient instructional sensitivity to reveal what has and has not happened over time when administered in 2028, and even quite possibly beforehand in 2024, is questionable.

Two broad recommendations consistent with these concerns and the related findings contained in this paper follow. For each recommendation, additional commentary is provided regarding matters that should be considered in acting upon each recommendation.

Recommendation 1

The NAEP Validity Studies Panel recommends that the NAEP science framework should be reviewed and revised to reflect contemporary changes in science standards, instruction, and assessment.

In reviewing and suggesting revisions to the science framework:

A. The panels should consider the distribution and focus of the content included in the framework regarding two factors. The first factor involves consideration about whether there should be continuity in the content foci within each domain of science across the grades, in ways similar but not necessarily identical to the disciplinary core ideas in life science, physical science, and Earth and space science described in the NRC framework. The second factor is related to the first and involves the specific set of topics included in each domain and across grades. A shift to this organization of content may allow the NAEP science assessment to provide important trend information across grades in the development of core knowledge in prioritized areas of each of the three major science disciplines.

B. The panels should consider NAEP's current science practices relative to a set of science and engineering practices that may be most important for students to understand and use. Such practices should be articulated in the framework as well as their implications for assessment at each grade level and across grades. Such a consideration includes the extent to which they emphasize active engagement with science and engineering practices, as articulated in the NRC framework, that is, the doing of science and engineering, when applied to science content rather than just knowing about those practices but not necessarily being able to use them.

C. The panels should consider the meaning of science proficiency and how that is expressed via performance expectations that integrate content and practice knowledge consistent with the separate but related considerations of science and engineering content and practices discussed above. Particular attention needs to be given to the demands of those performance expectations and how they could be represented in assessments that make use of the affordances of technology.

D. The panels should consider the inclusion of technology and engineering content and practices, similar to their inclusion in the NRC framework and NAEP TEL. Further comments on technology and engineering in the NAEP science framework are included below under Recommendation 2.

E. The panels should gather the most recent information on the status of implementation and impact of current state science standards and projections for the remainder of this decade. The panels should seek information on these matters from the Board on Science Education from NASEM, the National Science Teacher Association, the Council of State Science Supervisors, the Science SCASS of the Council of Chief State School Officers, and the American Association for the Advancement of Science.

Recommendation 2

The NAEP Validity Studies Panel recommends that in reviewing and revising the NAEP science framework, consideration should be given to the possible merger of aspects of the TEL framework with the science framework to create an integrated science and technology framework and assessment for administration in 2028.

The NAEP TEL framework and assessment have served useful purposes since their development and initial implementation in 2014. As noted earlier, NAEP TEL is due to be administered twice more at grade 8—in 2024 and again in 2028. Given the representation and integration of technology and engineering with science content domains in contemporary science frameworks and standards, as well as the partial overlap of the latter with the NAEP science and TEL frameworks and assessments, worth considering is whether the most important aspects of the NAEP TEL framework could be included in a revised NAEP science framework.

While the NAEP TEL Framework covers grades 4, 8, and 12, the TEL assessment has been developed only for grade 8. In addition to the limitation of the assessment to a single grade, the TEL construct representation and focus on technology literacy may have lost some of its currency and value in the intervening decade. A review of the complete grades 4–12 framework and the grade 8 assessment seems warranted especially considering existing state standards that include integrated content and practice knowledge focused on technology, engineering, and applications of science across grades 4–12.

A. In reviewing and suggesting revisions to the science framework, the panels should consider NAEP TEL's current content, practices, and forms of assessment for possible inclusion in an updated NAEP science framework and assessment.

B. In considering inclusion of NAEP TEL content and practices in an integrated science and technology framework and assessment, the panels should simultaneously consider what important aspects of the NAEP TEL framework and assessment would be lost if the assessment was discontinued after 2024 and whether continuation of NAEP TEL through 2028 is advisable even if a combined science and technology framework is developed for the 2028 NAEP science assessment.

Considerations of Trend

One hallmark of the NAEP program is its focus on monitoring progress over time and the analysis and reporting of trends in performance. The NAEP science trend extends back to 2009 and NAEP TEL to 2014. Assuming implementation of both current assessments in 2024, there will be 15 years of trend data for science and 10 years for TEL. Given the likely scope of a revision to the NAEP science framework and the implications for the 2028 assessment, as well as the possibility of incorporating aspects of TEL in the new framework and assessment, it seems highly likely that preserving the science or TEL trend through 2028 will not be feasible or advisable. Whether breaking trend in either case in 2028 is both warranted and necessary demands careful attention in deliberations that ensue in NAGB's decisions about revisions to both NAEP science and TEL and their futures. In such deliberations, priority should go to insuring the validity of the revised science framework and assessment for 2028 and beyond. Doing so should not be compromised in a possibly misguided effort to preserve trend at all costs.

REFERENCES

Alonzo, A. C., & Gotwals, A. W. (Eds.). (2012). *Learning progression in science: Current challenges and future directions.* Sense.

American Association for the Advancement of Science. (1990). Science for all Americans: A Project 2061 report on literacy goals in science, mathematics, and technology. *Bulletin of Science, Technology & Society*, 10(2), 93–101.

American Association for the Advancement of Science. (1993). *Benchmarks for science literacy.* Oxford University Press.

American Association of Medical Colleges. (2012). *MR5 fifth comprehensive review of the Medical Colleges Admission Test (MCAT): Final MCAT recommendations.* https://www.aamc.org/download/273766/data/finalmr5recommendations.pdf

Banilower, E. R., Smith, P. S., Malzahn, K. A., Plumley, C. L., Gordon, E. M., & Hayes, M. L. (2018). *Report of the 2018 NSSME+.* Horizon Research, Inc

Beggrow, E. P., Ha, M., Nehm, R. H., Pearl, D., & Boone, W. J. (2014). Assessing scientific practices using machine-learning methods: How closely do they match clinical interview performance? *Journal of Science Education and Technology, 23*, 160–182.

Behrens, J. T., DiCerbo, K. E., & Foltz, P. W. (2019). Assessment of complex performances in digital environments. *The Annals of the American Academy of Political and Social Science*, *683*(1), 217-232.

Bennett, R. E. (2008). *Technology for large-scale assessment.* ETS Report No. RM-08-10. Educational Testing Service.

Bergner, Y., & von Davier, A. (2019). Process data in NAEP: Past, present and future. *Journal of Educational and Behavioral Statistics, 44(6),* 706–732.

Berman, A. I., Haertel, E. H., & Pellegrino, J. W. (Eds.) (2020). *Comparability of large-scale educational assessments: Issues and recommendations*. National Academy of Education.

Bransford, J., Brown, A., Cocking, R., Donovan, S., & Pellegrino, J. W. (2000). *How people learn: Brain, mind, experience and school (expanded edition).* National Academy Press.

Bybee, R., McCrae, B., & Laurie, R. (2009). PISA 2006: An assessment of scientific literacy. *Journal of Research in Science Teaching, 46*(8), 865–883.

Bybee, R. W. (2010). Advancing STEM education: A 2020 vision. *Technology and Engineering Teacher, 70*(1), 30.

College Board. (2009). *Science: College Board standards for success*. https://secure-media.collegeboard.org/apc/cbscs-science-standards-2009.pdf

College Board. (2011a). *AP biology curriculum framework 2012–2013.* https://secure-media.collegeboard.org/digitalServices/pdf/ap/10b_2727_AP_Biology_CF_WEB_110128.pdf

College Board. (2011b). *AP chemistry curriculum framework 2013–2014.* https://secure-media.collegeboard.org/digitalServices/pdf/ap/IN120085263_ChemistryCED_Effective_Fall_2013_lkd.pdf

Corcoran, T. B., Mosher, F. A., & Rogat, A. (2009). *Learning progressions in science: An evidence-based approach to reform.* Columbia University, Teachers College, Consortium for Policy Research in Education, Center on Continuous Instructional Improvement.

Davey, T., Ferrara, S., Shavelson, R., Holland, P., Webb, N., & Wise, L. (2015). *Psychometric considerations for the next generation of performance assessment.* http://www.ets.org/Media/Research/pdf/psychometric_considerations_white_paper.pdf

Dickinson, E. R., Gribben, M., Schultz, S. R., Spratto, E., & Woods, A. (2021). *Comparative Analysis of the NAEP Science Framework and State Science Standards: Technical report.* Unpublished manuscript. Retrieved from the National Assessment Governing Board

Drasgow, F. (Ed.) (2016). *Technology and testing: Improving educational and psychological measurement.* Routledge.

Duran, R., Zang, T., Sanosa, D., & Stancavage, F. (2020). *Effects of visual representations and associated interactive features on student performance on National Assessment of Educational Progress (NAEP) pilot science scenario-based tasks.* NAEP Validity Studies Panel.

Ercikan, K., & Pellegrino, J. W. (Eds.). (2017) *Validation of score meaning for the next generation of assessments.* Taylor & Francis.

Gane, B. D., Zaidi, S. Z., & Pellegrino, J. W. (2018). Measuring what matters: Using technology to assess multidimensional learning. *European Journal of Education, 53*(2), 176–187.

Gerard, L. F., Ryoo, K., McElhaney, K. W., Liu, O. L., Rafferty, A. N., & Linn, M. C. (2016). Automated guidance for student inquiry. *Journal of Educational Psychology, 108*(1), 60–81.

Gorin, J. S., & Mislevy, R. J. (2013). *Inherent measurement challenges in the Next Generation Science Standards for both formative and summative assessment* [Paper presentation]. ITS Invitational Research Symposium on Science Assessment, Washington, DC, United States.

Harris, C. J., Krajcik, J. S., Pellegrino, J. W., & DeBarger, A. H. (2019). Designing knowledge-in-use assessments to promote deeper learning. *Educational Measurement: Issues and Practice, 38*(2), 53–67.

Lee, Y.-H., Hao, J., Man, K., & Ou, L. (2019). How do test takers interact with simulation-based tasks? A response-time perspective. *Frontiers in Psychology, 24*.

Mislevy, R. J. (2018). *Sociocognitive foundations of educational measurement.* Routledge.

Mislevy, R. J., & Haertel, G. D. (2006). Implications of evidence-centered design for educational testing. *Educational Measurement: Issues and Practices, 25*(4), 6–20.

Mislevy, R. J., & Riconscente, M. M. (2006). Evidence-centered assessment design: Layers, concepts, and terminology. In S. Downing & T. Haladyna (Eds.), *Handbook of test development* (pp. 61–90). Erlbaum.

Mullis, I. V. (2019). *White paper on 50 Years of NAEP use: Where NAEP has been and where it should go next.* NAEP Validity Studies Panel. https://www.air.org/sites/default/files/2021-06/50-Years-of-NAEP-Use-June-2019.pdf

National Assessment Governing Board. (2008). *Science framework for the 2009 National Assessment of Educational Progress.* https://www.nagb.gov/content/dam/nagb/en/documents/publications/frameworks/science/2009-science-framework.pdf

National Assessment Governing Board. (2013). *Technology and engineering literacy framework for the 2014 National Assessment of Educational Progress.* https://www.nagb.gov/content/dam/nagb/en/documents/publications/frameworks/technology/2014-technology-framework.pdf

National Assessment Governing Board. (2014). *Science framework for the 2015 National Assessment of Educational Progress.* https://www.nagb.gov/content/dam/nagb/en/documents/publications/frameworks/science/2015-science-framework.pdf

National Assessment Governing Board. (2018a). *Revision of framework development policy for NAEP assessments.* https://www.nagb.gov/content/dam/nagb/en/documents/what-we-do/quarterly-board-meeting-materials/2017-11/12-framework-policy-revision.pdf

National Assessment Governing Board. (2018b). *Technology and engineering literacy framework for the 2018 National Assessment of Educational Progress.* https://files.eric.ed.gov/fulltext/ED594359.pdf

National Research Council. (1996). *National science education standards*. The National Academies Press.

National Research Council. (2000). *Inquiry and the National Science Education Standards: A guide for teaching and learning*. The National Academies Press.

National Research Council. (2002). *Learning and understanding: Improving advanced study of Mathematics and Science in U.S. high schools*. The National Academies Press.

National Research Council. (2007). *Taking science to school: Learning and teaching science in grade K-8*. The National Academies Press.

National Research Council. (2009). *Learning science in informal environments: People, places, and pursuits*. The National Academies Press.

National Research Council. (2012). *A Framework for K–12 Science Education: Practices, crosscutting concepts, and core ideas*. The National Academies Press.

National Research Council. (2013). *Next Generation Science Standards: For states, by states*. The National Academies Press.

The Nation's Report Card. (2019). *See how U.S. fourth-, eighth-, and twelfth-grade students performed in science*. https://www.nationsreportcard.gov/highlights/science/2019/

Nehm, R. H., Ha, M., & Mayfield, E. (2012). Transforming biology assessment with machine learning: Automated scoring of written evolutionary explanations. *Journal of Science Education and Technology, 21*, 183–196.

Neidorf, T., Stephens, M., Lasseter, A., Gattis, K., Arora, A., Wang, Y., Guile, S., & Holmes, J. (2016). *A comparison between the Next Generation Science Standards (NGSS) and the National Assessment of Educational Progress (NAEP) Frameworks in science, technology and engineering literacy, and mathematics. Technical report.* National Center for Education Statistics. https://nces.ed.gov/nationsreportcard/subject/science/pdf/ngss_naep_technical_report.pdf

Pellegrino, J. W. (2013). Proficiency in science: Assessment challenges and opportunities. *Science, 340*, 320–323.

Pellegrino, J. W. (2016). *21st Century science assessment: The future is now*. SRI International.

Pellegrino, J. W., Chudowsky, N, & Glaser, R. (Eds). (2001). *Knowing what students know: The science and design of educational assessment*. The National Academies Press.

Pellegrino, J. W., & Hilton, M. L. (Eds.) (2012). *Education for life and work: Developing transferable knowledge and skills in the 21st Century*. The National Academies Press.

Pellegrino J. W., & Quellmalz, E. S. (2010). Perspectives on the integration of technology and assessment. *Journal of Research on Technology in Education, 43*(2), 119–134.

Pellegrino, J. W., Wilson, M. R., Koenig, J. A., & Beatty, A. S. (Eds.) 2014. *Developing assessments for the Next Generation Science Standards.* The National Academies Press.

Pohl, S., Ulitzsch, E., & von Davier, M. (2021). Reframing rankings in educational assessments. *Science, 372*(6540), 338–340.

Schmidt, W. H., McKnight, C. C., & Raizen, S. A. (1997). *A splintered vision: An investigation of U.S. science and mathematics education.* Kluwer Academic.

Smith, P. S. (2020). What does a national survey tell us about progress toward the vision of the NGSS? *Journal of Science Teacher Education, 31*(6), 601–609.

Way, D., & Strain-Seymour, E. (2021). *A framework for considering device and interface features that may affect student performance on the National Assessment of Educational Progress*. NAEP Validity Studies Panel.

Williamson, D. M., Xi, X., & Breyer, F. J. (2012). A framework for evaluation and use of automated scoring. *Educational Measurement: Issues and Practice, 31*(1), 2–13.

Wilson, M. R., & Bertenthal, M. W. (Eds.). (2005). *Systems for state science assessments*. The National Academies Press.

Zhai, X. (2021). Practices and theories: How can machine learning assist in innovative assessment practices in science education. *Journal of Science Education and Technology, 30*(2), 1–11.

Zhai, X., Haudek, K. C., Shi, L., Nehm, R., & Urban-Lurain, M. (2020). From substitution to redefinition: A framework of machine learning-based science assessment. *Journal of Research in Science Teaching, 57*(9), 1430–1459.

Zhai, X., & Pellegrino, J. W. (in press). Large-scale assessment in science education. *Handbook of research on science education* (Vol. III). Routledge.

Zhai, X., Yin, Y., Pellegrino, J. W., Haudek, K. C., & Shi, L. (2020). Applying machine learning in science assessment: A systematic review. *Studies in Science Education, 56*(1), 111–151.

From: Petersen, Anne
To: NAGB Queries
Cc: Waybright, Tyler; Shelley Loving-ryder; Melody Bushley; Gregory Macdougall; Thayer Myra zup95782
Subject: NAEP Science Framework
Date: Monday, September 27, 2021, 10:01:27 AM
Attachments: 2019-science-framework_tdw.pdf

Thank you for the opportunity to review the NAEP Science Framework. The assessment and instruction teams at the Virginia Department of Education have independently reviewed the document and a summary of the comments are provided both in the text below and embedded in the attached document.

Recommendations

The NAEP framework was changed so that it aligns to national standards and that alignment remains. The edit recommendations and concerns indicated below and in the attached document do not necessitate a rewrite of the framework by themselves. The framework appears sufficient to achieve the goals of the NAEP program.

Concerns

Virginia twelfth grade students have not participated in the grade 12 NAEP assessment; however, the inclusion of physics content typically covered in a first year high school physics course may cause a public relations issue to those states that do participate in the assessment. Student performance on the physics content of the NAEP may not be an indicator of student mastery of physics concepts; instead, it may reflect an equity issue. At this time, 59% of schools with 80% of the student population consisting of Black, Lantinola, and Indegienous students do not have first year physics coursework as part of their course options (National Academy of Science, 2021). In addition, 90% of schools that are considered high poverty do not offer physics (National Academy of Science, 2021).

A second concern with the inclusion of the physics content on the 12th grade assessment is that there is currently a critical shortage of physics teachers in the United States (EdSource, 2019).

The Virginia Department of Education recognizes that physics coursework should be accessible to all students and that a robust understanding of physics concepts can prepare students for higher education and future careers; however, reporting student performance on high school first year course physics concepts may cause public confusion as to the complex issues involved with K-12 physics education. Lower student performance on the physics content in 12th grade may be an indicator of a lack of opportunity versus poor performance.

Possible Edits to NAEP CF (see attached document for specific suggested edits)

The NAEP framework was reviewed by VDOE assessment staff and made 3 types of edits:

1. Simple grammatical edits like “Earth” or “the Earth.” (most of the edit suggestions made were this edit)

2. Content clarifications and changes in science through time. (there were only a few)

3. Notes for VDOE staff as to the degree of alignment with VA CF.

Please feel free to reach out to VDOE if you have any questions on the feedback provided.

Anne Petersen
Tyler Waybright

--
Anne Petersen, Ph.D.
Science Coordinator
Virginia Department of Education
101 N.14th St., Richmond

Individual comments provided in tracked changes – Anne Petersen, Virginia Department of Education

Page number	Location on Page	Excerpt from NAEP Science Framework	Recommended Edit / Comment
p. 19	Second paragraph (after bullet points), last sentence	However, using three broad content areas as an organizer helps ensure that key science content is assessed in a balanced way.	not done on NAEP grade 12
p. 20	Fifth bullet point	A deliberate attempt was made to limit the breadth of science content to be assessed so that some important topics could be measured in-depth. Once core content was identified in each science area, additional content statements could be added only if others previously included were eliminated.	not completely sure what accuracy means to VA here. We may teach things at different times.
p. 21	Exhibit 4 title	Exhibit 4. NAEP science content topics and subtopics	hope to see more content subtopics than this but the intro does state that NAEP have been "paired" down. NAEP seems similar to VA in this case. the "benchmark" expectation is quite high.
p. 22	Second paragraph, last sentence	The content statements form the basis for explaining or predicting naturally occurring phenomena. For example, the above content statement about objects in motion can be used to explain and predict the motions of many different specific objects (e.g., an ice skater, an automobile, an electron, or a planet).	i disagree NAEP will not explain (maybe partially) or predict movements of electrons or planets. "Benchmark" level could possibly do this.
p. 23	Exhibit 6 title	Exhibit 6. Commentary on a Physical Science content statement	I feel that VA is a bit more rigorous here than is shown by Exhibit ^
p. 24	Exhibit 6 title continued.	Exhibit 6 (continued). Commentary on a Physical Science content statement	seems to be on par with VA CF except for last bullet
p. 24	First bullet point, Exhibit 6	Some waves are transverse (water seismic) and other waves are longitudinal (sound, seismic).	water is both VA struggles with the same problem
p. 24	Second bullet point, Exhibit 6	In transverse waves, the direction of the motion is perpendicular to the disturbance.	"direction of wave propitiation" In transverse waves, the direction of the motion is perpendicular to the ~~disturbance~~.
p. 24	Third bullet point, Exhibit 6	In longitudinal waves, the direction of motion is parallel to the disturbance.	In longitudinal waves, the direction of motion is parallel to the ~~disturbance~~.
p. 24	Fourth bullet point, Exhibit 6	Waves (e.g., light waves) traveling from one material to another undergo transmission, reflection, and/or changes in speed.	Marked but no comment

Individual comments provided in tracked changes – Anne Petersen, Virginia Department of Education

Page number	Location on Page	Excerpt from NAEP Science Framework	Recommended Edit / Comment
p. 24	Third bullet point after Exhibit Box	Empty cells in the content statement tables denote that a particular subtopic is not recommended for assessment at that grade level.	Very true
p. 24	Fourth bullet point after Exhibit Box	Retention of foundational knowledge from one grade to the next is assumed; however, if the relevant content statement does not appear in a succeeding grade level, it should not be assessed.	This is no small point. VA folks do not believe in this notion. VA folks say this is not fair. Like the NAEP 12 grade test having LS and most VA kids took it in 10^{th}. I believe the test is designed to test student "residual" knowledge of the three content domains and it can do but VA may not participate in grade 12
p. 25	First paragraph under **Physical Sciences** heading	Familiar changes	
p. 25	First paragraph under **Physical Sciences** heading	Erosion of mountains	Not sure these are familiar
p. 28	Second paragraph in textbox	Understanding the substance of water requires knowledge across the Physical Science categories of Matter, Energy, and Motion.	Understanding ~~the substance of water~~ requires knowledge across the Physical Science categories of Matter, Energy, and Motion. "Properties of" probably ok as is
p. 28	First paragraph after textbox, last sentence	The Periodic Table demonstrates the relationship between the atomic number of the elements and their chemical and physical properties and provides a structure for inquiry into the characteristics of the chemical elements (grade 12).	The Periodic Table ~~demonstrates~~ the relationship between the atomic number of the elements and their chemical and physical properties and provides a structure for inquiry into the characteristics of the chemical elements (grade 12). illustrates
p. 30	First paragraph, last sentence	The Sun as the main energy source for the Earth provides an opportunity at all grade levels to make important connections between the science disciplines (see the following textbox).	The Sun as the main energy source for ~~the Earth~~ Earth provides an opportunity at all grade levels to make important connections between the science disciplines (see the following textbox).
p. 30	Last paragraph, second sentence	As the diver falls, her speed (kinetic energy) increases as her potential energy decreases.	their, they

Individual comments provided in tracked changes – Anne Petersen, Virginia Department of Education

Page number	Location on Page	Excerpt from NAEP Science Framework	Recommended Edit / Comment
p. 32	Fourth sentence	The Earth and an airplane do not need to be in contact…	~~The~~ Earth and an airplane do not need to be in contact…
p. 33	Exhibit 8 title	Exhibit 8. Physical Science content statements for grades 4, 8, and 12	these learning progressions are very familiar and similar to VA
p. 33	P4.5	P4.5 Magnets can repel or attract other magnets. They can also attract certain nonmagnetic objects at a distance.	not sure we stress this as much as they seem to do
p. 33	Footnote	Although this content statement generally holds true, some compounds decompose before boiling.	not needed for this audience but ok
p. 35	P12.8	P12.8 Atoms and molecules that compose matter are in constant motion (translational, rotational, or vibrational).	Holy cow, NMR this is organic
p. 35	P8.9	P8.9 Three forms of potential energy are gravitational, elastic, and chemical. Gravitational potential energy changes in a system as the relative positions of objects are changed. Objects can have elastic potential energy due to their compression, or chemical potential energy due to the nature and arrangement of the atoms.	much stronger than VA cf
p. 35	P8.10	P8.10 Energy is transferred from place to place. Light energy from the Sun travels through space to Earth (radiation). Thermal energy travels from a flame through the metal of a cooking pan to the water in the pan (conduction). Air warmed by a fireplace moves around a room (convection). Waves (including sounds and seismic waves, waves on water, and light waves) have energy and transfer energy when they interact with matter.	P8.10 Energy is transferred from place to place. Light energy from the Sun travels through space to Earth (radiation). Thermal energy travels from a flame through the metal of a cooking pan to the water in the pan (conduction). Air warmed by a fireplace moves around a room (convection). Waves (including sounds and seismic waves, waves on water, and light waves) have energy and transfer energy ~~when~~ as they interact with matter.
p. 36	P8.13	P8.13 Nuclear reactions take place in the Sun. In plants, light from the sun is transferred to oxygen and carbon compounds, which, in combination, have chemical potential energy (photosynthesis).	P8.13 ~~Nuclear~~ Fusion reactions take place in the Sun. In plants, light from the sun is transferred to oxygen and carbon compounds, which, in combination, have chemical potential energy (photosynthesis).

Individual comments provided in tracked changes – Anne Petersen, Virginia Department of Education

Page number	Location on Page	Excerpt from NAEP Science Framework	Recommended Edit / Comment
			Comment: This should probably be singular
p. 38	Exhibit 8 Continued title	Exhibit 8 (continued). Physical Science content statements for grades 4, 8, and 12	PS is way above level of VA CF
p. 38	P12.22	P12.22 Gravitation is a universal attractive force that each mass exerts on any other mass. The strength of the gravitational force between two masses is proportional to the masses and inversely proportional to the square of the distances between them.	way above VA cf
p. 38	P12.23	P12.23 Electric force is a universal force that exists between any two charged objects. Opposite charges attract while like charges repel. The strength of the electric force is proportional to the magnitudes of the charges and inversely proportional to the square of the distance between them. Between any two charged particles, the electric force is vastly greater than the gravitational force.	way above
p. 39	Second paragraph, first sentence	Understanding principles in Life Science is inextricably linked with understanding principles in Physical Science and Earth and Space Sciences.	theres that word again
p. 41	Text box, last sentence	Therefore, although synthesis and breakdown are common to both plants and animals, photosynthesis (the conversion of light energy into stored chemical energy) is unique to plants, making them the primary source of energy for all animals.	Anne is “primary” enough to allow inclusion of thermal vent chemotrophs?
p. 42	Second paragraph, third sentence	In these grand-scale cycles, the total amount of matter and energy remains constant, even though their form and location undergo continual change.	In these ~~grand-scale~~ cycles, the total amount of matter and energy remains constant, even though their form and location undergo continual change. Comment: not sure why this is here is it a technical or statistical term?
p. 44	First paragraph under Evolution and Diversity, third sentence	The modern concept of evolution, including natural selection and common descent, provides a unifying principle for understanding the history of life on	The pencil mark is over “principle” but no written comment.

Individual comments provided in tracked changes – Anne Petersen, Virginia Department of Education

Page number	Location on Page	Excerpt from NAEP Science Framework	Recommended Edit / Comment
		Earth, relationships among all living things, and the dependence of life on the physical environment.	
p. 45	L12.1	L12.1 Living systems are made of complex molecules (including carbohydrates, fats, proteins, and nucleic acids) that consist mostly of a few elements, especially carbon, hydrogen, oxygen, nitrogen, and phosphorous.	teach to this level in bl?
p. 45	L12.3	L12.3 Cellular processes are regulated both internally and externally by environments in which cells exist, including local environments that lead to cell differentiation during the development of multicellular organisms. During the development of complex multicellular organisms, cell differentiation is regulated through the expression of different genes.	this also sounds on level with VA CF
p. 46	Exhibit 10 (continued) title	Grade 12	much of this content is taught in VA
p. 46	Footnote	The statement "they use the energy from light" does not imply that energy is converted into matter or that energy is lost. See textbox "Crosscutting Content: Uses, Transformations, and Conservation of Energy," p. 42.	I really do not think this is needed
p. 47	Exhibit 10 continued title	Exhibit 10 (continued). Life science content statements for grades 4, 8, and 12	Table is very similar to VA in most respects
p. 47	L4.4	L4.4 When the environment changes, some plants and animals survive and reproduce; others die or move to a new location.	change. eg. seasons
p. 48	L8.10	L8.10 The characteristics of organisms are influenced by heredity and environment. For some characteristics, inheritance is more important; for other characteristics, interactions with the environment are more important.	VA goes into Mendel
p. 48	L12.9	L12. 9 The genetic information encoded in DNA molecules provides instructions for assembling	nice!

Individual comments provided in tracked changes – Anne Petersen, Virginia Department of Education

Page number	Location on Page	Excerpt from NAEP Science Framework	Recommended Edit / Comment
		protein molecules. Genes are segments of DNA molecules. Inserting, deleting, or substituting DNA segments can alter genes. An altered gene may be passed on to every cell that develops from it. The resulting features may help, harm, or have little or no effect on the offspring's success in its environment.	
p. 49	L8.11 (last sentence)	L8.11 Extinction of a species is common; most of the species that have lived on the Earth no longer exist.	L8.11 Extinction of a species is common; most of the species that have lived on ~~the~~ Earth no longer exist.
p. 49	L8.12 (last sentence)	L8.12 Biologists consider details of internal and external structures to be more important than behavior or general appearance.	this may not prove to be true in the see "canis" and "the species problem"
p. 49	L12.13	L.12.13 Evolution is the consequence of the interactions of (1) the potential for a species to increase its numbers, (2) the genetic variability of offspring due to mutation and recombination of genes, (3) a finite supply of the resources required for life, and (4) the ensuing selection from environmental pressure of those organisms better able to survive and leave offspring.	Interesting!
p. 50	First paragraph, third sentence	This concept of Earth as a complex and dynamic entity of interrelated subsystems implies that there is no process or phenomenon within the Earth system that occurs in complete isolation from other elements of the system.	This concept of Earth as a complex and dynamic entity of interrelated subsystems implies that there is no process or phenomenon within ~~the~~ Earth system that occurs in complete isolation from other elements of the system.
p. 50	Last paragraph, third sentence	Other Web-based programs allow students to view and process satellite images of Earth, to direct a camera on board the Space Shuttle, and to access professional telescopes around the world to carry out science projects.	a little dated at this point
p. 50	Footnote	Earth is capitalized, rather than referred to as "the earth," in order to recognize it as one of the planets in the solar system.	see gregg

Individual comments provided in tracked changes – Anne Petersen, Virginia Department of Education

Page number	Location on Page	Excerpt from NAEP Science Framework	Recommended Edit / Comment
p. 51	Second paragraph under *Earth in Space and Time* heading	Applies to entire paragraph	"the" earth is removed here, as it should be
p. 51	First paragraph under **Objects in the Universe** heading	"the Sun and the Moon"	remove "the" if one is going to capitalize the proper name?
p. 52	First paragraph, second sentence	However, it is now known that the Sun is the central and largest body in the solar system, which includes Earth and other planets and their moons as well as other objects such as asteroids and comets.	Ok no the here. this should be fixed one way or the other
p. 52	First paragraph, second sentence under **History of Earth** heading	Initially, there was no life and no molecular oxygen in the atmosphere.	or water
p. 52	Third paragraph, second sentence under **History of Earth** heading	Some changes are due to slow processes, such as erosion and weathering and others are due to rapid processes such as volcanic eruptions, landslides, and earthquakes (Grade 4).	cosmic impacts
p. 53	First paragraph under **Properties of Earth Materials** heading	Earth materials that occur in nature include rocks, minerals, soils, water, and the gases of the atmosphere. Natural materials have different properties that sustain plan and animal life (grade 4).	nice
p. 53-54	Last sentence on page 53 going into 54	The current explanation is that the outward transfer of Earth's internal heat propels the plates comprising Earth's surface across the face of the globe, pushing the plates apart where magma rises to form mid-ocean ridges, and pulling the edges of plates back down where the Earth materials sink into the crust at deep trenches (grade 12).	The current explanation is that the outward transfer of Earth's internal heat propels the plates comprising Earth's surface across the face of the globe, pushing the plates apart where magma rises to form mid-ocean ridges, and pulling the edges of plates back down where the Earth materials ~~sink~~ subducted into the ~~crust~~ mantel at deep trenches (grade 12).
p. 54	First paragraph, second sentence under **Energy in**	The Sun is the major source of energy for phenomena on Earth's surface.	we use "our" instead of "the" but we do not caps sun

Individual comments provided in tracked changes – Anne Petersen, Virginia Department of Education

Page number	Location on Page	Excerpt from NAEP Science Framework	Recommended Edit / Comment
	Earth Systems heading		
p. 55	First paragraph, last sentence under **Biogeochemical Cycles**	For example, carbon occurs in carbonate rocks such as limestone, in coal and other fossil fuels, in the atmosphere as carbon dioxide gas, in water as dissolved carbon dioxide, and in all organisms as complex molecules that control the chemistry of life (grade 12).	nice!
p. 56	Textbox heading	Crosscutting Content: Biogeochemical Cycle	This is great stuff
p. 56	Second paragraph, first sentence	Essentially fixed amounts of chemical atoms or elements cycle with the Earth system, and energy drives their translocation of matter(e.g., changes of state, gravity)	Essentially fixed amounts of chemical atoms or elements cycle with ~~the~~ Earth system
p. 56	Third paragraph	Biogeochemical cycles are described more fully in the Earth Systems section of exhibit 12, Earth and Space Science Content Statements for Grades 4, 8, and 12.	Biogeochemical cycles are described more fully in ~~the~~ Earth Systems section of exhibit 12, Earth and Space Science Content Statements for Grades 4, 8, and 12.
p. 58	E8.3	E8.3 Fossils provide important evidence of how life and environmental conditions have changed in a given location.	not sure we go this far
p. 58	E8.4	E8.4 Earth processes seen today, such as erosion and mountain building, make it possible to measure geologic time through methods such as observing rock sequences and using fossils to correlate the sequences at various locations.	pretty heavy into fossils here more so than VA CF
p. 59	Grade 12 header at top of table (note that comment refers to Grade 8)	Grade 12	the grade 8 material here is above VA CF
p. 60	Grade 8 header at top of table	Grade 8	pretty high level compared to VA CF
p. 61	E12.10	E12.10 Climate is determined by energy transfer from the Sun at and near Earth's surface. This energy transfer is influenced by dynamic processes such as	we should have this is VA CF

Individual comments provided in tracked changes – Anne Petersen, Virginia Department of Education

Page number	Location on Page	Excerpt from NAEP Science Framework	Recommended Edit / Comment
		cloud cover, atmospheric gases, and Earth's rotation, as well as static conditions such as the positions of mountain ranges, oceans, seas, and lakes.	
p. 62	Title of Exhibit	Exhibit 12 (continued). Earth and Space Sciences content statements for grades 4, 8, and 12	NAEP might be interpreted as being more rigorous in 12
p. 62	E4.10	E4.10 The supply of many Earth resources such as fuels, metals, fresh water, and farmland is limited. Humans have devised methods for extending the use of Earth resources through recycling, reuse, and renewal.	Nice!
p. 62	E12.11	E12.11 Earth is a system containing essentially a fixed amount of each stable chemical atom or element. Most elements can exist in several different chemical forms. Earth elements move within and between the lithosphere, atmosphere, hydrosphere, and biosphere as part of biogeochemical cycles.	nice
p. 68	First illustrative item	The Earth's Moon is A. always much closer to the Sun than it is to the Earth. B. always much closer to the Earth than it is to the Sun. C. about the same distance from the Sun as it is from the Earth. D. sometimes closer to the Sun than it is the Earth and sometimes closer to the Earth than it is to the Sun.	~~The~~ Earth's Moon is A. always much closer to the Sun than it is to ~~the~~ Earth. B. always much closer to ~~the~~ Earth than it is to the Sun. C. about the same distance from the Sun as it is from ~~the~~ Earth. D. sometimes closer to the Sun than it is ~~the~~ Earth and sometimes closer to ~~the~~ Earth than it is to the Sun.
p. 73	Footnote	In addition, 12th graders at the Advanced level are expected to be able to identify a scientific question for investigation. See appendix B for achievement level descriptions.	this seems odd shouldn't this be done at all levels
p. 75	Second paragraph, last sentence	After students have run the modeling software, they are asked a series of questions (e.g., the size of the hare population over time).	They have had this since 2009. VA should be ashamed…

Individual comments provided in tracked changes – Anne Petersen, Virginia Department of Education

Page number	Location on Page	Excerpt from NAEP Science Framework	Recommended Edit / Comment
p. 79	Comment is on the graphic	Forest succession graphic	we have this art
p. 83	Exhibit 14 title	Exhibit 14. Generating examples of grade 8 performance expectations	mailing the table to PEM and ETS
p. 85	E8.2	E8.2 Gravity is the force	Gravity is ~~the~~ a, or one of the forces
p. 86	First bullet point in Using Scientific Inquiry sectoin	Using scientific Inquiry: • Arrange a set of photographs of the Moon taken over a month's time in chronological order and explain the order in terms of a model of the Earth-Sun-Moon system.	Arrange a set of photographs of the Moon taken over a month's time in chronological order and explain the order in terms of a model of ~~the~~ Earth-Sun-Moon system.
p. 87	Second Items to Assess Using Science Principles	Items to Assess Using Science Principles Illustrative Item A space station is to be located between the Earth and the moon at the place where the Earth's gravitational pull is equal to the Moon's gravitational pull.	A space station is to be located between ~~the~~ Earth and the moon at the place where the Earth's gravitational pull is equal to the Moon's gravitational pull.
p. 89	Item Suggestion 1	NASA wants to launch a spacecraft with rockets from Earth so that it will reach and orbit Mars. Which of the following statements about this flight is WRONG: A. In the first phase of the flight, the forces acting on the spacecraft are the thrust of the rocket engine, gravity, and friction from the Earth's atmosphere. B. When the rocket engine shuts off, the only force acting on the spacecraft is the force of gravity. C. Once the spacecraft is above the Earth's atmosphere and the rocket engine is off, it will travel at a constant speed since there is no gravity in space.	Comment: falcon heavy (VDOE) is a better cluster than this Edits: A. In the first phase of the flight, the forces acting on the spacecraft are the thrust of the rocket engine, gravity, and friction from ~~the~~ Earth's atmosphere. B. When the rocket engine shuts off, the only force acting on the spacecraft is the force of gravity. C. Once the spacecraft is above ~~the~~ Earth's atmosphere and the rocket engine is off, it will travel at a constant speed since there is no gravity in space.
p. 104	Illustrative Items	Illustrative Items	What causes days and night? A. ~~The~~ Earth spins on its axis. (66%)

Page number	Location on Page	Excerpt from NAEP Science Framework	Recommended Edit / Comment
		What causes days and night? A. The earth spins on its axis. (66%) B. The earth moves around the Sun. (26%) C. Clouds block out the Sun's light. (0%) D. The earth moves into and out of the Sun's shadow. (3%) E. The Sun goes around the Earth. (4%) The main reason for its being hotter in summer than in winter is: A. The earth's distance from the Sun changes. (45%)	B. ~~The~~ Earth moves around the Sun. (26%) C. Clouds block out the Sun's light. (0%) D. ~~The~~ Earth moves into and out of the Sun's shadow. (3%) E. The Sun goes around the Earth. (4%) The main reason for its being hotter in summer than in winter is: ~~The~~ Earth's distance from the Sun changes. (45%)
p. 133	Last paragraph, first sentence	*In the Earth and space science*, students at the *NAEP Proficient* level should be able to explain how gravity accounts for the visible patterns of motion of the Earth.	*In the Earth and space science*, students at the *NAEP Proficient* level should be able to explain how gravity accounts for the visible patterns of motion of ~~the~~ Earth.
p. 135	Third paragraph	*In the physical sciences,* students at the *NAEP Basic* level should be able to … critique data that claim to show how gravitational potential energy changes with distance from the Earth's surface	*In the physical sciences,* students at the *NAEP Basic* level should be able to … critique data that claim to show how gravitational potential energy changes with distance from ~~the~~ Earth's surface
p. 137	First paragraph	…and evidence for human effects on the Earth's biogeochemical cycles	and evidence for human effects on ~~the~~ Earth's biogeochemical cycles

From: Moulding, Brett
To: NAGB Queries
Subject: Comments on the NAEP Science Assessment Framework
Date: Friday, August 27, 2021 9:12:56 AM

NAGB Leadership,
Comments on the future revision of NAEP Assessment Framework for Science

Whether the NAEP Science Assessment Framework needs to be updated.
The NAEP Science Assessment Framework needs to be revised.

If the framework needs to be updated, why a revision is needed.
The current Framework does not identify the science being taught in the majority of our schools. The science NAEP cannot be a report card on science education in the nation if it does not measure the current science being taught in our schools. The current NAEP framework is not consistent with the current research in how students learn.

What should a revision to the NAEP framework include?
The revision should include a clear alignment to the National Academies Framework for K-12 Science Education. The revision should include descriptions of the three-dimensional science performances that need to be assessed. The New NAEP Framework needs to include measurement of students using Practices, Crosscutting Concepts, and Core Ideas consistent with the NGSS approach to science performance expectations.

Thank you,
Brett

Brett Moulding
Retired
Utah State Office of Education Curriculum Director and Instruction
Former NAEP Science Advisory Committee Member

From: Cary Sneider
To: NAGB Queries
Subject: NAEP Science Framework
Date: Friday, August 20, 2021 2:36:35 PM
Attachments: A-Cary"s final Comments to NAGB 2019 re TEL&Science.docx

Hello Friends,

When I ended my tenure on NAGB I made the following plea for updating the NAEP Science Framework to be consistent with the Framework for K-12 Science Education: Practices, Crosscutting Concepts and Core Ides (NRC 2012) and the subsequent Next Generation Science Standards (NGSS Lead States 2013). These have now been adopted or adapted by 44 states. Such an updated would essentially be a merger of most of the TEL and an improved NAEP Science Framework. I have attached those comments to this email.

As I've also noted in some of my prior comments during my time on the Board, NAEP has been referred to as a "Gold Standard" and a "North Star." These qualities are not the same. The "Gold Standard" refers to NAEP as a "truth-teller," because of meticulous attention to scientific rigor and detail. The "North Star" means that NAEP also points to a future destination. In this case it means that the updated NAEP Science Framework should not just reflect the two existing documents now being used by most states to guide their own science standards, but blaze the trail for future improvements in what students should know and be able to do in the STEM fields.

Warm regards, Cary

Cary Sneider, PhD

Cary Sneider's parting comments to the full NAGB Board

Friday, August 2, 2019

I'm completing 8 years on the Board, but in a sense, it's been 16 years, since my friend and colleague, Alan Friedman rolled off the Board just before I joined. Alan was a friend and mentor for most of my career. Many of us were very sad when he passed away after a brief illness at age 72.

Part of Alan's legacy to the Board and to me has been the NAEP TEL. I want to spend a few minutes reflecting on that. As a fresh context I'd like to ask how many of you read the story of the New Navy that was referenced in a recent *Staying On Board* newsletter.

There were three parts of that story relevant to the TEL. They correspond to the three phases of the engineering design process, which is the cornerstone of engineering, which is deeply embedded in the Framework for K-12 Science Education (NRC 2012) and the subsequent NGSS (2013). In contrast to prior science standards, the Framework and NGSS emphasize not just what students should know about science, but what skills they need to develop to use what they know to solve meaningful problems.

1) Defining the Problem. In contrast to the old Navy, when the purpose of training was for sailors to learn to do their job right, today's sailors are trained in many different jobs. They have to ask themselves "Am I doing the right job?" Similarly, an essential aspect of engineering, which is now a part of the science standards in 44 states, is "Am I solving the right problem?

2) Generating Creative Solutions. There's an example of creative thinking in which sailors figure out how to secure the ship to the dock using only the materials that were in front of them. That's solving a problem under constraint—one critical aspect about problem solving that students have to learn during 12 years of schooling.

3) Optimize. Once you have met the criteria and constraints of a problem you are not done. You need to refine the solution. We learned from the article that things were going so well with the new Navy that the brass decided to end the experiment early and build more light ships and hire more of the right kinds of people. Then problems cropped up. Problems always crop up with new technologies. Continuing the experiment to refine the solution is an important part of the process. In engineering it's called "optimization."

PEOPLE. The upshot of the New Navy article is that the recruiters need to find "the right people." But as educators, we don't have the luxury of turning away 9 out of 10 kids that show up for our classes. We need to prepare all of them for a rapidly changing world.

They Learn Engineering in School. The data from the context variables on the TEL inform us that more than half of our students take courses in engineering—in addition to the science courses that will—as more schools adopt the new standards—help them learn to define problems, creatively solve them under constraint, and be persistent as they continue to refine and optimize solutions to persistent problems.

In future meetings you'll be considering revision of the Science Framework. When that work is done, if it measures what students are expected to learn, it will incorporate 50% to 80% of the TEL, depending on grade level. **Essentially, that means merging the Science and TEL frameworks.** When that happens, it is my hope that funds previously spent on separate administration of the TEL can be repurposed to support state and TUDA level assessment for science (now more appropriately referred to as STEM) so that educators across the country have a golden meter stick to see how well they're doing. That's the baton I'm passing along from Alan and from me.

Input regarding the NAEP Science Assessment

Cary Sneider, Former NAGB Member

September 4, 2021

In the following paragraphs I will argue that the NAEP Science Framework needs to be updated to include much of what is in the NAEP TEL Framework. Once that is done the TEL can be eliminated and funds saved can be used to conduct science assessments at grades 4, 8, and 12 at the state and TUDA levels.

Does the NAEP Science Framework need to be updated?

Yes.

If the framework needs to be updated, why is a revision needed?

1. The NAEP Science Framework is significantly out-of-date. The NRC's consensus *study A K-12 Science Education Framework: Practices, Crosscutting Concepts, and Core Ideas* (2012) and the subsequent *Next Generation Science Standards* (NGSS Lead States, 2013) has gained traction in 44 states that have adapted or adopted new standards based on these documents. Even states that claim not to base their standards on either of these documents are influenced by them.

An essential innovation of these new standards documents is the inclusion of engineering as a part of science. It is deeply woven into the fabric of the standards, as both a set of practices complementary to science, as well as crosscutting concepts, and even core ideas, which are listed at the same level as the traditional sciences. The reason for including engineering as an essential element of science is stated in the Framework as follows:

> We anticipate that the insights gained and interests provoked from studying and engaging in the practices of science and engineering during their K-12 schooling should help students see how science and engineering are instrumental in addressing major challenges that confront society today, such as generating sufficient energy, preventing and treating diseases, maintaining supplies of clean water and food, and solving the problems of global environmental change. (NRC 2012, p. 9).

Providing a foundation in engineering design allows students to better engage in and aspire to solve the major societal and environmental challenges they will face in the decades ahead. The same document also makes clear distinctions among the important terms science, technology, and engineering.

> In the K–12 context, "science" is generally taken to mean the traditional natural sciences: physics, chemistry, biology, and (more recently) earth, space, and environmental sciences . . . . We use the term "engineering" in a very broad sense to mean any engagement in a systematic practice of design to achieve solutions to particular human problems. Likewise, we broadly use the term "technology" to include all types of human-made systems and processes—not in the limited sense often used in schools that equates technology with modern computational and

> communications devices. Technologies result when engineers apply their understanding of the natural world and of human behavior to design ways to satisfy human needs and wants. (NRC 2012, p. 11-12)

2. NGSS performance expectations have substantial overlap with NAEP Science and NAEP TEL at the **8th and 12th grade levels**.

According to a study by AIR commissioned by NAGB:

> "Ninety percent or more of NGSS performance expectations at the middle school and high school levels covered content that overlaps with NAEP science or TEL at grades 8 and 12, respectively (Neidorf et al. 2016)."

This means that the great majority of students in middle and high school will increasingly have an opportunity to learn what is in the TEL Framework through science instruction. It will be important to monitor implementation of those standards over the next decade—and only a combined Science-TEL framework, administered across states, can do that. While administering NAEP Science and NAEP TEL in a coordinated fashion would provide useful information, a revised NAEP Science Assessment could improve the monitoring function. Also, the science assessment would be fairer to students and teachers, and of greater interest to educational leaders in cities and states if it were consistent with the new standards.

What should a revision to the framework include?

1. What states are currently advocating. The purpose of the NRC's Framework and NGSS, led by the National Governor's Association and Council of Chief State School Officers, was to help all states pull in the same direction. If NAGB is to be the North Star, its essential that a new Framework not attempt to lead in an entirely different direction. In addition to being guided by these two documents, however, it will be important to commission a study of state science standards to ensure that the six states that claim more independence in their science standards are included.

2. **Additional topics from the TEL.** The TEL consists of three parts: Design and Systems, Technology and Society, and Information and Communications Technology. The first two are very strongly represented in the NGSS and Framework, and therefore in the great majority of state standards. The third area is not taught explicitly in most schools. A consolidated framework would therefore consist, in broad strokes, of the first two areas of the TEL and an updated version of the Science Framework. What will be lost is some of the third part of the TEL, which may be more closely related to ELA than to science.

If these recommendations are followed, NAGB would be able to report on accomplishments of our nation's youth in their ability to solve problems, to analyze systems, and understand key issues at the intersection of technology and society as a part of the Science Report Card. NAGB has broken new ground by developing the TEL, the first fully DBA assessment in its portfolio. That was an important accomplishment, but now it's time to consolidate it with Science, so that we can have an efficient assessment that is maximally useful to the states, while at the same time increasing NAGB's efficiency.

3. New topics highlighted by recent world events. If NAGB is to serve as the North Star, the NAEP Framework should also lead, not just follow the states. So, it will be important to consult with a wide variety of experts. Among the considerations should be the experience of a highly stressful pandemic, and the possible inclusion of topics directly related to epidemiology, vaccinations, institutions such as the CDC and WHO, and the nature of science.

178 Albion St., Suite 210
Wakefield, MA 01880
781.245.2212
781.245.5212
cast.org
@CAST_UDL

National Assessment Governing Board
800 North Capitol Street, NW, Suite 825
Washington, DC 20002

RE: NAEP Science Framework

Submitted *via email* to nagb@ed.gov

Dear Governing Board,

Since 1984, CAST (originally the Center for Applied Special Technology) has worked relentlessly to ensure that our nation is one where learning has no limits for all individuals. CAST pioneered Universal Design for Learning (UDL), a set of principles and guidelines for inclusive design for learning—including curricula, learning goals, materials, instructional methods, and assessments. UDL is now incorporated in key federal education, career training, and workforce laws.[i] UDL provides the basis for innovation and success in expanding and strengthening education across all subject areas (e.g., reading, mathematics, science). When applied to assessments, UDL can ensure that accessible normative and summative assessments are available to all students regardless of any potential learning barrier they may experience whether it be due to socio-economic status, language, or disability status.

CAST is pleased to submit comments and recommendations to the National Assessment Governing Board (NAGB) query regarding the National Assessment of Education Progress (NAEP) Science Framework ("the Framework"). Because universal design is included as a minor reference in the current framework, CAST strongly urges the NAGB to update the Framework to make it consistent with current federal law and documented best practices in the application of inclusive design in student engagement, student learning, assessment design, and assessment application.

CAST leads work funded through grants provided by the National Science Foundation (NSF), U.S. Departments of Education (ED) and Labor (DOL), state education agencies, local education agencies, as well as the private sector. CAST seeks to ensure that the full power of UDL is applied to technology, instructional, and assessment design and practice in order to remove barriers to learning and assessment in digital as well as physical settings. Our UDL initiatives encourage and support the design of flexible learning environments that anticipate learner variability and provide alternative routes or paths to success, as well as provide flexible opportunities for learners to demonstrate their construct-relevant knowledge, skills, and abilities during summative, formative, and diagnostic assessment. UDL acknowledges that the variability of how people learn is the *norm* rather than the exception. UDL provides viable alternatives for *all* learners to access in-person, blended, and online education and assessment, providing a responsive framework to support students and educators in any academic subject, including in science.

In support of our recommendation that NAGB update the Framework, CAST has examined and compared NAEP participation data for students with disabilities and English Learners (ELs) in the science assessment for the years 2009, 2015, and 2019 respectively. While NAEP data show that participation rates do increase between 2009 and 2019 for both groups of students (NAEP Science Assessment data)[ii], the participation rates remain well below NAEP's own 95 percent requirement (NAEP Policy, 2014).[iii]

Additionally, the participation of students with disabilities falls between grades 8 and 12 (NAEP Participation Rate).[iv] Therefore, CAST strongly encourages NAGB to consider our recommendations, which intend to ensure that the [new] NAEP science assessment incorporates from the outset the most modern and inclusive design so that a variable and diverse student population can successfully access and complete the assessment in grades 4, 8, and 12 at a participation rate of at least 95 percent. To help NAGB accomplish these goals, we offer the following:

General Recommendations

- Incorporate the principles of UDL throughout the Framework to support and assure student access to the NAEP science assessment, regardless of literacy level, language, and/or disability status.
- Adopt a validity framework that promotes consideration of the broad range of construct-irrelevant factors learners bring to testing. This framework should be applied from the beginning of test and item design in an effort to reduce reliance on retrofitted accommodations that provide inadequate support and/or compromise construct integrity. Examples of such frameworks, based on principles of UDL, include Dolan et al. (2013)[v] and Almond et al. (2010)[vi], the former of which has been applied in development of next-generation science assessments (e.g., Quellmalz et al., 2016).[vii]
- Eliminate all references to No Child Left Behind and include in a new Framework references and citations consistent with current law, the Elementary and Secondary Education Act currently known as the Every Student Succeeds Act (ESSA).[viii]
- Eliminate use of the term 'special needs', replacing such term with 'students with disabilities' to ensure consistency with the ESSA and the Individuals with Disabilities Education Act (IDEA).
- Discuss how to include students with the most significant cognitive disabilities in NAEP assessments who take state-designed alternate assessments on alternate achievement standards. Currently these students are not included *in any* NAEP assessment. Recent research has demonstrated the promise of combining learning map model- and UDL-based approaches in evaluating the science knowledge, skills, and abilities of students with significant cognitive disabilities.

Recommendations for the Framework (based on current pages 2-5):

- Add new rationale to ensure the Framework and new NAEP Science assessment:
    - **Inclusive Design**: Incorporate the principles of UDL as an essential component to developing a robust assessment tool from inception and design to roll-out of the assessment.
    - **Student Diversity**: Respond to the growing and increasingly diverse student population in the nation, the inclusion of all types and ages of students in the general curriculum, and the growing emphasis and commitment to serve and be accountable for all students. Such diversity does include students with disabilities and English Learners (ELs); however, the Framework *must assure* the meaning of diversity is expanded [beyond students with disabilities and ELs] consistent with NAEP resources developed in recent years (NAEP Engineering Framework).
    - **Cultural Relevance**: Acknowledge that advances have been made in understanding cultural relevance and its impact on student engagement, learning and assessment.
    - **Access Features**: Include specificity in the need for the assessment to be designed with access features consistent with [WCAG 2.1] and UDL recommendations and provide built-in navigation and access supports (e.g., motoric supports, language/glossary, audio, fonts, text size, etc.) without altering the science construct. Such features are increasingly no longer considered 'accommodations' and instead are regularly available to all users. The Framework must require and acknowledge their incorporation and encourage/allow for their use for all students.

- **Accessibility and Accommodations**: Ensure full accessibility in the design of test items, including in the availability of standard accommodations for students with disabilities and ELs as required by federal laws (IDEA and Section 508).[ix] The Framework must assure accessibility specifically includes the use and interoperability with any external assistive technology [device/system] required by the student. Consistent with ESSA[x] such accessibility is specifically intended to increase inclusion of formerly excluded groups in assessments, including the NAEP (e.g., students with disabilities and English learners).
- **Computer Skills**: Clarify that recent events show that young students (e.g., grade 4 NAEP test takers) may have insufficient access to and training in computer use for fair inclusion in digital assessments.
- **Access to Broadband**: Make clear that many communities and schools that exist in digital deserts may have insufficient access to broadband services to support access to the assessment across grades 4, 8, and 12.

Recommendation for the Steering Committee (current page 5):

- Provide guidelines to the Steering Committee which clarifies the framework applies UDL in determining assessment content, access features and—when necessary—accommodations consistent with the objectives being assessed. (Rose et al., 2018)[xi]

Recommendations for the Model of Assessment Development and Methods:

- Ensure the methodology outlines how the assessment incorporates inclusive design and is built upon the principles of UDL, and also includes access features including in the use and interoperability with assistive technology
- Describe considerations for English learners and students with disabilities. In particular, that assessment design applies a UDL-based validity framework to help ensure full accessibility, including in the use and interoperability with assistive technology, consistent with ESSA.[xii]

Recommendation: Chapter 4: Students With Disabilities and English Language Learners (Current Pages 114-115)

- Make updates consistent with current research and practice, incorporating the principles of UDL throughout the Framework to support and assure student access to the NAEP science assessment, regardless of literacy level, language and/or disability status. (Rose et al., 2018)[xiii]

Recommendations: Chapter 4: Key Attributes of Effective Assessment (current page 124)

- Takes into account student diversity as reflected in gender, geographic location, language proficiency, race/ethnicity, socioeconomic status, and disability status consistent with NAEP policies (e.g., NAEP Engineering Framework, 2018).[xiv]
- Clarifies the design and implementation is guided by the best available research on assessment item design and delivery:
    - so that it is accessible to all students and whose design minimizes the need for any/standard accommodations for students with disabilities and English Learners.
    - so that students with disabilities and other diverse learners are considered during initial assessment design so they can fully participate and are provided adequate means to demonstrate their construct-relevant knowledge, skills, and abilities, including—but not limited to—the use and interoperability with any needed external assistive technology. (Almond et al., 2010; ESSA; Dolan et al., 2013)[xv]
    - Eliminate the use of the term 'special needs'.

CAST thanks the NAGB for the opportunity to provide these comments, to advocate for a revision to the NAEP Science Framework, and to provide thoughts on how the Framework can be updated to align with current federal policy and documented best practices in the application of inclusive design in assessment design and application. This will allow the nation to provide all learners the opportunity to demonstrate fairly and accurately their science knowledge, skills, and abilities regardless of any potential learning barrier they may experience, whether it be due to socio-economic status, language, or disability status.

Please contact CAST's Director of Federal Relations Sherri Wilcauskas at swilcauskas@cast.org with any questions or for additional information.

Sincerely,

David Gordon
Interim CEO

---

[i] P.L. 110-315, P.L. 113-28, P.L. 114-95, P.L. 115-224, National Education Technology Plan (2021), U.S. Department of Education.

[ii] National Center for Education Statistics Appendix Tables (2009) at: https://nces.ed.gov/nationsreportcard/pdf/main2009/2011451.pdf; Appendix Tables (2015) at: https://www.nationsreportcard.gov/science_2015/files/2015_Science_Technical_Appendix.pdf; Appendix Tables (2019) at: https://www.nationsreportcard.gov/science/supporting_files/2019_appendix_sci.pdf

[iii] National Assessment Governing Board *Testing and Reporting on Students with Disabilities and English Language Learners Policy Statement*, (2014) at: https://www.nagb.gov/content/dam/nagb/en/documents/policies/naep_testandreport_studentswithdisabilities.pdf

[iv] National Center for Education Statistics Appendix Tables (2009) at: https://nces.ed.gov/nationsreportcard/pdf/main2009/2011451.pdf; Appendix Tables (2015) at: https://www.nationsreportcard.gov/science_2015/files/2015_Science_Technical_Appendix.pdf; Appendix Tables (2019) at: https://www.nationsreportcard.gov/science/supporting_files/2019_appendix_sci.pdf

[v] Dolan, R.P., Burling, K., Harms, M., Strain-Seymour, E., Way, W. (Denny), & Rose, D.H. (2013) *A Universal design for Learning-based Framework for Designing Accessible Technology-Enhanced Assessments* at: http://images.pearsonclinical.com/images/tmrs/dolanudl-teaframework_final3.pdf

[vi] Almond, P., Winter, P., Cameto, R., Russell, M., Sato, E., Clarke-Midura, J., Torres, C., Haertel, G., Dolan, R., Beddow, P., & Lazarus, S. (2010). Technology-Enabled and Universally Designed Assessment: Considering Access in Measuring the Achievement of Students with Disabilities: A Foundation for Research. *The Journal of Technology, Learning and Assessment*, *10*(5) at: https://ejournals.bc.edu/index.php/jtla/article/view/1605

[vii] Quellmalz, E. S., Silberglitt, M. D., Buckley, B. C., Loveland, M. T., & Brenner, D. G. (2016). Simulations for Supporting an d Assessing Science Literacy. In Y. Rosen, Y., Ferrara, S., & Mosharraf, M. (Eds.). (2016). *Handbook of Research on Technology Tools for Real-World Skill Development*. IGI Global at: http://doi:10.4018/978-1-4666-9441-5

[viii] See: P.L. 114-95

[ix] See: P.L. 108-446, Sections 300.105 and 300.324; and 29 U.S.C. 794d

[x] See: P.L. 114-95, Section 1111, (b)(2)(B)(vii)(II)

[xi] Rose & Gravel, (2013); Daley & Rappolt-Schlichtmann, 2009; Rose & Meyer, (2006); Blascovich, Mendes, Tomaka, Salomon, & Seery,(2003); Csiksentmihalyi, (1991)

[xii] See: P.L. 114-95, Section 1111, (b)(2)(B)(vii)(II)

[xiii] Rose & Gravel, (2013); Daley & Rappolt-Schlichtmann, 2009; Rose & Meyer, (2006); Blascovich, Mendes, Tomaka, Salomon, & Seery,(2003); Csiksentmihalyi, (1991)

[xiv] The 2018 NAEP Technology and Engineering Literacy Framework at: https://www.nagb.gov/content/dam/nagb/en/documents/publications/frameworks/technology/2018-technology-framework.pdf

[xv] Almond, P., Winter, P., Cameto, R., Russell, M., Sato, E., Clarke-Midura, J., Torres, C., Haertel, G., Dolan, R., Beddow, P., & Lazarus, S. (2010). Technology-Enabled and Universally Designed Assessment: Considering Access in Measuring the Achievement of Students with Disabilities: A Foundation for Research. *The Journal of Technology, Learning and Assessment*, *10*(5) at: https://ejournals.bc.edu/index.php/jtla/article/view/1605; P.L. 114-95, Section 1111, (b)(2)(B)(vii)(II); Dolan, R.P., Burling, K., harms, M., Strain-Seymour, E., Way, W. (Denny), & Rose, D.H. (2013) *A Universal design for Learning-based Framework for Designing Accessible Technology-Enhanced Assessments* at: http://images.pearsonclinical.com/images/tmrs/dolanudl-teaframework_final3.pdf

From: Chester E. Finn, Jr
To: NAGB Queries
Subject: NAEP Science Framework
Date: Wednesday, September 8, 2021 3:54:22 PM
Attachments: 2012-State-Science-Standards-NAEP-6.pdf

In response to your request for comments on the current NAEP science framework, I'm pleased to weigh in, both on my own behalf and that of the Thomas B. Fordham Institute. We formally reviewed that framework in 2012 in connection with a wide-ranging Fordham examination of state K-12 science standards. This led to an A-minus grade for the NAEP framework from our reviewers (led by the distinguished biologist Paul Gross). This included a maximum score of 7 out of 7 for the framework's "content and rigor." You can see that review at
http://edexcellencemedia.net/publications/2012/2012-State-of-State-Science-Standards/2012-State-Science-Standards-NAEP.pdf
and I attach a copy with this note.

Here's how we explained our decision to review the NAEP framework side-by-side with the standards of 50 states and DC: "The National Assessment of Education Progress (NAEP) is the most-often used barometer of student learning in science. Results from NAEP are used to compare student achievement across states and to judge states' student proficiency levels. Because NAEP is so central to the conversation on state and national science achievement, we felt it was important to analyze the quality of its implicit standards—embodied in its assessment framework—to see how they compare with the quality of each state's standards."

I should note that most state standards fared dismally in that review--only a handful got top marks.

Which leads me both to underscore the singular importance of NAEP and its frameworks as pacesetters and academic gold standards, and to say that the document you're starting with is very, very strong in its present form. As the old saying goes, if it ain't broke....It may well need some updating but the National Assessment Governing Board should think long and hard before undertaking a wholesale overhaul or replacement.

Thanks for your consideration.

Chester E. Finn, Jr.
Distinguished Senior Fellow & President Emeritus, Thomas B. Fordham Institute, and Senior Fellow, Hoover Institution, Stanford University

1016 16th Street NW, 8th floor, Washington DC 20036
Website: www.fordhaminstitute.org

SCIENCE

NAEP

GRADE	SCORES		TOTAL SCORE
A-	Content and Rigor	7/7	9/10
	Clarity and Specificity	2/3	

Overview

The NAEP *Science Framework* for science is an extended statement of science learning expectations at grades four, eight, and twelve. The NAEP assessment is based on the science content, skills, and testing procedures outlined in the *Framework*. Sample questions show how learning expectations discussed in the *Framework* are actualized in the assessment.

Although the *Framework's* design and organization are complex and in a few places difficult to understand, in general the document works well, providing a useful epitome of K-12 science knowledge and related skills.

There are two main issues to be addressed in evaluating this *Framework*. One is length—the number of content expectations that it includes is substantial, even though limited to three grade levels. The second is purpose: How may we evaluate this *Framework*, which is conceived as a design for testing, as a set of standards that can guide curriculum making? Early in its 155 pages, the *Framework* makes this important distinction between content and curriculum:

> **Key *principles as well as facts, concepts, laws, and theories* that describe regularities in the natural world are presented…as a series of *content statements* to be assessed at grades 4, 8, and 12…[T]hese statements comprise the NAEP science content. *They define only what is to be assessed by NAEP and are not intended to serve as a science curriculum framework.* (emphasis added)**

The writers are to be congratulated for having taken the trouble thus to define "content" as used by them. Yet although the *Framework* is not intended as a comprehensive set of standards for K-12 science, it clearly does *imply* such a set. In fact, it is unlikely that state education officials, district administrators, and teachers will ignore its plentiful science content and proposed achievement levels, particularly in light of the strong influence that NAEP and its assessment results carry in American primary and secondary education. Thus, we treat the NAEP *Science Framework* here as a set of expectations for K-12 science knowledge—a.k.a. science content standards.

Organization of the Framework

NAEP sidesteps enduring debates over how to define scientific relationships among themes, principles, content, practices, scientific reasoning, inquiry, and so forth by

Document(s) Reviewed

- *Science Framework for the 2009 National Assessment of Educational Progress.* 2009. Accessed from: http://www.nagb.org/publications/frameworks/science-09.pdf
- *NAEP Science Sample Questions: Grade 4.* 2009. Accessed from: http://nces.ed.gov/nationsreportcard/pdf/demo_booklet/09SQ-O-G04-MRS.pdf
- *NAEP Science Sample Questions: Grade 8.* 2009. Accessed from: http://nces.ed.gov/nationsreportcard/pdf/demo_booklet/09SQ-G08-MRS.pdf
- *NAEP Science Sample Questions: Grade 12.* 2009. Accessed from: http://nces.ed.gov/nationsreportcard/pdf/demo_booklet/09SQ-G12-MRS.pdf

Figure 1. Crossing content and practices to generate performance expectations

		Science Content		
		Physical Science Content Statements	**Life Science Content Statements**	**Earth and Space Sciences Content Statements**
Science Practices	Identifying Science Principles	Performance Expectations	Performance Expectations	Performance Expectations
	Using Science Principles	Performance Expectations	Performance Expectations	Performance Expectations
	Using Scientific Inquiry	Performance Expectations	Performance Expectations	Performance Expectations
	Using Technological Design	Performance Expectations	Performance Expectations	Performance Expectations

dividing science knowledge into just two broad categories: principles and practices. The various principles comprise what is usually called science content: facts, concepts, theories, and laws. They are organized into the now-familiar content areas: physical, life, and earth and space sciences.

Next, NAEP identifies four science practices: identifying science principles, using science principles, using scientific inquiry, and using technological design.

Finally, the *Framework* designers assemble all three areas of general content (principles and their expansions) and all four general areas of practice into a matrix. Each resulting cell of this matrix is a potentially large set of performance expectations (see Figure 1). Thus for every general content area, there are four possible (and testable) practices corresponding to the -*ing* actions listed: 1) recognizing, naming, or describing the content; 2) employing the content correctly in one of its contexts; 3) showing skills needed to use that content in answering a scientific question, and 4) applying the content in a design or engineering problem.

Organization of Content Topics

Within the three main content domains (physical, life, and earth and space), how many standards do K-12 students really need to meet? In science education, at present, this is a vexed question. Some say "very few." Others say "enough to display, at least, the *range* of modern science." Still others would answer "a whole lot." NAEP settles somewhere in the middle by expanding its three content areas into eighteen foundational statements: six on physical science, five on life science, and seven on earth and space science. These are then further specified by various detailed explanations encompassing most of the basics at each assessed grade level (four, eight, and twelve), but increasing in number, sophistication, and detail from fourth grade through twelfth grade.

The physical science content area illustrates this complex structure. It is divided into six basic principles: properties of matter, changes in matter, forms of energy, energy transfer and conservation, motion at the macroscopic level, and forces affecting motion. These six principles are represented by fifteen actual content statements in fourth grade, by sixteen statements in eighth grade, and by twenty-three statements in twelfth grade. Therefore, all assessable physical science is represented in this *Framework* by fifty-four short statements of science content.

Moreover, these content statements are amplified at each grade. For example: One of the six principles of physical science is "changes in matter." In fourth grade, this principle is represented by one explicit content standard—that cooling and heating can convert matter from one recognizable state (solid, liquid, or gas) to another. In eighth grade, "changes in matter" expands to two representations, one on the molecular organization of matter and the other on chemical reactions and the conservation of mass in the course of reaction. And by twelfth grade, this principle expands to three (carefully crafted) statements, one on the energetics of state change, a second on atomic structure and electrons in atoms, and a third on chemical bonds and reactions.

In addition to the fifty-four content statements for physical science, there are thirty-two for life science and thirty-nine for earth and space science—a total of 125 explicit content statements. Since all the assessable content of K-12 science is supposed to be covered, that is not an unreasonable number.[1]

Content and Rigor

Physical Science

Content statements for fourth-grade physical science are comprehensive and emphasize properties, states, and transformations of matter. They address adequately the basics of energy and motion in grade-appropriate terms. Content statements for eighth-grade physical science—concerned with physical and chemical change—are more specific and comprehensive than are our own criteria (see Appendix A). For twelfth grade, content is strong except for light treatment of some important advanced topics of twelfth-grade chemistry (reaction mechanisms, acid-base chemistry, chemical bonds in important classes of macromolecules). Overall, the physical science content presented covers the necessary ground with neither critical omissions nor trivialities.

Earth and Space Science

The earth and space science content is well chosen. Content and sequencing concerning Earth's internal structure and plate tectonics—including the key geological evidence from seafloor spreading—are analytical and sufficiently comprehensive. For the principle "earth in space and time," the single fourth-grade expectation appropriately concerns the distinction between slow and catastrophic change. Fossils appear in eighth grade, as do mountain building and erosion. Twelfth-grade expectations expand to include, among other topics, the scale and magnitudes of geologic time. Perfect science standards would give more attention to the earth's age and to stellar evolution (as exemplified in the Hertzsprung-Russell diagram). The *Framework* gives weather and climate unusual prominence, but at the expense of astronomy and cosmology. That said, the development of scientific ideas is generally appropriate throughout the grades, and the few omissions are compensated for by careful presentation of the included content.

Life Science

Life science coverage is broad and reasonably inclusive. Basic themes—such as the mechanisms of heredity—are represented (as they should be) at all three grade levels. But "evolution and diversity," central to modern biology, does not appear until eighth grade—and some even of its simplest elements not until twelfth grade. Even then, there is no mention of the now-indispensable molecular and population genetics relevant to evolution. Somewhat disproportionate attention is paid to ecology and ecosystems (here under the thematic head of "interdependence"), and that comes at the expense—inter alia—of physiology, control systems, and developmental biology. Basic cell biology, on the other hand, is very well covered and is sequenced thoughtfully by grade.

The *Framework's* principles and detailed content statements cover virtually all the expectations spelled out in our review criteria and introduce no significant peripheral matter. A full-credit score of seven out of seven for content and rigor is justified. (See Appendix A: Methods, Criteria, and Grading Metric.)

Clarity and Specificity

This *Framework* document concedes—as it must—that distinctions among its four basic practices are anything but sharp. They are nevertheless convenient for communicating skill expectations and for representing the underlying standards that must guide writers of test questions. The authors are evidently comfortable with the residual ambiguities, perhaps judging that they do not damage the implied standards. They make possible, presumably, the construction of fair and comprehensive tests, which is of course what the *Framework* is about. Nevertheless, while the total number of principles is appropriate, the potentially dense intersections of them and the practices (that is, the total number of principles as expanded grade by grade, multiplied by the four broad and not sharply distinguishable practices) make it difficult for a reader to comprehend a bounded set of expectations. Thus clarity is to some extent compromised by complexity; as such, the *Framework* is awarded a score of two out of three for clarity and specificity. (See Appendix A: Methods, Criteria, and Grading Metric.)

[1] The *Framework* reports that content selection was guided primarily by two national sources: the *Benchmarks for Science Literacy* of the American Association for the Advancement of Science (1993) and the *National Science Education Standards* of the National Research Council (1996), plus follow-up documents. The authors note, however, that those documents do not limit or prioritize content in the form of assessable units. (In fact they are often concerned with history, philosophy, and sociology of science.) The NAEP *Science Framework* concerns itself with "science" as commonly understood. And its tabulated content is justified and supported by clarifications and discussions of "crosscutting"—content relevant to more than one of the three science domains.

9115 Westside Parkway
Alpharetta, GA 30009
888.413.3669
cognia.org

Lesley Muldoon
Executive Director
National Assessment Governing Board
U.S. Department of Education

Dear Ms. Muldoon,

These comments are submitted by Cognia, a global non-profit education company, in response to the request for preliminary public comments for the Science Assessment Framework for the 2028 National Assessment of Educational Progress (NAEP). The comments submitted by Cognia focus on science frameworks and equity in the development of assessments.

Cognia has served as a trusted partner for over 125 years, aiding education providers in providing and advancing the pathways of success for all learners, supporting continuous improvement and accreditation. In addition, for nearly forty years, Cognia has delivered high-quality assessment services in support of student learning and growth, and accountability for both general education students and students with significant cognitive disabilities. Cognia is a leading provider of custom-designed assessments, specializing in a full range of text test development activities.

Cognia's team is diverse and expansive with expertise and experience in assessment, accreditation, certification, systems thinking, continuous improvement, school turnaround, and professional learning to provide comprehensive, aligned, and innovative services. We serve education organizations at every level from state agencies and large school systems to individual schools, leaders, and teachers. Cognia is committed to ensuring every child has equal access to learning opportunities and resources. This process begins with helping our institutions address the complex issues related to diversity, equity, accessibility, and inclusivity through quality of education.

Cognia is leading efforts to address the history and legacy of racism in educational assessment through development of *A Call to Action: Confronting Inequity in Assessment (Lyons, Johnson, and Hinds, 2021)*. Working closely with Lyons Assessment Consulting, several authors from Cognia contributed to this paper, which provides a strong foundation for the work Cognia is doing with respect to diversity, equity, accessibility, and inclusion. *A Call to Action: Confronting Inequity in Assessment* offers deep dives into five opportunities for centering the principles of diversity, equity, accessibility, and inclusion in the design and use of educational assessments. Problems related to equity are not limited to those of racial injustice, but the authors focus this document primarily on race-related issues in the hope that dismantling such structures will provide pathways for addressing other marginalized communities in our society generally and in educational assessment specifically. The Call to Action is designed to foster meaningful conversation and innovative ideas for advancing practice in educational measurement and improving our assessments to help move us toward a more equitable future. As an organization, we are dedicated to supporting our institutions in their improvement of what they do to help students learn.

9115 Westside Parkway
Alpharetta, GA 30009
888.413.3669
cognia.org

The comments below have been compiled from our experts in content development, measurement services, and equity and transformation learning services.

Cognia Recommendations for Revisions to the NAEP Science Framework

As a "key measure in informing the nation on how well the goal of scientific literacy for all students is being met," the NAEP Science Assessment should be based upon the standards, instruction, and research in science education most immediately influencing the nation's science classrooms. It should also embody culturally relevant assessment practices, to ensure representation and fair evaluation of all student groups. While we have several clear recommendations for necessary revisions of the content elements of the current NAEP Science Framework (National Assessment Governing Board, 2019), we feel it imperative to begin our recommendations on the point of equity, diversity, accessibility, and inclusion. The necessity of attention first and foremost being placed on creating an equitable science assessment framework cannot be overstated in order to support all students in learning science.

Rationale for an Equitable Science Assessment Framework

A new equitable science framework would emphasize diversity, equity, accessibility, and inclusion to support learning, increase engagement, and provide visible representation in content with a goal to improve diversity in representation of underrepresented groups in science fields of study and the workplace. This framework would consider students as the focal point and include meaningful interactions and feedback loops with the community as reflected by the students' contexts and communities.

An equitable science framework is a commitment to serving *all* students throughout the assessment design, development, and implementation process. This framework would ensure that underrepresented students are visible in curriculum and assessment content and would provide opportunities to create culturally relevant approaches for students from marginalized groups, particularly students of color, students living in poverty, and non-male identified students. Increased student (and community) engagement, especially from underrepresented groups, will expand opportunities for equitable representation in advanced studies in science fields and the workplace.

Culturally relevant assessment practices are supported by the sociocultural perspective on how students learn. Making sense of new learning concepts is developed and maintained by mental schema, and we integrate new knowledge by searching for meaning and relevance, building on our prior understandings organized in mental structures informed by our lived experiences and social interactions (National Academies of Sciences, Engineering, and Medicine, 2018). Culturally sustaining assessment validates the cultural embeddedness of learning and explicitly attends to the sociopolitical reality of students in marginalized populations. It affirms their cultures and identities, creates counter-narratives, and ultimately builds student agency for understanding, critiquing, and confronting systems of social

9115 Westside Parkway
Alpharetta, GA 30009
888.413.3669
cognia.org

injustice *(Lyons, Johnson, and Hinds, 2021)*. When students are at the center of assessment, students are reflected in the curriculum and assessment content.

Creating a practice for understanding diverse learners and connecting them to science activities includes outreach and engagement with families and community members. This begins with the assessment development process, curriculum integration, and solving real problems. A community issue and/or problem can be framed within the context of an informal or formal learning community that includes multiple stakeholders such as learners, educators, local community members, businesses, and other nonprofit organizations. Embedding this within an equitable framework will increase community connection to scientific practice and data, and support the inclusion of participation from communities that have not had an adequate voice in the scientific educational process.

Growth Mindset Approach

A growth mindset is the belief that learning skills and qualities are on a continuum and can be developed through effort and support from others. A growth mindset can be cultivated in the classroom environment with students and educators, as well as with parents and guardians.

In a recent growth mindset study by PISA (2021), students who present a growth mindset score higher than their peers with a fixed mindset. People who consider their ability to be malleable (a growth mindset) will strive to develop it by setting challenging learning goals. They consider effort an inherent part of the learning process and setbacks to be fruitful experiences to assimilate…This leads them to stretch and expend efforts to reach their full potential whereas people with a fixed mindset are more likely to develop a hunger for approval that restricts them to their comfort zone (Dweck and Yeager, 2019).

Growth mindset can be leveraged as a strategy to support students of color and underrepresented students by reflecting growth mindset approaches in the language used in the framework in order to increase learner self-efficacy and motivation to learn from mistakes, and expand scientific skills centered on real world/life problem solving and knowledge. This also supports centering an approach for encouraging students to engage with science within the context of the framework.

Revising Development Processes to be Centered on Equity

In operationalizing an equity science assessment framework, the development process must be updated to include the long-overdue centering of students in assessment and meaningful engagement of stakeholders who are representative of student populations served by NAEP. Exhibit 1 illustrates an updated process of equitably generating assessment items and tasks and interpreting student responses that includes these commitments. Stakeholders include parents/caretakers, community members, and perhaps high school students and younger students.

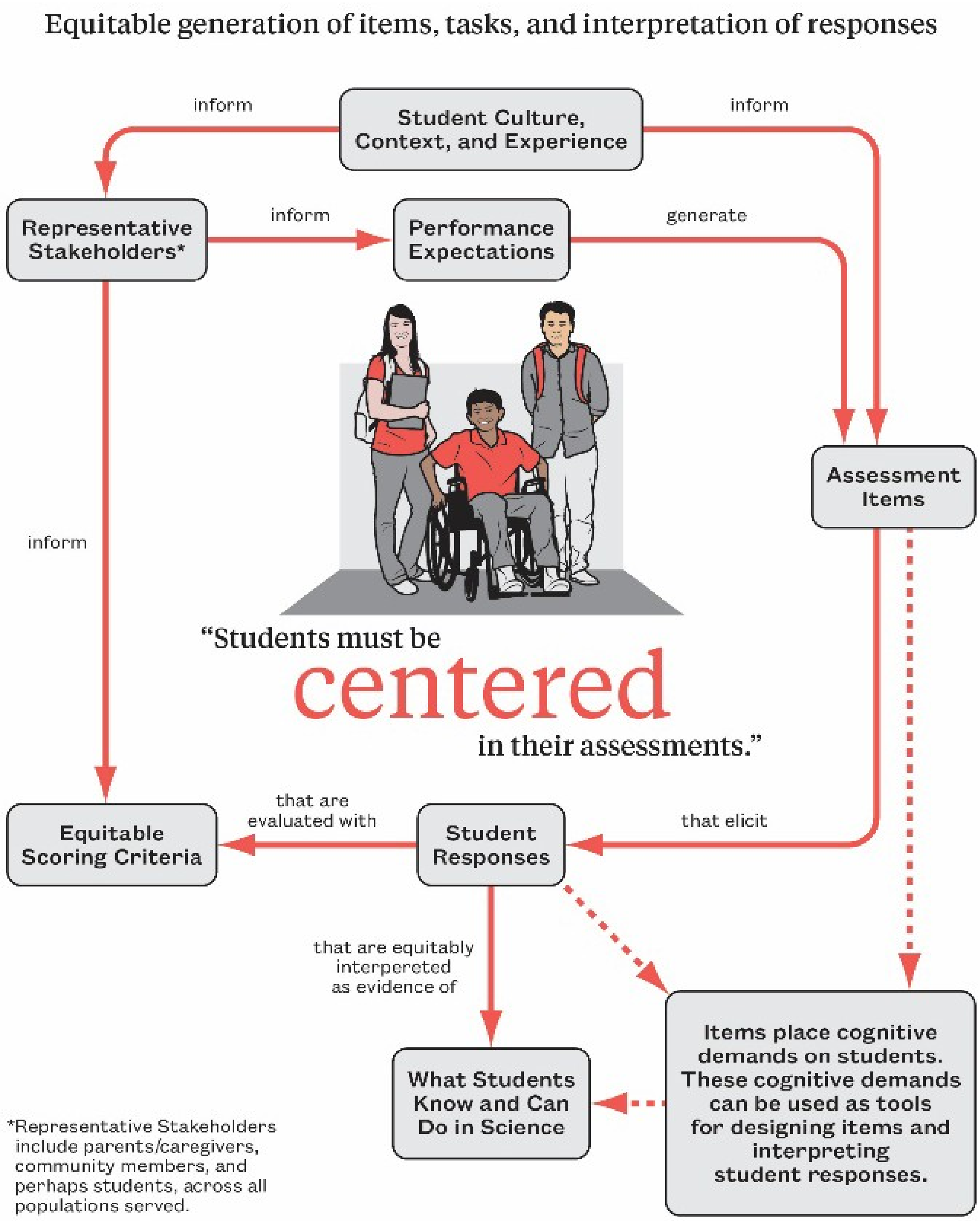


Exhibit 1: Student centered assessments.

An item or task is an individual question or exercise on the NAEP Science Assessment and is used to gather information about students' knowledge and abilities. Items and tasks are anchored in well-informed performance expectations, which describe in observable terms what students are expected to know and do on the assessment.

cognia™

9115 Westside Parkway
Alpharetta, GA 30009

888.413.3669
cognia.org

As shown in Exhibit 1, students must be at the center of any assessment of their progress. Their cultures, contexts, and experiences must inform the development of assessment items and tasks and the understanding and actions of representative stakeholders who are involved in the development process. In turn, representative stakeholders are involved in the creation of performance expectations by providing input regarding the cultural relevance and responsiveness of the expectations, including how to connect the performance expectations to students' lived experiences (e.g., relevant phenomena). These equitable and inclusive performance expectations guide the development of assessment items and tasks.

The cognitive demands and cultural relevance of assessment items and tasks can then be used to interpret students' responses as evidence of what students know and can do in science and how science concepts and skills relate to students' lives. Educators Shane Safir and Jamila Dugan cite the importance of developing assessments that reflect the mindsets and habits of professionals in the field and that "this shift from students as consumers of information to practitioners of field knowledge is especially significant for Black, brown and Indigenous students, signaling that they belong to a larger intellectual community (Safir and Dugan, 2021). The assessments that students encounter should include tasks that elicit authentic student performance to the extent practicable.

The development of scoring criteria for all student-constructed responses to items and tasks also actively involves representative stakeholder engagement, in order to ensure that all student populations are considered and represented in the scoring criteria. Exhibit 1 suggests that assessment development is both a multifaceted and iterative process, with significant consideration given to examining the equitable performance of assessment items across all tested populations as a compulsory part of the piloting process.

In evaluating item performance, in the Call to Action we suggest that examining differential item functioning (DIF) separately by gender, socioeconomic status, and race is now not only insufficient, but counter-productive in that cross-sectional views of item DIF are washing out the within-group intersectional effects (e.g., low SES Black females) (Russell, 2020). Class, race, ethnicity, language, and gender diversity are all possible influences on the manner in which knowledge is acquired and demonstrated on an assessment (Gordon, 1995). The field should be able to quickly move to detecting intersectional effects in estimates of cumulative test bias, or differential test functioning, particularly with the large sampling that NAEP is able to perform (Lyons, Johnson, and Hinds, 2021).

In summary, it is no longer enough to point to diversity, equity, accessibility, and inclusivity solely based on traditional approaches such as universal design, accommodation features, and classic DIF categories. While these approaches have their place, a true shift that starts with and maintains students at the center of the assessment is required for the NAEP Science Assessment to measure and reflect the science achievement of our nation's current students.

cognia™
9115 Westside Parkway
Alpharetta, GA 30009
888.413.3669
cognia.org

Constructs to be Assessed

The conditions that necessitated the revisions resulting in the *Science Framework for the 2019 National Assessment of Educational Progress* – namely publication of new science standards, advances in research, growth in innovative assessment approaches, and the need for increased inclusivity – are the same conditions that point to the need to revise the framework at present. While we assert that prioritization of diversity, equity, accessibility, and inclusion must be the driver of a new framework as the most critical lens for revisions, we have also identified several aspects of the assessed content that need to be reviewed and revised as well.

Since the publication of *A Framework for K-12 Science Education* and the *Next Generation Science Standards* (NGSS), almost all states have adopted the NGSS as their science standards or have developed science standards that are *Framework*- or NGSS-adapted. As was the case with the *Science Framework for the 2019 National Assessment of Educational Progress*, a change in the standards driving science curriculum and instruction clearly necessitates revisions to the framework again. The NAEP Science Framework needs to be updated to reflect the constructs presented in the NGSS, structured around the philosophy of three-dimensional performance expectations. Content, practices, and crosscutting concepts need to be redefined and aligned to match the way they are operationalized in the NGSS. We will elaborate on the considerations for each dimension more specifically in the following paragraphs.

Content (Disciplinary Core Ideas)

In this case, "content" refers to traditional disciplinary-based knowledge. The content in the NAEP Science Framework needs to be crosswalked with the Disciplinary Core Ideas (DCI) presented in the NGSS to redefine the appropriate set of content for the NAEP Science Assessment going forward.

While there is significant overlap for some concepts between the NAEP Science Framework and the NGSS, there are also many differences. Some content in the current NAEP Science Framework is not emphasized to the same degree in the NGSS, and likewise there are some concepts in the NGSS that are missing or sparse in the NAEP framework. As an example, in Physical Science, wave concepts and the connections between speed and energy are two content topics more prominent in the NGSS DCIs than the NAEP Science Framework; as another example, there is a heavy emphasis on motion graphs in the NAEP framework, whereas in the NGSS, motion graphs are not specifically codified into separate DCIs but are a part of the tools for evidence used by students to make claims about an object's motion or forces on an object. Similar examples appear in Life Science and Earth and Space Sciences as well.

Those revising the framework will also need to attend to any shifts in grade levels for content. Learning progressions should continue to underpin the content statements across grades in each domain, just as both the NGSS DCIs and the current NAEP Science Framework have done. To better reflect this

cognia™
9115 Westside Parkway
Alpharetta, GA 30009
888.413.3669
cognia.org

in the new framework, we recommend considering a coding scheme that does account for these progressions rather than the sequential numbering currently used in the NAEP Science Framework. Additionally, developers must be mindful in applying those learning progressions in item development to ensure there is understanding of the effect of cognitive complexity, practice, and crosscutting concept influences at each node of content along the progression, such that assessment items measure constructs as appropriate and intended for the grade level.

A very significant additional consideration related to grade levels is whether the NAEP elementary assessment grade should be changed from grade 4 to grade 5. While the *National Science Education Standards* organized the elementary grade band K-4, the NGSS created elementary standards by grade for grades K-5 and designated the middle school grade band standards for grades 6-8. A large number of states have redesigned their elementary science assessment to assess students at grade 5 instead of grade 4 in adopting NGSS or NGSS-like standards, and NAEP assessment designers should give serious consideration to doing the same as they examine the content to include in the framework.

In addition to the three traditional content areas of Physical, Life, and Earth and Space Sciences, the NGSS includes Engineering Design as a content domain. While the NAEP Science Framework addresses elements of engineering and technological design, it has been more so through the practices, and the framework revision will need to look at recategorizing and elevating Engineering Design as *A Framework for K-12 Science Education* and the NGSS do.

While the nationwide shift to NGSS-based instruction is argument in and of itself for revising the NAEP Science Framework, the NGSS are also internationally benchmarked standards. In preparing to develop the K-12 Science Framework and the NGSS, Achieve completed an international benchmarking study of ten countries' science standards, including those countries who are consistent high performers on PISA and TIMMS. The current NAEP Framework acknowledges the importance of comparing expectations against international science education achievement expectations.

Practices

In defining the Science and Engineering Practices, the writers of *A Framework for K-12 Science Education* intentionally defined several targeted practices "to better specify what is meant by inquiry in science and the range of cognitive, social, and physical practices that it requires" (National Research Council, 2012). While the current NAEP Science Framework includes "practices," they are simply too broad to focus towards the specific expectations of current science instruction, and new practices need to be defined, aligned to the eight practices of the NGSS.

Some of the expectations within the four NAEP practices overlap with various NGSS practices, e.g., explaining observations and proposing and evaluating alternative explanations within Using Science Principles align with concepts for Constructing Explanations and Engaging in Argument from Evidence;

cognia™
9115 Westside Parkway
Alpharetta, GA 30009
888.413.3669
cognia.org

proposing and critiquing solutions, considering criteria and constraints, and identifying tradeoffs within Using Technological Design align with concepts for Defining Problems and Designing Solutions. However, there is much more interpretation and generality associated with the NAEP practices, which renders them insufficiently aligned to the expectations of current science instruction. Further, the first practice, Identifying Science Principles, would not be considered a practice according to the NGSS, and in fact should not be assessed. The NGSS set expectations for knowledge in use, and simply being able to recognize or recall facts is no longer sufficient for demonstrating proficient science achievement. Also, in regard to engineering practices, the NAEP Science Framework restricts assessment of design to only the science principles associated with the problem and does not include other considerations (e.g., economic, social) for the problem. This, however, contradicts the current need to build more relevant, equitable assessments that do engage students based on their lived experience and social justice. Some other assessments, such as PISA, seem to be more fully engaging with social and global problems, and NAEP assessment designers should do the same for equity, putting students at the center of the assessment.

Crosscutting Content (Crosscutting Concepts)

In the current NAEP Science Framework, “crosscutting content is not represented by abstractions such as ‘models,’ ‘constancy and change,’ or ‘form and function,’ but is anchored in the content statements themselves” (National Assessment Governing Board, 2019). This approach is quite opposite that of *A Framework for K-12 Science Education* as well as the *National Science Education Standards* and *Benchmarks for Science Literacy*, which defined crosscutting concepts (or unifying concepts and processes, common themes in *NSES* and *Benchmarks*, respectively) as more schematic approaches to science thinking, i.e., concepts having explanatory value via “an organizational framework for connecting knowledge from the various disciplines into a coherent and scientifically based view” (National Research Council, 2012). The NAEP Science Framework needs to pivot back to defining theme-based crosscutting concepts, which in fact was how they were represented in the 1996-2005 Framework. This shift is required to provide coherence and consistency between NAEP and current NGSS-based instruction, bringing the third dimension of the performance expectations into alignment. *A Framework for K-12 Science Education* defines seven crosscutting concepts, which should be the basis for redefining crosscutting concepts in the new Science Framework. If for some reason NAEP framework developers choose not to align to this definition of crosscutting concepts, they should name this concept something else in the new framework in order to avoid confusion for the field.

Additional Recommendations for Revising the Science Constructs to be Assessed

As the next set of framework constructs are created, the wording of each statement needs to be carefully reviewed to detect and eliminate bias and to ensure inclusivity. Some current content statements

9115 Westside Parkway
Alpharetta, GA 30009
888.413.3669
cognia.org

are biased and not inclusive – for example, “manmade,” “heavenly body,” etc. The new framework needs to clearly avoid such phrasing.

In tandem with updating the constructs to be assessed in the next framework, we encourage NAEP assessment developers to be thorough in updating the accompanying specifications documentation. We recommend including a significant amount of explicit information around clarifications and assessment boundaries, as this level of detail is in our experience extremely useful in ensuring assessment items measure the constructs as intended. Further, we recommend including examples of grade-appropriate phenomena for the assessed content in the specifications, although it should be made clear that the examples are not an exhaustive list and analogous phenomena should also be used in assessment development. Many of those examples, or similar examples, as well as assessment items should continue to be included in the framework itself, to provide direct illustration of how the framework constructs and assessment design will be operationalized.

The framework and specifications should also document clear methodology around the creation of performance expectations for NAEP assessments, given that the crosses of DCIs, SEPs, and CCCs (assuming they are adopted) will yield a far greater number of possible combinations than the crosses of content and practice in the current NAEP framework. At present, states vary on the approach of assessing any possible combination of the foundational dimensions of the standards versus assessing only the specifically crossed performance expectations defined in the NGSS. Given that NAEP has a different purpose than a state accountability assessment does, we propose that continuing to be more generalized may better reflect the variety in format and instruction of the standards across the nation, as well as the holistic way instruction should occur, and would provide the opportunity to measure a range of applied performances that students can do. Whatever methodology is chosen, clear definition of the blueprint that any given NAEP assessment’s performance expectations must meet will be paramount in the design and interpretation of the assessment and results. NAEP developers must be extremely transparent and explicit about the interpretations – and non-interpretations – of the assessment results based on the defined methodology in comparison to each particular state’s standards and approach.

It will be important for NAGB to select an organization well-versed in the NGSS and the advances in science education research to do the work around construct revisions, and this organization should be continually executing on a strong mission in support of diversity, equity, accessibility, and inclusion. NAGB should also connect with members of the National Research Council of the National Academies for advisement on the status of NGSS implementation and any revision considerations for the NGSS. The time lag between framework revisions and the first NAEP assessment to be aligned to a new framework is significant and given that the NGSS are almost nine years old already, any effort to ensure the NAEP Science Framework is not outdated before it even comes into use, both in terms of science content and student representation, will be extremely important.

9115 Westside Parkway
Alpharetta, GA 30009
888.413.3669
cognia.org

Item Types and Assessment Design

Based on the changes we have recommended to the constructs to be assessed, we offer additional recommendations relative to the NAEP assessment design to best support these proposed changes, beginning with overall assessment design principles and progressing to specific blueprint and item type feedback.

The very first steps in a principled approach to assessment design and development are to clearly define the assessment targets (for which we have made recommendations in the previous section) and to define intended score interpretations and uses (SIUs). We recommend, based on the proposed construct revisions for the new NAEP Science Framework and the known variations in the structure and implementation of NGSS-based standards and curriculum across the nation, that NAEP assessment designers take the time to very intentionally and explicitly define the SIUs for the forthcoming NAEP Science Assessments based on the new framework. There must be a clear, common understanding of what the new NAEP assessment is really telling the nation about its students and their achievement in science – accompanied then by transparent, emphasized, public messaging of the SIUs – in order for assessment results to be meaningful and actionable.

An associated piece in these first design steps, which follows defining the assessment targets and coordinates with a model of cognition or learning to guide the assessment design, is considering the framework to be used for cognitive complexity. Achieve has published ideas for reconceiving cognitive complexity for the NGSS (Achieve, 2019), which depart from Webb's Depth of Knowledge model (used by many states, though not by NAEP in science) and press for more depth than the four-level scheme used by NAEP for science. As previously noted, the lowest complexity level that focuses on identification and recall really no longer meets the bar for adequate science literacy and achievement. Items that only assess declarative knowledge should not be included in the assessment, or only included to the most limited extent. Given these considerations, we encourage framework developers to explore new schemes for cognitive complexity. We would also encourage conducting cognitive labs to probe the validity of the chosen new scheme as applied to science assessment items.

After these foundational design steps are completed, we offer the following additional recommendations for more detailed designing of the new framework and assessment:

- Continue to ground all assessment items in science phenomena and engineering design problems. The focus on sense-making around phenomena and designing solutions to problems is the heart of the vision for science education in *A Framework for K-12 Science Education* and is what we now aspire to for our students. Associated with this, there is abundant opportunity to continue to integrate, and even more fully integrate, the Nature of Science into assessment items. Intentional care must be taken to represent this lens and all phenomena in items authentically, however, rather than simply provide "window dressing" to declarative items. The illustrative item on page 97 of the current NAEP Science Framework is a prime example; the response demands of the item are completely separate from the framing of the history and nature of science. The

9115 Westside Parkway
Alpharetta, GA 30009
888.413.3669
cognia.org

new framework and the assessment items that it directs should require application and sense-making of the stimulus material for the response.

- In adopting recommendations made in this commentary, the distribution of content areas and cognitive complexity in the assessment will have to be revised as well. The NGSS has a different weighting of content in the standards by grade level, and we have already provided reasoning around revising cognitive complexity schemes and weighting in the assessment.
- The item types being used, and the distribution of those item types, must also be reevaluated. Given the increased complexity of the NGSS, a significant reliance on multiple-choice items may no longer be sufficient to fully assess the science constructs as intended. We anticipate the need to place greater emphasis on constructed-response items and leverage more item clusters, POE items, and performance tasks, as well as introduce technology-enhanced items (e.g., drag-and-drop items, graphing interactions). Some additional elaborations on recommendations for various item types are as follows:
  - POE items have significant relevance to NGSS with their strong emphasis on evidence and reasoning. We recommend utilizing POE items to a greater degree.
  - Item clusters, or even two-part items, can be used to assess constructs in greater depth, supporting valid measurement of students' sense-making. Branching items may also be useful to further pursue for this purpose, with potential to gauge depth of understanding and ability to sense-make around a phenomenon. Leveraging the ability online to lock responses and then update those students who cannot move far into a branching set with correct information and allow them to continue on to additional questions may also be an area of measurement innovation to study.
  - We question the utility of concept mapping to some degree, relative to other item types, when considering the demands of the NGSS. Perhaps concept maps can be applied to specific phenomena presented, but we have concerns around the degree of inference that can be made without requiring students to provide evidence and reasoning for the links between concept terms in the map. More research on this item type may be necessary to support continued use.
  - Performance tasks are generally agreed upon as a necessity for authentic assessment of the NGSS. We see value in both hands-on performance tasks and interactive computer tasks. There may be ways to leverage technology to enhance what can be measured with hands-on performance tasks, by controlling what information students provide and when they get additional information to respond to (e.g., students design and carry out an investigation, record information online about their procedure and results, and then responses are locked before students are presented with a correct procedure and result to interpret). Hands-on tasks will be well-suited to assess both scientific investigation and engineering design. Interactive computer tasks will continue to allow assessment of constructs that can't be investigated in a hands-on manner and/or with reasonable economy. We would recommend changing the assessment design parameters to include a task for *all* students in the new science assessment, however, given the highly authentic match to the new constructs that need to be assessed. We also recommend carrying out the previously proposed study to compare the hands-on performance tasks and interactive computer tasks.

9115 Westside Parkway
Alpharetta, GA 30009
888.413.3669
cognia.org

   - In considering equity, assessment developers may want to explore what affordances there are for more response modes relative to the item types. Is it possible to leverage technology and administration to support more students in providing responses in a mode that best allows them to show what they know and can do, for example, allowing recording of a spoken response rather than a typed response for a constructed response item?
- Ensure assessment development practices are aligned to the latest industry standards, as updated in the 2014 edition of the *Standards for Educational and Psychological Testing*.

As cited in the current framework, "The NAEP Science Assessment signals the kinds of responses to tasks, problems, and exercises, along with the kinds of knowledge and reasoning, that should be expected of students as a result of what is taught in the science curriculum." We agree that the NAEP assessment has this impact, and we believe that the next revision of the science framework must therefore reflect the current efforts to center science instruction around *all* students through the NGSS. Throughout the current framework, there are elements that already resonate with and reflect principles that ground the content of the K-12 Science Framework and the NGSS, and the requirement now is to update the framework to be in clear alignment and thus measure science achievement relative to the new vision for science education being implemented across the nation.

Sincerely,

Stephen Murphy

Chief Learning Officer

References

Achieve. (2019). A Framework to Evaluate Cognitive Complexity in Science Assessments. https://www.achieve.org/files/Science%20Cognitive%20Complexity%20Framework_Final_093019.pdf

Dweck and Yeager. (2019). “Mindsets: A View From Two Eras”, *Perspectives on Psychological Science*, Vol. 14/3, pp. 481-496 http://dx.doi.org/10.1177/1745691618804166.

Gordon. (1995). Toward an equitable system of educational assessment. Journal of Negro Education, 360–372.

Lyons, Johnson & Hinds. (2021). A Call to Action: Confronting Inequity in Assessment. https://www.lyonsassessmentconsulting.com/assets/files/Lyons-JohnsonHinds_CalltoAction.pdf

National Academies of Sciences, Engineering, and Medicine. (2018). How people learn II: Learners, contexts, and cultures. National Academies Press.

National Assessment Governing Board. (2019). Science Framework for the 2019 National Assessment of Educational Progress. https://www.nagb.gov/content/dam/nagb/en/documents/publications/frameworks/science/2019-science-framework.pdf

National Research Council. (2012). A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas. Washington, DC: The National Academies Press. https://doi.org/10.17226/13165

Russell, M. (2020). Personal communication.

Safir, S. and Dugan, J. (2021). Street Data: A Next Generation Model for Equity, Pedagogy, and School Transformation. Thousand Oaks, CA: Corwin.

From: Heinz, Michael
To: NAGB Queries
Cc: Heinz, Michaekil
Subject: NAEP Science Framework
Date: Friday, October 15, 2021 9:33:25 AM

CAUTION: This email originated from outside of the organization. Do not click links or open attachments unless you recognize the sender and know the content is safe.

Thank you for the opportunity to provide comments and recommendations relative to the Science *Framework for the 2019 National Assessment of Educational Progress* (hereafter referred to as the *NAEP Science Framework*). I am submitting this document on behalf of the Board of Directors and the members of the Council of State Science Supervisors.
The Council of State Science Supervisors (CSSS) provides leadership in advancing excellence in P-12 science education at the local, state, and national levels. Our members include state science supervisors who are responsible for academic standards in science and/or statewide science assessments in 48 states. In addition to our state members, our organization includes researchers from institutions of higher education, experts from federal STEM mission-based agencies, and leaders from informal education organizations. Our members work both independently and collaboratively to ensure widespread, consistent, coherent opportunities for high-quality science learning is available to all students across K-12 and that people of all backgrounds are welcomed in science learning environments.

As science education leaders working at the intersection of local, state, and federal policies, we are most aware of the systemic value of coherence between state and federal assessment and the ability of CS3 to facilitate such coherence. Assessment tends to drive instruction and it can drive us forward or backward. Coherence between state and federal assessment will provide state leaders with another tool to improve science instruction for all students.

Recognizing the important role that NAEP science assessment data plays in decision making in states, territories, and at the Department of Defense Education Activity, **CSSS advocates for updating the *NAEP Science Framework*.** In this document we provide evidence to support our recommendation and describe some of the key components that should be a part of the revised framework.

In the announcement soliciting comments and recommendations, we were asked to focus on three questions. In the following section, we provide our responses.

Whether the NAEP Science Assessment Needs to be updated.

CSSS is a proponent for updating the *Science Framework for the 2019 National Assessment of Educational Progress*. Just as previous *NAEP Framework*s have been based on the latest research, so should be the *2028 NAEP Science Framework*. Two consensus studies of the National Academy of Sciences are most relevant to this include *Taking Science to School: Learning and Teaching Science in Grades K*-8 (2007), and *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas* (2013). A consensus study results from extensive research and deliberations by diverse groups including scientists, engineers, mathematicians, learning scientists, educational practitioners, and educational policymakers. The National Academies of Sciences, Engineering, and Medicine are acknowledged as the "Advisors to the Nation."

As of this writing, forty-four states (representing 71% of U.S. students) have science standards influenced by the *Framework for K-12 Science Education.* Quite simply, since the National Academies of Sciences, Engineering, and Medicine are acknowledged as "Advisors to the Nation', these reports are the best information available for how best to instruct our youth. And with a statistic of over 70% of U.S. students being taught using standards influenced by the *Framework for K-12 Science Education,* it makes sense as a focal point of measurement for coherency with American trends in science education.

If the Framework needs to be updated, why is a revision needed?

The current *NAEP Science Framework* has two separate components, science content and science practices. *Framework for K-12 Science Education* also defines distinct practices, core ideas, and crosscutting concepts—the difference is the expectation that they are integrated in instruction and assessment.

The current NAEP Framework is focused on research from the 1990's, upon which we have built considerable information. New research outlined in research like How People Learn II: Learners, Contexts, and Cultures (2018) provides further input regarding integration of content and practice for improved and more equitable outcomes. Students do not use their knowledge of content, practice and cross-cutting concepts in isolation of one another. The knowledge interacts in ways that provide scaffolding for recall, integration and problem solving in the context of a novel or repeat phenomenon(a). As noted by the Achieve Framework for evaluating cognitive complexity, artificially separating these cognitive processes in assessment does not provide us with an accurate or equitable measure of student proficiency in science. It is in our best interest to align our measures with instructional practice.

A second reason that a revision is needed is that *A Comparison Between the Next Generation Science Standards (NGSS) and the National Assessment of Educational Progress (NAEP) Frameworks in Science, Technology and Engineering Literacy, and Mathematics* found that differences in the depth, breadth, detail, or focus of that content resulted in low to moderate levels of content alignment, with differences by grade and content domain (2015).

Alignment with practices was strong, but the emphasis of NGSS performance expectations across NAEP science and TEL practices differed from the emphases specified in the NAEP frameworks.

What should a revision to the framework include?

Recommendation 1: Increased attention to equity. A new framework should include a renewed look at how science assessments reflect and includes features of equitable assessment. The COVID-19 pandemic shined a spotlight on inequities and unjust public education practices. As a result, many states have reconceptualized how they are working to make teaching, learning, and assessments more equitable for all students, including reconceptualizing how assessments are constructed, how diverse student experiences are represented in assessment tasks, and how students are able to make their thinking visible. The NAEP assessments have a long-standing history of representing the best of what is known about disciplinary assessment practices and revising the science framework to better represent equitable science assessment provides NAGB with the opportunity to continue to play this leadership role. As an organization that is not constrained by limitations created by statewide policies, NAGB should position itself to take up that work and to exemplify how large-scale assessments can provide equitable opportunities for all students to make their thinking visible.

Recommendation 2: Align to current shifts in state science standards. A new framework should also be responsive to, and a reflection of what states are doing with academic standards and statewide assessments. For example, there is a low level of alignment between the *NAEP Science Framework* and the disciplinary core ideas for grades K-5 defined in the NRC's *Framework*.

In Closing, a revised *NAEP Science Framework* should provide the nation with data that can be used to evaluate the effectiveness of states' efforts to make science education more equitable and meaningful for each of our approximately 48 million students.

CSSS stands ready to offer our considerable expertise and experience to assist with soliciting stakeholder feedback and to participate on an expert panel to support revisions to the *NAEP Science Framework*., as we did for the 1996-2005 and 2009-2015 NAEP Frameworks. As President of CSSS, I would be pleased to provide names and contact information for individuals to serve the NAGB.

Respectfully,

Michael Heinz
President
Council of State Science Supervisors

From: DANIELLE MURPHY
To: NAGB Queries
Subject: NAEP Science Framework
Date: Monday, September 27, 2021 2:59:18 PM

CAUTION: This email originated from outside of the organization. Do not click links or open attachments unless you recognize the sender and know the content is safe.

Hello,

I definitely think that the updated framework needs to include authentic reference to equity and justice. There is enough research showing that typical science knowledge and standards unfairly favor certain races and genders. To ignore research, and the public cost of doing so, is doing a disservice to students and cannot be considered a fair assessment.

I hope you consider ALL students when designing this assessment.

Sincerely,
Danielle

NCSE National Center for Science Education

September 9, 2021

National Assessment Governing Board
by e-mail

Dear colleagues,

I am writing on behalf of the National Center for Science Education, a non-profit organization affiliated with the American Association for the Advancement of Science and the National Science Teaching Association, with comments on the current NAEP Science Assessment Framework.

In NCSE's view, the NAEP Science Assessment Framework, while valuable in its time, needs to be updated now.

The primary reason to update the NAEP Science Assessment Framework is that its content was largely based on the National Science Education Standards and the AAAS Benchmarks for Scientific Literacy, which were then the most authoritative guides to science education. They have since been supplanted by the NRC's *A Framework for K–12 Science Education* (2012) and the Next Generation Science Standards (2013), both of which are considerably more up-to-date with regard both to science content and pedagogical methods. By now, twenty states (plus the District of Columbia) have adopted the NGSS, which are based on the NRC Framework, and a further twenty-four states have adopted state science standards that are based on the NRC Framework: it is fair to say that a majority of the nation's public school students are learning science more or less in the way envisioned by these documents.

A revision to the Framework should thus align it to the content and structure of the NRC Framework and the NGSS.

In addition, NCSE recommends that special attention be given to socially but not scientifically controversial topics—evolution, climate change, and vaccination in particular—and to the nature of science. For a variety of reasons, these topics are often neglected or inadequately treated in American science education, even in authoritative documents such as the NRC Framework and the NGSS. It would therefore be helpful to consult state science standards that improve on the NGSS's treatment of these topics, such as Massachusetts's with regard to evolution and Wyoming's with regard to climate change, and position statements from relevant professional scientific societies such as the Society for the Study of Evolution and the American Meteorological Society. While it is not realistic to expect students across the country to receive instruction conforming to best practices, it is counterproductive to make allowances for states that have chosen to undereducate or miseducate their students.

Sincerely,

Ann Reid
Executive Director, NCSE

230 Grand Avenue, Suite 101, Oakland CA 94610 | 510-601-7203 | info@ncse.ngo | www.ncse.ngo

September 20, 2021

Dear National Assessment Governing Board,

Please find below comments relevant to the potential update of the NAEP Science Assessment Framework. I am comfortable with my name and affiliation being included with my comments.

I submit these comments based on my experience as a former state STEM leader at the Massachusetts Department of Elementary and Secondary Education. During my almost 12 years at the state agency, I was a member of a design team for the *Committee on a Conceptual Framework for New K–12 Science Education Standards*, was a Writing Team member for the *Next Generation Science Standards* (NGSS), was a state representative to the Lead State NGSS review process (facilitated by *Achieve*), and I led state STEM standards development and contributed to state assessment development. I also participated in several rounds of alignment reviews between NAEP and emerging or current science standards, including as a member of the NAEP/NGSS Comparison Panel in 2014, facilitated by the National Center for Education Statistics, and more recently between NAEP Science and selected state science standards, facilitated by HumRRO in 2020.

At a broad level, I would encourage a future iteration of NAEP science to maintain and/or enhance the following elements:

- Hands-on performance tasks. Such performance tasks are fundamental to doing science and necessary to provide opportunities to demonstrate the application of science concepts and practices. While a logistical challenge, these are critical and should be continued and even expanded as possible.
- Interactive computer tasks. The tasks have provided for a wider variety of innovative scenarios and contexts for students to apply their knowledge and skills. They are also helping to advance state-level assessment through proven examples of interactive assessment items. These too should be continued and expanded as possible.
- Integration of science content and practices. Science requires integration and application of both science concepts and practices together, not individually. The assessment of these two dimensions within individual items and across assessments is critical. Even as content or practices may be adjusted, and the practical implementation of assessing both dimensions may change, the measure and integration of both these dimensions should be continued.

Based on my experiences with science standards and assessment development in the recent past, I would encourage an update of the NAEP Science Assessment Framework for the following reasons:

- Since the last NAEP science revision, the National Research Council published the Framework for K-12 Science Education, and many states have adopted or adapted NGSS. Both efforts provide an updated framework of what is important to learn in science education, including the set of science concepts and a significantly different set of science practices.
- The NRC and NGSS documents attend to recent research on progression of learning in science education. An updated NAEP assessment framework can both attend to those and potentially contribute to the further study and articulation of science progressions of learning through the generation of data useful to researchers.
- There is a significant need for additional attention to equity, both from a racial perspective and to account other diversity within student populations. We must ensure that future NAEP assessments do not unintentionally disadvantage anyone from demonstrating their ability to perform science.
- An updated Framework provides an opportunity to advance multi-dimensional assessments that account for both concept and practice proficiencies in innovative items, assessment structures, and statistical analyses. More explicit guidance or specifications on item and assessment development should be produced to guide future NAEP administration. In my opinion the integration of the two dimensions of science concepts and practices is a substantial accomplishment; the integration of three dimensions at once (the third being cross-cutting practices, as defined by NRC and NGSS) is confounding to designers and users alike.

The work undertaking with NAEP Science is hugely influential to states across the country, and ultimately to curriculum and classroom practice. As such, I highly encourage an update to the NAEP Science Assessment Framework, and I am very interested in supporting and participating in work to achieve such an update.

Jacob Foster
Founder, STEM Learning Design, LLC (www.stemlearningdesign.com)

From: Jacqueline Huntoon
To: NAGB Queries
Subject: NAEP Science Framework
Date: Tuesday, September 28, 2021 2:11:30 PM

Please find below responses to the questions that were posed to the science education community:

- Q: Whether the NAEP Science Assessment Framework needs to be updated.
- A: Yes

- Q: If the framework needs to be updated, why a revision is needed.
- A: It focuses too heavily on content and tends to exclude the science and engineering practices and the crosscutting concepts. It should place greater emphasis on students' ability to use tools (which may include data presented to them) to investigate phenomena and design solutions to problems. A different way of saying this might be that it needs to focus on determining whether or not students can USE science as a tool to develop their own understanding.

- Q: What should a revision to the framework include?
- A: It should place more emphasis on applying the practices and crosscutting concepts in a variety of situations. I would also like to see less disciplinary differentiation because the interesting and challenging problems in science are less and less likely to be confined to one particular discipline. Even the example given for 8th grade earth science (gravity and planetary motions) has as much to do with physics as with earth science. I am an admittedly strong proponent of problem-based instruction in which science is taught as an integrated whole rather than as a series of separate disciplines. I am certain the leadership is aware of the National Academies reports on designing assessments in support of the Framework for K-12 Education and the NGSS. Documents such as these could provide good guidance.

Dr. Jacqueline E. Huntoon, PhD, PG
Provost and Senior Vice President for Academic Affairs
Michigan Technological University
www.mtu.edu

From: Kelly Barber-Lester
To: NAGB Queries
Subject: NAEP Science Framework
Date: Wednesday, September 29, 2021 5:26:33 PM
Attachments: image004.png

CAUTION: This email originated from outside of the organization. Do not click links or open attachments unless you recognize the sender and know the content is safe.

Good afternoon,

In response to the request for feedback that was elicited via email, I am writing to share some input into the updating of the NAEP Science Framework.

Upon reviewing the document found here (https://www.nagb.gov/naep-frameworks/science/science-framework-feedback.html), I was struck and deeply concerned by the fact that the following words and phrases were completely absent:

- Equity
- Equality
- Inequality
- Racism
- Bias
- Scientific racism
- Prejudice
- Sexism
- Ethics

The term "race" is only present insofar as it is used to refer to student demographics for tracking sub-group assessment performance. "Culture" is only found once in the document, in reference to "the role science has played in various cultures"(p. 96). The term "harm" is used almost exclusively to refer to harm that could be caused to environments or ecosystems, and never in reference to the harm that has been caused by scientific pursuits (for example, the ways in which science has been "advanced" by experimenting entirely unethically on specific minoritized populations).

Furthermore, there is no discussion of bias or the mitigation of bias (cultural or otherwise) in terms of assessment, which is a well-established and ongoing concern in the field of education.

As it stands, the framework presents a vision and version of science as objective, neutral, and divorced from context and its unquestionably troubled history (and present) as it pertains to issues of inequity broadly, and specifically racism and sexism.

I hope that you will take these observation into account when updating the framework. Issues of equity must be explicitly included and addressed within this framework. Continuing to teach science devoid of its messy and often uncomfortable intertwining with issues of inequity and oppression may be attractive in its simplicity, especially to those that already see themselves and those like them represented positively in textbooks and in the discipline; that approach, however, ensures that we will continue to struggle with these same issues in science as we move forward.

Best wishes,
Dr. Kelly J. Barber-Lester

Kelly J. Barber-Lester, Ph.D.
Assistant Professor
Pronouns: she/her/hers **Learn more about pronouns.**
School of Education- Office 345
1 University Drive I P.O. Box 1510 I Pembroke, NC 28372O

"The world is before you, and you need not take it or leave it as it was when you came in."
- James Baldwin

From: Wray, Kraig
To: NAGB Queries
Subject: Science framework feedback
Date: Monday, September 20, 2021 11:51:19 AM

CAUTION: This email originated from outside of the organization. Do not click links or open attachments unless you recognize the sender and know the content is safe.

In reading the executive summary the things that stand out as important to the NAEP committee are: applying science to students lives, science literacy for all students, participation in society and work, and addressing local, national, and global challenges. IF this is truly the purpose and primary driving factors for science education and therefore science assessment, I can not see how making sure phenomena, explanation, and understanding of science can exclude cultural and community ways of knowing and applying science. No where does the executive summary mention equity and making the practices relevant to local communities and students. Yet when you think deeply about the items listed above, they necessitate cultural relevance. Having members of the board and other team members that are knowledgeable about multiple ways of knowing, the history of marginalization, and by having these goals be explicit in the mission are essential to the success of the program. If we want students to be successful in science learning and for that learning to be reflected in the NAEP assessment, the development of an assessment with an equity focus is imperative.

Kraig Wray

Kraig A. Wray, Ph.D. (he/him/his)
Postdoctoral Scholar
Pennsylvania State University
Department of Curriculum & Instruction

From: Mark Looy
To: NAGB Queries
Subject: FW: NAEP Science Framework
Date: Saturday, September 25, 2021 3:30:07 PM

CAUTION: This email originated from outside of the organization. Do not click links or open attachments unless you recognize the sender and know the content is safe.

Greetings. I represent a non-profit organization with several staff holding earned doctorate degrees in science from prestigious institutions (e.g., Harvard and Brown). We appreciate the opportunity to suggest revisions to the science framework, especially in building the critical-thinking skills of students when they examine both sides of a scientific debate.

We submit that state and local educators should ensure that their teachers recognize that discussion about controversial subjects can lead to a more robust learning experience. For one, this approach helps hone the critical thinking ability of students. Unfortunately, there is false belief that it is unconstitutional to teach criticisms of topics such as evolution, the earth's age, the reliability of dating methods, etc. In reality, the constitutional approach would not prohibit the censoring of scientific ideas that run contrary to accepted belief, especially when credentialed scientists have opposing views. The teaching of controversial ideas held by dissenting scientists is both legal and beneficial—and with historical success as time and time again the status quo in science has been challenged.

Now, do we believe teachers should be *required* to teach creation science or ideas thatsupport a younger age of the earth? No. Such a policy would be counter-productive, for those positions would likely be taught poorly by most evolutionary instructors. But teachers shouldat least have the academic freedom to teach alternative ideas that are being presented by scientists, even if they happen to be in the minority.

--Mark Looy, CCO, Answers in Genesis

Mark Looy
CCO/Co-founder, Executive Department

From: Michael Lowry
To: NAGB Queries
Subject: Re: NAEP Science Framework
Date: Thursday, September 23, 2021 2:31:35 PM

Dear Sir/Madam:

I agree the framework should be updated to better reflect where we are as science educators, specifically as it relates to incorporating engineering practice (as found in NGSS) and the cross fertilization that is happening in STEM. The problems we face as scientists and engineers require more than the usual silos of "life science, physical science and earth science." The urgency of climate change should also play a more prominent role in the framework.

Regards,

Michael Lowry

--

Michael J. Lowry, NBCT, PAEMST
Science Department Chair
The McCallie School
500 Dodds AV
Chattanooga, TN 37404

Ancora imparo

Corrections to the NAEP Framework

1) In **E.12.8**, the statement “Plates are pushed apart where magma rises to form midocean ridges, and the edges of plates are pulled back down where Earth materials sink into the crust at deep trenches” is incorrect. The rise of magma at mid-ocean ridges is a passive effect, and not an active one. This statement incorrectly implies that the magma is rising up from the mantle and is actively pushing the two sides of the oceanic plates apart. The opposite is true. Other forces are pulling the plates apart, creating a low-pressure zone along the axis of the spreading center, and this pulls up mantle rock from below to fill the void. Because of the phenomenon of *pressure release*, as the hot rock is pulled up from below, certain minerals exceed their solidus temperature and exsolve from the solid mantle peridotite rock, rising up to the surface as more fluid magma with a gabbro/basalt composition, and either erupts on the seafloor as basalt or crystallizes within the crust as gabbro. The evidence for this passive, rather than active, upwelling of mantle rock beneath midocean ridges is multiple. First, there are no deep roots to the thermal anomalies beneath ridges; these are shallow features. Second, the state of stress within oceanic lithosphere is indicative of a significant “ridge-push” force, but this name is somewhat misleading because the magnitude of the ridge-push force is actually zero at the ridge itself and in fact increases away from the ridge, a result of the thermal topographic swell of the warm mid-ocean ridge rock (essentially, the ocean lithosphere “surfs” down the thermal swell from the ridge). Third, repeated geodynamic computer convection modeling has shown that the circulation of mantle convection, of which plate tectonics is the surface expression, is nearly entirely driven by the sinking into the deep mantle of subducted ocean lithosphere, also known as the “slab-pull” force. Basically, because heat is generated internally within the earth through diffuse radiogenic production from a small number of long-lived radioactive isotopes (K-40, U-235, U-238, Th-232), the actual patterns of mantle convection, and therefore plate tectonics, is a result of the cooling and sinking of Earth’s surface and not the heating of Earth’s interior.

So, to fix this, please change this sentence to:
“Old oceanic plates sink into the mantle at the deep trenches of subduction zones, creating a patterns of tectonic plate movements. Oceanic plates are pulled apart at mid-ocean ridges, allowing magma to rise to form new oceanic crust.”

2) In **E12.3**: Change “Stars, like the Sun,” to “E12.3: Stars, such as the Sun,”
The word “like” means “similar to,” but similes are generally exclusive. A flashlight might appear “like” a star at night, but it is not a star. Here, we want to use “such as” to reiterate that our sun is a star.

3) In **E.8.10**: Change “Earth’s magnetic field is similar to the field of a natural or man-made magnet with north and south poles and lines of force” to
“Earth’s magnetic field is approximately similar to the field of a natural or man-made magnet with north and south poles and lines of force.”

In fact, a quick glance at maps of the actual inclination and declination of Earth's magnetic field will show you that, in fact, Earth's magnetic field is actually not at all like the dipolar magnetic field from a simple north-south magnet. This is because Earth's magnetic field actually has significant contributions from higher-order magnetic terms (quadrupole, octupole, etc.). In fact, these terms dominate near the core-mantle boundary, but because they decay more rapidly with distance than the dipolar field, the dipole is more than 90% of the field at Earth's surface. Nonetheless, Earth's magnetic field is MUCH more complex than a bar magnet or solenoid, so we need to qualify this statement with something like "approximately."

4) In **E.12.9** Change "Earth systems have internal and external sources of energy, both of which create heat" to
"Earth systems have internal and external sources of energy, both of which provide heat."

It is misleading to say "create" heat for two reasons. First, heat is the ***transfer*** of energy, distinct from the thermal energy that is a material property of Earth substances. Second, we repeatedly say that energy/mass is conserved, neither created nor destroyed, so it could generated misconceptions to say "create heat."

5) In **E.12.10** Change "This energy transfer is influenced by dynamic processes such as cloud cover, atmospheric gases, and Earth's rotation, as well as static conditions such as the positions of mountain ranges, oceans, seas, and lakes" to
"This energy transfer is influenced by short-term processes such as cloud cover, Earth's rotation, ocean circulation changes, and the distributions of atmospheric gases, as well as long-term processes such as changes in the positions of continents, mountain ranges, ocean basins, and lakes."

This statement is very misleading. There is nothing "static" about the positions of mountain ranges, oceans, seas, and lakes! A good portion of geology addresses how these are all constantly changing over time. Likewise, a large part of research and understanding of climate examines how climate responds and changes to the occurrence and locations of mountain ranges (which increase erosion, removing carbon dioxide from the atmosphere and pushing global climates to be cooler) and ocean basins (which control how heat is circulated around Earth's surface, and is therefore dominant in controlling regional climates). Also, the ocean science community is quite adamant that there is just one ocean (as there is just one atmosphere), so you should avoid saying "oceans" when you really mean "ocean basins." Also, given the prominence of ocean circulation in controlling both regional and global climate changes, you should call out ocean circulation as distinct from the locations of ocean basins.

6) In **E.8.14** Change "Water, which covers the majority of Earth's surface, circulates through the crust, oceans, and atmosphere in what is known as the water cycle" to

"Water, which covers the majority of Earth's surface, circulates through the geosphere, ocean, and atmosphere in what is known as the water cycle."

Again, there is only one ocean. More significantly, most of Earth's water (estimated to be about 5 ocean's worth) is in the rock of Earth's mantle. This water is constantly being pumped into the mantle along with the subducting ocean lithosphere. This water in the mantle is critical to Earth's geology; it significantly lowers the viscosity of mantle rock, actually allowing the mantle to convect. Venus does not have plate tectonics, and this is likely because it is dry and does not have water. This water is constantly reentering the atmosphere and ocean at subduction zone volcanoes after it dehydrates from the sinking lithosphere at depths that begin about 100 km down.

7) **pp. 87-88**: Good gracious! Your whole example of finding a location between the earth and moon that has the same value of gravity is TOTALLY WRONG! The gravity at the surface of the moon is about 1/6 of that at Earth's surface, but this has LITTLE to do with the equipotential location between them! This is significantly influenced by the different densities within the two bodies (which determines the location of the radius of the surface, which therefore determines the values of gravity at that particular location!) All that matters for the equipotential is their masses!

If we let the distance from the center of Earth to point C be "R," then we can define the distance from the center of the moon to point C to be some fraction of that, called k*R. The total distance from the earth to the moon is therefore R+kR, or (1+k)R.

To find point C, we need to equate the values of g:

gE = gM
so
GMe / r^2 = GMm / (kr)^2

The G's and r's cancel, so we have:
k^2 = Mm/Me = 7.35e22 kg / 5.97e24 kg = 0.0123
so k =0.11 and therefore the distance from the center of the moon to point C is:
= k/(1+k) = 0.11 / 1.11 = 0.10
So point C is very close to being 1/10 of the way from the moon to the earth and NOT 1/6!!!!

So, on page 88, change the "Interpretation" to:
"Interpretation: The correct answer is C. Because the Moon has a mass that is about 1.2% of the mass of Earth, a body that experiences an equal gravitational force from Earth and the moon should be much closer to the moon. Point C is the only point that is closer to the moon. Note: Point C is about one-twelfth of the way between the moon and Earth; it should be one-tenth of the distance."

[Also note: “the moon” should not be capitalized, just as “the earth” is not capitalized (although “Earth” correctly ***is*** capitalized).]

8) Why is this framework intentionally obsolete? There are lots of references to old and outdated NRC reports, but nothing from the 21st century? Why is the NRC’s *Framework for K-12 Science Education* omitted? Why are the *Next Generation Science Standards* omitted? A total of 45 U.S. states and D.C. are now using K-12 science standards that are adopted or adapted from the NGSS, but the rest (Florida, Texas, etc.) are using the eight Science and Engineering Practices (SEPs) of the NGSS. Why are the NGSS’s eight SEPs omitted and not even mentioned? It is almost as if you are intentionally trying to have this framework be irrelevant upon arrival?

Michael Wysession
Chair, NSF’s Earth Science Literacy Initiative
Chair, Earth and Space Science for the NRC’s Framework for K-12 Science Education
Chair, Earth and Space Science for the Next Generation Science Standards

Professor of Geophysics, Department of Earth and Planetary Sciences
Executive Director, Center for Teaching and Learning
Washington University in St. Louis
St. Louis, MO 63130

Dear Committee,

I have concerns that are listed below.

EXECUTIVE SUMMARY

In the rapidly changing world of the 21st century, science literacy is an essential goal for all of our nation's youth. Through science education, children come to understand the world in which they live and learn to apply scientific principles in many facets of their lives. In addition, our country has an obligation to provide young people who choose to pursue careers in science and technology with a strong foundation for their postsecondary study and work experience. The nation's future depends on scientifically literate citizens who can participate as informed members of society and as a highly skilled scientific workforce, well prepared to address challenging issues at the local, national, and global levels. Recent studies, including national and international assessments, indicate that our schools still do not adequately educate all students in science.

Science seeks to increase our understanding of the natural world through empirical evidence. Such evidence gathered through observation and measurement allows for an explanation and prediction of natural phenomena. Hence, a scientifically literate person is familiar with the natural world and understands key facts, concepts, principles, laws, and theories of science, such as the motion of objects, the function of cells in living organisms, and the properties of Earth materials. Further, a scientifically literate person can connect ideas across disciplines; for example, the conservation of energy in physical, life, Earth, and space systems. Scientific literacy also encompasses understanding the use of scientific principles and ways of thinking to advance our knowledge of the natural world as well as the use of science to solve problems in real-world contexts, which this document refers to as "Using Technological Design."

The National Assessment of Educational Progress (NAEP) and its reports are a key measure in informing the nation on how well the goal of scientific literacy for all students is being met. The *Science Framework for the 2019 National Assessment of Educational Progress* sets forth the design of the NAEP Science Assessment. The 2019 NAEP Science Assessment will use the same framework used in 2009. The 2009 NAEP Science Assessment started a new NAEP science trend (i.e., measure of student progress in science), and the 2019 NAEP Science Report Card will include student performance trends from 2009 to 2019. Trends in student science achievement were reported from 1996 to 2005 as well. However, the trend from 1996 to 2005 was not continued due to major differences between the 2005 and 2009 frameworks. The 2009 – 2019 framework represents a unique opportunity to build on key developments in science standards, assessments, and research. This document is

The most important component of Scientific Literacy is to understand,

reflect upon issues critically and explicitly, empowers the future citizens to engage in critical deliberation on science-based social issues

Scientific literacy for democratic decision-making, Hagop A. Yacoubian, *Pages 308-327 | Received 18 Jun 2016, Accepted 19 Dec 2017, Published online: 29 Dec 2017*

"in a year-long TCA program, researchers administered attitudinal surveys to understand the program's impact on two important aspects of scientific literacy: students' perceptions of science as important to society and personal decision-making, and student ability to carry out scientific practices." https://eric.ed.gov/?id=EJ1228452

Engels, Mary; Miller, Brant; Squires, Audrey; Jennewein, Jyoti S.; Eitel, Karla
Electronic Journal of Science Education, v23 n3 p33-58 2019

Comments must be submitted via email to nagb@ed.gov with the email subject header *NAEP Science Framework* no later than 5:00 p.m. Eastern Time on Thursday, September 30.

When providing comment, please indicate if you are not comfortable with your name and affiliation being included with your comments, which may be shared and discussed publicly in upcoming Governing Board meetings and materials.

If the Governing Board decides that an update is needed, the charge to launch the revision process for the NAEP Science Framework is anticipated at the March 2022 quarterly Board meeting. 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.

Michelle

Michelle M McCarthy, M.Ed.
Science Instructional Coordinator
Department of Teaching and Learning
Montana Office of Public Instruction

- Phone: 406.444.3537
- Mobile: 406.860.6619
- Email: MMcCarthy5@mt.gov
- OPI Science Website

Department of Middle and Secondary Education
College of Education & Human Development

Mailing Address:
P.O. Box 3978
Atlanta, GA 30302-3978

In Person:
College of Education & Human Development
6th Floor
30 Pryor Street
Atlanta, GA 30303

Phone: 404-413-8060
Fax: 404-413-8063

October 15, 2021

To the NAEP Governing Board,

We wish to offer our collective feedback for the upcoming revision of the NAEP Science Framework. As a group of colleagues, we represent a diverse range of disciplinary expertise and research interests while also sharing a commitment to the continued improvement of K12 education and teacher preparation. Further, we also share a strong commitment to the ever-increasing importance of both considerations of and actions towards developing equitable classrooms for learners from all communities, prioritizing minoritized communities that for too long have been underserved or relatively abandoned by many elements of the national K12 infrastructure. In light of this shared vision, we offer several broad considerations and relevant literature for the board to review and incorporate into the new NAEP framework.

To begin, we will directly address the three guiding questions offered by the board in their call for public comment during this process. Yes, the NAEP Science Assessment Framework needs updating since the visions, standards, and curricular approaches for science education across the nation have undergone significant restructuring and reorienting in their emphases since the last version was developed. Why should the framework be revised? Although the NAEP Science Assessment must remain “curriculum-neutral”, the shift in focus across much of the nation towards visions and standards that emphasize “three-dimensional learning” (National Research Council, 2012; NGSS Lead States, 2013). These significant shifts involve prioritizing students in active learning experiences where they engage in various scientific practices while using important and broadly applicable science concepts to make sense of various real-world phenomena. As such, this fundamental view of learning that is grounded in science practice necessitates assessments that reflect that emphasis as well. The current NAEP framework and the assessment structures that have resulted from it involving mostly conceptual recall multiple-choice questions do not align well with these more active visions of science education nor do the various conceptualizations of ‘inquiry’ in the previous framework.

Further, the forced nature of assessments that rely heavily on multiple-choice questions does not reflect the wealth of knowledge that has developed over the past few decades regarding Universal Design for Learning (Meyer, Rose, & Gordon, 2014). We agree with broad considerations offered by the CAST organization (2015) that all assessments should “support learner variability through flexible assessments using UDL guidelines” which would also include more variety and flexibility in NAEP assessment item structure and ways of accessing the NAEP Science Assessment for different learners. Following UDL guidelines, assessments should “eliminate unnecessary barriers in assessments” including, for example, thick reading passages that may present greater challenges for multilingual learners and not connect the lived experiences of many groups of learners (AERA/APA/NCME, 2014; CAST, 2015). Finally, assessments should also “assess engagement as well as content knowledge”, which remains necessary for the previously described visions for science education and for developing more equitable assessments (Wiliam, 2010; CAST, 2015).

The final question posed in the call is “What should a revision to the framework include?” The remainder of this letter will offer a broad overview of two critical areas for consideration that any meaningful Science framework revision will include in significant and explicit ways. We also point to several national-level reports and texts along with more specific empirical research and perspective articles that could support the revision teams’ work and the growth of the NAEP Science Framework in beneficial ways.

Science Learning through Science Practices

Reviewing the previous NAEP Science Framework, an obvious but critical change that is necessary involves extensive revision of the language, foci, and structure of the framework and assessment items in ways that more accurately reflect the current visions for science education that guide most districts and states in the country, including the *Framework for K12 Science Education* (National Research Council, 2012) and the *Next Generation Science Standards* (NGSS Lead States, 2013), as well as corollary texts that focus on assessment at all levels (National Research Council, 2014; Schweingruber, Beatty, & NASEM, 2017). Science teachers, researchers, and administrators tirelessly but thoughtfully work to shift the nature of instruction and learning experiences offered to students in science classrooms throughout the country. These shifts emphasize the foundational role of engaging students in a collection of specific practices that reflect the work of scientists as they endeavor to develop and refine scientific understandings of the world and universe.

As emphasis on these practices continues to grow, the distribution of item types and guidance language in the NAEP Science Framework needs to reflect those shifts as well. Such change requires the inclusion of more performance tasks and simulation-based tasks and less knowledge or conceptual recall items (NRC, 2014). Further, efforts in science classrooms and standards aim for students to not simply engage in these practices, but to also learn about how they function in the development of scientific knowledge (Ford, 2015). Therefore, the practices should also be viewed as science “content” so that items could be developed that assess students’ understanding of the function of the practices. For helpful reviews of the nature of these practices and how science education continues to emphasize their role in learning, we recommend Crawford (2014) and Osborne (2014) as supportive reading for the revision team.

Department of Middle and Secondary Education
College of Education & Human Development

Research and curricular innovation of the last decade heavily emphasized two explanatory practices in science, modeling and argumentation. Modeling as one of the central sensemaking processes in science has been well established over the past decade (Miller & Kastens, 2018; Wade-James, Demir, Qureshi, 2018). The development and use of scientific models set the foundation for students to construct scientific understandings of systems as well as predictions about new but related systems, while also affording explicit opportunities to expand students' learning about the nature of science as they engage in modeling (Schwartz, 2019). Many different curricular interventions that have gained popularity in classrooms across the country are grounded in this major scientific practice (Windschitl, Thompson, & Braaten, 2020; Windschitl & Thompson, 2013). Thus, modeling is a primary practice that constitutes an important component of the "content valued by the nation". As such, the development of new assessment items should be heavily connected to the modeling practices. These items can have students engage in interpreting representational and mathematical models while also using developed models to make predictions about systems.

For argumentation, much research and development work has established several considerations for how students and teachers learn through and about the practice (Henderson, McNeill, Gonzalez-Howard, Close, & Evans, 2018; Osborne, 2014). The goal for learning through argumentation involves supporting learners' ability to develop evidence through the analyses of various types of data collected from a range of investigations and phenomena and use core science concepts to reason with that evidence and develop claims in response to compelling investigative questions (Grooms, Sampson, & Enderle, 2018; Sampson, Enderle, & Grooms, 2013). The robust scholarship around scientific argumentation led to the development of several prominent curricular resources and instructional models (Hand, Wallace, & Yang, 2004; McNeill & Krajcik, 2011; Grooms, Enderle, & Sampson, 2015) that have been taken up by districts and schools across the nation, establishing this fundamental practice of science as further "content valued by the nation". Items aimed at assessing students' grasp of argumentation and their ability to engage in it could address evaluating the quality of evidence provided for a claim, evaluating the coordination between evidence and claims, describing appropriate science concepts that have been used to reason through evidence in support of a claim, and considerations of confirmation bias and other fallacies when engaging in arguing from evidence.

Other practices have not received as much research attention but are at the forefront of many science learning experiences, such as computational thinking (Enderle, King, & Margulieux, 2021). Although much debate exists around holistic conceptualizations of this practice, some common elements exist across all of them, including abstraction, algorithmic thinking, and decomposition (Grover & Pea, 2013). These shared conceptual elements could serve as the focus for items that target students' understanding of computational thinking. Although the NAEP Science Framework aims to be "curriculum neutral", the framework does need to be designed in ways that make it flexible and applicable for the next ten years of growth in science education. To achieve this flexibility, attention must be given to the total assemblage of scientific practices being implemented in classrooms across the country, from major ones to those less emphasized.

Department of Middle and Secondary Education
College of Education & Human Development

Equity and the Assessment of Diverse Learners

Reviewing the previous NAEP Science Framework, there is a striking silence when it comes to considerations of diverse learners and equitable assessment. Several of the "Special Studies" identified in the previous framework do take steps towards considering the needs of diverse populations of learners. However, most of these studies focus on technical comparisons of formats and capabilities. The scholarship surrounding the significant influence of students' cultures and communities on their learning has grown tremendously since the publication of the previous framework. A recently published text from the National Academies of Science, Engineering, and Medicine (NASEM, 2018) provides an excellent introduction to this work by highlighting the important role that culture and learning contexts play in every student's learning trajectory, including the influence of culture on learners' biological, motivational, and reasoning development.

The *Standards for Educational and Psychological Testing* developed jointly by American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (2014) should be prominent in the revision of the NAEP Science Framework. These guidelines synthesize a vast body of literature regarding assessment and provide critical insights into many aspects of assessment development, including those of the size and scope of the NAEP. Concerning equity, the *Standards* offer great detail and consideration of the concept of "fairness" in assessments. This particular section of the *Standards* underwent significant expansion in response to the rapidly developing knowledge base surrounding equity and supporting diverse populations of learners, including recognizing this work as foundational to assessments as considerations of validity and reliability. A major tenet of fairness, as conceptualized in this text, is that assessment administrators must provide access for all examinees in various populations, particularly in allowing for accommodations and modification for learners with different cognitive, linguistic, and physical abilities (AERA/APA/NCME, 2014).

Behizadeh (2014) offers examples of how to align large-scale writing assessments with fundamental knowledge generated through sociocultural theories of learning, lenses that elucidate the construct of 'fairness' while highlighting the many challenges assessments, including NAEP, present for students, particularly those from marginalized communities. This work also draws attention to the consequential validity of such assessments. Consequential validity considers the intended and unintended impacts of large-scale assessments on all learners, and such considerations must acknowledge the detrimental impacts that assessment scores have had in the ways they have been used to characterize minoritized communities as deficient. To understand more nuanced concerns about how assessment scores, including NAEP across several disciplines, have been used in oppressive ways towards these communities of learners, we also recommend Love (2019), Muhammad, Ortiz, & Neville (2021), and Stinson (2015). For considerations of fairness and equity in science education across a range of student populations and learning environments, we also strongly recommend the committee consider the seven chapters in Section III: Diversity and Equity in Science Education of the *Handbook of Research on Science Education II* (Lederman & Abell, 2014). Finally, we recommend some of our work to provide insight into ways that students from minoritized populations, including Black girls and students who are deaf or hard of hearing, can be denied access to science through various aspects of the education system (Enderle, Cohen, & Scott, 2020; King & Pringle, 2018; Wade-James, King, & Schwartz, 2021).

Another element of the AERA/APA/NCME construct of fairness emphasizes the need to minimize barriers in accessing assessments, including aligning the design and development of assessment items using the tenets of UDL. As mentioned previously, UDL highlights the need to provide students taking assessments with multiple means of engagement, expression, and representation. Applying these principles to the design of assessment items entails the development of multiple question formats and response options, providing students with choices to enhance access for diverse learners. Further, in the design of all item types, issues that might restrict an examinee's ability to demonstrate what they know (AERA/APA/NCME, 2014) must be removed. Examples in the *Standards* (AERA/APA/NCME, 2014) provide ways to address these issues for various populations of learners. The work of Fine and Furtak (2020) offers insight into ways science assessments can be developed to support, rather than restrict, multilingual learners. Even the most straightforward consideration of minimizing barriers should include commitments to offering the assessment in multiple languages, rather than just English, and supporting students who are deaf or blind with additional video interpretations and audio recordings of assessment items so they all have the opportunity to represent their full understanding of the content.

The final and perhaps most critical element of 'fairness' explored in the *Standards* entails promoting fair test score interpretations. A requirement for fair test score interpretation involves the inclusion of data points and metrics that characterize students' "opportunity to learn" (OTL). Indeed, the *Standards* emphasize the importance of incorporating OTL metrics as causal factors in score interpretations. Such usage necessitates that the new NAEP Science Framework explicitly commits to avoiding traditional and staid comparisons of outcomes across learners from communities varying greatly in OTL metrics. Rather, the new framework should endeavor to focus on interpretations within communities and populations based on OTL metrics while also maintaining an 'asset' orientation in all interpretations (NASEM, 2018), rather than traditional 'deficit' views that have been associated with large-scale assessments, such as NAEP, and the reporting of outcomes.

Haertel, Moss, Pullin, and Gee (2008) assert that thoughtful consideration of OTL metrics extends beyond basic considerations of content resources and instructional practices. OTL metrics must consider how students are given opportunities to personally connect to their science learning experiences through "forms of knowledge and ways of using language [from their] everyday experiences in families and communities" (Haertel et al., 2008, p. 8) and funds of knowledge (Gonzalez, Moll & Amanti, 2005). Practically, to achieve this aspect of fairness, the NAEP Science Framework revision team must work to broaden the collection of OTL data from participating districts, administrators, communities, and schools. We encourage the revision team to consider an example of such nuanced quantitative analyses around a community-based science learning effort

Department of Middle and Secondary Education
College of Education & Human Development

offered by King and colleagues (2021). Further, as an example of a thoughtful and broad data collection effort around science education, including community OTL factors, the revision team should also review the work of Banilower and colleagues (2018), who produced the *Report of the 2018 National Survey of Science and Mathematics Education*, as well as the OTL instruments developed for the ATLAST (Assessing Teacher Learning About Science Teaching) project from the same organization, Horizon, Inc.

The United States continues to live through an acute inflection point as a society that highlights the sincere need for continued and sustained discussion and efforts that work to support *ALL* of its citizens, particularly young people. Such support cannot advance the communities where these learners come from without transparent and thoughtful reckoning with how large-scale assessments have shaped those learners' experiences within the national education system and been used to their detriment. Further, such reflection *must* be coupled with deliberate actions that work in direct opposition to the continued use of such harmful practices while also working to expand opportunity, fairness, and equity in our science classrooms. At a minimum, the NAEP Governing Board and those working on the revision of the NAEP Science Framework must explicitly and emphatically assert the importance of equity and fairness throughout the various elements of the framework and the design of the next NAEP Science Assessment.

We provide a full list of references cited throughout our letter in the hopes that the various revision teams will take time to read and reflect on their connections to the new framework. We hope the NAEP Governing Board and those working on the revision teams of the NAEP Science Framework sincerely reflect on the two major issues we have elaborated on above, science learning through science practices, and equity and assessment of diverse learners. Both warrant considerable attention and explicit inclusion in any new assessment framework for science education, particularly if the goal is to "maintain NAEP as the gold standard", including "ensuring that the NAEP frameworks are updated for modern expectations for students" *and* the country's entire K12 education system.

Sincerely,

Patrick J. Enderle, Ph.D.
Associate Professor of Science Education

Renee Schwartz, Ph.D.
Professor of Science Education

Kadir Demir, Ph.D.
Associate Professor of Science Education

Natalie King, Ph.D.
Assistant Professor of Science Education

David Stinson, Ph.D.
Professor of Mathematics Education

G. Sue Kasun, Ph.D.
Associate Professor of Language Education

Christine D. Thomas, Ph.D.
Professor of Mathematics Education

Paula Garrett-Rucks, Ph.D.
Associate Professor of World Languages Education

Department of Middle and Secondary Education
College of Education & Human Development

Nadia Behizadeh, Ph.D.
Associate Professor of Adolescent Literacy

Caroline C. Sullivan, Ph.D.
Associate Clinical Professor, Teacher Preparation and Social Studies

Gertrude Tinker Sachs, Ph.D.
Chair, Department of Middle & Secondary Education

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). Standards for educational and psychological testing. Washington, DC: American Educational Research Association.

Banilower, E., Smith, P., Malzahn, K., Plumley, C., Gordon, E., & Hayes, M. (2018). Report of the 2018 NSSME+. Chapel Hill, NC: Horizon Research, Inc.

Behizadeh, N. (2014). Mitigating the dangers of a single story: Creating large-scale writing assessments aligned with sociocultural theory. *Educational Researcher*, *43*(3), 125-136.

CAST. (2015). *Top 10 UDL tips for assessment.* CAST Professional Publishing. Retrieved from https://slds.osu.edu/posts/documents/top-10-udl-tips.pdf on October 11, 2021.

Crawford, B. (2014). From inquiry to scientific practices in the science classroom. In *Handbook of research on science education, volume II* (pp. 529-556). Routledge.

Enderle, P., King, N., & Margulieux, L. (2021). What's in a Wave?. *The Science Teacher*, *88*(4), 24-28.

Enderle, P., Cohen, S., & Scott, J. (2020). Communicating about science and engineering practices and the nature of science: An exploration of American Sign Language resources. *Journal of Research in Science Teaching*, *57*(6), 968-995.

Ford, M. J. (2015). Educational implications of choosing "practice" to describe science in the next generation science standards. *Science Education*.

Fine, C. G. M., & Furtak, E. M. (2020). A framework for science classroom assessment task design for emergent bilingual learners. *Science Education*, *104*(3), 393-420.

Gonzalez, N.E., Moll, L., & Amanti, C. (2005). *Funds of knowledge. Theorising practices in households, communities and classrooms.* Routledge.

Grooms, J., Sampson, V., & Enderle, P. (2018). How concept familiarity and experience with scientific argumentation are related to the way groups participate in an episode of argumentation. *Journal of Research in Science Teaching*, *55*(9), 1264-1286.

Grooms, J., Enderle, P., & Sampson, V. (2015). Coordinating scientific argumentation and the Next Generation Science Standards through argument-drivenTips inquiry. *Science Educator*, *24*(1), 45-50.

Grover, S., & Pea, R. (2013). Computational thinking in K–12: A review of the state of the field. *Educational researcher*, *42*(1), 38-43.

Hand, B., Wallace, C. W., & Yang, E. M. (2004). Using a Science Writing Heuristic to enhance learning outcomes from laboratory activities in seventh-grade science: quantitative and qualitative aspects. *International Journal of Science Education*, *26*(2), 131-149.

Henderson, J. B., McNeill, K. L., González-Howard, M., Close, K., & Evans, M. (2018). Key challenges and future directions for educational research on scientific argumentation. *Journal of Research in Science Teaching*, *55*(1), 5-18.

King, N. S., Collier, Z., Johnson, B. G., Acosta, M., & Southwell, C. N. (2021). Determinants of Black families' access to a community-based STEM program: A latent class analysis. *Science Education*.

King, N. S., & Pringle, R. M. (2019). Black girls speak STEM: Counterstories of informal and formal learning experiences. *Journal of Research in Science Teaching*, *56*(5), 539-569.

Lederman, N. G., & Abell, S. K. (Eds.). (2014). *Handbook of research on science education, volume II* (Vol. 2). Routledge.

Love, B. L. (2019). *We want to do more than survive: Abolitionist teaching and the pursuit of educational freedom*. Beacon Press.

McNeill, K. L., & Krajcik, J. S. (2011). Supporting Grade 5-8 Students in Constructing Explanations in Science: The Claim, Evidence, and Reasoning Framework for Talk and Writing. *Pearson*.

Meyer, A., Rose, D. H., & Gordon, D. T. (2014). *Universal design for learning: Theory and practice*. CAST Professional Publishing.

Miller, A. R., & Kastens, K. A. (2018). Investigating the impacts of targeted professional development around models and modeling on teachers' instructional practice and student learning. *Journal of Research in Science Teaching*, *55*(5), 641-663.

Muhammad, G. E., Ortiz, N. A., & Neville, M. L. (2021). A Historically Responsive Literacy Model for Reading and Mathematics. *The Reading Teacher*, *75*(1), 73-81.

National Academies of Sciences, Engineering, and Medicine. (2018). How People Learn II: Learners, Contexts, and Cultures. Washington, DC: The National Academies Press. doi: https://doi.org/10.17226/24783.

National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. National Academies Press.

National Research Council. (2014). Developing assessments for the next generation science standards.

NGSS Lead States. (2013). The next generation science standards. Achieve, Inc.

Osborne, J. (2014). Scientific practices and inquiry in the science classroom. In *Handbook of research on science education, Volume II* (pp. 593-613). Routledge.

Sampson, V., Enderle, P., & Grooms, J. (2013). Argumentation in science education. *The Science Teacher*, *80*(5), 30.

Schwartz, R. S. (2019). Modeling Competence in the Light of Nature of Science. In *Towards a Competence-Based View on Models and Modeling in Science Education* (pp. 59-77). Springer, Cham.

Schweingruber, H., Beatty, A., & National Academies of Sciences, Engineering, and Medicine. (2017). *Seeing students learn science: Integrating assessment and instruction in the classroom*. National Academies Press.

Stinson, D. W. (2015). The journal handbook of research on urban mathematics teaching and learning: A resource guide for the Every Student Succeeds Act of 2015.

Wade-Jaimes, K., King, N. S., & Schwartz, R. (2021). "You could like science and not be a science person": Black girls' negotiation of space and identity in science. *Science Education*.

Wade-Jaimes, K., Demir, K., & Qureshi, A. (2018). Modeling strategies enhanced by metacognitive tools in high school physics to support student conceptual trajectories and understanding of electricity. *Science Education*, *102*(4), 711-743.

Wiliam, D. (2010). What counts as evidence of educational achievement? The role of constructs in the pursuit of equity in assessment. *Review of Research in Education*, *34*(1), 254-284.

Windschitl, M., Thompson, J., & Braaten, M. (2020). Ambitious science teaching. Harvard Education Press.

Windschitl, M., & Thompson, J. (2013). The modeling toolkit: Making student thinking visible with public representations. The Science Teacher, 80(6), 63-69.

From: Christa Marie Haverly
To: NAGB Queries
Cc: Stefanie Marshall; Shakhnoza Kayumova; tcheuk@calpoly.edu; Vincent.Basile@colostate.edu; smcdonald@psu.edu; Dr. Jonte" C. Taylor (JT)
Subject: NAEP Science Framework
Date: Friday, October 15, 2021 10:51:07 AM

To Whom It May Concern:

We write to you as a collective to urge you to update the NAEP Science Assessment Framework, taking into consideration key points as described below. These recommendations account for the dynamic relationship between theories of learning and practice and how approaches to assessment become consequential to what is made (in)visible as knowledge in the classroom. Therefore, we urge NAEP to pay attention to the evidence that has emerged in equity-based scholarship that interrogates dominant ways of knowing in science education, towards recognizing and making visible the epistemological pluralisms that racially and linguistically diverse youth enact in classrooms.

Equitable science education is critical given the increasing racial, cultural, and linguistic diversity in our country; the potential for the fields of science to benefit from the varied perspectives and lived experiences of our current and future PK-12 populations; and the obligation of our country's education system to rigorously prepare all of our students to be scientifically literate. This obligation has become more stark as we watch citizens across our country reject wearing masks or receiving vaccines against COVID-19, actively denying wide scientific consensus of the importance of these measures to protect personal and public health. This obligation has also become more stark as we have watched Black, Indigenous, and other citizens of Color in this country fighting to be treated humanely, with dignity, and equitably, emphasized in the months following George Floyd's murder, but representative of centuries of struggle. Further, this obligation has become more stark as we have watched communities and species be decimated by increasingly harsh natural disasters and habitat loss caused by over a century of preventable and mitigatable changes to our climate.

Now more than ever we need a science education program that serves to broaden participation in the fields of science and consequently broaden the epistemological dimensions of the sciences themselves. We need a science education program that prepares our youth to make critical, life- and planet-saving decisions that are rooted in evidence, not conspiracies. It is therefore essential that the NAEP standards and assessments that measure outcomes of our work with students reflect the research and recommendations that we share with you here recognizing that teaching and learning practices are often shaped by assessments and accountability measures.

We offer four sets of recommendations:

1. Interrogate the assumptions about science knowledge embedded in the standards (i.e., whose histories and narratives are and are not included in this body of

knowledge and practices).

  a. For example, see Morales-Doyle, D., Childress Price, T., & Chappell, M. J. (2019). Chemicals are contaminants too: Teaching appreciation and critique of science in the era of Next Generation Science Standards (NGSS). *Science Education, 103*(6), 1347-1366, and
  b. Rodriguez, A. J. (2015). What about a dimension of engagement, equity, and diversity practices? A critique of the Next Generation Science Standards. *Journal of Research in Science Teaching*, *52*(7), 1031-1051.

2. Update the technical aspects of the assessments themselves to be more inclusive of historically marginalized student populations.
   a. Consider implications and limitations of administering the test solely in English (see work from Guillermo Solano-Flores, Alison Bailey, and Jamal Abedi)
   b. Fund the special studies on “innovative assessment tasks, testing special needs students, and computer adaptive testing” (p. 121 of current NAEP framework).
   c. Develop assessment tools that can guide teachers and researchers to critically examine whether or not the assessments they are using or developing are sensitive to the instruction and the diverse ways students' thinking and knowledge can be embodied and represented.

3. Invite people to participate in this review process, including on the expert panel, who are multilingual, of Color, differently abled, and so on; leverage their expertise and lived experiences; and provide them with authority and agency to make substantive changes to the program.

4. Seek recommendations from the National Academies' Committee on Equity in PreK-12 STEM Education, which will be announced in the coming months.

Sincerely,

Christa Haverly, Ph.D., Northwestern University
Stefanie Marshall, Ph.D., University of Minnesota- Twin Cities
Shakhnoza Kayumova, Ph.D., University of Massachusetts, Dartmouth
Tina Cheuk, Ph.D., California Polytechnic State University, San Luis Obispo
Vincent Basile, Ph.D., Colorado State University
Scott McDonald, Ph.D., Pennsylvania State University
Jonte’ C. Taylor, Ph.D., Pennsylvania State University

Comments to the NAEP Science Assessment Framework
Submitted by the National Science Teaching Association
October 14, 2021

Thank you for the opportunity to address the National Assessment of Educational Progress (NAEP) Science Assessment Framework. The National Science Teaching Association (NSTA) is the world's largest organization promoting excellence and innovation in science teaching and learning for all. We are committed to best practices in teaching science and its impact on student learning. NSTA offers high-quality science education resources and continuous opportunities for learning that help science educators grow professionally and excel in their career.

As requested, we have focused our response on these three questions:

- Does the NAEP Science Assessment Framework need to be updated?
- If the framework needs to be updated, why is a revision needed?
- What should a revision to the framework include (or exclude)?

Working with a group of practitioners from several NSTA standing committees, we have answered these questions through the lens of what science and engineering could look like in 10 years and how technology can and should support more complex and meaningful assessments that reflect how people have been documented to learn science.

Does the NAEP Science Assessment Framework need to be updated?

NSTA strongly believes the NAEP Science Assessment Framework must be updated.

The current framework is extremely outdated. It is antiquated regarding standards for science education and science education research and is predicated on standards that originated before 2005.

Currently, states, districts, and schools are focusing their science curricula and instructional programs on *The Framework for K–12 Science Education: Practices, Crosscutting Concepts, and Core Ideas,* released by the National Academies of Sciences, Engineering, and Medicine in 2012. Twenty states have adopted the *Next Generation Science Standards* (2013). As outlined in the recent report *Call to Action for Science Education: Building Opportunity for the Future,*

> "The *Framework* catalyzed an ongoing transformation of elementary and secondary science education across the United States. The *Framework* provides guidance for improving science education that builds on previous national standards for science education and reflects research-based advances in learning and teaching science. As of April 2020, 44 states and the District of Columbia had developed and adopted science standards that are informed by or

> directly based on the *Framework.* This represents approximately 70% of K–12 public school students. The vision for science education outlined in the *Framework* differs in important ways from how science has traditionally been taught. It emphasizes engaging students in using the tools and practices of science and engineering and providing them with opportunities to explore phenomena and problems that are relevant to them and to their communities."

In conclusion, we emphatically state that the current NAEP Science Framework is woefully outdated, designed for a specific purpose that has largely ceased to exist, and incompatible with contemporary science curricular frameworks.

If the NAEP Science Assessment Framework needs to be updated, why is a revision needed?

Science education in the United States is currently in a state of transition as we move to align classroom teaching practices with *A Framework for K–12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. Teachers are shifting toward helping students employ science and engineering practices rather than solely familiarizing them with scientific principles.

Currently the NAEP science framework has the following item distributions: Science Content, Science Practices, and Items by Type (interactive computer tasks, hands-on performance tasks, and specific question types).

The next NAEP science framework should reflect how we currently teach and project the development of science teaching over the next decade.

The current NAEP science assessment framework does not adequately reflect the computational thinking required for grasping complex scientific issues, as well as the use of large databases. The current framework does not support the explicit nature of science pedagogy, nor does it reflect the shift to three-dimensional thinking needed for sensemaking. Each of these are found in the National Research Council's (NRC) *Framework for K–12 Science Education*. Furthermore, science and engineering design thinking and application should be added to the NAEP framework.

Illustrative NAEP questions are too narrow in scope and tend toward the mere acquisition of principles and facts. For example, representative NAEP questions in 12th-grade physics focus on familiarity with gravitational force equations and relationships between variables, which tend to reward memorization. This type of question should be replaced with a broader assessment of a student's understanding of how gravitational fields can store and transfer energy.

The NAEP range of topics also seems very broad in nature and less in-depth, which results in rewarding memorization and familiarity with specific concepts, but not their application or extension. The NAEP framework should more accurately reflect the depth of learning and application that is now expected of students.

What should a revision to the framework include (or exclude)?

According to the *Call to Action for Science Education: Building Opportunity for the Future,*

> “Science assessments and accountability systems need to be aligned with the vision for high-quality science instruction. Assessing science learning in ways that are aligned to our vision will require approaches that go beyond single tests of factual knowledge. Traditional, large-scale, multiple-choice tests cannot capture the ability of students to engage in the practices of science and reason about evidence. An advantage of the new approach to science instruction is that it provides many opportunities for assessing learning informally (formative assessment) as students engage in investigations, create representations, and discuss evidence. However, designing useful and meaningful formal assessments such as tests will require careful articulation of the desired learning goals and how students can demonstrate that they have achieved them.”

To genuinely be forward-looking, future science assessment based on the NRC *Framework* should capture a student’s ability to behave like a scientist and to engage in scientific practices to deconstruct and make sense of a situation or phenomenon.

The revision should include the following:

- *Modeling as a practice.* Students should be asked to create, evaluate, and/or revise models, and use them to predict the result of changes to system components. The development of explanatory models can help students make their thinking visible and can be an equalizer for English Language Learners.
- *Planning investigations.* Students should be able to identify independent and dependent variables and to design scientifically valid investigations.
- *Analyzing data.* Students should be able to analyze complex, real-world data using graphing and graphing analysis tools.
- *Engaging in argument from evidence.* Students should be assessed on their ability to use evidence to construct and justify a scientific claim.

Each of these elements should be approached with a recognition that the science experiences of many students are not equitable, inclusive, or reflective of our expanding diversity as a nation.

It is important to note that the recent pandemic has facilitated the shift in science teaching that is unprecedented in its scope and duration. The use of simulations, along with hands-on experiential learning, is much more common than when the current NAEP science framework was adopted. Subsequently, the scope of science teaching has changed to better reflect three-dimensional sensemaking. As a result, the NAEP framework should be modified to include novel approaches that incorporate shifts in science practices that are observed. To this end, a revised NAEP Science Assessment Framework would increase validity by reflecting the shifts that form the foundation of students' sensemaking through the practices, inquiry, nature of science, science content, and crosscutting concepts.

In addition to these ideas, we offer some specific suggestions for changes to the current NAEP science framework:

While the Science and Engineering Practices and the Disciplinary Core Ideas expressed in *NGSS* are evident in the framework, the Crosscutting Concepts need to be more explicitly represented. Hence, summary charts should be included to reflect the current three-dimensional sensemaking supported by the nature of science. Less emphasis should be placed on identifying science principles, and more emphasis should be placed on higher order of reasoning skills. However, the current sample questions focus more on rote knowledge and do not give students opportunities to demonstrate the application of that knowledge to novel situations.

Scientific and Engineering Practices, rather than principles, should be reflected. Science Practices should be expanded to include analyzing and interpreting data; using mathematics and computational thinking; constructing explanations (for science) and designing solutions (for engineering); engaging in argument from evidence; and obtaining, evaluating, and communicating information. When these practices are added, students should be able to demonstrate their science literacy based on performance expectations.

In conclusion, it can be said that the value of any assessment is rooted in the purpose for which it is intended. If one purpose of NAEP is to provide a longitudinal trajectory of how American students are learning science across their compulsory education, then its science assessment framework must reflect the dramatic shifts in the mode of instruction, as well as the curricula upon which that instruction is based.

This statement has been endorsed by the Council of State Science Supervisors and the National Science Education Leadership Association.

October 15, 2021

Dear National Assessment Governing Board,

I am writing to communicate my professional perspectives in response to requests for commentary about the NAEP Science Frameworks. As background, I am a professor of STEM teacher education at the University of Connecticut and Co-Editor of the journal Science Education. As I examine the 2019 Science Assessment Framework document, several aspects caused great concern. Especially given the unique times in which we find ourselves, I want to earnestly communicate the need for major shifts to the NAEP Science Frameworks. In their current condition, I found few positive advances over previous iterations. Given the sea changes in society, and in light of considerable research gains in the learning science, school leadership, and instructional delivery, without dramatic improvements to the NAEP Science Framework, we will miss an opportunity to respond to contemporary challenges. Any efforts to maintain the status quo with the NAEP Science assessment will effectively neglect this unique chance to make positive changes to K-12 science education throughout our nation. Below are several concerns which need your attention:

A. <u>Perils of Supporting Deficit Explanations via NAEP Science Results</u>. Even with the Coleman Report clearly demonstrating racial differences in student performance were much stronger within rather than between schools, NAEP continues its pattern of feeding information to the contrary. Decision-making purported to inform policy and practice to support school is overshadowed by data “gaps” that compares states and school urbanicity. For those who accept inequities as challenge worth resolving, the unit of change is known to be at the school level. Responses to questions about WHY science performance gaps exist are greatly influenced by HOW such data are collected and reported. I would submit that NCLB data powerfully influenced achievement gap discourses simply by disaggregating school level data. Seeing disparities in outcomes within specific schools and communities made it much harder ignore the reality the inequities lurk within the places where we send our children and for which we pay taxes. Rather than support deliberations about the presence of science achievement gaps as artifacts of institutional and organizational factors – with an eye toward remedying those disparities – NAEP data will instead perpetuate beliefs about gap inevitability and progress toward closing those gaps is only likely as scores by White students come. Absent from the design is information that might indicate how non-White student performance could be improved. More than recognizing complicity with fostering such narratives, I would submit that NAEP should proactively develop data reporting approaches that could redirect media, political, and layperson discussions in ways that disrupt widespread beliefs that demographics dictate destinies.

B. <u>Supporting Equity and Diversity Research in Science Education</u>. Although the framework expresses the ambition of collecting data suitable for informing policymaking and support secondary forms of research, to date there has been very little research about the results from NAEP Science. We can attribute this to shortcomings of the data collection – weaknesses which have frustrated those of us who would like to do this research. For example, the intersection of student gender, race, and social class are very relevant to building better understandings of science achievement. NAEP Science data has the potential to advance understandings of a variety of equity concerns (and to in turn shape instructional practices) only if more thought is given to making such data available. NAEP's own report cards reduce "Score Gaps" to singular designations without revealing whether Black females and Black males perform similarly. OR similarly multidimension features for NSLP eligibility, English learner status, etc. While some might suggest such analyses are possible (via special access to data), that approach has not proven to be fruitful. There are few to no examples within the demonstration material for NAEP Data Explorer. But the absence of such secondary research for the NAEP Science cannot be blamed on the research community. Instead, the NAEP system itself is not supportive of those types of studies – despite expressed claims that secondary research studies are a goal.

What I hoped to communicate in this letter is the immense potential for NAEP to shape, inform, and improve science education with a potentially national scope. My frustrations are rooted in the fear that such possibilities will be missed. As a consequence, not only would potential advancements be lost but also the likelihood that outdated perceptions of school science would be perpetuated by dubious information. In addition to the concerns about marginalizing equity as expressed above, I am deeply troubled by how outdated the resources are the are being used to shape the NAEP Science Frameworks. Included in this list is the absence of research published with the past ten years, the failure to acknowledge the substance of NGSS, and even the presence of retired and deceased members on your various committees. In some respects, I would advocate that the NAEP Science Frameworks begin with fresh people and perspectives rather than continue moving forward with such a dilapidated foundation. There are admittedly many dimensions of the NAEP Framework process that I cannot fully appreciate. On the other hand, as a research journal editor and participant in national communities of science education research, I can only hope that the NAGP will recognize the real possibility of missing a vital opportunity to improve science education by continuing with the current strategies.

In closing, the NAEP Science Assessment Framework is in profound need of updating. The materials used as the basis for this framework are outdated and fail to make effective use of contemporary understandings of science teaching and learning. Further, the framework's updating must attend to the shifting demographics of America's schools. More than acknowledge the existence of students who are traditionally marginalizing from science learning opportunities as consequences of their race, social class, English fluency, disabilities, gender, and immigration status, such awareness must accompany a strong centering of equity as a singular goal – in the design of the assessments, the structure of the data collection, and the release and reporting of results. Otherwise, it seems inevitable that the status quo procedures will further reify discriminatory assumptions and actions as by-products of the subsequent Science Report Cards.

Respectfully submitted,

John Settlage, Professor
University of Connecticut

From: Renee Schwartz
To: NAGB Queries
Subject: NAEP SCIENCE FRAMEWORK
Date: Friday, October 15, 2021 6:42:04 PM

Comments re the NAEP Science Framework revision:

The Board of Directors and Executive Committee of NARST [A global organization for improving science education through research] submit the following suggestions regarding the upcoming revision of the NAEP Science Framework:

The NAEP science framework faces a precarious challenge: standardizing the instrument across time to identify longitudinal patterns while accommodating changes in science education. The document thoughtfully addressed the tensions created by these competing goods. Even though some aspects of the framework reflect more current reform in science education (e.g., crosscutting concepts), it is difficult to ascertain the extent to which the NAEP science framework aligns with the more recent emphases put forth by the Framework for K-12 Science Education (NRC,2012) and the Next Generation Science Standards [NGSS]. There are notable differences between how the current NAEP framework and the NGSS define, focus, and recommend science concepts and science and engineering practices. A misalignment may prove problematic when using NAEP science achievement data to better inform decisions in policy and practice. It would be more advantageous for the advancement of K-12 science learning if more items corresponding with current science education reform are developed and included in the forthcoming assessment.

On one hand, the importance of context and its role in learning were primarily absent in the framework. Examples of prospective assessment items were abstract. On the other hand, in the cases in which concepts were embedded in context, the contexts (e.g., hares in state park) featured the lived experiences of dominant groups in U.S. society (e.g., upper middle class). It seemed the science framework did not incorporate decades of sociocultural research on cultural responsiveness and inclusivity in learning and assessment. Additionally, while noting the framework spoke to the need to consider the language demands of test items for English language learners, there were no explicit actions related to considerations of item development responsive to language. Indeed, the sample items shared were laden with dense language and vocabulary, particularly in context-driven items.

Because of the prevalent inequities in the quality of science education in K-12 education, it would be very useful for NAEP to develop equity indicators with respect to achievement and school and community factors, like those used in international assessments. Intentional attention to equity and social justice within science curriculum and instruction are essential for developing scientific literacy.

Sincerely,
Renee' Schwartz, President of NARST
Eileen Parsons, Immediate Past President of NARST
Gillian Roehrig, President-Elect of NARST
Jerome Shaw, Secretary/Treasurer of NARST
Lisa Martin-Hansen, Executive Director of NARST
NARST Board of Directors:
Scott McDonald, Leon Walls, Noemi Waight, Christina Schwarz, Malcolm Butler, Theila Smith, Bhaskar Upadhyay, Knut Neumann, Brooke Whitworth, Sonya Martin
Troy Sadler and Felicia Moore Mensah: Editors of the Journal of Research in Science Teaching
Michael Bowan: NARST Liaison to NSTA
Cynthia Crockett: NSTA Liaison to NARST

Renee' Schwartz, PhD

Professor, Science Education

Georgia State University

President *NARST: A global organization for improving science education through research* [narst.org]

Program Coordinator: *PhD Teaching and Learning, Science Education,* Georgia State University

Department of Middle and Secondary Education

College of Education and Human Development

Office: CEHD, 30 Pryor St. #629

NSELA Response to: Seeking Initial Public Comment Prior to Updating the Science Assessment Framework for the 2028 National Assessment of Educational Progress (NAEP)

The National Science Education Leadership Association (NSELA) is an organization of approximately 600 members in science leadership roles either at the school, district, university, informal science, or state level. Our mission is to catalyze leadership to maximize effective science teaching and learning in a complex and changing environment. We connect and support emerging and experienced leaders by providing high-quality professional development, a collegial network, access to research and resources, and a voice for leaders in science education. As requested by the National Assessment Governing Board, our members have provided feedback to address three questions about the current NAEP Science Assessment Framework:

- Does the NAEP Science Assessment Framework need to be revised?
- If the NAEP Science Assessment Framework needs to be revised, why is a revision needed?
- What should a revision to the NAEP Science Assessment Framework include?

NSELA recommends that yes, the NAEP framework does need to be revised. There have been many new findings from research in science education since the writing of the last NAEP Science Assessment Framework in 2005. The publication *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas* (2012) includes more current research in science education than does *The National Science Standards* (1996) with which the 2005 NAEP Framework is aligned.

The current NAEP Science Assessment Framework is heavily focused on science content knowledge rather than the integration of science content with crosscutting concepts and science and engineering practices. With 44 states having revised their science standards to align with *A Framework for K-12 Science Education,* including 20 states that have adopted the Next Generation Science Standards, the 2005 NAEP Science Assessment Framework does not accurately assess what today's science students know, understand, can do and apply. This creates a misalignment in what is being assessed on the NAEP science assessment and the current research and best practices for students. Although the 2019 NAEP report is very comprehensive and recognizes how science can change, it is still based on antiquated science education research with the intent to create a snapshot of what is being taught in American schools. The following proposed changes will better align the NAEP Assessment Framework with current science education research and practices.

Rather than aligning science content with *The National Science Standards* (1996) and *Benchmarks for Science Literacy* (1993), align the content with the recommendations of *A Framework for K-12 Science Education* (2012) and the *Next Generation Science Standards* (2013). In developing performance expectations and performance assessment items, consider merging not only science content with science practices, but also integrating crosscutting concepts, as recommended in *A Framework for K-12 Science Education*. This change would create a need for a section of the NAEP Science Assessment Framework that focuses on the Crosscutting Concepts to be assessed.

For the Science Content section of the NAEP Science Assessment Framework, consider focusing less on nuggets of knowledge and more on application of that knowledge to make sense of phenomena. To better align with the recommendations of *A Framework for K-12 Science Education* and the *Next Generation Science Standards*, consider aligning the content section of the NAEP Framework with the disciplinary core ideas within these documents.

For the Science Practices section of the NAEP Science Assessment Framework, rather than using the former broad science practices "identifying science principles, using science principles, using science inquiry, and using technological design", instead use some of the science and engineering practices listed within *A Framework for K-12 Science Education* and the *Next Generation Science Standards.* Possible science practices to be assessed might include: Developing and Using Models, Planning and Carrying Out Investigations, Analyzing and Interpreting Data, Using Mathematics and Computational Thinking, and Engaging in Argument from Evidence. The focus should be on using the science and engineering practices to determine whether students can "do" science.

The Assessment Design section of the NAEP Science Assessment Framework needs to be updated to include performance expectations where science content, crosscutting concepts, and science and engineering practices intersect. Assessing all three dimensions (content, concepts, and practices) will require a greater number of performance-type assessment items, either hands-on or computer simulation-based, where students might use multiple data sources to construct reasonable explanations, analyze data, develop scientific arguments, or develop conclusions. Give students a scenario to make sense of that they may actually see in their lives. Look for a development of student thinking to make sense of the scenario - consider multiple questions around this scenario to scaffold and get at student ability to work and think like a scientist.

For the Science Achievement Level Descriptors section of the NAEP Science Assessment Framework, the descriptors need to align with the changes in content strands recommended in *A Framework for K-12 Science Education.* Use the *Next Generation Science Standards* to review appropriate descriptors. Use the grade band endpoints given for 6-8 and 9-12 as no matter what content sequence may be utilized within a state, by the end of grade 8 and 12 all students should have learned the content being assessed.

For the section of the NAEP Science Assessment Framework focused on English Language Learners and Students with Disabilities, first consider changing the term ELLs to Multilingual Students as is more widely utilized today. Ensure grade appropriate language is utilized to assess student proficiency of grade level standards. Provide the opportunity for the test to be read aloud as an option for any child who takes the NAEP to ensure we are offering a level playing field and reading does not hinder the ability to respond. Align NAEP assessment modifications or accommodations with those that are utilized by states across the country.

The purpose of the NSELA recommendations is to better align the NAEP assessment with the current expectations for student learning within science classrooms across the country. Having relevant, meaningful assessment data is important to science education leaders. Aligning the NAEP Science Assessment Framework with current science education research and practice will result in a NAEP assessment that more accurately measures student understanding and application of science.

LEARNING POLICY INSTITUTE
Research. Action. Impact.

To Members of the National Assessment Governing Board:

Thank you for the opportunity to provide feedback regarding the development of the next framework for the National Assessment of Educational Progress in Science. This document shares feedback collected by the State Performance Assessment Learning Community (SPA-LC) from science education communities across the nation in response to the three questions posed by NAGB:

1. Whether the NAEP Science Assessment Framework needs to be updated.
2. If the framework needs to be updated, why a revision is needed.
3. What should a revision to the framework include?

SPA-LC, coordinated by the Learning Policy Institute, represents over 25 states and 10 national partners committed to the development and implementation of meaningful and balanced assessment systems, beginning with science. SPA-LC's members include state commissioners, curriculum and instruction directors, assessment directors, and science leadership within state education agencies as well as local communities. Together, SPA-LC supports within- and cross-state efforts to develop meaningful assessment systems in science through support for better instruments, effective capacity building, and meaningful policies. As such, we find ourselves distinctively positioned to offer relevant input regarding the country's distinguished assessment of scientific learning.

A careful review of the current NAEP science framework and progress in science education--including state standards, foundational research, contextual and environmental shifts, and recent advances in science teaching, learning, and assessment practice was completed by convening three focus groups and collecting information via survey. As a result of this review, **SPA-LC recommends that the NAEP Science Framework be updated in targeted ways to better reflect both the current state of science education across the country, as well as the direction in which we expect science education efforts to shift in the next decade. Specifically, we recommend:**

1. **Update the content and practice included in the framework to align with the most recent research on what students should know and be able to do.**
2. **Prioritize integration of content, practice, and conceptual elements in service of sense-making.**
3. **Ensure the NAEP science framework supports and addresses needed advances in assessment design and use.**

LEARNING POLICY INSTITUTE
Research. Action. Impact.

Below, we outline key shifts that should be addressed in the next science framework. The SPA-LC community stands ready to support any efforts to make these and other needed shifts to ensure that NAEP remains a relevant cornerstone of science assessment systems nationwide.

The need for an update.

According to the National Assessment Governing Board, NAEP frameworks are updated for modern expectations for students and to "address recent standards, curricula, and instruction, research on cognitive development, and the latest perspectives on what students should know and be able to do" (NAGB, 2021). Since the last substantial review of the NAEP science framework, there have been sufficient shifts in science education research and practice to recommend a review and revision of that framework.

Advances in research on how students learn and demonstrate science understanding and practice. Since the NAGB last made substantial changes to the NAEP science framework, the following developments in science education and assessment have initiated a great deal of adaptation in the field:

- Release of the publications How Students Learn Science in the Classroom and Taking Science To school, which together began to push the community to think, "beyond the artificial dichotomy between content and process in science" (TSTS, p viii)
- Development, publication and release of "A Framework for K-12 Science Education: Practice, Crosscutting Concepts, and Core Ideas."
- Supporting cognitive research such as How People Learn II: Learners, Contexts, and Cultures (2018) provide further input regarding integration of content and practice for improved and more equitable outcomes.
- Assessments begin to use sensemaking and cognitive complexity models that incorporate multi-dimensional analysis of student interaction with phenomena such as those illustrated in "A Framework to Evaluate Cognitive Complexity in Science Assessments."
- Substantial efforts to support research-based instructional models that prioritize students' active engagement in phenomena and sense-making ("figuring out") as the mechanism for science teaching, learning, and assessment. This includes materials themselves (e.g., OpenSciEd, inquiryHub, Multiple Literacies Project Based Learning, etc) as well as within criteria for high quality materials (EQuIP, EdReports) and assessment (e.g., Science task screeners, Task Annotation Project in Science, New Meridian Science Assessment Framework, Harris et. al. work to focus assessments on knowledge-in-use)

LEARNING POLICY INSTITUTE
Research. Action. Impact.

Substantial shifts in the science standards landscape. The most recent versions of the NAEP science framework have largely attended to and reflected the 1996 National Science Education Standards (NSES). While these standards provided a strong foundation for science education and assessment, the release of *A Framework for K-12 Science Education* led to the development and widespread adoption of new standards such as the Next Generation Science Standards (NGSS) and other, similar standards. These standards, currently adopted in over 45 states and the District of Columbia, reflect key conceptual shifts in standards, teaching, learning, and assessment. Given the widespread use of new standards, a review and revision of blueprint content/practice alignment may be warranted to ensure that what is tested by NAEP is reflective of what students are given the opportunity to learn in their classrooms .

Advances in equitable science assessment design and implementation. As states, districts, and teachers have worked to implement new science standards, there has been a call to redesign science assessments such that they 1) better reflect what we expect students to understand and be able to do in science, and 2) attend to equity in assessment in ways that move beyond traditional conceptions of bias and sensitivity. This includes:

- Centering sense-making and knowledge-in-use as essential elements of aligned science assessment items and tasks
- Leveraging advances in simulations, item sets/clusters, scoring algorithms, and test design to better approximate performance-based tasks and approaches that more authentically represent science learning and mastery
- Attending to features of equity within assessment design and use, including racial equity; culturally responsive assessment practices; and attending to student interest, identity, and agency within assessment design.

Many of these advancements reflect both a desire to develop more valid assessment instruments and reports as well as an effort to ensure that assessments are coherent with instructional and professional learning components of the science educational system. It will be important that the NAEP science framework attend to these shifts in assessment understanding, design, and practices to produce assessment results that both represent the state of science learning in the country as well as serve to lead the way for assessment work of the future.

LEARNING POLICY INSTITUTE
Research. Action. Impact.

What revisions should include.

While there is endless nuance and details that could be addressed, SPA-LC makes three central recommendations for revisions to the NAEP science framework:

1. **Update the content and practice included in the framework to align with the most recent research on what students should know and be able to do.**
2. **Prioritize integration of content, practice, and conceptual elements in service of sense-making.**
3. **Ensure the NAEP science framework supports and addresses needed advances in assessment design and use.**

Recommendation 1: Update the content and practice included in the framework to align with the most recent research on what students should know and be able to do.

Rationale. As described above, science teaching, learning, and assessment have been deeply influenced by *A Framework for K-12 Science Education* and the shifts represented by new standards based on it (e.g., NGSS). Recent analyses of content alignment between current state standards and the NAEP science framework have found substantial differences, including differences in targeted science ideas and how scientific practice is represented. For example, A Comparison Between the Next Generation Science Standards (NGSS) and the National Assessment of Educational Progress (NAEP) Frameworks in Science, Technology and Engineering Literacy and Mathematics (Neidorf et. al., 2016) found:

- At grades 3-5 only 38% of performance expectations were aligned to the [NAEP] Science framework, with 44% alignment at both middle and high school.
- Considering only grade 4 NGSS performance expectations for the grade 4 NAEP 36% of performance expectations were aligned.
- Across all grades the highest degree of alignment was in life sciences (from 48-54%) with the lowest degree of alignment in physical science (29-42%)

Additionally, the existing overlap between the NGSS practices and the practices outlined in the current NAEP framework provides a strong foundation for a meaningful framework and related assessment. A revision to the framework provides an opportunity to consider how the practices are represented in ways that are coherent with other science education efforts. Questions to consider include:

- In what ways can the practices be better integrated as an essential part of sense-making--either through making sense of phenomena or designing solutions to problems?
- In what ways should the existing practices be clustered to both reflect and complement how the practices are used together in instruction and assessment nationally?

- Are the measures used to assess scientific practice in alignment with the goals of science educational practice across the country?

With 20 states (and the District of Columbia and Department of Defense Educational Agency) aligned directly with NGSS and 24 states aligned with the *Framework for K-12 Science Education*, it may be appropriate to revise the NAEP science framework to better align with current state activities. This will ensure NAEP is able to appropriately monitor science learning across states and over time, remaining a vital element of our understanding of how science education is progressing.

Recommendation 2: Prioritize integration of content, practice, and conceptual elements in service of sense-making.

Rationale. According to the *Framework for K-12 Science Education* (p. 218; emphasis added), "Standards and performance expectations that are aligned to the framework must take into account that **students cannot fully understand scientific and engineering ideas without engaging in the practices of inquiry and the discourses by which such ideas are developed and refined [1-3]. At the same time, they cannot learn or show competence in practices except in the context of specific content.**" Research suggests that surfacing student understanding and ability in science requires that they are able to show both the depth of their conceptual understanding of science ideas as well as their ability to engage in scientific practice together. Recent work focused on how to assess student mastery of widely adopted science standards requires a shift toward assessments that ask students to actually engage in using science ideas and practice together in service of sense-making; conversely, assessing students for understanding outside of the context of the integration of content and practice would provide incomplete-- and potentially even inaccurate--information about true student facility with science expectations.

While the current NAEP Science framework and associated assessment specify and assess important aspects of science content and science practice, these are often done separately. Moving forward, it may be appropriate to consider more intentional integration of science core ideas, practices, and crosscutting concepts in both framework and assessment design.

LEARNING POLICY INSTITUTE
Research. Action. Impact.

Recommendation 3: Ensure the NAEP science framework supports and addresses advances in assessment design and use.

A primary way the NAEP science framework influences the national science education community is through the NAEP science assessment, which has had a long history of setting the standard for high-quality assessment design in science. For the NAEP science assessment to continue to be both immediately compelling and forward-leading, it will be important for NAGB to consider how revisions to the science framework are accompanied by revisions to the assessment, including:

- Items and forms that can appropriately engage sense-making at the nexus of multiple dimensions, including effective use of performance tasks and technology enhanced items and scoring paradigms.
- Ensure proper alignment to updated framework goals.
- Develop tasks that center making sense of appropriate and compelling phenomena as their foundational basis.
- Attend to advances in equitable assessment that include and expand beyond attention to bias and sensitivity considerations.
- Consider alternative cognitive complexity models to address multidimensionality of items and item sets.

As a measure of educational trends, the NAEP assessment would need to address continuity across tests, requiring innovation in terms of equating and development of linking items from form to form. While this may be a complex undertaking, it is not impossible, and given the large-scale, non-accountability model of the NAEP assessment, the creative use of matrix blocks to achieve the desired outcomes may offer a useful solution.

Conclusion.

The NAEP science framework, and associated assessment, are strong components of current science assessment systems. With key revisions, they stand to continue shining a light on how we can continue supporting effective and meaningful science learning for all students. We stand ready to assist NAEP and the National Assessment Governing Board in support of this effort.

Warm regards,

Aneesha Badrinarayan, Senior Advisor, Learning Policy Institute

on behalf of the State Performance Assessment Learning Community.

Seeking Initial Public Comment Prior to Updating the Science Assessment Framework for the 2028 National Assessment of Educational Progress (NAEP)

Deadline Extended to Oct. 15

Comments should specifically address three things:

Comments must be submitted via email to nagb@ed.gov with the email subject header NAEP Science Framework no later than 5:00 p.m. Eastern Time on Friday, October 15.

1. **Does the NAEP Science Assessment Framework needs to be updated?**
2. **If the framework needs to be updated, why a revision is needed?**

In general, no. The principles and frame work are sound, stressing empirical knowledge and testing. As is appropriate with a general framework, discussion of scientifically disputed or politically charged issues such as anthropogenic climate change or embryonic stem cell research are avoided.

However, given the current political and educational climate, this may change. If it does and climate change becomes a specific focus of discussion in the framework, below we offer a few suggestions to provide a balanced discussion of theories of climate change, and an accurate assessment of climate data versus model projections.

3. **What should a revision to the framework include?**

Any discussion of Climate Change within the framework should be focused on helping students learn how to think through the issue and weigh different types of information. For example, any climate-specific material should teach students the difference between verified objective observations and data versus predictive models.

Regarding specific components of the climate change issue, any climate-specific framework should include:

1. The theory that anthropogenic greenhouse gas emissions are causing catastrophic changes to the climate is not settled science, and this should be acknowledged.
2. Science does not proceed by consensus (which is a political term tantamount to vote counting) but rather be experimentation and discovery, grounded in verifiable data, and independent testing.
3. Myriad factors, many only poorly understood, drive climate changes over the short, medium, and long-terms.
4. Climate model projections of temperature fail to accurately mimic actual temperatures and temperature trends as measured by ground-based weather stations, global satellites, and weather balloons.
5. Projections of climate change impacts are driven by computer model simulations of temperature responses to greenhouse gases and speculative assumptions about climate feedback mechanisms. Simple models that don't include feedback mechanisms better track actual temperature measurements and project less warming with each additional unit of carbon dioxide.
6. Statements regarding worsening weather conditions should note that there have been few if any observed worsening global trends for extreme weather despite decades of speculation that such worsening is imminent. Objective data and measurements show each of these weather phenomenon are well within the range of natural historic variation and most types of extreme weather events show no recent change or a trend of less frequency and severity.

7. Additional carbon dioxide in the atmosphere has contributed to a substantial greening of the earth and record crop production, which has resulted in declining rates of starvation and hunger.
8. Cold conditions result in more premature deaths each year than warm conditions. As the Earth has warmed modesty, the number of deaths attributable to extreme temperatures has substantially declined.

Specific issues in the current Text:

On Pg. 42 (62 incl. preface) box under life sciences should state, "Plants also require light and carbon dioxide to grow."

Pg. 54/55 (74/75) mentions climate, but doesn't discuss the difference between weather and climate. Climate changes aren't measured or determined over the short term of just a few years, but rather over 30-year periods. Modest changes between periods don't signal climate change for a region, only substantial changes do.

Pg. 61/62 (81/82) Boxes discussing changes in earth system and biogeochemical cycle are accurate.

If climate change is discussed in the updated NAEP assessment, it should note the long-term decline in carbon dioxide in the atmosphere prior to the Industrial Revolution. Most plants evolved before the long-term decline began, when carbon dioxide levels were considerably higher than today. It would also note that if carbon dioxide levels dip below 150 ppm, plants can't photosynthesize and begin to die. The Earth came perilously close to that prior to the Industrial Revolution.

Avoid controversial and overly politicized topics related to energy systems, but if it is discussed ensure that students are provided a balanced view of the virtues and drawback of each source of energy generation. All forms of energy have environmental impacts.

Possible Design experiments:

Set up three plants (sets of plants) in greenhouse-like conditions, one with ambient carbon dioxide levels, a second with elevated carbon dioxide, a third with even more elevated carbon dioxide. Study growth rates, mass, fruiting, etc…

Use GIS system to map the greening of the earth.

Pg. 117 (137) Hands-on-Performance vs. Interactive Computer Investigations

Make clear that computer model simulations are only as good as the assumptions built into them. The more complex the phenomenon to be simulated and the farther out in time projections are made, history, research, and data show the less accurate the model simulations are. For climate, many of the factors or forcing mechanisms that impact climate are only poorly understood, and thus attempts by modelers to mathematically capture them are very speculative and error-prone. In the end, when models are run, their outputs should be compared to hard data for phenomenon for which data is available, and if the data and the model outputs conflict, the model outputs are not to be trusted and either the model must be adjusted, or the hypothesis reexamined.

Suggested reading material or supplementary classroom material:

Short pieces or Monographs:

Craig Idso, et al., "Why Scientists Disagree About Global Warming," Non-Governmental International Panel on Climate Change, 2015; https://www.heartland.org/publications-resources/publications/why-scientists-disagree-about-global-warming

Anthony Watts and James Taylor, "Climate at a Glance: Facts for Climate Realists," The Heartland Institute, 2021, **(insert link here)**

A Global Warming Primer, H. Sterling Burnett (ed), The National Center for Policy Analysis, 2013, http://www.ncpathinktank.org/pdfs/Global-Warming-Primer-updated-reduced-size.pdf

Book Length Discussions:

Gregory Wrightstone, **Inconvenient Facts: The science that Al Gore doesn't want you to know** (Silver Crown Productions, 2017); for purchase on Amazon: https://www.amazon.com/Inconvenient-Facts-science-that-doesnt/dp/1545614105

Bjorn Lomborg, **False Alarm: How Climate Change Panic Costs Us Trillions, Hurts the Poor, and Fails to Fix the Planet** (Basic Books, 2020); for purchase on Amazon: https://www.amazon.com/False-Alarm-Climate-Change-Trillions/dp/1541647467/ref=pd_lpo_3?pd_rd_i=1541647467&psc=1

Steven E. Koonin, **Unsettled: What Climate Science Tells Us, What It Doesn't, and Why It Matters** (BenBella Books, 2021); for purchase on Amazon: https://www.amazon.com/Unsettled-Climate-Science-Doesnt-Matters/dp/1950665798/ref=pd_bxgy_img_2/140-1238615-9822725?pd_rd_w=E89Hq&pf_rd_p=c64372fa-c41c-422e-990d-9e034f73989b&pf_rd_r=G36RP2E13RENSEN00W4W&pd_rd_r=81f9f61d-5348-4d8d-a548-46774737b653&pd_rd_wg=K9EBl&pd_rd_i=1950665798&psc=1

From: Susan Codere
To: NAGB Queries
Subject: NAEP Science Framework
Date: Wednesday, August 18, 2021 12:33:13 PM

CAUTION: This email originated from outside of the organization. Do not click links or open attachments unless you recognize the sender and know the content is safe.

Dear NAGB Science Framework Committee,
Please accept my comments regarding

Solicitation of Public Comments for Updating the Science Assessment Framework for the 2028 National Assessment of Educational Progress (NAEP)

As requested, my comments specifically address:
(a) Whether the 2019 NAEP Science Framework needs to be updated and (b) if the framework needs to be updated, why a revision is needed.
Comment - Yes, the NAEP Science Framework needs to be revised. The current NAEP Science Framework was developed before T*he Framework for K-12 Science Education* and the *Next Generation Science Standards* were completed, and thus does not reflect the focus of the most recent standards considered as the current 'national level' standards guidance documents in the US K-12 system.
and
(c) what should a revision to the framework include?
Comment - The revision should include a restructuring to place value on all 3 dimensions of science learning -- Science and Engineering Practices, Disciplinary Core Ideas, and Crosscutting Concepts in an integrated way and NOT as individual constructs and should not focus on technology applications.
The National Academies Board on Science Education has conducted numerous study sessions and produced publications to guide science assessment. This guidance should be reflected in the new NAEP Science Framework.

Thank you for this opportunity to provide public comment.

Susan Codere
ML-PBL Project Director
ML-PBL website
https://mlpbl.open3d.science/

From: Tom Keller
To: NAGB Queries
Subject: NAEP Science Framework
Date: Monday, September 13, 2021 10:33:34 AM

Thank you for this initial opportunity to provide comments and recommendations regarding the updating the Science Assessment Framework for the 2028 National Assessment of Educational Progress.

I have been active in science education at the state and national level for thirty years, as a classroom teacher, school leader, state science supervisor in Maine and senior program officer at the National Academy of Sciences. While at the National Academy, I co-directed development of the *Framework for K-12 Science Education,* with a committee of 18 scientists, engineers, educational researchers, cognitive scientists and educational practitioners, including 2 Nobel laureates.

This document is the most recent record of current research on science education, and makes some important advances that are being implemented across the country.

For this reason alone, the NAEP Science Education Framework must be reviewed and updated. The last NAEP Framework was completed prior to the findings listed in the National Academy of Sciences' Framework. The major step forward described in the National Academy's Framework is the melding of science and engineering practices, crosscutting standards and disciplinary core ideas as the fundamental unit of instruction. Separating these three dimensions reverts to past thinking on process versus content.

It is vital that the review of the NAEP Framework include significant participation by members of the Council of State Science Supervisors. As science education leaders working at the intersection of local, state, and federal policies in each state and jurisdiction, they are most aware of the systemic value of coherence between state and federal assessment and have the ability to facilitate such coherence. Assessment tends to drive instruction and it can drive us forward or backward. Coherence between state and federal assessment will provide state leaders with another tool to improve science instruction for all students.

The Council of State Science Supervisors played an outsized role in gathering and collating feedback for the 2005 NAEP Framework. I am sure that they would be happy to once again work with the Framework committee to collect meaningful feedback that represents the nation.

Relative to the three questions posed by the NAGB communication:

- *Whether the NAEP Science Assessment Framework needs to be updated.*

Clearly the NAEP Framework requires updating. The last updating was done in 2005 and this was prior to both the National Academy of Sciences' Framework and other seminal science education consensus studies reported by the Academy.

The National Academy's *Framework for K-12 Science Education* cites the need and power of instructing students in science and engineering practices, crosscutting concepts and disciplinary core ideas as a whole rather than separating science into content and practices as does the current NAEP Framework. This is a major difference for which the current NAEP Framework looks back and the National Academy's Framework looks forward.

- *If the framework needs to be updated, why a revision is needed.*

The current NAEP Framework has two separate components, science content and science practices. This leads to teaching them separately. And we know assessment tends to drive instruction. Many older textbooks have a first chapter on 'the scientific method' and never return to that topic. Science and engineering practices, a much better conceptualization of 'the scientific method', should be experienced repeatedly and the skills to do so should be constantly improved.

Also consistency between the NAEP Framework and what and how science and engineering are taught in schools, most of whom are using standards influenced by the Academy's Framework also makes the case for a revision.

- *What should a revision to the framework include?*

An important consideration is to know how the results will be used. If this truly is the Nation's Report Card and is not intended for any use by states, that brings up a different set of considerations. But if it is to be taken seriously by states, there has to be some value in it for them. So aligning as much as possible to the current science educational frameworks in use – and for most, that is the National Academy of Sciences' Framework, makes the results useful.

It is important that input of state science education leaders who work in this area daily be included in a revision.

Certainly a revision must include the three dimensions described in the National Academy's Framework. NAEP has the capacity to create assessment scenarios and bundles that assess these dimensions in an authentic and reliable way.

In summary, a revision to the NAEP Framework is necessary and I am willing to assist in any process to make that a reality.

Thank you.

Tom Keller

--
Thomas E. Keller, Ed. D.

Founder & Director

STEM Education Strategies LLC
208 N Newcastle Rd
Newcastle, ME 04553

www.stemeducationstrategies.org

Dr. Thomas R. Tretter
Professor of Science Education
Director Center for Research in Mathematics and Science Teacher Development
Director Gheens Science Hall & Rauch Planetarium
University of Louisville

Page xii (executive summary) and throughout document uses the label "***Science Practices***" in a way not completely aligned with NGSS "Science and Engineering Practices" – recommend updating these to the NGSS practices (8 of them, instead of 4) which also part of NGSS vision for the practices to cross science content and "…generate student performance expectations, and assessment items can then be developed based on these performance expectations". NOTE: this implies that all of chapter 3 will need to be revised.

Page xii (executive summary) and throughout. Need to incorporate the ***third dimension of NGSS*** as well – crosscutting concepts. These 3 dimensions (content, practices, crosscutting) then are used to generate performance expectations (detailed in NGSS) which can guide development of assessment items that measure all 3 dimensions. NOTE: May need to add an additional chapter focused on crosscutting concepts (parallel to science practices) OR add this as a primary new section in the updated "science practices and crosscutting concepts"

Page xii (executive summary) "***distribution of items***" needs to be reconsidered in light of NGSS. Both in terms of content emphases (or not) at each grade band, and if any NGSS practices should be emphasized or not.

Page xiIi (executive summary). Consider expanding the formats/types of ***interactive computer tasks***; see examples of what various states are doing in their science assessments. For example, building/modifying scientific models (different from existing 'empirical investigation' or 'simulation'). Also consider making interactive computer tasks a standard part of the assessment for all testtakers rather than a subset, given the widespread availability of computers and/or internet access (especially post-COVID pandemic when school systems across the world had to figure out how to instruct online – and make those resources accessible to all students).

Page 5 (and elsewhere) – update to indicate "***framework informed by NGSS***" (which have replaced the prior Benchmarks and NSES). Aligned with many of the comments above about updates to align with NGSS.

Will need to update "***Descriptions of NAEP Basic, NAEP Proficient, and NAEP Advanced*** must be as clear as possible" so that the NAEP levels are aligned with all 3 dimensions of NGSS thinking that would be assessed… so that for example 'basic' still includes descriptions about the level of skill/understanding that students bring to using practices, or using crosscutting concepts as a sense-making lens.

WISCONSIN DEPARTMENT OF
Public Instruction

Jill K. Underly, PhD, State Superintendent

September 30, 2021

Lesley Muldoon, Executive Director
National Assessment Governing Board
800 North Capitol Street, NW, Suite 825
Washington, DC 20002

Document Number: 2021-17676

Dear Ms. Muldoon:

The Wisconsin Department of Public Instruction (WDPI) appreciates the opportunity to provide public comment on preliminary guidance by the National Assessment Governing Board (NAGB) in updating the Assessment Framework for the 2028 National Assessment of Educational Progress (NAEP) in Science. Please find the WDPI's feedback in response to the NAGB's updates to the Science Assessment Framework for the 2028 NAEP below.

The 2019 NAEP Science Framework does not need to be updated.

The stated purpose of the NAEP in Science is to evaluate trends in scientific literacy overall and by demographic group. The current content, practices, and test design adequately accomplish this goal. The focus on phenomena and content linked to practice mirror the National Research Council's (NRC) Framework for K-12 Science and the Next Generation Science Standards (NGSS). While that mirroring is not a strong alignment, that is not the purpose of the NAEP.

Further, a review would likely result in relatively small changes that will not significantly change the impact this framework and test have on the field. Changes are unlikely to affect student learning. Instead, they are more likely to perpetuate the unhelpful focus on a practice referred to as gap gazing[1], which highlights achievement gaps instead of focusing on real systems change.

If a committee is formed, this could be an opportunity to expand innovative approaches to the NAGB's work. The WDPI suggests that the NAGB dedicate some time and capacity to developing materials and guidance that support systems of assessment and effective implementation of those systems.

If a revision is going to happen, a few ideas should be considered.

The WDPI believes that if the NAGB updates the 2019 NAEP Science Framework, the following suggestions must be taken into consideration:

1. Replace the Depth of Knowledge - Level One items that rely on memorization skills with items that test the student's skills in application, evaluation, and analysis of concepts.

[1] https://www.researchgate.net/publication/227252559_Beyond_Gap_Gazing_How_Can_Thinking_About_Education_Comprehensively_Help_Us_Reenvision_Mathematics_Education

2. Allow for deeper exploration of phenomena by having sets of multiple items digging into a particular phenomenon.
3. Create phenomena or contexts that would interest students and engage them in a real-life scenario that requires critical societal thinking and would better reflect scientific literacy instead of looking at phenomena that are disconnected from any meaningful context (e.g., random food webs).
4. Involve learners by engaging them in the practices of modeling, asking questions, and critiquing evidence or scientific practice, which could support more effective sensemaking and prompt scientific literacy development.
5. Align the NAEP Science Framework completely to the 2012 NRC Framework for K-12 Science and the NGSS, which would provide a more coherent signal and system for the field.

Thank you for the opportunity to comment. If you have any questions, please contact Viji Somasundaram, Director, Office of Educational Accountability, at visalakshi.somasundaram@dpi.wi.gov.

Sincerely,

Carl J Bryan

Carl Bryan
Administrative Rules Coordinator

CB:vs

Attachment B

IES Institute of Education Sciences

550 12th Street SW
Washington, DC 20202
Office: 202 123 4567

Assessment Development Committee
Item Review Schedule
November 2021 – May 2022
October 8, 2021

Review Package to Board	Board Comments to NCES	Survey/ Cognitive	Review Task	Approx. Number Items	Status
November 2, 2021	November 30, 2021	Survey	Mathematics (4, 8) Existing Item Pool Review	52*	
November 2, 2021	November 30, 2021	Survey	Reading (4, 8) Existing Item Pool Review	48*	
November 24, 2021 (Off-cycle)	December 21, 2021 (Off-cycle)	Cognitive	Reading (4, 8) Passage Clearance & Concept Sketch Review *2026 Operational (2024 Pilot)*	40-50 passages 21-25 concept sketches	
January 3, 2022 (Off-cycle)	January 27, 2022	Cognitive	Mathematics (4, 8) Concept Sketch Review *2026 Operational (2024 Pilot)*	12	
February 17, 2022	March 11, 2022	Survey	Mathematics (4, 8) Pre-Cog lab review *2026 Operational (2024 Pilot)*	10-50	
February 17, 2022	March 11, 2022	Survey	Reading (4, 8) Pre-Cog lab review *2026 Operational (2024 Pilot)*	10-50	

*Items that appear at multiple grades are only counted once.

Committee on Standards, Design and Methodology

November 16, 2021

2:00 – 4:00 pm ET (Virtual)

CLOSED

AGENDA

2:00 – 2:10 pm	**Welcome, Introductions, and Agenda Overview** *Suzanne Lane, Chair*	
2:10 – 2:50 pm	**Update: 2022 NAEP Administration (CLOSED)** *Enis Dogan, NCES*	Sent under separate cover
2:50 – 4:00 pm	**Briefing and Discussion: Improving Measurement and Reporting of Lower-Performing Students (CLOSED)** *Suzanne Lane* *Enis Dogan*	Attachment A
Information Item	**Review and Revision of NAEP Mathematics and Reading Achievement Level Descriptions**	Attachment B

Improving Measurement and Reporting of Lower-Performing Students

During the most recent Governing Board meeting, Ebony Walton of the National Center for Education Statistics (NCES) gave a presentation entitled, “A Decade of Monitoring Study Progress (or Lack Thereof) Through the Lens of NAEP.” One of the main findings was that the scores of lower-performing students in math, reading, and science at grades 4, 8, and 12 declined nearly across the board from 2009 to 2019, while the scores of higher-performing students generally improved or held constant during the same time period. These patterns are sometimes referred to as divergent trend lines.

The August COSDAM discussion focused specifically on the growing percentage of students in the category below NAEP Basic and the need to better understand what lower-performing students know and can do. Some COSDAM members have expressed interest in exploring whether the Board should reconsider the current policy of defining only three achievement levels: *NAEP Basic*, *NAEP Proficient*, and *NAEP Advanced*. The policy specifies that, “The remaining region that falls below the *NAEP Basic* cut score shall be identified as ‘below *NAEP Basic’* when a descriptor is necessary.” The percentage of students in this lowest category is reported but there is no accompanying achievement level description (ALD).

One current challenge to reporting additional information for students performing below NAEP Basic is the relatively lower number of items at the lower end of the scale for some assessments. During the August COSDAM meeting, Enis Dogan, NCES, presented some item-person maps to demonstrate the alignment between student score distributions and the difficulty level of test items. Although there was variation by grade and subject, in general the item-person maps showed that there were more students than items at the lower end of the scale and more items than students at the upper end of the scale.

The NAEP frameworks generally contain a lot of rigor and cognitive complexity. NCES has several efforts underway to increase the number of items that lower-performing students can access. For example, in NAEP Mathematics a special effort was undertaken several years ago to provide additional clarifications and constraints for measuring basic skills in some of the framework objectives in the Mathematics Assessment and Item Specifications to produce easier items; the resulting items are known as “KaSA items,” or Knowledge and Skills Appropriate items. Similar efforts could be applied to other subjects.

An alternative approach to understanding what lower-performing students can do was demonstrated by a special study recently released by NCES about the knowledge and skills possessed by students who performed below the *NAEP Basic* achievement level on the grade 4 reading assessment. This study focused on oral reading fluency, which is not part of the NAEP Reading Framework, but is a prerequisite skill to reading comprehension as defined by the framework.

Given the increasing divergence in performance, it may not be sufficient to develop more items at the lower end of the scale unless those items can also be administered to lower-performing students; this would require an adaptive testing model. The NAEP program has engaged in extensive research on adaptive testing, but it is not yet part of the operational assessment.

During the November COSDAM meeting, Enis Dogan will describe current NCES efforts to increase the number of items at the lower end of the scale and will briefly review previous and current efforts on adaptive testing for NAEP. COSDAM Chair Suzanne Lane will lead the Committee in a discussion of other efforts that the Board may want to consider to achieve the goal of improving measurement and reporting for lower-performing students.

This session is closed due to the inclusion of secure data on NAEP item pools.

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, and 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 or NAEP Reading Framework currently scheduled for implementation in 2026. 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, making 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.

During the May 2021 COSDAM meeting, the final Design Document for the NAEP ALD Review study was discussed by the Committee members and there were no recommended changes.

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.

In response to previous COSDAM discussions, the project schedule was modified to account for attempting to conduct the panel meetings in person (if public health conditions allow) in late 2021 and early 2022. The pilot meeting took place virtually during the week of October 25, 2021, and the operational meeting will take place during the week of February 21, 2022. The resulting 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.

Project Update (October 2021)

Following the August 2021 COSDAM meeting and additional discussion with COSDAM leadership about the risks of conducting in-person panel meetings in the wake of the Delta variant of COVID-19, a contract modification was issued to revert the panel meetings back to a virtual format and extend each panel meeting from 4 days to 5 days to account for the change in format.

Project work since the last COSDAM update in early August has focused on adapting to the virtual format and preparing for the pilot study, including: recruiting panelists, preparing materials, training facilitators, organizing meeting logistics, and conducting quality control checks. The TAC met on August 23 and September 28-29 to review and discuss materials and other preparations for the pilot study.

The pilot study took place from October 25-29; following review by the TAC in November and early December, lessons learned and any proposed modifications for the operational meeting will be discussed with COSDAM via a focused webinar in mid-December or early January.

Reporting and Dissemination Committee

November 4, 2021
3:30 – 5:00 pm
Zoom

AGENDA

3:30 – 4:45 pm	**Anticipating the Next NAEP Release: Lessons from State Assessments and Questions to Consider** *Tonya Matthews, Chair* *Damian Betebenner, Center for Assessment*	*Attachment A*
4:45 – 5:00 pm	**General Updates** *Marty West, Vice Chair*	*See Executive Committee tab for SV update*

Anticipating the Next NAEP Release:
Lessons from State Assessments and Questions to Consider

The Reporting and Dissemination Committee meeting on November 4th will focus on anticipating issues likely to arise from reporting the 2022 Nation's Report Card in reading and mathematics. Presumably, the public will compare 2022 results to 2019 results and will try to determine how COVID-19 shaped NAEP scores and trends.

Reporting and Dissemination Committee members know that NAEP should not be used to draw causal conclusions or to evaluate impacts of policies or disruptions to education. However, the data will provide valuable insights. Thus, the central question for the meeting will be: ***How can we prepare now to help NAEP's stakeholders and the public use and interpret next year's results?***

The discussion at R&D will be informed by several sources. At the November 4th meeting, Damian Betebenner from the Center for Assessment will present about his work interpreting state assessment data within the context of COVID, explaining potential pitfalls to anticipate. To prepare for his presentation, please read this post.

Note that this meeting's discussion will center on next year's reporting specifically. Yet, over the last two years, stakeholders have recommended ways to improve NAEP reporting generally. These ideas may hold relevance to reporting issues for NAEP 2022.

First, in March 2021, the Governing Board hosted Gerunda Hughes to discuss equity and NAEP. Early indicators point to COVID's differential effects on student learning by subgroup and by prior performance. Delving into how Dr. Hughes' recommendations intersect with post-COVID results on NAEP may prove useful.

Gerunda Hughes asked the Governing Board to make each aspect of NAEP more equitable, including its reporting. Hughes urged the Board to:

- Expand NAEP's subgroup comparisons to capture and report differences created by societal inequity, cultural inequity, familial inequity, staffing and instructional inequity, assessment inequity, among others.
- Report all comparisons between subgroups defined by race/ethnicity (as opposed to the current approach of using white students as the reference group). Highlight differences in ways that foster new insights to help improve student achievement, not highlight already known disparities.

- Broaden the set of contextual variables and demographics which NAEP reports.
- Facilitate secondary analyses with contextual data.

In addition, the TUDA Task Force, a partnership with the Council of the Great City Schools which comprises a dozen leaders from districts participating in the Trial Urban District Assessment program (TUDA), emphasized the importance of setting expectations for NAEP 2022 results and raised critical questions about NAEP reporting:

- What are more useful and actionable ways of disaggregating data to understand how COVID shaped performance?
- How can reporting contextualize the Importance of NAEP within the pandemic and consider the impact of *changed learning opportunities* on results?

The TUDA Task Force members counseled the Governing Board to ensure communications distinguish what NAEP results can say from what they cannot. The Task Force also urged the Board to publicize NAEP's track record in helping states and districts focus on skills where students are improving and where they need additional support. The Board should not offer policy solutions but should amplify facts and statistics.

Ian Rowe, a scholar at the American Enterprise Institute and advocate for using NAEP data, has implored the education community to change the reporting of disparities in academic outcomes. Rowe questions why reports frame outcomes primarily through performance gaps by poverty and race. Even if gaps between black students and white students narrow by a few points, the majority of black students and the majority of white students remain below *NAEP Proficient* on reading and mathematics. The gap shrinks; the deficit persists. If NAEP collects and reports different data—disaggregating results so that racial and income categories are no longer viewed as monolithic—stakeholders may arrive at a more nuanced and accurate understanding of academic disparities, which may lead to change.

In an email to the Board, committee member Tyler Cramer praised recent examples of the Board's efforts to highlight ways that NAEP data inform conversations on educational equity. He urged the Board to extend this work by connecting NAEP data to school finance data and/or to civil rights data, clarifying what NAEP can and cannot do, especially when compared to state assessments.

In sum, as the committee deliberates on how to report and message NAEP 2022 results, stakeholders from within and beyond the Board urge NAEP reporting to:

1. Consider how *changed learning opportunities* may shape results
2. Highlight all subgroup comparisons
3. Disaggregate data beyond monolithic categories of race/ethnicity and NSLP-eligibility
4. Report on differences created by other types of inequity - societal, cultural, familial, financial (i.e., school resources), etc.

5. Clarify NAEP’s unique contributions, specifying what NAEP can and cannot do, to enable accurate interpretations of results

In the discussion after Dr. Betebenner’s presentation, committee members will consider how these recommendations could be incorporated within reporting and messaging NAEP 2022.

Joint Meeting of Assessment Development and Reporting and Dissemination Committees

Friday, November 12, 2021
2:30 – 3:30 pm

Zoom:
https://us02web.zoom.us/webinar/register/WN_VfB0D8cWSjub5HnmzMzx2g

AGENDA

2:30 – 2:35 pm	**Welcome** *Dana Boyd, Chair, Assessment Development Committee* *Tonya Matthews, Chair, Reporting and Dissemination Committee*	
2:35 – 2:40 pm	**Overview of Contextual Questionnaire Review** *James Deaton, National Center for Education Statistics*	
2:40 – 3:30 pm	**Review of Contextual Questionnaire Items** *Dana Boyd* *Tonya Matthews*	*Link available in email*

Nominations Committee

November 15, 2021

5:00 – 6:00 pm ET (Virtual)

CLOSED MEETING

AGENDA

5:00 – 5:15 pm	**Welcome and Agenda Overview** *Paul Gasparini, Chair*	
5:15 – 5:30 pm	**Preview 2021 Nominees** *Paul Gasparini*	Attachment A
5:30 – 5:40 pm	**Review Rating System and Timeline** *Munira Mwalimu, Nominations Committee Staff*	Attachment B
5:40 – 5:50 pm	**Discussion: Rating Process** *Paul Gasparini*	
5:50 – 6:00 pm	**Next Steps** *Paul Gasparini*	

ACTION: 2026 NAEP Reading Assessment and Item Specifications

The Board discussed the 2026 NAEP Reading Assessment Framework at several meetings beginning in July 2020 and unanimously approved the framework during the August 2021 quarterly meeting. The final step in the framework development process is the creation of Assessment and Item Specifications to provide direction to the National Center for Education Statistics (NCES) on implementing the framework. Board adoption of the Reading Assessment and Item Specifications at the November Board meeting will ensure that NCES has adequate time to implement the assessment for the 2026 NAEP administration.

Following Board approval of the Assessment and Item Specifications, the Board's role in developing the framework is complete; responsibility then shifts to NCES to implement the item development process. NCES has provided a short description (attached) of how they use a framework and the accompanying specifications for item development. The Board policy on Item Development and Review describes general principles for the development and review of all NAEP assessments.

According to the Board policy on Framework Development, NCES is the primary audience for the Assessment and Item Specifications and the document is intended to include the following information:

- types of items;
- guidelines for stimulus material;
- types of response formats;
- scoring procedures;
- achievement level descriptions;
- administration conditions;
- ancillary or additional materials, if any;
- considerations for special populations;
- sample items, including a substantial number and range of sample items with scoring guidelines for each grade level; and
- any unique requirements for the given assessment

The Reading Assessment and Item Specifications include much of the same text from the framework (often verbatim) with additional elaboration on the information listed above. The achievement level descriptions (ALDs) are included as appendices.

In addition to providing some further details intended to inform item development, the Reading Assessment and Item Specifications include some edits made to the original version of the ALDs that appeared in Appendix B of the framework document. The purpose of these recent edits was to address references to knowledge and skills that are not feasible to assess on NAEP (e.g., rewriting a story from a different character's perspective). These edits were made based on feedback from Board staff, NCES staff, and the Technical Advisory Committee and are provided in tracked changes as an additional attachment.

Since NCES is the primary audience for the specifications, they reviewed and provided feedback on several drafts of this document and have confirmed that they have no outstanding concerns.

A joint Assessment Development Committee (ADC) and Committee on Standards, Design and Methodology (COSDAM) review took place on September 21; as a result of that review and follow up conversations with the ADC and COSDAM Chairs and Vice Chairs, the appendix that included recommendations for special studies was removed from the document. To address a request from the joint meeting, language was added to be more explicit about the need for there to be a distribution of comprehension targets at each achievement level. A summary of the review and resolution of issues discussed was sent to ADC and COSDAM members on October 6; no additional questions or concerns were raised.

Role of the Frameworks, Specifications and ALDs in the Item Development Process

Once the Frameworks and Specifications have been adopted by the National Assessment Governing Board, it is NCES's responsibility to use those documents to guide item development. The Achievement Level Descriptors (ALDs) are included in the guidance provided by the two larger documents. ALDs provide a context for crucial knowledge and skills described in the Frameworks.

Test design (including passage selection in Reading), item-writing, and review are iterative processes that involve staff from NCES, its contractors, and members of external advisory committees. Each of these entities use the guidance provided by the Frameworks and Specifications, including the ALDs, when fulfilling their respective roles. Alignment, coverage, relevance, and difficulty are considered at each point by the full complement of reviewers. In this process, the ALDs provide a means to evaluate whether the collection of items and blocks reflect the range of levels of student ability. Prior to moving on to the next step of development, there is consensus among the groups of reviewers that the content is consistent with the guidance provided. The ongoing reviews act as a system of checks and balances, where NCES's interpretation and application of these documents is either confirmed or challenged in each round of review.

Item Development Process

NCES uses the following item development cycle to develop every assessment item and block carefully before they are integrated into the operational NAEP assessment. Each phase of the cycle includes review by one or more organizations familiar with the subject area Framework, Specifications and ALDs.

1. **(For Reading only) Reading Text Selection.** Texts and text sets are identified by the ETS reading item development team. Proposed texts are reviewed by the ETS bias and sensitivity review team and the ETS editorial staff before they are submitted to NCES item development staff and contractors, Reading Standing Committee members, and the Governing Board Assessment Development Committee (ADC).
2. **Initial Item Reviews.** For Reading, after passages are approved, items are developed by the ETS reading item development team. ETS subject area teams with appropriate and deep content expertise develop draft items for the various subject area assessments. For all subject areas, once draft items are completed and any stimulus (e.g., images, videos) are selected, ETS reviewers conduct framework alignment and construct

measurement reviews, as well as editorial and bias and sensitivity reviews. Items and stimuli are then reviewed by NCES item development staff and contractors and subject area Standing Committees.

3. **Pretesting.** Following initial item reviews, items and support features are pretested, using: 1) cognitive interviews with individual students to determine how they respond to proposed items and stimuli (including texts for reading); 2) tryouts under “live” testing conditions with groups of 50–200 students from the target population; and 3) usability studies. Data from these trials inform ETS item development team’s item revisions.
4. **Revised Item Reviews**. After items and stimuli are pretested and revised by item developers, ETS reviewers conduct editorial, bias and sensitivity reviews, cold read reviews, and language accessibility reviews. Items are then reviewed by NCES item development staff and contractors and subject area Standing Committees. 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.
5. **Piloting.** New blocks of items are folded into the administration of operational blocks of a live assessment. By comparing student and item performance across the new and old items and blocks, NAEP developers can determine whether items effectively scale together with the old, measuring the same underlying subject area construct.
6. **Post-pilot Reviews**. Following the collection of pilot data (up to 3,000 students per item), the following groups review pilot data, item level analyses, and items (as well as texts for Reading):
    - ETS item development team
    - ETS data analysis and reporting team
    - ETS Differential Item Functioning 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 subject area Standing Committee
    - Governing Board ADC

Following post-pilot reviews, items (and blocks in Reading since items travel with their passages) are selected for operational use, based on their performance in the pilot, committee reviews, contractor recommendations, and NCES’s decisions. All of these are informed by the Frameworks, Specifications and ALDs.

Reading Assessment and Item Specifications

for the

2026 National Assessment of Educational Progress

National Assessment Governing Board

U.S. Department of Education

November 3, 2021

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

TABLE OF CONTENTS

What Is In This Assessment and Item Specifications Document?

This document is a companion to the *Reading Framework for the 2026 National Assessment of Educational Progress* (NAEP). The 2026 NAEP Reading Framework informs NAEP assessment development, describing the subject matter to be assessed and the questions to be asked, as well as the assessment's design and administration. This Assessment and Item Specifications document extends the Framework, providing greater detail about development of the items for the 2026 NAEP Reading Assessment. The primary audience for this document is the National Center for Education Statistics (NCES) and its contractors, who will use both the Specifications and the Framework to develop the 2026 NAEP Reading Assessment.

Background on NAEP

The National Assessment of Educational Progress (NAEP) has measured student reading achievement nationally since 1971, and state-by-state since the early 1990s, providing the nation with a snapshot of what students in this country know and can do in reading. Starting in 2002, urban school districts that meet certain selection criteria could volunteer to participate in the Trial Urban District NAEP Assessment (TUDA).

There are two distinct assessments administered by NAEP. The NAEP Long-Term Trend assessment has measured trends in achievement among 9-, 13-, and 17-year-old students nationally since 1971, and the assessment's content has been essentially unchanged ever since. The second assessment, referred to as "main NAEP," is adjusted over time to reflect shifts in research, policy, and practice. The content and format of the main NAEP Reading Assessment are the focus of the Framework and this Assessment and Item Specifications Document.

The main NAEP Reading Assessment is administered at the national, state, and selected urban district levels every two years, by Congressional mandate. In reading, NAEP results are reported on student achievement in grades 4, 8, and 12 at the national level, and for grades 4 and 8 at the state level and for large urban districts that volunteer to participate.

Overview of the 2026 NAEP Reading Assessment and Item Specifications

To develop the 2026 NAEP Reading Framework, WestEd, under contract to the Governing Board, engaged in a comprehensive process that involved extensive review of the scientific research literature on reading; consultation with three committees of national and state policymakers, state assessment staff, reading educators, and others who use data from the NAEP Reading Assessment; and wide public review of successive drafts of the Framework. The 2026 NAEP Reading Assessment will be developed to represent the content emphasis, complexity of reading, item format guidelines, and other requirements of the NAEP Reading Framework.

The 2026 NAEP Reading Framework and these accompanying Assessment and Item Specifications were developed in a time of intense interest in the improvement of reading achievement and an attunement to the scientific literature on the acquisition and growth of reading skills. NAEP's purpose has always been "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 subjects[s]" (National Assessment of Educational Progress Reauthorization Act, 2002) and the updates to the Framework reflect this purpose.

The structure of the Assessment and Item Specifications mirrors the structure of the Framework. As with the Framework, the Assessment and Item Specifications document is divided into four chapters, the contents of which are briefly described below. Following the chapters is a set of appendices.

Chapter 2 of the Framework defines reading comprehension and describes major components of the 2026 NAEP Reading Assessment. Chapter 2-S of the Assessment and Item Specifications frames reading comprehension as a meaning-making process characterized by diverse texts and varied reading purposes.

Chapter 3 of the Framework describes the design and development of the 2026 NAEP Reading Assessment. Chapter 3-S of the Assessment and Item Specifications provides illustrations of how the assessment design can be enacted.

Chapter 4 of the Framework explains how the results of the 2026 NAEP Reading Assessment will be reported. Chapter 4-S of the Assessment and Item Specifications 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 the NAEP Reading Assessment in the Nation's Report Card.

NAEP Administration and Student Samples

As currently planned, the 2026 NAEP Reading Assessment will be administered to students in grades 4, 8, and 12, and results will be reported at national, regional, and state levels. To be able to provide accurate estimates of student reading achievement, schools throughout the country are randomly selected to participate in the assessment in a process that includes stratification to provide adequate representation of the broad population of U.S. students, the populations of students in each state participating in NAEP, and the populations of students from participating large urban districts. Accordingly, the assessed sample includes schools of various types and sizes from a variety of community and geographical regions. The student populations of these schools represent different levels of economic status; racial, ethnic and cultural backgrounds; and instructional experiences. Students with disabilities and English language learners are included, and accommodations are provided.

The test design also considers the need to obtain reliable estimates for the population of students at each assessed grade level. Therefore, a large pool of assessment items is developed and used to build multiple test forms using a matrix sampling design. That is, many items are administered, but each student takes only a subset of the items. In addition to the reading items, contextual questionnaires accompany each test form and are administered in separately timed sessions. Each student will spend approximately one hour taking the 2026 NAEP Reading Assessment.

NAEP reporting provides comprehensive information about what U.S. students know and can do in reading. In addition, NAEP provides comparative subgroup data according to gender, race/ethnicity, socioeconomic status, and geographic region; describes trends in performance over time; and reports on relationships between student achievement and contextual variables.

Reporting Results of the NAEP Reading Assessment

The NAEP Reading Assessment is an assessment of overall achievement, not a tool for diagnosing the needs of individuals or groups of students. Therefore, reported scores are always at the aggregate level. By law, scores are not produced for individual schools or students. Results are reported for the nation as a whole, for regions of the nation, for states, and for large districts that volunteer to participate in the NAEP Trial Urban District Assessment (TUDA).

Under the provisions of the Every Student Succeeds Act legislation, states receiving Title I grants must include assurances in their state assessment plans that they will participate in the NAEP Reading and Mathematics assessments at grades 4 and 8. Local school districts that receive Title I funds must agree that they will participate in biennial NAEP administrations at grades 4 and 8, if they are selected. However, participation in NAEP is not considered a substitute for the federally mandated state-level assessments in reading and mathematics at grades 3 to 8.

National and state level results are reported in terms of scale scores, achievement levels, and percentiles. Average scores for groups of students are given on the NAEP 0–500 scale and as percentages of students who attain each of the three achievement levels established and defined by the National Assessment Governing Board (Governing Board). These policy definitions can be found in the Governing Board's *Developing Student Achievement Levels for the National Assessment of Educational Progress Policy Statement* (2018a) and in Exhibit 1.1.

Since 1990, the Governing Board has used student achievement levels for reporting results on NAEP assessments. The achievement levels represent an informed judgment of "how good is good enough" in the various subjects that are assessed.

Policy-level definitions describe, in general terms, what students at each grade level should know and be able to do on the NAEP assessment to perform at the NAEP Basic, NAEP Proficient, or NAEP Advanced levels. Achievement level descriptions of student performance at each grade 4, 8, and 12 for the 2026 NAEP Reading Framework are provided in Appendix A. These updated reading-specific achievement level descriptors will replace those from the previous Framework to guide item development and initial stages of standard setting (if necessary) for the 2026 NAEP Reading Assessment.

Exhibit 1.1: Generic Achievement Level Policy Definitions for the National Assessment of Educational Progress

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.

Comparison of the 2009–2019 and the 2026 NAEP Reading Frameworks and Assessment and Item Specifications

This Assessment and Item Specifications document reflects several major changes, both those made to the Framework and those made to support item development. The changes are summarized in the following section and in Exhibit 1.2.

Building from the Framework used for the 2009–2019 NAEP Reading Assessments and following from digital innovations, updates to the Framework include consideration of 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 and associated assessment and item specifications are presented in Exhibit 1.2. While the new documents are an update to the current documents, the continuity between the existing Framework and Assessment and Item Specifications and the 2026 NAEP Reading Framework and Assessment and Item Specifications is substantial.

Exhibit 1.2. Similarities and Differences Between the 2009–2019 and 2026 NAEP Reading Frameworks and Assessment and Item Specifications

	Current Framework and Assessment	**2026 Framework and Assessment & Item Specifications 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, 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 (UDEs)	Digitally based assessment as of 2017 includes tools and support features: ● Highlighting and note-taking ● Text-to-speech on Directions and Help screens	Types of UDEs and possible examples: ● Task-based UDEs – Highlighting and note-taking – Text-to-speech on Directions and Help Screens

	Current Framework and Assessment	2026 Framework and Assessment & Item Specifications Update
	• 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 • Multipart 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	– Zoom-in and selection of color schemes – Sequential directions and transitions for reading a collection of texts – Look-back buttons to return to relevant section of text – Graphic organizers – Item foreshadowing – Multipart response frames – Samples of student writing as examples – Resetting by providing correct response to answered questions • 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 • Informational UDEs – Text providing brief topic previews – Pop-up notes for definitions of obscure words or phrases that are not part of the Comprehension Target being tested
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 the following: • disciplinary contexts • socioeconomic status within race/ ethnicity, whenever feasible • 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

Aligning with the Framework and the Assessment and Item Specifications

The assessment should be developed so that it is aligned with the guidelines as set forth in the Framework and in these Assessment and Item Specifications. More specifically:

- The content of the assessment should be matched with the content of the Framework and the Assessment and Item Specifications. The assessment as a whole should reflect the content emphasis, complexity of reading, item format guidelines, and other requirements outlined in the Framework.
- While it is not possible to cover all possible combinations of content and complexity for each achievement level on one assessment, appropriate alignment between the assessment and the Framework and Assessment and Item Specifications at each grade should be maintained in the item pools. The developer should avoid under- or overemphasizing particular content or complexity levels.
- The assessment should report and interpret scores based on the Framework, the Assessment and Item Specifications, and the NAEP Achievement Level Descriptions (ALDs). That is, the assessment should be developed so that scores will reflect both the guidelines in the Framework and Assessment and Item Specifications and the range of performances illustrated in the NAEP Reading Framework ALDs.
- The assessment design should match the characteristics of the targeted assessment population. That is, the assessment should give all students tested the opportunity to demonstrate their knowledge and skills of reading as covered by the Framework and the Assessment and Item Specifications.

A valuable resource for learning more about NAEP can be found on the Internet at http://nces.ed.gov/nationsreportcard/. This site contains reports describing results of recent assessments, as well as a searchable tool for viewing released items. The items can be searched by many different criteria, such as grade level and content area. Information about the items includes student performance data and any applicable scoring rubrics. NAEP released items that are used as examples and nonexamples in this document are marked with the designation that matches the item name or identified by the question ID from the NAEP Questions Tool website (NCES, n.d.).

The NAEP Reading Assessment measures what students know and can do in reading. This chapter focuses on the reading content of the 2026 NAEP Reading Assessment and the connections between the content and the NAEP Definition of 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 Framework (see Chapter 2).

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 comprehension items students respond to as they take the test. Students' responses to the questions provide the observable data that NAEP uses to describe 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. 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 NAEP Reading Assessment also attends to structures of language within each Comprehension Target.

Locate and Recall

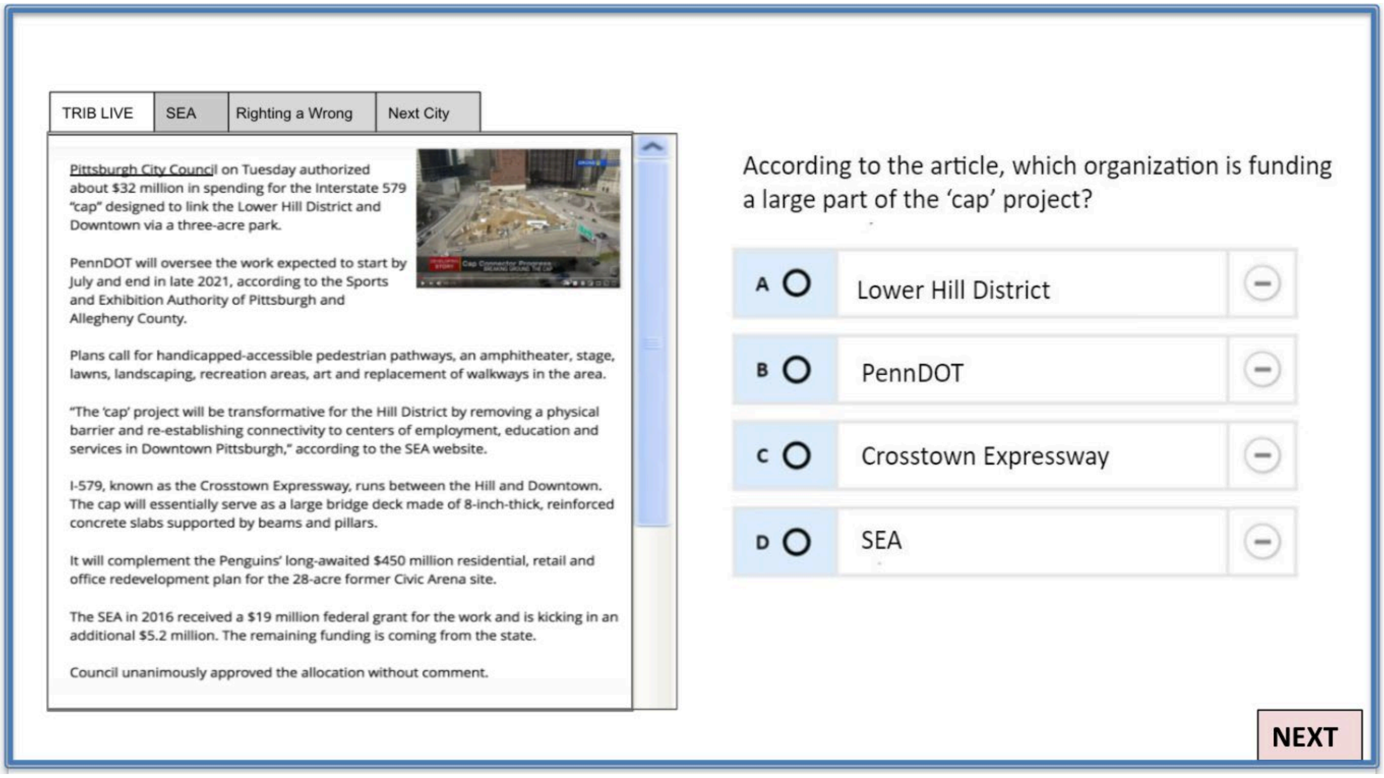
As readers locate or recall information from what they read, they may identify clearly stated main ideas or supporting details, or they may find essential elements of a story, such as characters, time, or setting. In order to comprehend, readers need to identify important information and form connections 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.

In this Grade 12 Social Studies item (Exhibit 2.1) demonstrating the Locate and Recall Comprehension Target, students locate explicit information in the text.

Exhibit 2.1. Sample Item: Locate and Recall

Comprehension Target: Locate and Recall; **Item Format:** SR-SSMC; **Key:** D
Skill: Locate explicit information in a text.



Integrate and Interpret

When readers engage in behaviors involving integrating and interpreting, they make 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.

Items assessing the Integrate and Interpret target 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.

In this Grade 8 Science item (Exhibit 2.2) demonstrating the Integrate and Interpret Comprehension Target, students determine which source texts support stated claims.

Exhibit 2.2. Sample Item: Integrate and Interpret

Comprehension Target: Integrate and Interpret; **Item Format:** SR-Grid; **Key:** See below the item.
Skill: Determine whether textual evidence supports claims.

Click on the boxes to show the claim(s) that each source supports. Some sources will have more than one box selected.

	Source #1: How Do We Remember?	Source #2: Memory Masters	Source #3: Interpreters: Silver-Tongued Masters of Memory
Find out how your memory systems process information.	☐	☐	☐
Learn how to improve your memory skills.	☐	☐	☐
Learn about the kinds of challenges presented at a memory competition.	☐	☐	☐

KEY:

Find out how your memory systems process information.: Source #1: How Do We Remember?, Source #3: Interpreters: Silver-Tongued Masters of Memory

Learn how to improve your memory skills.: Source #2: Memory Masters

Learn about the kinds of challenges presented at a memory competition.: Source #2: Memory Masters

This item appeared in the 2019–20 Grade 8 Smarter Balanced Sample Items published by The Regents of the University of California with Item ID 61235.

In this Grade 4 Literature item (Exhibit 2.3) demonstrating the Integrate and Interpret Comprehension Target, students are asked to determine the reason a story's plot is resolved based upon their analysis of the story's plot and character interactions.

Exhibit 2.3. Sample Item: Integrate and Interpret

Comprehension Target: Integrate and Interpret; **Item Format:** SR-SSMC; **Key:** A
Skill: Recognize reason for plot resolution in a story

Why does the judge decide that the merchant does not have to pay?

A	Nasreddin Hodja shows that the innkeeper's demand is silly.	
B	The innkeeper finally agrees that the merchant is right.	
C	The amount of money the innkeeper wants is much too high.	
D	Nasreddin Hodja proves that he is a good friend of the judge.	

This item appeared in the 2017 NAEP grade 4 Reading administration with NAEP Item ID 2017-4R5 #10.

Analyze and Evaluate

Readers who Analyze and Evaluate engage in 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; Goldman 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.

Items assessing the Analyze and Evaluate target might ask readers 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, 2020; 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.

In this Grade 4 Literature item (Exhibit 2.4) demonstrating the Analyze and Evaluate Comprehension Target, students are asked to provide an opinion about who is the most important character in a story and explain the opinion using specific information from the story as support.

Exhibit 2.4. Sample Item: Analyze and Evaluate

Comprehension Target: Analyze and Evaluate; **Item Format:** SCR; **Key:** N/A
Skill: Produce new text for argumentative purpose, based upon analysis of a text.

Who do you think is the most important character in the story?
Explain your opinion using specific information from the story.

This item appeared in the 2017 NAEP grade 4 Reading administration with NAEP Item ID 2017-4R5 #9.

Use and Apply

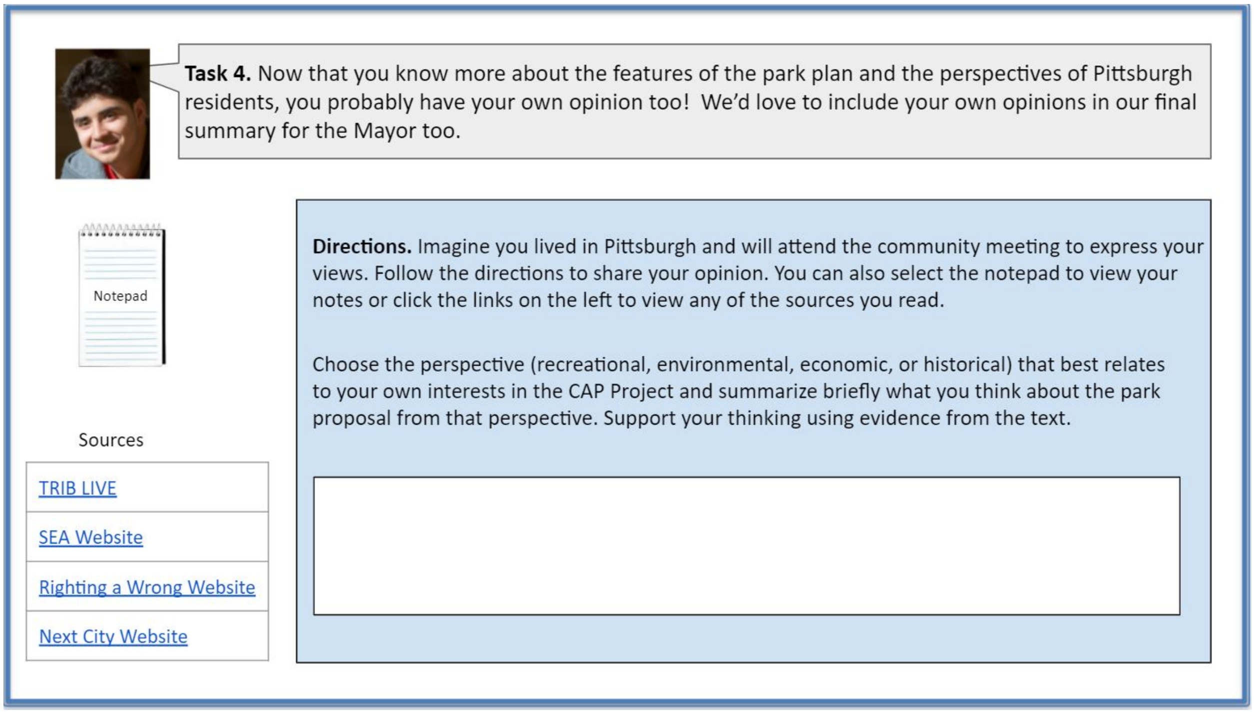
The Use and Apply target reflects the culmination of comprehension, in which understandings acquired during reading are used in new situations or applied in the development of novel ideas and products (Goldman et al., 2019; Pearson, Palincsar, Biancarosa, and Berman, 2020). This target reflects contemporary understandings that comprehension involves a series of processes that leads to readers taking some kind of action in the world outside of text. In doing so, 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 actions, they take into account the reading purposes, the norms established by genre and disciplinary conventions, and the expectations about what is deemed appropriate and compelling to members of the target audience (Gee, 2001; Goldman et al., 2016; Moje, 2015).

Items assessing the Use and Apply target 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 generating 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.

In this Grade 12 Social Studies item (Exhibit 2.5) demonstrating the Use and Apply Comprehension Target, students are asked to write an opinion, incorporating evidence from a text, that aligns to one of multiple perspectives that students had read as part of an assessment block regarding the proposed development of a city park.

Exhibit 2.5. Sample Item: Use and Apply

Comprehension Target: Use and Apply; **Item Format:** ECR; **Key:** N/A
Skill: Produce new text for argumentative purpose, based upon analyses of multiple texts and perspectives.



The photograph of Moises is sourced from [https://images.all4ed.org/high-school-boy-in-hallway (photographer Allison Shelley/The Verbatim Agency for EDUimages)](https://images.all4ed.org/high-school-boy-in-hallway).

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. By targeting a range of knowledge and skills under conditions that replicate many aspects of authentic reading, the NAEP Reading Assessment provides a more ecologically valid measure of students' reading comprehension.

Contexts and Purposes

As stated in the 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. (See the section "Organizational Features and Structures of the Reading Construct" in Appendix A-S for additional details.)

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 taking the 2026 NAEP Reading Assessment will read across three disciplinary contexts: Reading to Engage in Literature, Reading to Engage in Science, and Reading to Engage in Social Studies. Their performance will be reported by disciplinary context, along with an aggregate score for performance across all three contexts. Reading will involve texts that are drawn from the range that students encounter when reading about literature, science, and social studies. Examples of types of text to be used are provided in the text selection section of Chapter 3 and Exhibit 3.10 (also see “Selecting Texts” section in Framework Chapter 3).

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 literature 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. Individuals read a variety of literature texts to appreciate elements of craft and to reflect on point of view, varied perspectives and experiences, and human dilemmas relevant to solving personal, social, and ethical problems. 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 other authors and literary traditions; and to situate problems in contemporary and historical contexts.

Science Contexts. Science contexts focus primarily on observing and explaining the natural world. Although scientific activities do not depend exclusively on reading, texts play an important role in learning about and communicating science ideas in school and out-of-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 for 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, 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 (National Council for the 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 (e.g., trade-offs, how perspective impacts representation) and understandings about how claims are developed and supported. Individuals read a variety of social studies texts to understand historical and contemporary issues and to solve community, national, and world problems. Reading in social studies also requires the application of a broad range of the reading processes described in the Comprehension Targets.

Purposes

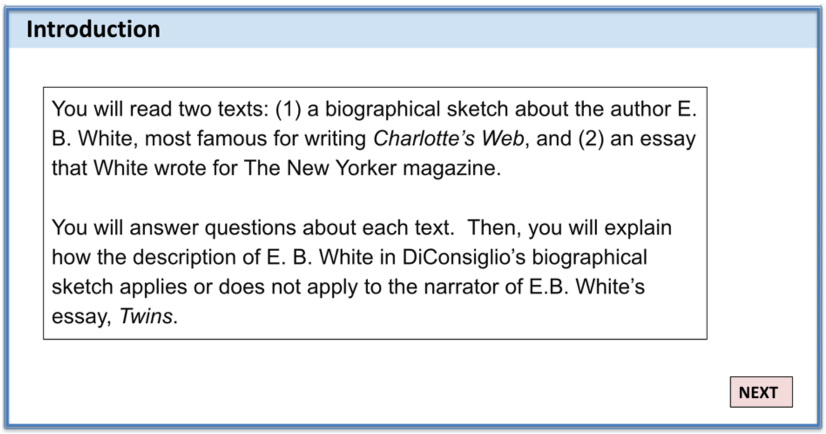
Purposes reflect a commitment on the part of NAEP to ensure that readers know why they are engaging in every part of the assessment, and to reflect the fact that all reading is done in relation to specific purposes. Within the disciplinary contexts described above, the assessment will be oriented toward 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 or Reading to Solve a Problem.

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. RSP 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.

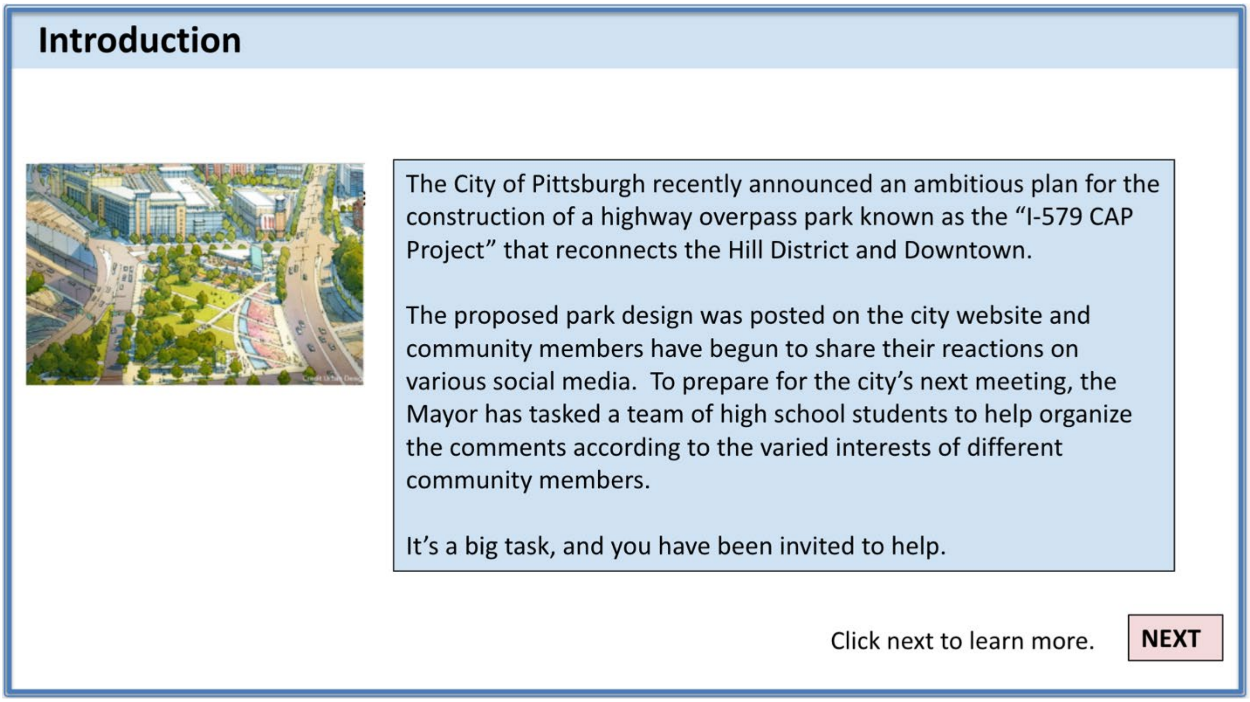
RDU items require students to read texts carefully and respond to comprehension test items generated from the four Comprehension Targets. For example, as in Exhibit 2.6, items may assess students' understanding of concepts described in a science text or the development of a literary theme. These purposes 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.

Exhibit 2.6. Example of a Reading to Develop Understanding Purpose for a Grade 8 Literature Block



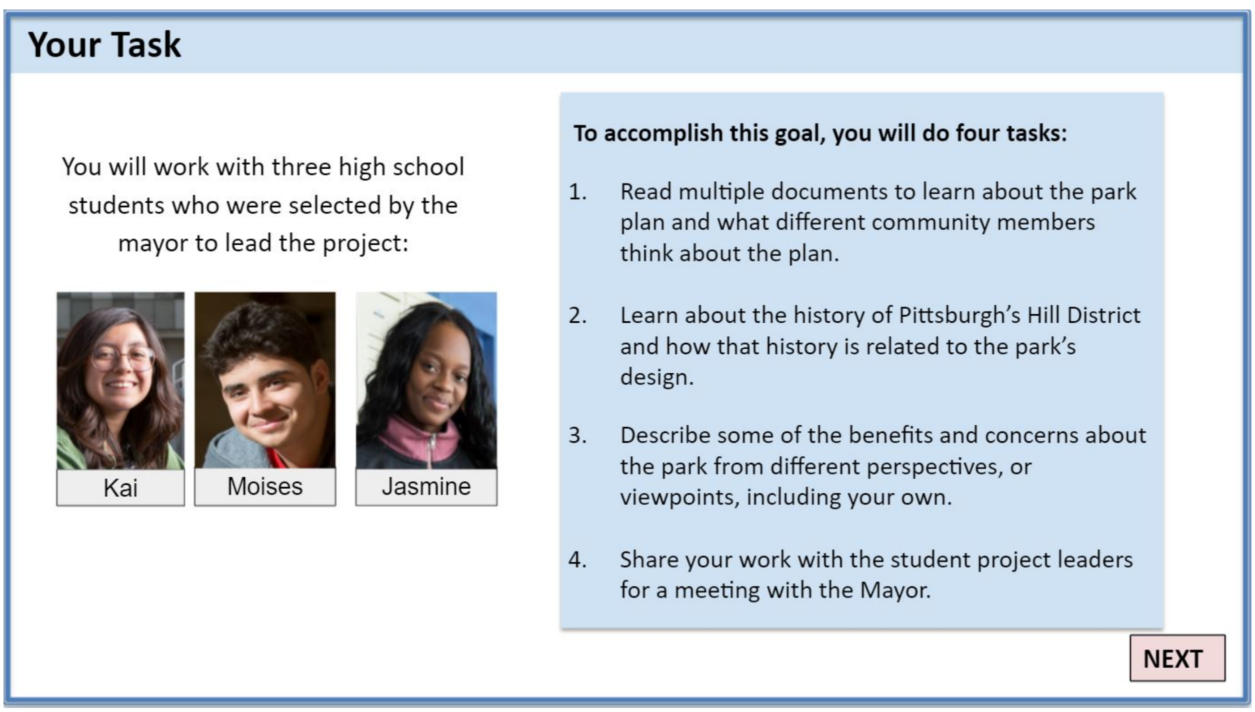
RSP items require that students work across multiple texts and perspectives in order to solve a problem. These activities entail using information gained during text comprehension in the service of a specific action or in the creation of a product. For example, as in Exhibit 2.7, readers might be asked 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.

Exhibit 2.7. Example of a Reading to Solve a Problem Purpose for a Grade 12 Social Studies Block



Specific Purposes. In addition to these broad purposes, more specific purposes for reading particular texts or engaging in particular tasks will 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. (See Exhibit 2.8.)

Exhibit 2.8. Example of four task-specific purposes in a Grade 12 Social Studies Block



The photograph of Kai is sourced from [https://images.all4ed.org/high-school-boy-and-girl-near-playground (photographer Allison Shelley for EDUimages)](https://images.all4ed.org/high-school-boy-and-girl-near-playground). The photograph of Moises is sourced from [https://images.all4ed.org/high-school-boy-in-hallway (photographer Allison](https://images.all4ed.org/high-school-boy-in-hallway)

[Shelley/The Verbatim Agency for EDUimages)](https://images.all4ed.org). The photograph of Jasmine is sourced from https://images.all4ed.org/high-school-boy-and-girl-drive-robots [(photographer Allison Shelley/The Verbatim Agency for EDUimages)](https://images.all4ed.org).

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 are intended to provide guidance about the purpose for reading and comprehending text, providing explicit connections to activities 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

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 nonschool settings. Sampling recommendations are described in more detail in Chapter 3. This portfolio of texts ranges from classic to contemporary forms that characterize reading within and across varied disciplinary contexts. These texts will reflect multiple and diverse criteria: cultural diversity, disciplinary representation, and developmental appropriateness with regard to complexity, topic, and modality.

Disciplinary Texts

The 2026 NAEP Reading Assessment will utilize texts 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. (See Exhibit 3.10.)

Literature Texts. NAEP will draw on literature texts to reflect the range of classic and contemporary genres, text structures, literary language, and traditions that students experience in their classrooms and communities. Literature texts may reflect long-standing traditions, like myths, short stories, novels, drama, and poetry. They might also include current evolving forms, such as fan fiction, author interviews, book reviews, and graphic novels. Variety in reading literature might also be reflected in specific discourse patterns, including word choice, sentence structure, and use of figurative language. Literary language can also situate narratives in time and place as well as in cultural traditions and may draw on archetypal characters typical of those traditions. Texts in literature may also cue non-literal interpretations by using irony, satire, or other literary elements and devices (Appleman, 2017; Lee, Goldman, Levine, & Magliano, 2016; Rabinowitz, 1987).

Science Texts. Science texts will reflect the formats, language, and structural elements germane to pedagogical, public, and professional science discourse. This discourse conveys 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 may draw on varied text structures, such as cause and effect, correlation, problem and solution, sequence, comparison, exemplification, descriptive classification, extended definition, and analogy. In addition to description, exposition, and narrative genres, science texts may also include many kinds of visuals, including tables, graphs, equations, diagrams, models, and flowcharts (Cromley et al., 2010; Lemke, 1998; van den Broek, 2010). Several challenging language structures common to these texts include 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 social studies including a wide array of text types, forms of representation, sources of information, and perspectives. These texts document human activity across societies and time periods and may include newspaper articles, diaries, letters, speeches, records of sale, advertisements, official government documents, photographs, cartoons, maps, artwork, music, and video and audio recordings. They may also include classroom textbooks and interpretive books and articles about events, time periods, or people. 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.

Texts in a Digital Platform

As initiated in 2017, the NAEP Reading Assessment will continue to be presented entirely 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 literature, 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 may include simulations that, for example, 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 to respond appropriately to a particular reading purpose.

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 additional considerations. The application of measures used to assess text complexity are described more fully in Chapter 3 of this document.

Text and the NAEP Definition of Reading Comprehension.

Texts used in the 2026 NAEP Reading Assessment align with the NAEP Definition of Reading Comprehension. They reflect the three disciplinary contexts, multiple genres and modalities used in both school and out-of-school settings, as well as the many kinds of digital and multimodal texts that make up the textual repertoires of most students. Utilizing a broad array of texts increases the chances that all readers will encounter texts that connect to their experiences and identities as well as 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).

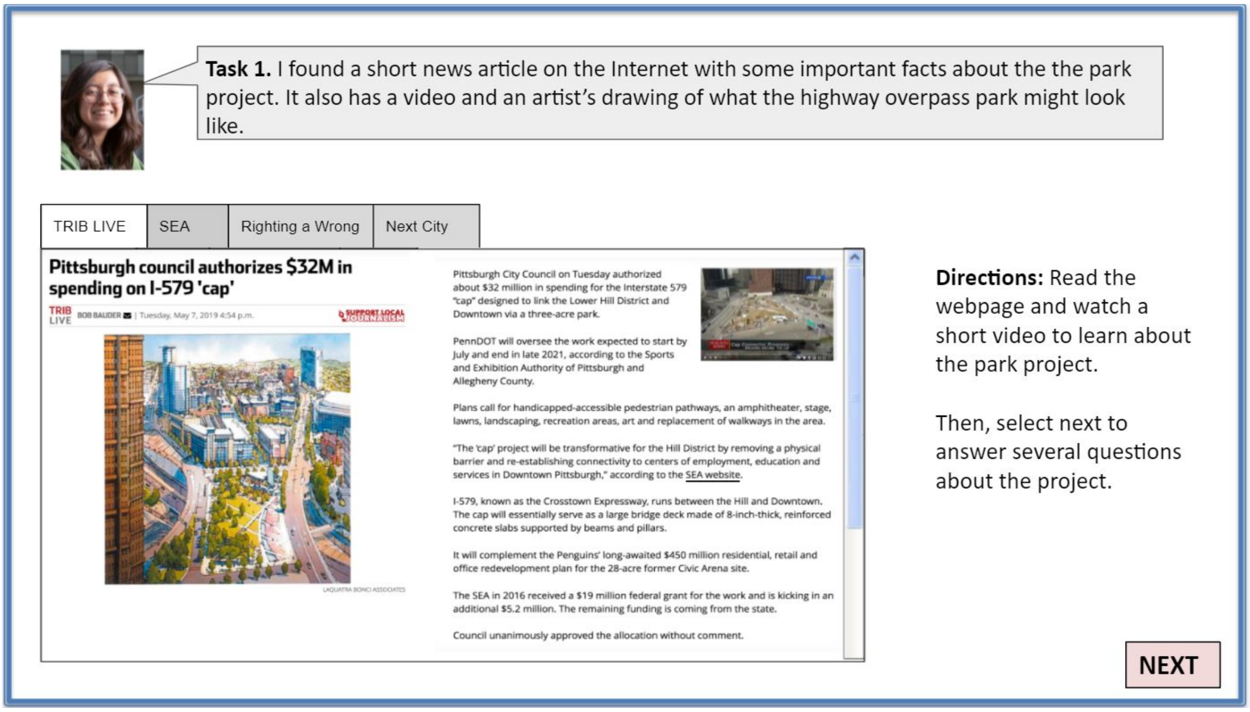
UDEs are design elements of the assessment environment intended to help all test-takers access, organize, analyze, and express ideas when engaging in complex tasks, such as reading comprehension (Johnstone, 2003; Johnstone, Altman, & Thurlow, 2006). As such, UDEs aid students' ability to engage with 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.

The 2026 NAEP Reading Assessment uses three expanded categories of UDEs: task-based, motivational, and informational.

Task-based UDEs

Task-based UDEs are designed to clarify requirements and guide readers in their use of available resources. They increase access and sustain readers' attention as they take the assessment. They clarify the expectations for readers and help them examine and use available resources within the assessment blocks (CAST, 2020; Dejong, 2006; Zhang & Quintana, 2012). They maximize the likelihood that readers are able to cognitively engage with complex NAEP-designed reading experiences within the compressed time frame of an assessment. They might include a sequential set of directions to communicate expectations for how and why readers should engage with a collection of texts; they can also help readers plan and monitor their work across multiple texts and tasks (de Jong, 2006). They might also include graphic organizers that allow readers to record and revisit their ideas, reduce time spent on searching and scrolling, and, thus, provide more time for students to read, evaluate, and engage with text content. These UDEs might also include simulated student work examples that offer models of approaches to tasks before students complete similar tasks independently (e.g., Sparks & Deane, 2014). Task-based UDEs may also include the kind of resetting feature, described earlier, which has been part of NAEP since 2019.

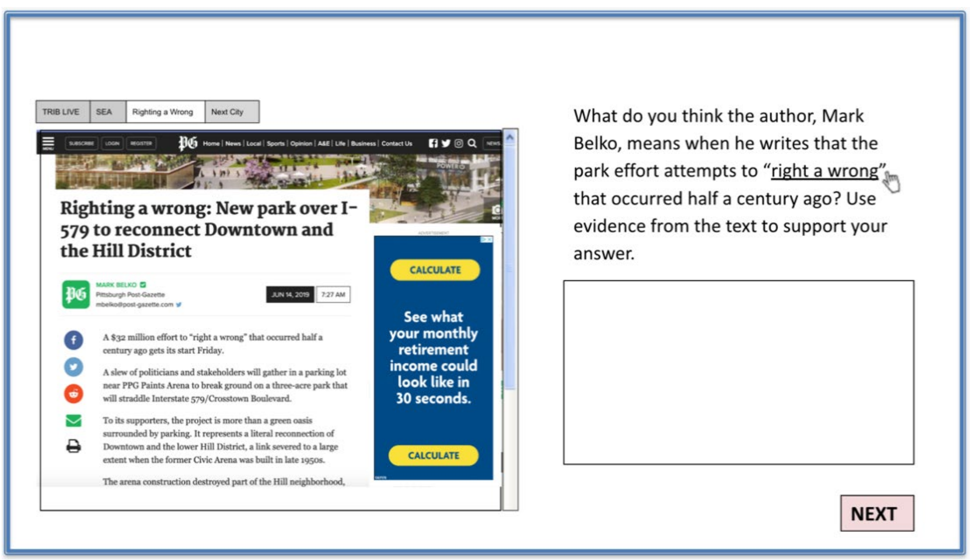
Exhibit 2.9. Example of a Task-based UDE from a Grade 12 RSP Block



The photograph of Kai is sourced from https://images.all4ed.org/high-school-boy-and-girl-near-playground (photographer Allison Shelley for EDUimages).

This task-based UDE (Exhibit 2.9) includes 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 2.10. Example of a Task-based UDE from a Grade 12 RSP Block



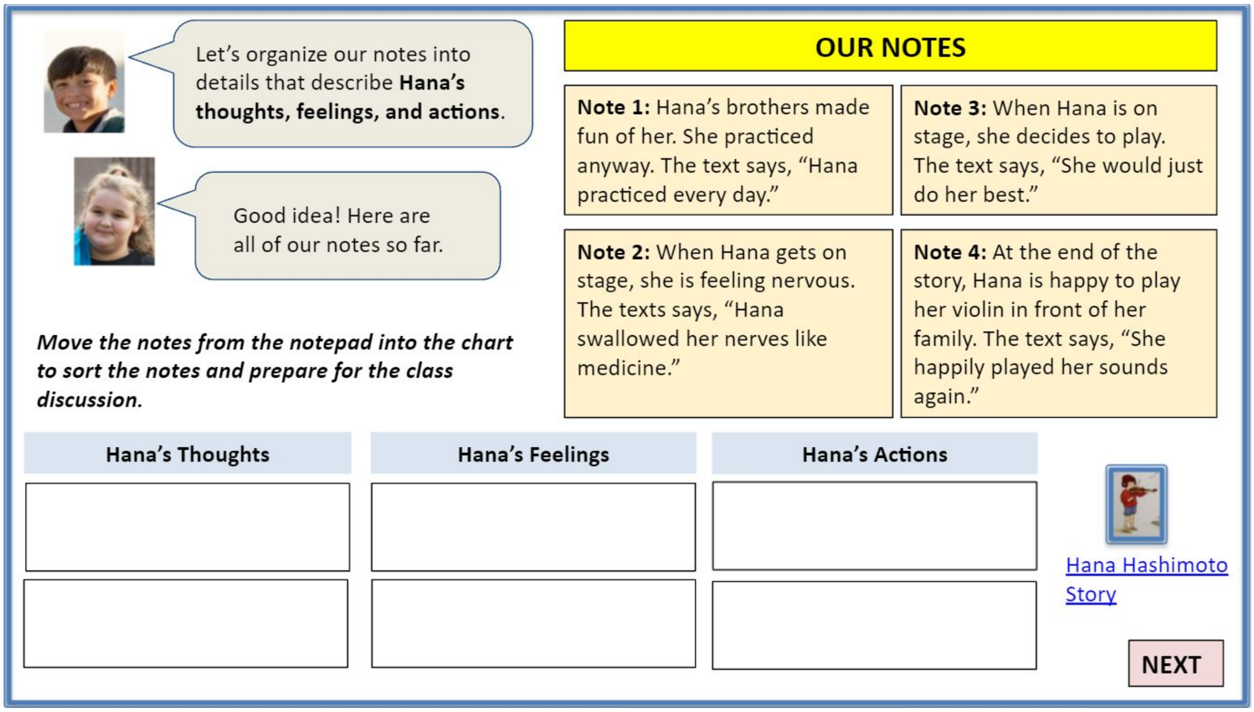
Incorporated into this short constructed-response item (Exhibit 2.10) is a task-based UDE as a look-back button that asks readers to integrate and interpret information in an online newspaper article about the historical significance of the park's design.

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 (Dalton & Proctor, 2008; Buehl, 2017; CAST, 2020; Guthrie & Klauda, 2016). 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).

Exhibit 2.11. Example of a Motivational UDE from a Grade 4 RDU Block



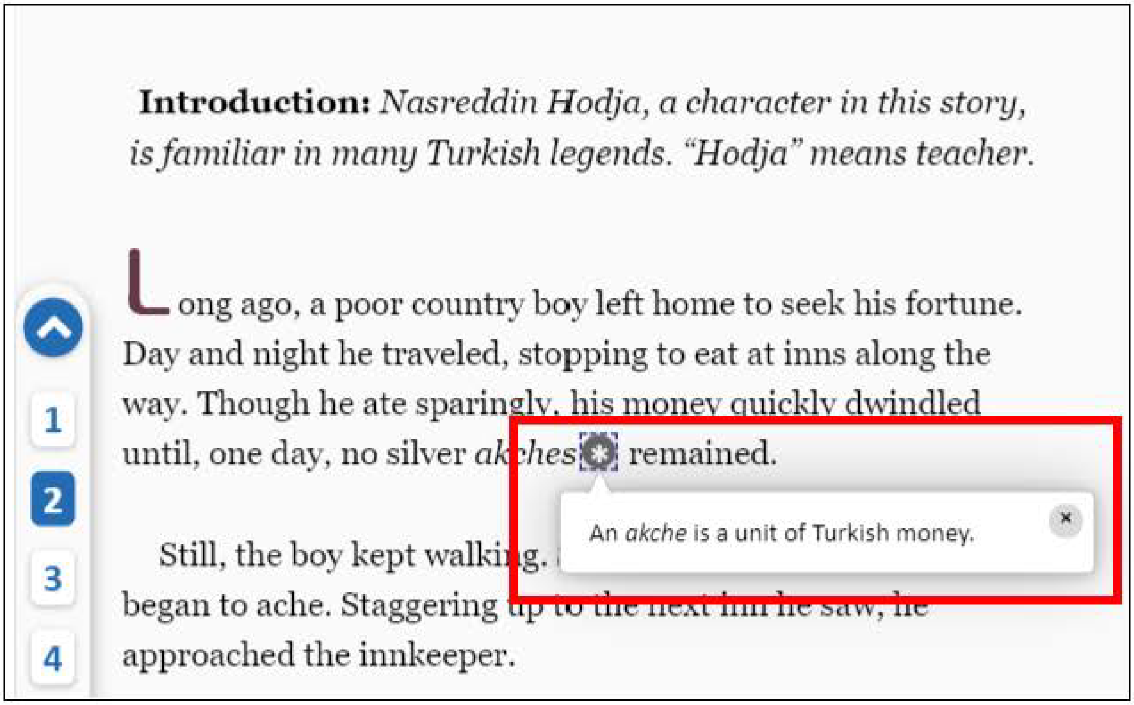
The photograph of Gabe is sourced from https://images.all4ed.org/third-grade-boy-with-backpack-outside/. The photograph of Luisa is sourced from https://images.all4ed.org/fifth-grade-girl-mask-break (photographer Allison Shelley for EDUimages).

In this example (Exhibit 2.11), 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" (as an additional task-based UDE) 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.

Informational UDEs

Informational UDEs are designed to maximize students' ability to engage with the content that is being tested by providing relevant context. Informational UDEs do not reduce the difficulty level of assessment items but rather they provide orientations to topics, concepts, or obscure vocabulary that students may need to make meaning from text as they read (Kintsch, 1998; McNamara, 2021; van den Broek & Helder, 2017). Informational UDEs consist of brief passage introductions (e.g., a short description of the author or text) to provide context about what the student is reading and vocabulary pop-ups to offer on-demand definitions of obscure words that are not part of the content being assessed. Unless video, image, or other kinds of introductions are already part of an authentic source text, topic previews may take the form of written texts only.

Exhibit 2.12. Example of an Informational UDE from a Grade 4 RDU Block



This example (Exhibit 2.12) from a NAEP Grade 4 block illustrates two informational UDEs. The first informational 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 informational UDE appears in the form of a vocabulary pop-up box defining the Turkish word "akche."

UDEs and the NAEP Definition of Reading Comprehension.

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 include lookback buttons and definitions of some words 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.

Summarizing the Relationship Between the Definition and Assessment Components

This chapter has described the reading content of the 2026 NAEP Reading Assessment and the connections between the content and the NAEP Definition of Reading Comprehension. Exhibit 2.13 summarizes these connections.

Exhibit 2.13. Relationships Between the NAEP Definition of Reading Comprehension Definition and the 2026 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	Address an array of skills and strategies related to comprehension,	Query different types of comprehension within and across texts and	Attend to disciplinary contexts, purposes, and text challenges

	and outside of the assessment.	including literal, inferential, analytical, and critical responses along with items that ask students to apply ideas in the texts.	different aspects of the texts, including local and global features and meanings.	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 interactive 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.
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 obscure words that are not the focus of the assessment items, and instructions on how to complete assessment tasks.	Increase broad access to texts, such as providing definitions of obscure words not measured on the assessment and offering lookback buttons.	Provide information that clarifies the nature and order of tasks and expected responses.

Chapter 3 describes the structure of the assessment and illustrates the use of key design principles and development practices that will support NAEP test developers to create an assessment that includes the components described.

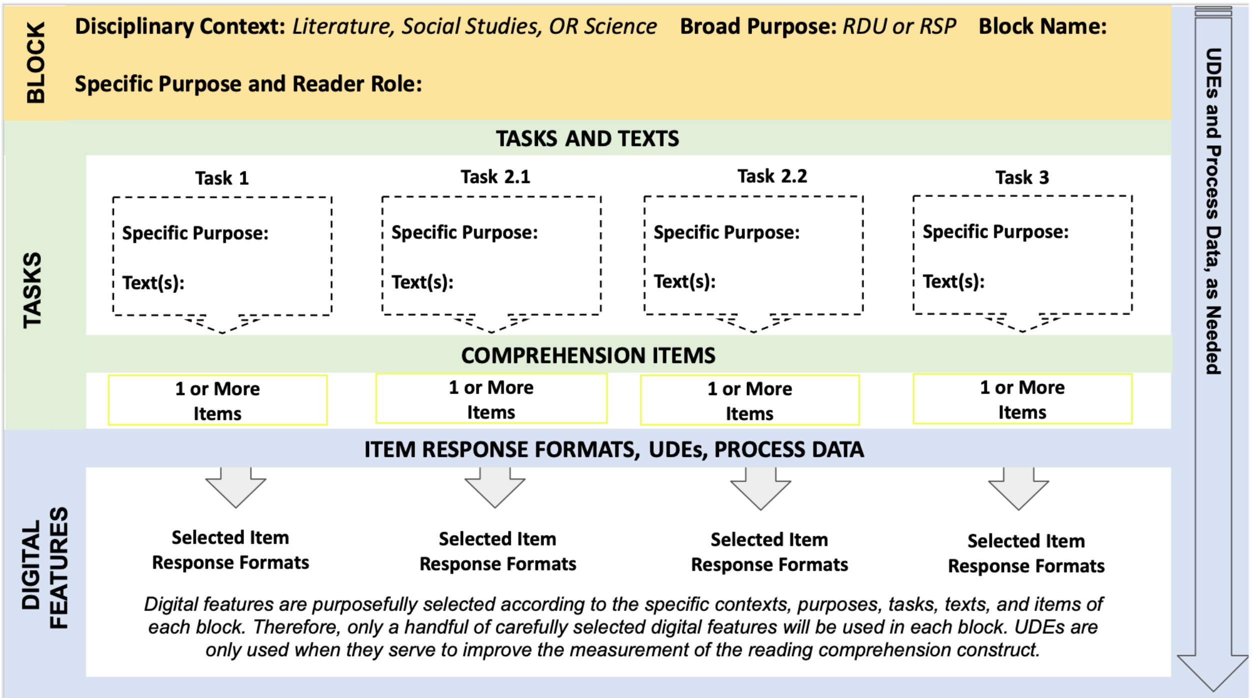
This chapter describes the assessment design components that contribute to best educational measurement practices, as outlined by the National Research Council (Pellegrino, et al., 2001; AERA/APA/NCME, 2014), and that were used in previous NAEP Reading assessments (National Assessment Governing Board, 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, including text and comprehension items, 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. The third section details considerations for leveraging digital assessment features, including item response formats, Universal Design Elements (UDEs), and process data in line with principles of validity, fairness, and inclusivity (AERA/APA/NCME, 2014).

Situating Readers Within Assessment Blocks

A block is the largest organizational unit for the 2026 NAEP Reading Assessment. In a typical NAEP Reading Assessment session, test-takers engage in two grade-appropriate blocks. The design of every block is intended to situate readers within a *disciplinary context*, a *broad purpose for reading*, and a *specific purpose* and *role* for the reader working through the block, as shown in Exhibit 3.1 This exhibit provides one sample approach to an assessment block; other approaches are possible that would have variations in the components (e.g., the number of tasks and texts). As developers develop a block, they make decisions about each of the components. In the following section, we describe some of the different design principles for consideration as decisions are made about the assessment components illustrated. Refer to Appendix C for additional considerations and guidelines for block development, along with grade-level block sketches with more examples.

Exhibit 3.1. Design Components of a 2026 NAEP Reading Assessment Block



Designating Disciplinary Context

All blocks will sample from a range of grade-appropriate texts within one of three disciplinary contexts—literature, science, or social studies. 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, with increasing emphasis on informational contexts as the grades progress. Exhibit 3.2 shows the design principle and provisional distribution targets for sampling disciplinary contexts at each grade level.

Exhibit 3.2. Principle and Provisional Distribution Targets for Sampling Disciplinary Contexts by Grade Level

<table>
<tr><td colspan="5">Principle for Sampling Disciplinary Contexts: The percentage of Literature contexts decreases across grades as the percentages of Science and Social Studies contexts increase.</td></tr>
<tr><td colspan="2">Grade Level</td><td>4</td><td>8</td><td>12</td></tr>
<tr><td rowspan="3">Disciplinary Context</td><td>Reading to Engage in Literature</td><td>50%</td><td>40%</td><td>33%</td></tr>
<tr><td>Reading to Engage in Science</td><td>25%</td><td>30%</td><td>33%</td></tr>
<tr><td>Reading to Engage in Social Studies</td><td>25%</td><td>30%</td><td>33%</td></tr>
</table>

Designating a Broad Reading Purpose

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). Therefore, 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.

As described in Chapter 2, 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. 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. RSP activities entail developing understanding, or comprehending text but are 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 3.3.

Exhibit 3.3. Principle and Provisional Distribution Targets for Sampling Broad Reading Purposes by Grade Level

<table>
<tr><td colspan="5">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.</td></tr>
<tr><td colspan="2">Grade Level</td><td>4</td><td>8</td><td>12</td></tr>
<tr><td rowspan="2">Broad Reading Purpose</td><td>RDU</td><td>60%</td><td>50%</td><td>40%</td></tr>
<tr><td>RSP</td><td>40%</td><td>50%</td><td>60%</td></tr>
</table>

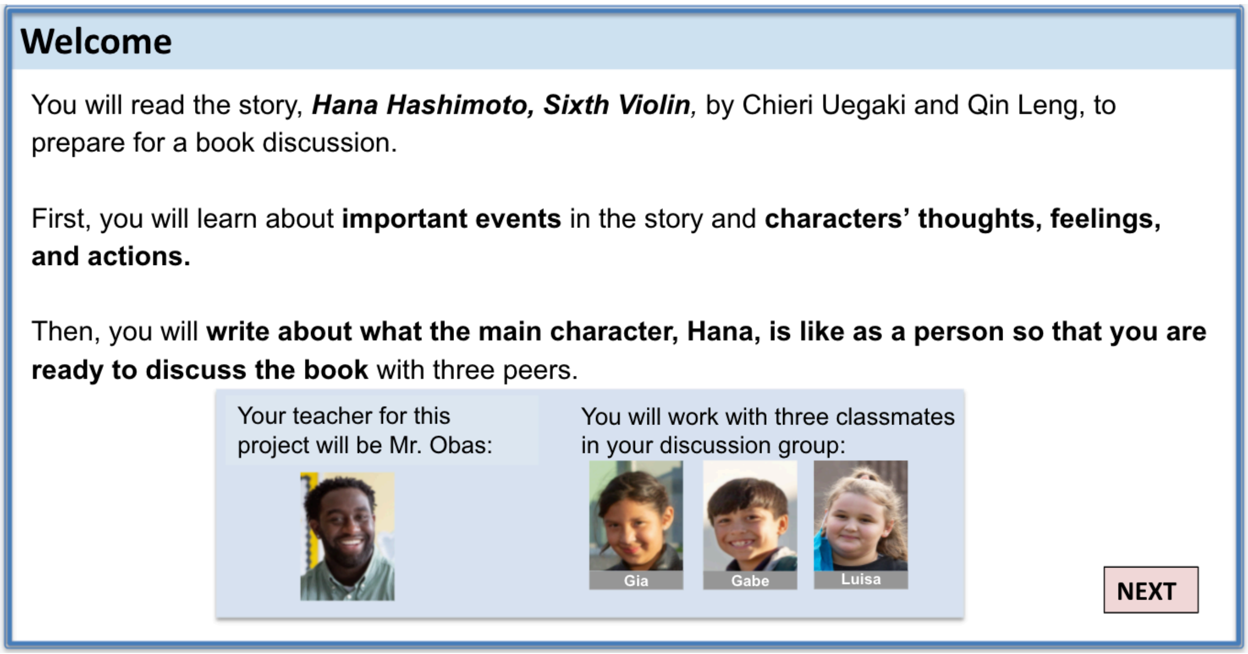
Identifying Block-Specific Purposes and a Reader Role

Apart from the identified broad purpose, each block also has its own specific purpose and reader role that shape how and why readers engage with the tasks, texts, and comprehension items in one of the three disciplinary contexts. These block-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 should 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.

A goal of the 2026 NAEP Reading Framework is to design an assessment that immerses readers in discipline-specific blocks for which both reading purpose and reader role are transparent. By making purpose and role clear to test-takers, the assessment better simulates the situations in which most readers find themselves in school, workplace, and community situations. Block-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 3.4 depicts an example of what readers might see when they begin a 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*[1] by Chieri Uegaki and Qin Leng (2014) 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.

Exhibit 3.4. Block-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



The photograph of Mr. Obas is sourced from https://images.all4ed.org/male-sixth-grade-math-teacher-with-protractor (photographer Allison Shelley for EDUimages). The photograph of Gia is sourced from https://images.all4ed.org/elementary-boy-with-backpack-and-girl-with-notebook/ (photographer Allison Shelley for EDUimages). The photograph of Gabe is sourced from https://images.all4ed.org/third-grade-boy-with-backpack-outside/. The photograph of Luisa is sourced from https://images.all4ed.org/fifth-grade-girl-mask-break (photographer Allison Shelley for EDUimages).

Developing Assessment Tasks: Texts and Items

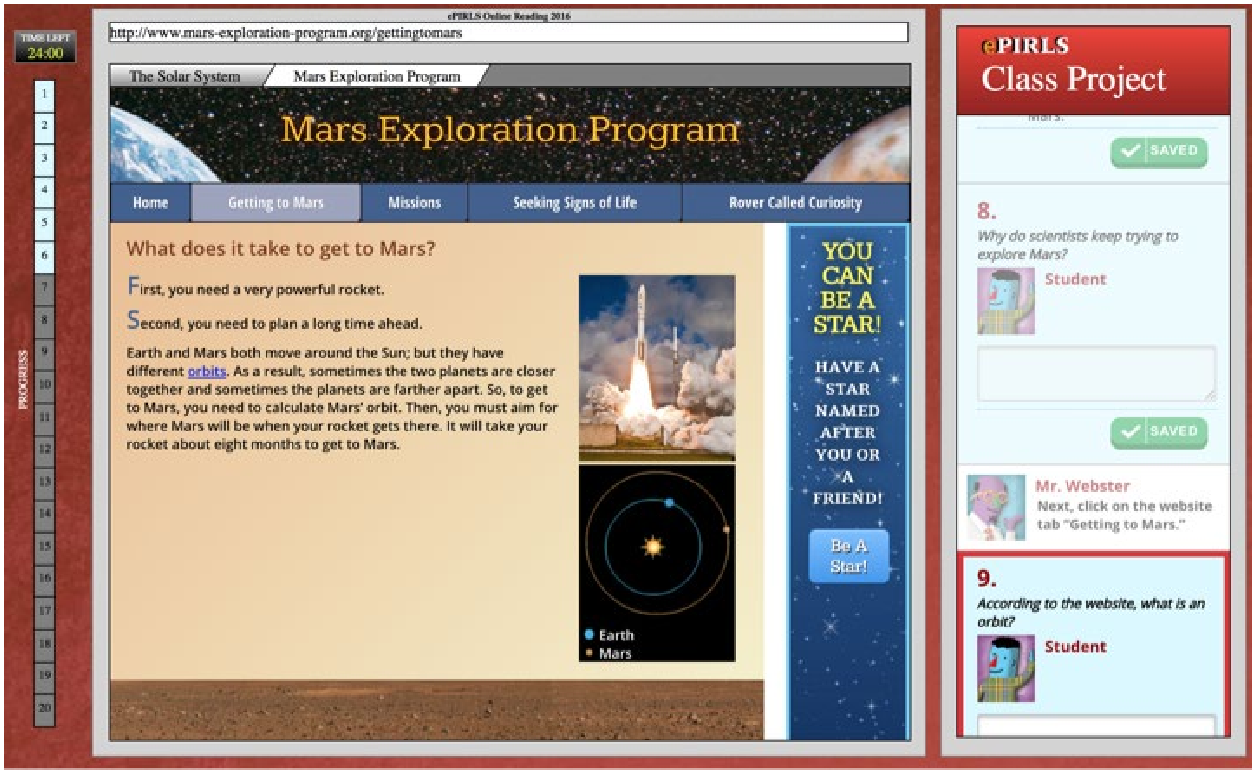
After readers are situated in an assessment block, they encounter two or more tasks, each with its own specific purpose. A task is a subunit within each block on the 2026 NAEP Reading Assessment. Each NAEP Reading Assessment block has two or three tasks, one or more texts, and related comprehension items. Developers should 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 social media post might result in more than one text for a particular task.

For example, Exhibit 3.5 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),

[1] Material from *Hana Hashimoto, Sixth Violin* written by Chieri Uegaki and illustrated by Qin Leng is used by permission of Kids Can Press Ltd., Toronto, Canada. Text © 2014 Chieri Uegaki. Illustrations © 2014 Qin Leng.

a photo of a rocket, a photo of the surface of Mars, and a dynamic image of two planets spinning around the sun. The item is intended to assess 4th graders' understanding of how to use embedded hyperlinks to locate and recall important information about the passage.

Exhibit 3.5. Example of multiple texts readers encounter as part of one task on the ePIRLS (2016) Grade 4 reading assessment



Besides consideration of grade appropriateness and the subject material in the creation of a task or multiple tasks in an assessment block, developers should pay close attention to and make careful decisions regarding the time demands placed on students to successfully accomplish the task in a given time limit, including the total passage length of all passages presented in the task (along with consideration of text complexity demands) and the total number of items, in association with their formats, that must be answered.

Selecting Texts

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. Exhibit 3.6 provides guidance to developers about sampling different kinds of texts (where texts include multimodal forms of representation).

Exhibit 3.6. 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 3.7 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 3.7. 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 are found in Exhibit 3.10.	*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

Text Selection Criteria. Passages selected for the 2026 NAEP Reading Assessment should adhere to rigorous criteria that include the following:

- ***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 nonfiction 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. An authentic text is defined as communication or composition produced by an author for publication purposes. Most texts included in the 2026 NAEP Reading Assessment should be presented in their entirety, as students would typically encounter them. However, some texts may be excerpted from, for example, a novel, a play, or a long essay. Excerpted material should be carefully analyzed, and minimally altered if necessary, to ensure that it is coherent in structure. Texts should be selected to evoke the range of reading comprehension processes, or targets. 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 of a certain genre that is suitable for a specific grade level for a RSP purpose. In these 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 presented items focused on evaluating the credibility or accuracy of such texts.

Exhibit 3.8 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 3.8. 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 regarding form and purpose • Vetted for accuracy, authenticity, and appropriateness by experts in the discipline, NCES's text selection panel, and the Assessment Development Committee • 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 the NAEP Reading Assessment

Developmental Appropriateness of Texts. Texts included in the assessment will be of different lengths. Exhibit 3.9 provides ranges for the total number of words in the text(s) within a given block. 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. This word count 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. Differing passage lengths are employed for several reasons, including the broad purpose of a block and 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).

Exhibit 3.9. 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

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 text in NAEP assessments built from the 2026 NAEP Reading Framework may continue to include video elements, consistent with previous NAEP Reading Assessments administered since 2017, 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 be kept relatively short, especially in consideration of the length of written texts within the task. Video length should typically remain in the range of one to three minutes, with some flexibility allowed to account for the density of information in the video and for the specific requirements of the task. The developer should obtain or create a transcript of a video to aid item development and ancillary materials development.

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. Exhibit 3.10 provides a list of the text types and elements that test developers should 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 words). While it is impossible in NAEP to represent the entire range, these elements define the portfolio of possibilities that developers should consult when 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 3.10. 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 • Coming of age stories • Satires • Science fiction • Magical realism • Fantasy • Comic books • Graphic novels • Manga • Fanfiction Poetry • Haiku, sonnet, ballad, dirge, epic, etc. Related Nonfiction • Memoirs	• Plot types • Character types • Narrative elements (character, setting, plot, conflict, rising action, climax, resolution) • Figurative language (symbolism, imagery, simile, metaphor, personification, satire) • Point of view • Theme • Soliloquy, dialogue, and monologue • Diction, word choice • Repetition, exaggeration • Flashback • Foreshadowing • Mood, tone, irony, paradox, and sarcasm • Visual and graphical elements such as illustrations and photographs

	• Biographies and autobiographies • Literary analyses • Reviews and recommendations • Author profiles	• Multimodal elements such as narrative soundscapes • Description • Narrative and expository text structures
Science	• Science reports • Press releases • Science news and features • Science magazine articles • Reference materials and field guides • Discovery narratives • 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 • Procedures • Published research articles • Personal communications	• 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) • Embedded 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 simulation, time lapse photography and animations
Social Studies	• Historical and contemporary documents such as newspaper articles, editorials, political cartoons, broadsides, blogs, census data, diaries, letters, speeches, inventories and records of sale, advertisements, archival documents • Biographies and autobiographies • Historical and contemporary photographs and video • Data (tables, charts, graphs, infographics) conveying information such as demographic, employment and	• 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 • Historical expressions and terminology • Ideological markers of language and rhetorical devices (word choices, emotional appeals, hyperbole)

	education levels, voter registration and turnout statistics, Gross Domestic Product and other economic measurements, etc. • Interpretive explanations or arguments about historical, social, and cultural phenomena and trends. • Procedural texts, public service announcements	• Visual and graphical elements such as maps, timelines, political cartoons, photographs • Multimodal elements such as digital stories • Event models (how historical events are described) • Spatial (place, location) and temporal indicators (era, time, sequence, and tense)

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, and complexity and "considerateness," including both qualitative and quantitative measures (e.g., conventional readability criteria, reader-text connections, language structures and vocabulary considerations) (Armbruster & Anderson, 1985). Selections should also reflect 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 should attend to variations in organization and cohesion in line with text structures and text features that are found across disciplinary contexts.

To gauge the grade-level appropriateness (i.e., in terms of the conceptual and linguistic challenge) of a text for development, the 2026 NAEP Reading Assessment will rely on a combination of quantitative, qualitative, and reader attributes. Quantitative approaches rely on an algorithm to create either a single score or a small set of scores to estimate the difficulty readers might have understanding a particular text. The most common single scores are a Lexile (Stenner, 1996) or grade level designations, such as the popular Flesch-Kincaid (Kincaid, et al., 1975). A Lexile provides a point on a scale running from "Beginning Reader" to 2000L for obscure scientific or legal documents that can be understood by only a handful of experts. Readability formulas like the Flesch-Kincaid usually convert their numerical scales to a grade level scale (from 1.0 to 20+, for example) to convey the idea of the typical student who would be able to understand a text that scaled at a particular grade level. Increasingly, readability systems provide both an overall score and a small set of scores (e.g., Graesser, et al., 2014; Sheehan, et al., 2014).

For the 2026 NAEP Reading Assessment, NAEP will investigate the validity and utility of various quantitative indicators, including several of the more recent, more complex, and nuanced measures (see Hiebert & Pearson, 2014; Nelson, Perfetti, Liben, & Liben, 2012) indicators, such as TextEvaluator (Sheehan, Kostin, Napolitano, & Flor, 2014) and the Coh-Metrix Text Easability Assessor (Graesser, McNamara, Cai, Conley, Li, & Pennebaker, 2014), to select one (or more) that best fits the needs of NAEP—and that complement the approaches that NAEP uses to examine the qualitative facets of text complexity.

Similarly, NAEP will expand the range of qualitative tools currently in use (NAGB, 2009)—to include even more careful examination of the language used to render key concepts and the relationships among them accessible to readers. This is particularly important in light of greater emphasis in the 2026 NAEP Reading Framework on discipline-specific texts, settings in which language exerts substantial influence on the accessibility of texts for the general population of students as well as for specific groups, such as English learners and students with disabilities. The general approach employed in applying qualitative analyses of complexity is to train analysts to use specific criteria to unearth linguistic (largely vocabulary, syntax, or discourse) features that serve either as barriers or bridges to comprehension. Barriers can include rare words, obscure syntax (e.g., negative conditional clauses), or complex rhetorical frames for large sections of text (e.g., a conflict-resolution scenario). Bridges, by contrast, might include a diagram, an internal definition of a rare word, an explicit clue word like "unless" to signal the relationship among ideas, or explicit naming of the parts of a conflict-resolution frame.

Passage mapping is routinely conducted as a part of a text selection process. Mapping procedures result in a graphic representation of a possible stimulus selection that clearly highlights the hierarchical structure and the interrelatedness of the components of the texts. Story mapping, for example, shows how the setting of a story is related to and contributes to the development of plot and theme. Concept mapping shows the structure of informational text, along with the concepts presented and the relational links among concepts. Organizing information hierarchically within a text allows identifying the various levels of information within a text so that items can target the most important aspects of what students read.

For the 2026 NAEP Reading Framework, these successful practices from the previous NAEP Reading Assessment development should be supplemented with more recent developments, particularly those deployed by PARCC and SBAC in developing their assessments (Hain & Piper, 2016). For example, the qualitative text complexity rubrics published by the State Collaborative on Assessment and Student Standards (SCASS) are useful tools to determine qualitative text complexity. There are two rubrics, one for literary texts and one for informational texts. Both rubrics incorporate four traits: Meaning; Structure; Language; and Knowledge Demands. Each trait has one to three criteria to determine if the qualitative text complexity falls into one of four text complexity levels: Low; Middle Low; Middle High; and High. Similar qualitative text complexity rubrics are employed by many state assessments.

Finally, NAEP will conduct analyses for what have been called reader-task considerations (NGA-CCSSO, 2010) or reader attributes or text-task scenarios (Valencia et al, 2014). All three of these approaches ask the question, "for whom, in what specific contexts, and with what levels of support are specific texts more or less accessible, i.e., harder or easier to comprehend?"

Exhibit 3.11 describes considerations regarding the distribution of selected texts, especially now that many of the texts within NAEP will bring digital affordances along with those of print texts. Ideas within each cell are likely to change and expand as new kinds of texts and technologies continue to emerge.

Exhibit 3.11. Text Structures and Features Within and Across Single Static and Dynamic Texts and Multilayered Digital Text Environments

SINGLE STATIC TEXT	SINGLE DYNAMIC TEXT
Text structures are comparable to those in a printed format for texts designed to inform, entertain and/or persuade. **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.	**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). **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.

MULTILAYERED DIGITAL TEXT ENVIRONMENT
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 that may contain either related or conflicting ideas. Multimodal elements (words, moving images, animations, color, music and sound) may appear in any or all texts.

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 should 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, Wixson & Pearson, 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 (e.g., 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., Texts also often appear and are used in sets. Thus, it is important to determine the grade-appropriate number of texts in a block, 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 multilayered digital text environments (Cho & Afflerbach, 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 videos or 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.

For an example of digital texts that are embedded in a webpage or other texts, please see Exhibit 3.5 above. Sometimes a digital text will be one of the stimulus passages in an assessment block. Fourth-grade students encountering a literary disciplinary context assessment block, for example, would first read a printed short fictional piece and answer some comprehension items. Then the students would watch a short animation that would elicit its own comprehension items, as well as comparisons of character(s), setting, plot, and theme with the printed text. As another example, in the social studies disciplinary context, 8th-grade students would be presented with a recent print article, accompanied with an embedded map, about how the stones assembled at Stonehenge may have been sourced and moved from their original location in England to the location at Stonehenge. After answering a few reading comprehension items about the article and its map, the students would then be presented with a stimulus digital text, watching a news segment about the same topic with similar and differing details. Students would answer a few more reading comprehension items about the digital text and then encounter more items that incorporate both texts to assess evaluation and synthesis skills. In the science disciplinary context, 12th-grade students may first be presented the stimulus material of a digital text of a National Oceanic and Atmospheric Administration (NOAA) video demonstrating how plastics enter the environment and end up in the oceans and create trash gyres, such as the Great Pacific Garbage Patch. After answering items regarding the digital text, students are then presented with a scientific journal excerpt focused upon the Great Pacific Garbage Patch and how it is affected by seasons and currents. Students then would answer items about the journal excerpt, followed by items assessing their skills of analysis and synthesis related to both texts.

Engaging experts in selecting texts that reflect authentic social and cultural traditions in a range of disciplinary contexts, without placing students at a disadvantage based on their particular social and cultural context.. 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. 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. In accordance with the Board's legislative mandate to "ensure that all items selected for use in the National Assessment are free from racial, cultural, gender, or regional bias," experts should represent diverse cultures and languages in order to identify texts that reflect the broad range of student readers' knowledge and experiences. The passages that are selected should themselves be drawn from texts that reflect a diverse range of cultures, regions, and experiences.

Bias and Sensitivity Considerations. Along with the consideration of disciplinary appropriateness of texts and of the standards for coherence and complexity of texts within and across disciplines, assessment developers, in their selection of texts, must analyze the texts for any bias and sensitivity issues (e.g., topics to avoid) that could negatively affect a student's testing experience.

Topics to Avoid. In addition to certain authors, publications, and publishers, there are a number of subjects and contexts that, while suitable for classroom use, would be considered inappropriate for assessment purposes. A story about a child dealing with death may be read as a classroom assignment; however, the teacher in the classroom has a chance to prepare students before they read the selection, and students have the opportunity to talk through their reactions. No such opportunities are available in a testing situation. In general, a topic might be unacceptable for any of the following reasons:

1. The text evokes an emotional response that might affect test performance. Examples include texts that are frightening or very humorous.
2. The topic is too controversial, such as abortion, gun control, and evolution.
3. The topic has been used extensively in standardized tests or textbooks, making it overly familiar or boring to students.
4. The topic could be biased against a particular demographic, for example, due to socioeconomic level.

Some of the bias and sensitivity issues to avoid include:

- gender bias (i.e., neither gender would have an advantage due to prior knowledge or interest)
- inappropriate content, including age, ethnic, and cultural bias
- stereotyping
- derogatory statements
- violence, sex, or objectionable language
- expressions of religious belief
- highly emotional themes (death, divorce, abuse, terrorism, etc.)
- emotionally charged historical events

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 the following:

- ***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. Exhibit 3.12 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.

Exhibit 3.12. Distribution of Cognitive Comprehension Targets Across Grade Level and Broad Purposes

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 Integrate and Interpret, 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. Thus, 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: Reading to Develop Understanding and Reading to Solve a Problem Blocks (% Target Ranges per Block)**
Grade 4	
Locate and Recall	15–40%
Integrate and Interpret	10–40%
Analyze and Evaluate	10–25%
Use and Apply	0–30%
Grade 8	
Locate and Recall	10–25%

Integrate and Interpret	20–35%
Analyze and Evaluate	20–35%
Use and Apply	0–30%
Grade 12	
Locate and Recall	10–25%
Integrate and Interpret	25–35%
Analyze and Evaluate	25–40%
Use and Apply	0–45%

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 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 explanations regarding how their responses will be evaluated.
- ***Alignment with an array of skills of navigation and inference.*** 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). Items may require readers to draw on information contained in audio or visual features.
- ***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 from 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 knowledge, without even reading the passage.

Planning the Distribution and Characteristics of Comprehension Items. The four Comprehension Targets do not represent a hierarchy of strategies or skills; rather, 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-S contains examples of what test items might ask readers to do with respect to each of the four Comprehension Targets. Appendix A-S: Achievement Level Descriptions provides for each grade some of the possible disciplinary context-specific skills that are associated with the Comprehension Targets and which may appear in an administered NAEP assessment block. Items must be developed to address the range of Comprehension Targets with the expectation that there will be a distribution of Comprehension Targets at each achievement level.

The guidelines for distributing items mapped to Comprehension Targets across grade levels and blocks presented in Exhibit 3.12 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 item 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 (i.e., the *type* of inference [a local, straightforward inference within a paragraph versus 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]). These factors introduce variations in task and item demands that impact the difficulty of a particular comprehension item on the reading assessment.

Language Structures and Vocabulary in the Comprehension Items. The phrase "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.

Exhibit 3.13 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. 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 obscure words of limited application across grade-appropriate texts, and idiomatic expressions characteristic of particular cultural and idiosyncratic discourse practices.

Exhibit 3.13. Inclusion and Exclusion Criteria for Connected Language and Vocabulary

Language Structures & Vocabulary Included / Excluded 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 ts (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	• Obscure words of limited application across grade-appropriate texts and discipline-specific concepts (e.g., fiduciary, apotheosis) • Idiomatic expressions (e.g., spill the beans, up in the air)

A total 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 relationship 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 and 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 development for the 2026 NAEP Reading Assessment should include scoring rubrics and trainings 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 by having all test components 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).

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). 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, UDEs, and process data, as described in each of the next three sections. See Exhibit 3.14 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 3.14. 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 (see, for example, Lee, 2020; Solano-Flores & Nelson-Barber, 2001). 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. • Informational UDEs preview untested topic knowledge and provide definitions for obscure vocabulary not intended to be assessed. This maximizes the extent to which the assessment can measure the same, intended construct 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 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 experience, knowledge, language use, or 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 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. 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. 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. Selected-response items have a variety of formats, some of which allow for more than one correct response. The listed formats reflect a subset of those with the potential to be developed for the 2026 NAEP Reading Assessment.

- **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).

- **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.
- **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.

The table in Exhibit 3.15 lists and describes selected response item formats, indicates other names by which an item format might be known, and provides the location of exhibits within the *Assessment and Item Specifications* of examples. At the beginning of the table are guidelines to assist with the development of selected response items.

Exhibit 3.15. Selected Response Item Information

Selected Response (SR) Development Guidelines
In a well-designed selected-response item, the stem clearly presents the question to the student. The stem may be in the form of a question, a phrase, or an expression, as long as it conveys what is expected of the student. Selected response items should have the following characteristics:

- The stem includes only the information needed to make the student's task clear.
- Options are as short as possible and are parallel in length.
- Options are parallel in structure, syntax, and complexity.
- Options do not contain inadvertent cues to the correct answer, such as repeating a word from the stem in the correct answer or using specific determiners (e.g., all, never) in the distractors (incorrect options).
- Distractors are plausible, but not so plausible as to be possible correct answers.
- Distractors are designed to reflect the measurement intent of the item, not to trick students into choices that are not central to the idea being assessed.

NAEP Item Formats	**Similar Item Formats/ Abbreviations**	**Student Interaction**	**Location(s) of Example Item(s)**
single-selection multiple choice (SSMC)	multiple choice (MC)	Student selects one of four given response options.	Exhibit 3.16 Exhibit 3.17
multiple-selection multiple choice (MSMC)	multiple select (MS)	Student selects two of five given response options.	Exhibit 3.18
matching	drag and drop; gap match	Student inserts one or more source elements (e.g., graphics) into target fields (e.g., cells of a table).	Exhibit 3.19
zone	hot spot (HS)	Student responds by selecting one or more regions on a graphic stimulus.	Exhibit 3.20 Exhibit 3.21
grid	matching table (MT; MTG)	Student evaluates reading analyses (e.g., central ideas shared or not shared between two texts) with respect to certain criteria. The response is entered by selecting cells in a table in which	Exhibit 3.22

		rows typically correspond to the statements and columns to the properties checked.	
In-line choice (IC)	in-line dropdown	Student responds by selecting one option from one or more drop-down menus that appear in various sections of an item.	Exhibit 3.23
select in passage	hot text (HT); text highlight	Student selects (or selects and drags to a target field) one or more highlighted pieces of text (options) in the reading passage.	Exhibit 3.24

Single-Selection Multiple Choice. Multiple choice items are an efficient way to assess knowledge and skills, and they can be developed to measure various levels of rigor. In a well-designed multiple-choice item, the stem clearly presents the problem to the student. The stem may be in the form of a question (i.e., closed stem) or an incomplete sentence (i.e., open stem) where the options each complete the sentence, as long as the stem conveys what is expected of the student. The stem is followed by four answer choices, or options, only one of which is correct. A generic scoring rubric can be used for all single-selection multiple choice items, as they are scored dichotomously:

- 1 = Correct: This response represents the one correct option.
- 0 = Incorrect: These responses represent one of the three incorrect options.

The item in Exhibit 3.16 illustrates a straightforward stem with a direct question. The distractors are plausible, but only one response option is correct.

Exhibit 3.16. Selected Response Example: **Single-Selection Multiple Choice Item from NAEP Grade 4 Literature**

In this item, students are given options of how a main character in the story became successful.

What did the boy do to become successful?

A ○ He raised hens from the eggs the innkeeper gave him.

B ○ He became a sea merchant and traveled to many places.

C ○ He learned from the innkeeper how to make his fortune.

D ○ He borrowed money to buy a new sailing ship.

This item appeared in the 2017 grade 4 NAEP Reading administration with NAEP Item ID 2017-4R5 #1.

Exhibit 3.17. Selected Response Example: Single-Selection Multiple Choice Item from Smarter Balanced Grade 12 Science

In this Smarter Balanced Grade 12 item associated with a Science context, students are asked to determine an author's point of view based upon the author's inclusion of conflicting information in the text.

What does the conflicting information about the effects of oil on blue crab larvae reveal about the author's point of view?

Ⓐ It reinforces the author's belief that scientists do not yet know how the oil will affect the blue crab population.

Ⓑ It suggests that the author disagrees with scientists who predict long-term damage to the blue crab population.

Ⓒ It reinforces the author's feeling that scientists may never know the true effects of oil on the blue crab population.

Ⓓ It suggests that the author feels scientists have not devoted enough attention to the effects of oil on blue crab larvae.

This item appeared in the 2019–20 Grades 11–12 Smarter Balanced Sample Items published by The Regents of the University of California with Item ID 183143.

Multiple-Selection Multiple Choice. As with single-selection multiple choice items, the stem of a well-designed multiple-selection multiple choice item clearly presents the problem to the student. The stem may be in the form of a question (i.e., closed stem) or an incomplete sentence (i.e., open stem) where the options each complete the sentence, as long as the stem conveys what is expected of the student. To avoid confusion for students, it is common in assessment development that the stem in multiple-selection items is followed by five response options with two correct response options (when single-selection items on the same assessment have four options with exactly one option correct). Directions for this item format should indicate the number of correct responses that students should select. Due to the selection of multiple responses, items allow for partial credit. A generic scoring rubric can be used for all multiple-selection multiple choice items:

- 2 = Correct: This response represents the two correct selections and no incorrect selections.
- 1 = Partial: These responses represent one correct selection and one incorrect selection.
- 0 = Incorrect: These responses represent no correct selections.

Correctly responding to items using the multiple-selection format is more challenging than single-selection multiple choice items, as students must determine not only the relationship between a response and the item stem, but also the relationships among the response options (Baghaei & Dourakhshan, 2016). The item in Exhibit 3.18 specifically asks students to select two correct response options that are textual evidence supporting an inference stated in the stem. Using a multiple-selection multiple choice item format allows for the assessment of student recognition that multiple pieces of evidence support key ideas in a text.

Exhibit 3.18. Selected Response Example: Multiple-Selection Multiple Choice Item from Smarter Balanced Grade 12 Science

In this Grade 12 Smarter Balanced item associated with a Science context, students are asked to determine which two pieces of textual evidence support an inference provided in the stem.

Select the **two** sentences from the text that **best** support the inference that blue crabs may be less impacted by the oil spill than some scientists predict.

☐ Tiny creatures might take in such low amounts of oil that they could survive, Thomas said.

☐ "In my 42 years of studying crabs I've never seen this," Perry said.

☐ She told the magazine there are two encouraging signs for the wild larvae—they are alive when collected and may lose oil droplets when they molt.

☐ "Crabs are very abundant. I don't think we're looking at extinction or anything close to it," said Taylor, one of the researchers who discovered the orange spots.

☐ Still, crabs and other estuary-dependent species such as shrimp and red snapper could feel the effects of remnants of the spill for years, Perry said.

This item appeared in the 2019–20 Grades 11–12 Smarter Balanced Sample Items published by The Regents of the University of California with Item ID 183102.

Matching. Matching items take many forms, but each involves the dragging and dropping of one or more objects. For example, a matching item may require the dragging of text or graphics into indicated spaces; the ordering of presented text (e.g., in an item assessing summary of a text); or the matching of a subset of objects from one set of information to objects in another set.

Matching items can quickly become quite complicated, based on the number of dragging and dropping actions required. In addition to accessibility concerns, item writers should consider the number of actions in light of the measurement intent of the item—that is, how much information students need to provide to demonstrate evidence of understanding of the assessed objective. Additionally, when possible, the development of more objects to drag than locations in which to drop them tends to allow students to make an error in one placement without impacting the other placements.

This selected response item format allows for dichotomous or partial credit scoring, dependent upon the item construct. Directions for this item format should indicate either the number of correct responses or that students should select all of the correct responses. Due to the selection of multiple responses, some items allow for partial credit. For these items, scoring guides are developed to indicate how the partial credit is allocated.

The item in Exhibit 3.19 asks students to drag one piece of evidence from each of two Literature text sources. The student must evaluate which piece of textual evidence from each reading passage supports a provided shared theme of the two texts.

Exhibit 3.19. Selected Response Example: Matching Item from PARCC Grade 8 Literature

The table shows a shared theme of the passage from *The Black Pearl* and the poem "The Last Bargain."

Complete the table with **one** piece of evidence from **each** text that **best** supports the shared theme. Drag and drop the pieces of evidence that **best** support the shared theme into the appropriate rows of the table. Not all pieces of evidence will be used.

- *The Black Pearl*: "'They often die or become dull before a year passes.'" (paragraph 7)
- *The Black Pearl*: "'And the price, gentlemen, remains twenty thousand pesos.'" (paragraph 22)
- "The Last Bargain": "But his power counted for nought . . ." (line 4)
- "The Last Bargain": "Her smile paled and melted into tears . . ." (line 12)

Shared Theme
It is important to know what is truly valuable.
Evidence from *The Black Pearl*
Evidence from "The Last Bargain"

This item appeared in the 2019 Grade 8 Released Items published by Partnership for Assessment of Readiness for College and Careers (PARCC) with Item ID FF429345509.

Zone. Zone items involve the selection of a graphic or graphics or the selection of a location or locations on a graphic. As with matching items, writers should consider the number and type of student actions required in light of accessibility and the measurement intent of the item. When developing an item that requires the selection of graphics, consideration should be given to the number of graphics presented and the number of correct graphics. When developing an item that requires the selection of a location or locations on a graphic, consideration should be given to the size and clarity of the graphic, the number of locations that are selectable, and the number of correct locations. For zone items, the selectable locations should be purposeful and clearly defined.

This selected response item format allows for dichotomous or partial credit scoring, dependent upon the item construct. Directions for this item format should indicate either the number of correct responses or that students should select all of the correct responses. Due to the selection of multiple responses, some items allow for partial credit. For these items, scoring guides are developed to indicate how the partial credit is allocated.

Exhibit 3.20. Selected Response Example: Zone Item from ePIRLS' Grade 4 Social Studies

This item 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.

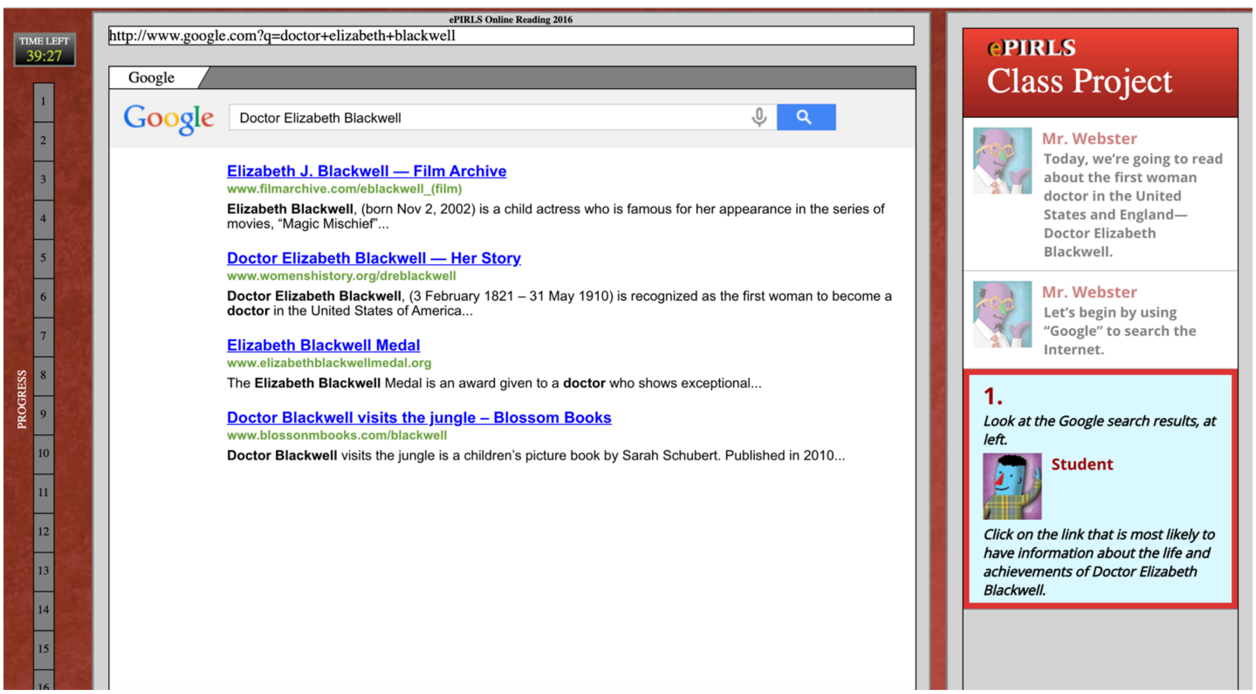
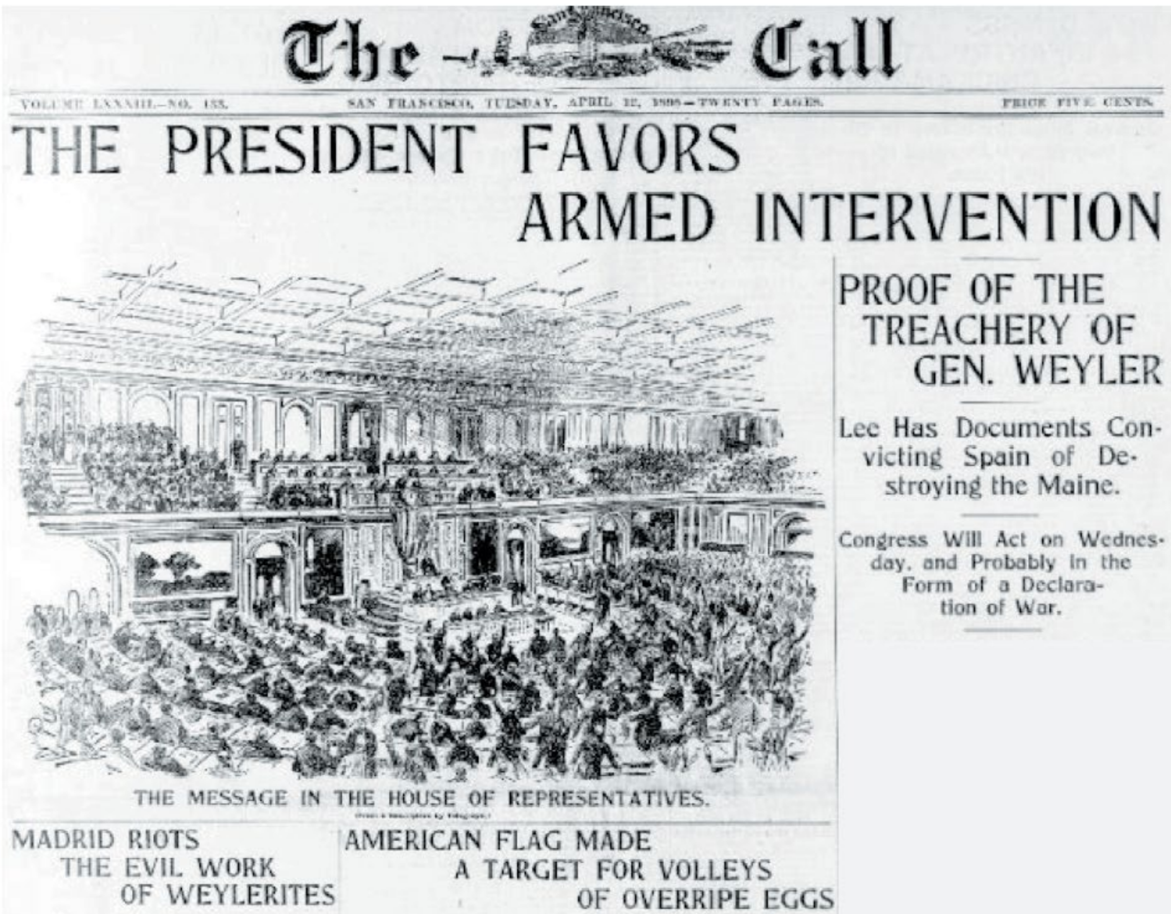


Exhibit 3.21. Selected Response Example: Zone Item for Grade 8 Social Studies

This item is a Grade 8 zone format item associated with the Social Studies context that asks students to use knowledge acquired from the task-based reading to discern three examples of yellow journalism in the components of newspaper headlines.

Read the newspaper headlines from *The Call*. Using what you have learned about yellow journalism, highlight **three** examples of sensational words or phrases in the headlines.



The Call

VOLUME LXXXIII—NO. 133. SAN FRANCISCO, TUESDAY, APRIL 12, 1898—TWENTY PAGES. PRICE FIVE CENTS.

THE PRESIDENT FAVORS ARMED INTERVENTION

PROOF OF THE TREACHERY OF GEN. WEYLER

Lee Has Documents Convicting Spain of Destroying the Maine.

Congress Will Act on Wednesday, and Probably in the Form of a Declaration of War.

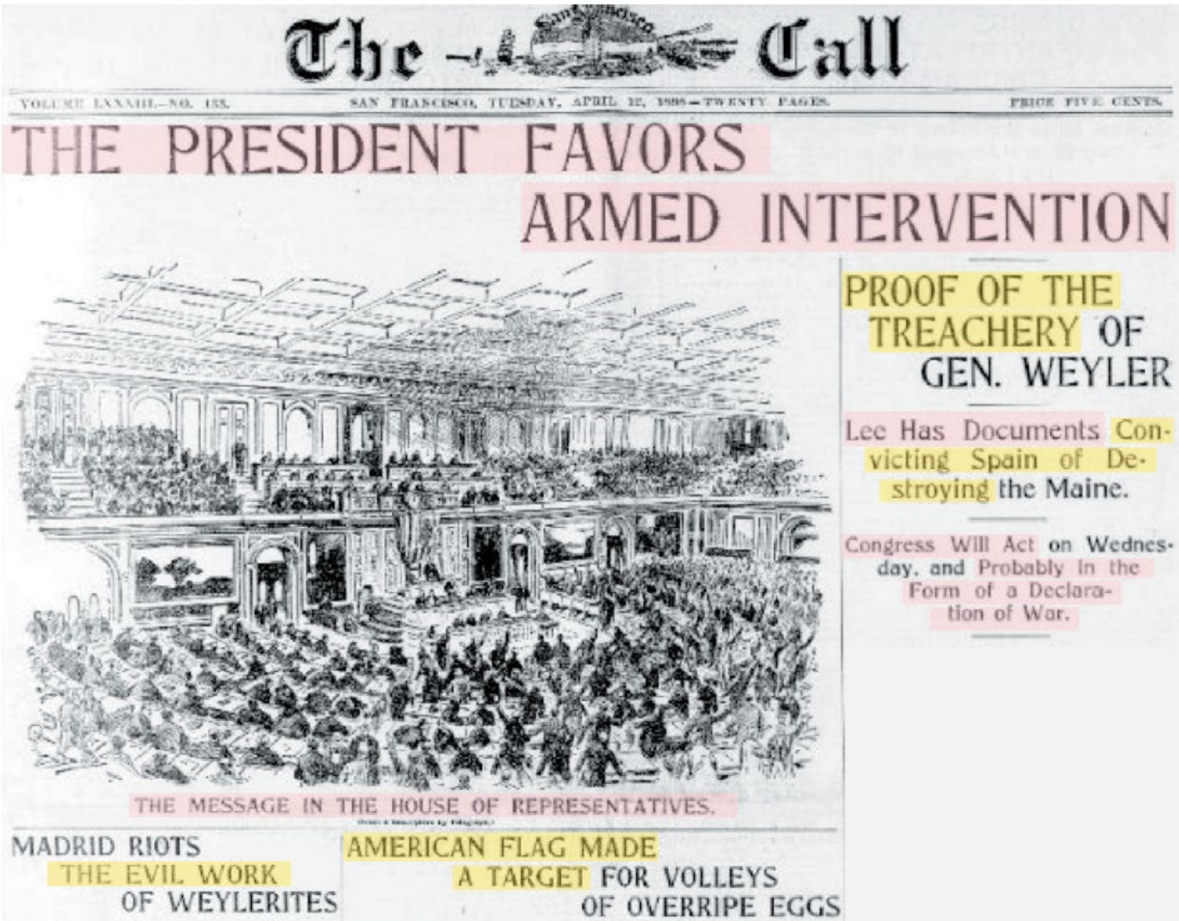
THE MESSAGE IN THE HOUSE OF REPRESENTATIVES.

MADRID RIOTS THE EVIL WORK OF WEYLERITES

AMERICAN FLAG MADE A TARGET FOR VOLLEYS OF OVERRIPE EGGS

Scoring Information

The highlighted portions of text will be selectable. Correct answers are highlighted in yellow; distractors are shown in pink. Three correct answers are required for full credit.



The Call

VOLUME LXXXIII.—NO. 133. SAN FRANCISCO, TUESDAY, APRIL 12, 1898—TWENTY PAGES. PRICE FIVE CENTS.

THE PRESIDENT FAVORS ARMED INTERVENTION

PROOF OF THE TREACHERY OF GEN. WEYLER

Lee Has Documents Convicting Spain of Destroying the Maine.

Congress Will Act on Wednesday, and Probably in the Form of a Declaration of War.

THE MESSAGE IN THE HOUSE OF REPRESENTATIVES.

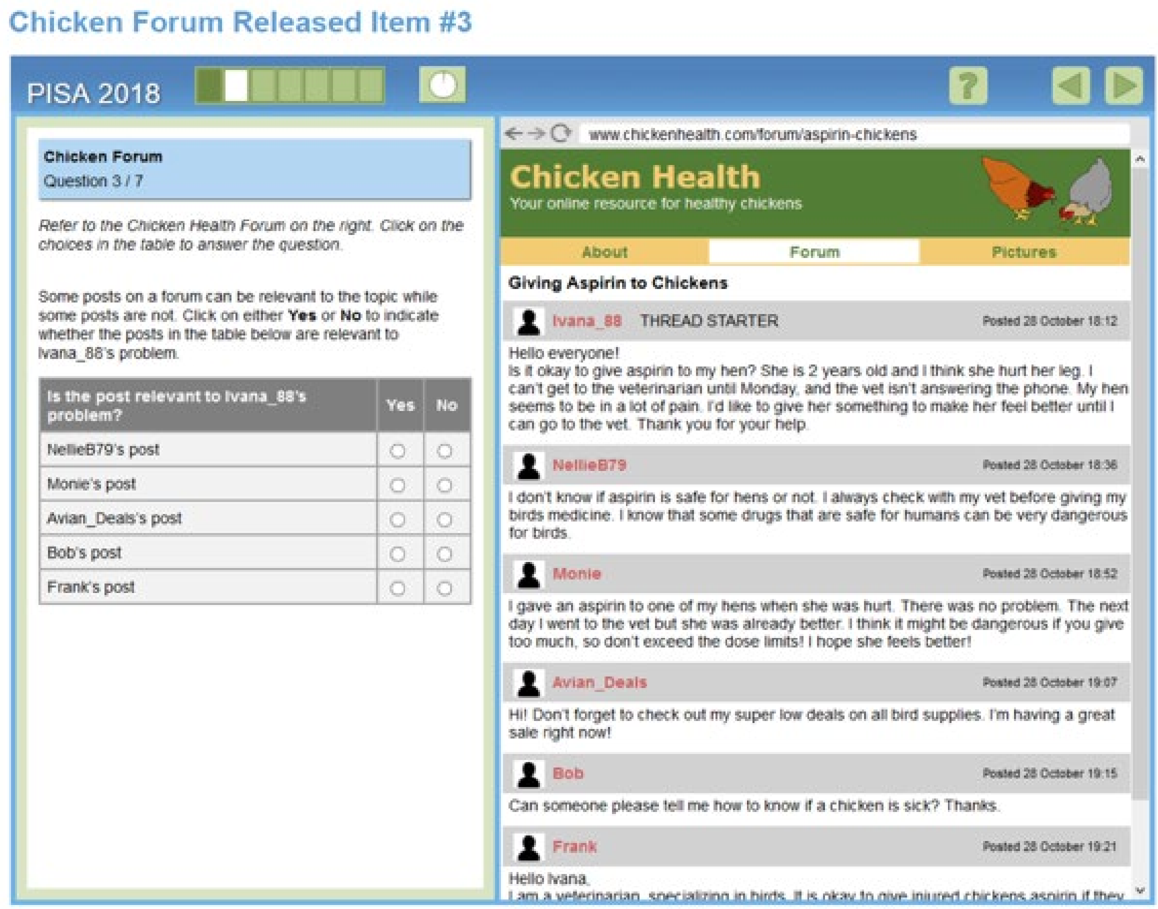
MADRID RIOTS THE EVIL WORK OF WEYLERITES

AMERICAN FLAG MADE A TARGET FOR VOLLEYS OF OVERRIPE EGGS

Grid. Grid items involve the selection of cells in a table to indicate a response. The rows of the table contain stimuli to be considered. The stimuli should be related. The first cell in each column of the table lists the options from which students choose. The options should be plausible for each stimulus. As with previously discussed item formats, writers should consider the number and type of student actions required in light of accessibility and the measurement intent of the item – that is, how much information students need to provide to demonstrate evidence of understanding of the assessed objective. This should inform the number of rows and columns included in an item.

This selected response item format allows for dichotomous or partial credit scoring, dependent upon the item construct. Directions for this item format should indicate either the number of correct responses or that students should select all of the correct responses. Due to the selection of multiple responses, some items allow for partial credit. For these items, scoring guides are developed to indicate how the partial credit is allocated.

Exhibit 3.22. Selected Response Example: Grid Item from PISA



In-line Choice. In-line choice items require students to select text that correctly completes a statement. Typically, the item stem presents information relevant to the completion of one or more statements. The statements are written beneath the stem, with drop-down menus that present plausible options for sentence completion. Item writers should take care when determining the number of options for each drop-down menu, as the total number of response options has the potential to impact the amount of reasoning required for students to complete the item. Additionally, in terms of accessibility, a student taking the test with a screen reader must listen to every potential answer, so the number of options in each drop-down menu impacts the number of combinations that the student must hear and manage.

This selected response item format allows for dichotomous or partial credit scoring, dependent upon the item construct. Directions for this item format should indicate either the number of correct responses or that students should select all of the correct responses. Due to the selection of multiple responses, some items allow for partial credit. For these items, scoring guides are developed to indicate how the partial credit is allocated.

Exhibit 3.23. Selected Response Example: In-line Choice Item from ePIRLS' Grade 4 Mars Block

This item from ePIRLS' assessment for grade 4 asks students to use the digital diagram to answer questions by selecting responses from a drop-down menu.

Select in passage. This item type, also commonly referred to as hot text or text highlight, requires students to select one or more pieces of text that have been highlighted as options in sections (in one or more paragraphs) of a reading passage or excerpted and presented in the item itself. Typically, the item stem presents a statement (e.g., a central idea) that requests the student to select the appropriate supporting textual evidence from the text. The item can also request the student to identify which piece(s) of text identify or support a literary element, such as character traits, or help the reader derive meaning of vocabulary or figurative language via context. Options should be near each other and should not force the student to scroll across many paragraphs of a passage.

This selected response item format allows for dichotomous or partial credit scoring, dependent upon the item construct. Directions for this item format should indicate either the number of correct responses or that students should select all of the correct responses. Due to the selection of multiple responses, some items allow for partial credit. For these items, scoring guides are developed to indicate how the partial credit is allocated.

Exhibit 3.24. Selected Response Example: Select in Passage Item from Smarter Balanced Grade 4 Literature

In this Smarter Balanced Grade 4 Literature example of a select in passage (hot text) item, the student is asked how the word "disappointed" used elsewhere in the text to describe the narrator's reaction to an event is supported by surrounding textual evidence. This item positions excerpted text as a functional hot text element in the item itself rather than using fuller sections of the text (or the full text) in a side pane; this strategy eases the complexity of the task for younger students. The key is highlighted in this example, but other phrases in the excerpted text, operating as distractors, are selectable (i.e., "a cloudy Saturday"; "skip the Farmer's Market"; "what I wanted").

The author uses the word disappointed in the passage. Click on the group of words in the sentence that **best** shows that idea.

It was a cloudy Saturday, and I thought we would be able to skip the Farmer's Market. I had hoped to do what I wanted today, but no such luck.

This item appeared in the 2018–19 Grade 4 Smarter Balanced Sample Items published by The Regents of the University of California with Item ID 182898.

Constructed Response Options. Constructed response items for the NAEP Reading Assessment also include a variety of formats. 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 word, a phrase, or a sentence or two. The fill-in-the-blank (FIB) item type is also considered a short constructed response format.
- **extended constructed response** – Students respond by entering an extended text in a response box that consists of multiple sentences (typically one or more paragraphs).
- **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.

The table in Exhibit 3.25 describes constructed response item formats, indicates abbreviations by which an item format might be known, and provides the locations of exhibits within this document of examples. At the beginning of the table are guidelines to assist with the development of constructed response items.

Exhibit 3.25. Constructed Response Item Information

<table>
<tr><td colspan="4">Constructed Response (CR)
Best used when student communication of the correct response and/or support for a response provides greater evidence than use of selected response item types. Constructed response items should have the following characteristics:
• The stem or prompt clearly and concisely sets up the task and conveys what is expected of the student.
• As part of describing the task, the prompt may indicate the text(s) or source(s) that the student should use to produce the response, whether the student must include supporting details or evidence from the text(s), and if applicable the genre of the writing to be produced.
• The task should be feasible for the average student in the amount of time allotted for the item type in the assessment block.</td></tr>
<tr><td>NAEP Item Formats</td><td>Abbreviations</td><td>Description</td><td>Location(s) of Example Item(s)</td></tr>
<tr><td>short
constructed response</td><td>SCR</td><td>Student provides a written response as a word, phrase, sentence, or brief explanation to a question or a prompt. Fill-in-the-blank items are also considered SCR items.</td><td>Exhibit 3.26
Exhibit 3.27
Exhibit 3.28</td></tr>
<tr><td>extended constructed response</td><td>ECR</td><td>Student provides a written response that typically is at least one paragraph, but may be a multi-paragraph essay, to a prompt that demands a higher degree of analysis and/or application of knowledge.</td><td>Exhibit 3.29
Exhibit 3.30</td></tr>
<tr><td>hybrid constructed response</td><td>HCR</td><td>Student selects one or more choices in one part of an item and then provides a brief explanation of their choice(s).</td><td>Exhibit 3.31</td></tr>
</table>

Every constructed response item has a scoring guide that defines the criteria used to evaluate students' responses. Some short constructed response items can be scored according to guides that permit partial credit, while others are scored dichotomously as either correct or incorrect. All constructed response scoring guides are refined from work with a sample of actual student responses gathered during pilot testing of items. All constructed-response items will be scored using rubrics unique to each item. General principles that apply to these rubrics follow:

- Students will not receive credit for incorrect responses.
- All scoring criteria will be text based; students must support statements with information from the reading passage.
- Partial credit will be given for responses that answer a portion of the item but do not provide adequate support from the passage.
- Student responses will be coded to distinguish between blank items and items answered incorrectly.
- Responses will be scored on the basis of the response as it pertains to the item and the passage, not on the quality of writing.
- As part of the item review, the testing contractor will ensure a match between each item and the accompanying scoring guide.

Students are provided information, via task-based UDEs, on elements required for a complete response in individual item stems and/or in overviews of writing prompt items. This information provides all students with greater access to the item and defines the parameters for their response, honoring their time and energy as they engage in the work.

All constructed-response items should communicate clearly to the student how the response to the item will be evaluated, for example whether they must justify their response with reference to the text. In developing the scoring rubric for an item, writers should think about what kind of student responses would show increasing degrees of knowledge and understanding (e.g., as outlined in the ALDs). Writers should provide answer information and sketch sample responses for each score category, even before pilot use. Doing so scaffolds development of a clear scoring rubric and provides guidance for those scoring the item. Item writers should refer to additional directions for developing scoring guides, provided by Governing Board policy and the assessment development contractor, when constructing scoring information for an item. Additionally, the use of passage maps support the development of scoring rubrics.

Short Constructed Response. To provide more reliable and valid opportunities for extrapolating about students' approaches to problems, NAEP assessments include items referred to as short constructed response (SCR) items. These are short-answer items that require students to provide a word, phrase, sentence, or possibly write a brief explanation. SCR items may be scored as correct, incorrect, or partially correct, depending on the complexity of the response required and the information gained from students' responses.

Some short constructed-response items are written to be scored dichotomously. Short constructed response items with two scoring categories should measure knowledge and skills in a way that selected response items cannot or provide greater evidence of the depth of students' understanding. They are also useful when there is more than one possible correct answer, when there are different ways to explain an answer, or when a brief justification is required. Item writers should take care that short constructed response items would not be better or more efficiently structured as selected response items—there should be real value in having students actually constructing a response, rather than selecting the right answer from among wrong answers.

Other short constructed response items are written to be scored on a three-category scale. Short constructed response items with three scoring categories should measure knowledge and skills that require students to go beyond giving a simple acceptable answer that can obviously be scored correct or incorrect. Items scored with a 3-point rubric allow degrees of accuracy in a response so that a student can receive some credit for demonstrating partial understanding of a concept or skill.

As stated previously, item writers must draft a scoring rubric for each short constructed response item. For dichotomous items, the rubrics should define the following two categories:

- 1 = Correct: These responses represent an understanding of the text and a correct response to the item.
- 0 = Incorrect: These responses represent a lack of understanding and an incorrect response to the item.

For items with three score categories, the rubrics should define the following categories:

- 2 = Correct: These responses represent an understanding of the text and a correct response to the item.
- 1 = Partial: These responses represent a partial understanding of the text and a partially correct response.
- 0 = Incorrect: These responses represent little or no understanding of the text and an incorrect response.

Exhibit 3.26. Constructed Response Example: Short Constructed Response Item from Grade 4 NAEP Social Studies

In this Grade 4 NAEP item associated with a Social Studies context, students are asked to explain how a key detail supports the main idea of an article.

Why do you think Marian Anderson began her concert by singing the words, “My country, ‘tis of thee, sweet land of liberty, of thee I sing”? Use information from the article to support your answer.

This item appeared in the 2011 grade 4 NAEP Reading administration NAEP Item ID 2011-4R10 #8.

Exhibit 3.27. Constructive Response Example: Short Constructed Response Item from Grade 8 Smarter Balanced Science

In this Grade 8 Smarter Balanced item associated with a Science context, students are asked to determine which source is most relevant to a specified topic and provide written justification with evidence in their response.

All of the sources provide information about memory. Which source would be **most** relevant to students researching ways to help remember information? Justify your answer and support it with details from the source.

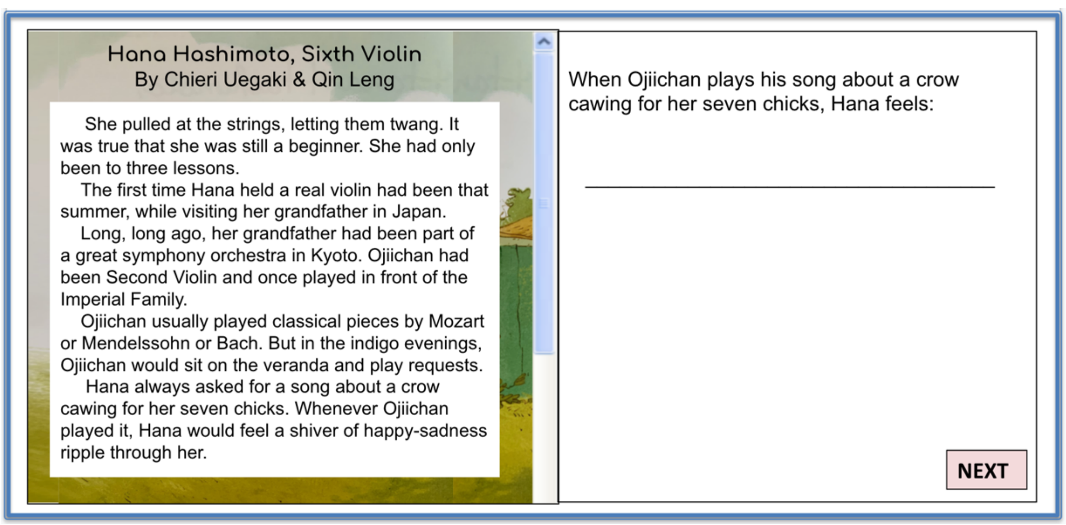
This item appeared in the 2019–20 Grade 8 Smarter Balanced Sample Items published by The Regents of the University of California with Item ID 55409.

Fill-in-the-blank (FIB) items with one response box are SCR items. FIB items require students to enter short text (e.g., a character's name or a text-derived fact). Some FIBs are written to be scored dichotomously with two scoring categories: correct or incorrect. FIBs with two scoring categories should measure knowledge and skills in a way that multiple choice items cannot, or be designed to elicit greater evidence of students' understanding. Such FIBs might be appropriate, for example, to measure reading skills tied to locating explicit textual evidence to avoid guessing (which could be a factor if a multiple choice item were used). FIB items are also useful when there is more than one possible correct answer or when there are different ways to display an answer. Item writers should take care that FIB items would not be better or more efficiently structured as selected response items (such as multiple choice or in-line choice); there should be a purpose for the use of the item type, based on the measurement intent of the item.

Item writers should draft a scoring rubric for each FIB. A writer will not necessarily need to determine the scoring categories for an item, as this depends on the robustness of the item as determined in an iterative item development process.

Exhibit 3.28. Constructed Response Example: Short Constructed Response Fill-in-the-Blank Item from Grade 4 Literature

In this Grade 4 item associated with a Literature context, students are asked to identify how a character feels based on another character's actions and provide the answer as a short text entry.


Hana Hashimoto, Sixth Violin
By Chieri Uegaki & Qin Leng

She pulled at the strings, letting them twang. It was true that she was still a beginner. She had only been to three lessons.

The first time Hana held a real violin had been that summer, while visiting her grandfather in Japan.

Long, long ago, her grandfather had been part of a great symphony orchestra in Kyoto. Ojiichan had been Second Violin and once played in front of the Imperial Family.

Ojiichan usually played classical pieces by Mozart or Mendelssohn or Bach. But in the indigo evenings, Ojiichan would sit on the veranda and play requests.

Hana always asked for a song about a crow cawing for her seven chicks. Whenever Ojiichan played it, Hana would feel a shiver of happy-sadness ripple through her.

When Ojiichan plays his song about a crow cawing for her seven chicks, Hana feels:

NEXT

Extended Constructed Response. Extended constructed response items entail a greater amount of rigor than short constructed response items. In general, extended constructed response items ask students to think deeply about what they have read, to integrate concepts, to analyze a situation, or to explain a concept. These items should be developed so that the knowledge and skills they measure are worth the additional time and effort that they take the student to respond and the time and effort that scoring the response takes. Extended constructed response items typically have four scoring categories: Extensive, Essential, Partial, and Incorrect.

In developing the scoring rubric for an extended constructed-response item, writers should think about the kind of student responses that would show increasing degrees of knowledge and understanding. Writers should sketch condensed sample responses for each score category.

Item writers must develop a draft scoring rubric specific to each extended constructed response item. The rubric should clearly reflect the measurement intent of the item. Item writers also should include a justification or explanation for each rubric category description. Doing so will allow the writer to document the scoring rubric, as well as provide guidance for scoring the item. Extended constructed response items will usually have four scoring categories (with the possibility for additional score categories as appropriate):

- 3 = Extensive: These responses represent an in-depth, rich understanding of the text and a correct response supported by multiple pieces of information from the passage.
- 2 = Essential: These responses represent a solid understanding of the text and a correct response supported by some information from the passage.
- 1 = Partial: These responses represent some understanding of the text and little or no information from the text as part of the response.
- 0 = Incorrect: These responses represent little or no understanding of the text and an incorrect response.

Exhibit 3.29. Constructed Response Example: Extended Constructed Response Item from Grade 4 NAEP Literature

In this Grade 4 NAEP item associated with a Literature context, students are asked to evaluate how a character does or does not change over the course of a story.

Do you think that the innkeeper changes in the story? Use specific information from the beginning and end of the story to support your opinion.

This item appeared in the 2017 grade 4 NAEP Reading administration with NAEP Item ID 2017-4R5 #11.

Exhibit 3.30. Constructed Response Example: Extended Constructed Response Item from Grade 12 NAEP Social Studies

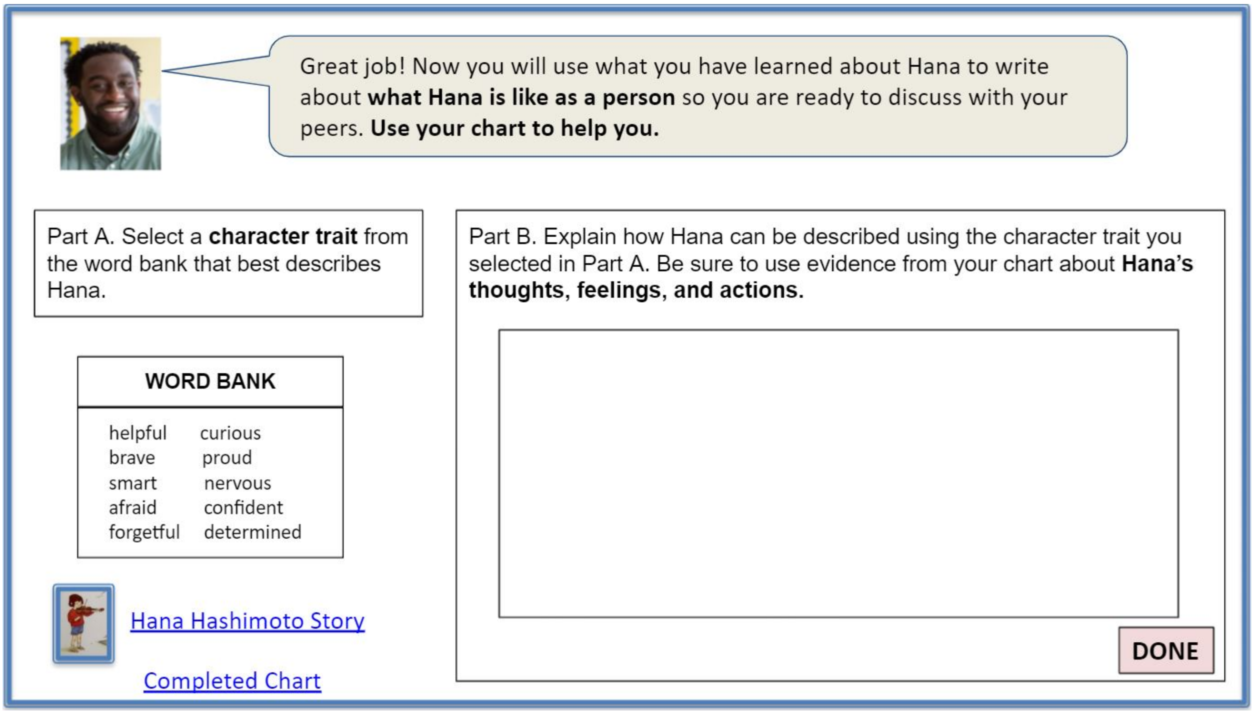
In this Grade 12 NAEP item associated with a Social Studies context, students are tasked to write a an explanatory essay delineating an argument and its claims in a speech.

Roosevelt emphasizes "responsibility" and "duty" throughout his address. According to Roosevelt, why should the nation take responsibility? What are two responsibilities or duties that Roosevelt believed were important?

This item appeared in the 2013 grade 12 NAEP Reading administration with NAEP Item ID 2013-12R11 #6.

Hybrid Constructed Response. As depicted in Exhibit 3.31, in a Grade 4 Literature two-part item with a hybrid constructed response format, students are given a word bank (also acting as a task-based UDE) from which to select a relevant character trait (these could be hot spots where a reader clicks on a word, the word is highlighted and gets recorded as the student's answer to Part A) when asked to describe 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, in Part B, to focus their limited time and cognitive resources on applying evidence from the text about Hana's thoughts, feelings, and actions to a constructed response analysis of the kind of person Hana is.

Exhibit 3.31. Constructed Response Example: Hybrid Constructed Response Item from Grade 4 Literature



The photograph of Mr. Obas is sourced from https://images.all4ed.org/male-sixth-grade-math-teacher-with-protractor (photographer Allison Shelley for EDUimages).

Scoring rubrics for this item type will have more than two scoring categories. Scoring guides are developed to indicate how credit is allocated based on the item construct. For items with three score categories, the rubrics should define the following categories:

- 2 = Correct: These responses represent an understanding of the text and a correct response to the item.
- 1 = Partial: These responses represent a partial understanding of the text and a partially correct response.
- 0 = Incorrect: These responses represent little or no understanding of the text and an incorrect response.

For items with four scoring categories (with the possibility for additional score categories as appropriate):

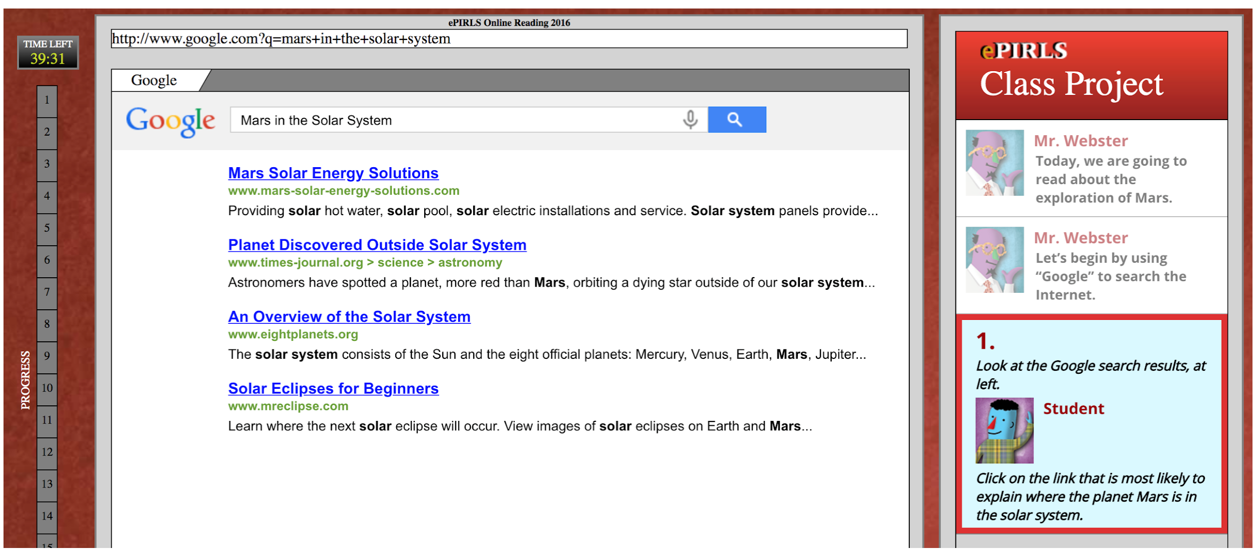
- 3 = Extensive: These responses represent an in-depth, rich understanding of the text and a correct response supported by multiple pieces of information from the passage.
- 2 = Essential: These responses represent a solid understanding of the text and a correct response supported by some information from the passage.
- 1 = Partial: These responses represent some understanding of the text and little or no information from the text as part of the response.
- 0 = Incorrect: These responses represent little or no understanding of the text and an incorrect response.

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 consortia assessments such 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).
- 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 3.32. Constructed Response Example: Dynamic Search Engine Item from ePIRLS 2016 for Grade 4 Students



At each grade level, across the item pool and within an assessment block, developers should develop the range of item response types so that every student experiences answering short constructed response, extended constructed response, and selected response types. A flexible distribution of each item response type is allowed so that the developer can create a set of items for an assessment block that is best suited for the types of texts presented and to the texts' content. Flexible distributions of item response types across each grade level are presented in Exhibit 3.33.

Exhibit 3.33. 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%

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. They are designed to mirror typical (non-testing) reading situations to improve the validity of the assessment.

All readers have access to UDEs. UDEs, or the "built-in features of computer-based assessments," have been included in NAEP since the introduction of the digital platform in 2017, and are available for *all* students (NCES, 2021). 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 (NCES 2019a) Examples of accommodations available on some assessments include extended time, options for responses in braille or Sign Language, or having test-items read aloud. UDEs that venture into this territory and by design target the performance of one group (e.g., students with a visual impairment) are by definition no longer UDEs, as they cannot be applied universally. Distinguishing an appropriate UDE will not always be straightforward —for example, decisions about what exactly makes a vocabulary term obscure or idiosyncratic, or about when introductory text inadvertently provides interpretations that test-takers are supposed to generate on their own. Throughout this *Specifications* document, examples are provided to help assessment developers quickly identify tools and features that could introduce bias.

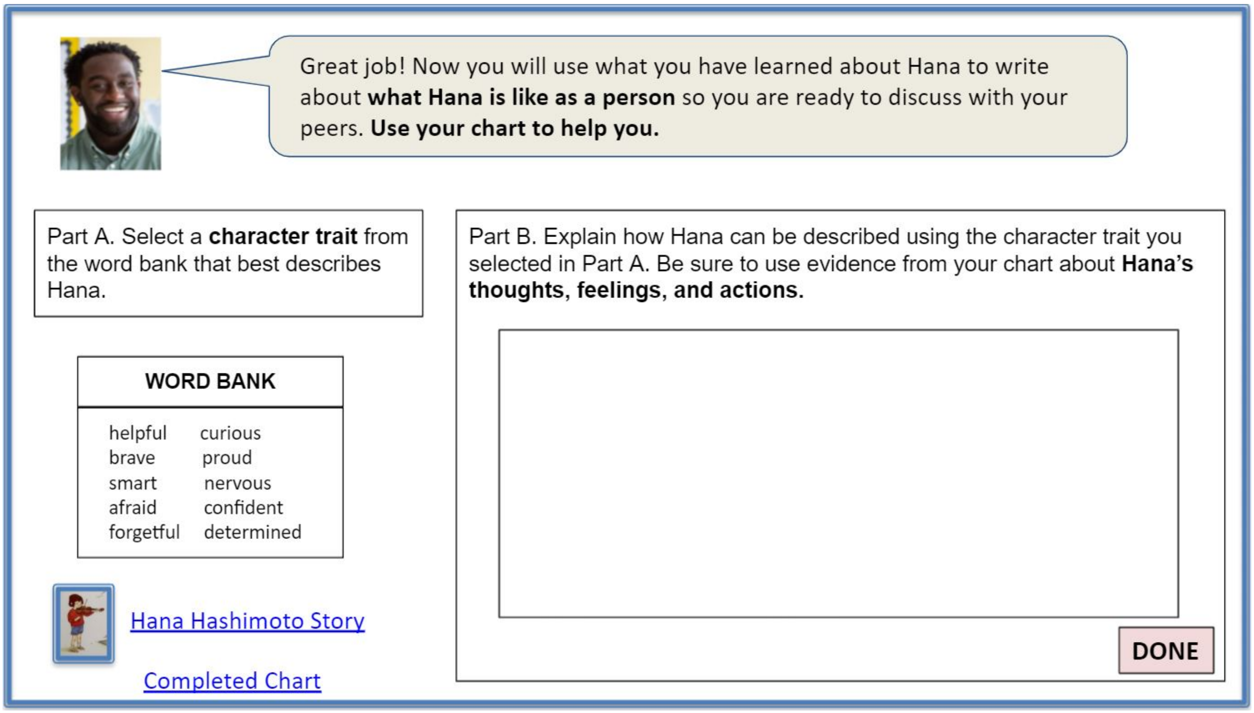
Types of UDEs. Examples of UDEs already exist in the operational NAEP Reading Assessment (e.g., highlighters and look-back buttons) to reflect real-world experiences and how readers use technology. Amidst the use of these digital supports by all test-takers, NAEP has effectively maintained the ability to capture trends over time (NCES, 2021). There are increasingly complex reading purposes and more dynamic texts in today's society. The 2026 Framework calls for a modest expansion of UDEs to reflect increasingly complex reading purposes and more dynamic texts that students encounter in school (Mislevy, 2016). The 2026 NAEP Reading Framework includes three broad categories: task-based UDEs, motivational UDEs, and informational UDEs. The three categories of UDEs are designed to accomplish three different functions, as described below.

Task-based UDEs. In the 2026 NAEP Reading Assessment, task-based UDEs are used to clarify requirements, guide readers in their use of available resources in the testing space, and 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 3.34 illustrates an example of an Analyze and Evaluate 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 a character in depth, drawing on specific details in the text. To demonstrate this skill, the student needs to identify a character trait that is relevant and then 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 Analyze and Evaluate Comprehension Target 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.

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.

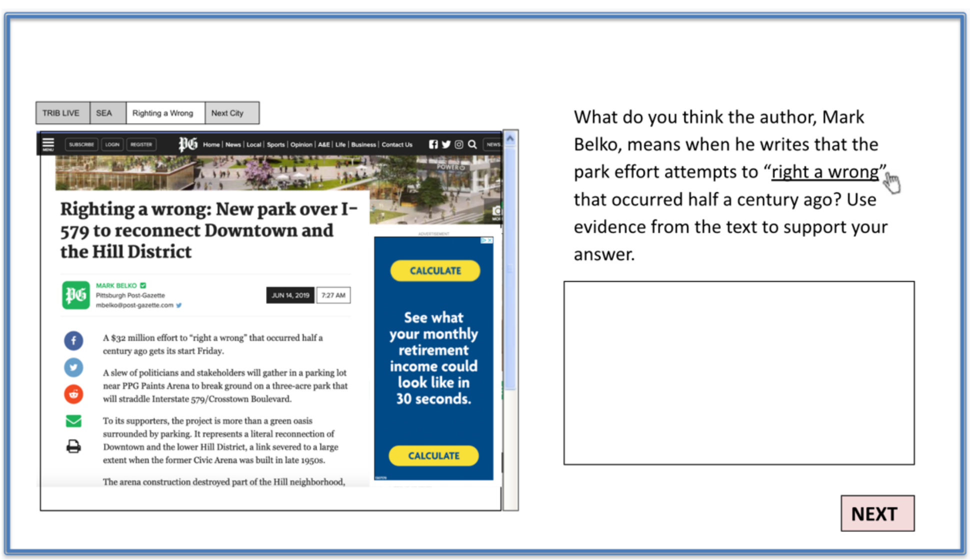
Exhibit 3.34. A Grade 4 Analyze and Evaluate item illustrating a task-based UDE in the form of a word bank providing a set of character traits from which readers can select their choice and then use as part of their constructed response



The photograph of Mr. Obas is sourced from https://images.all4ed.org/male-sixth-grade-math-teacher-with-protractor (photographer Allison Shelley for EDUimages).

Exhibit 3.35 shows a task-based UDE in the form of a look-back button. This type of UDE decreases construct-irrelevance by assisting all students in quickly locating content in a text relevant to the assessment item, without wasting valuable test-taking time and cognitive engagement by the student searching multiple texts that are part of the task.

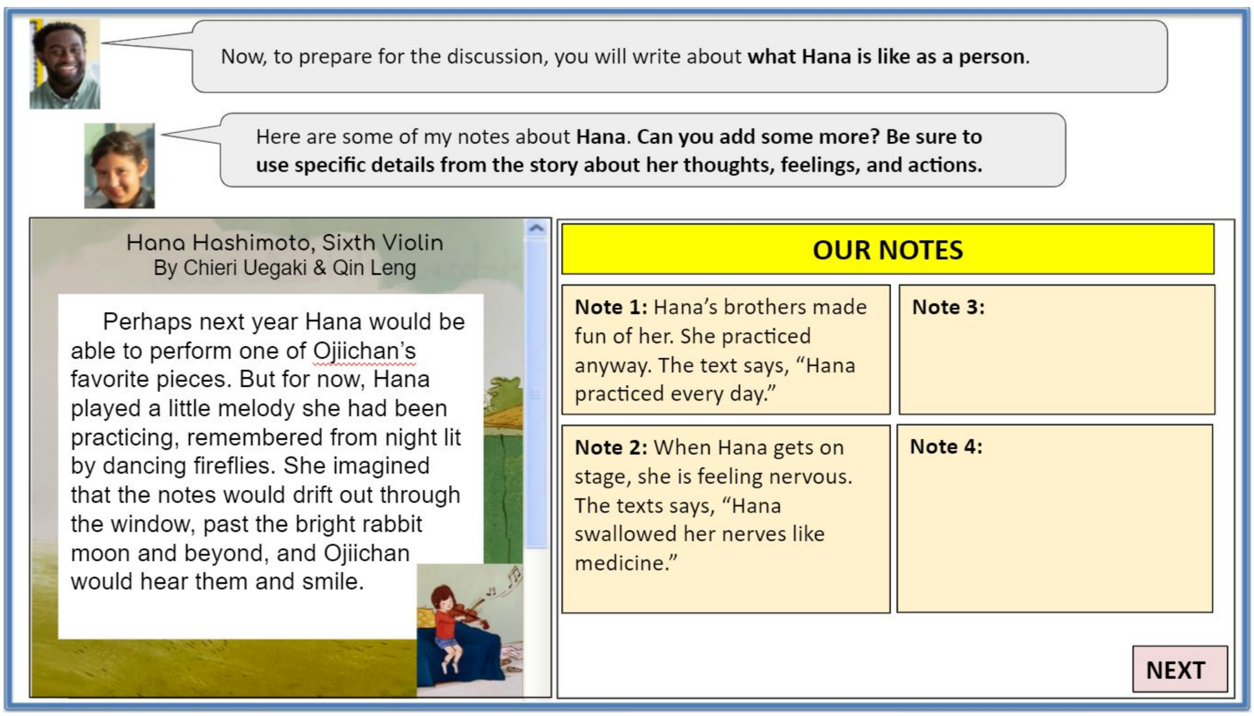
Exhibit 3.35. A Grade 12 short constructed-response item with a look-back button (task-based UDE) that asks readers to integrate and interpret information in an online newspaper article about the historical significance of a park's design



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 that helps to generate a minimal amount of interest in an assessment block. 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 block.

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 3.36). 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. 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, & White, 2011).

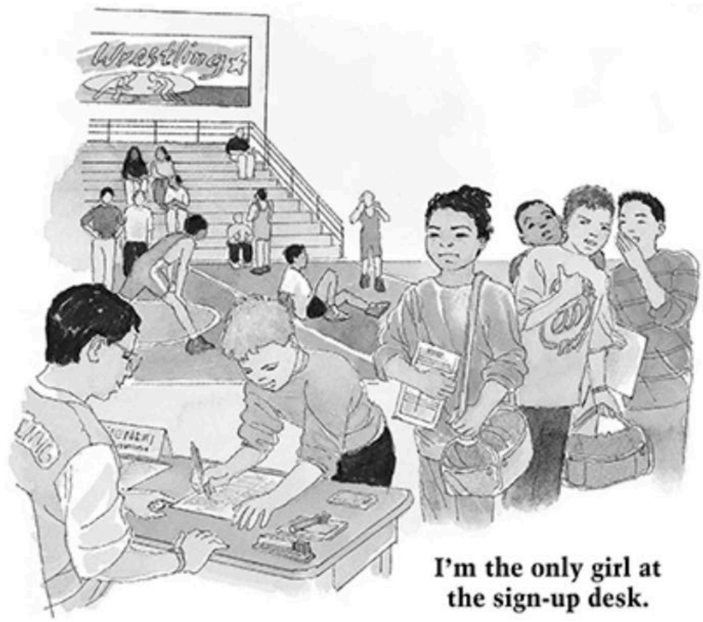
Exhibit 3.36. A Grade 4 motivational and task-based UDE with teacher and student task characters reminding the reader of the task goal for the second task.



The photograph of Mr. Obas is sourced from https://images.all4ed.org/male-sixth-grade-math-teacher-with-protractor (photographer Allison Shelley for EDUimages). The photograph of Gia is sourced from [https://images.all4ed.org/elementary-boy-with-backpack-and-girl-with-notebook/ (photographer Allison Shelley for EDUimages)](https://images.all4ed.org/elementary-boy-with-backpack-and-girl-with-notebook/).

Exhibit 3.37, 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 informational UDE because they briefly introduce the context of the story students are about to read (a girl signing up for something, among only boys).

Exhibit 3.37 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.

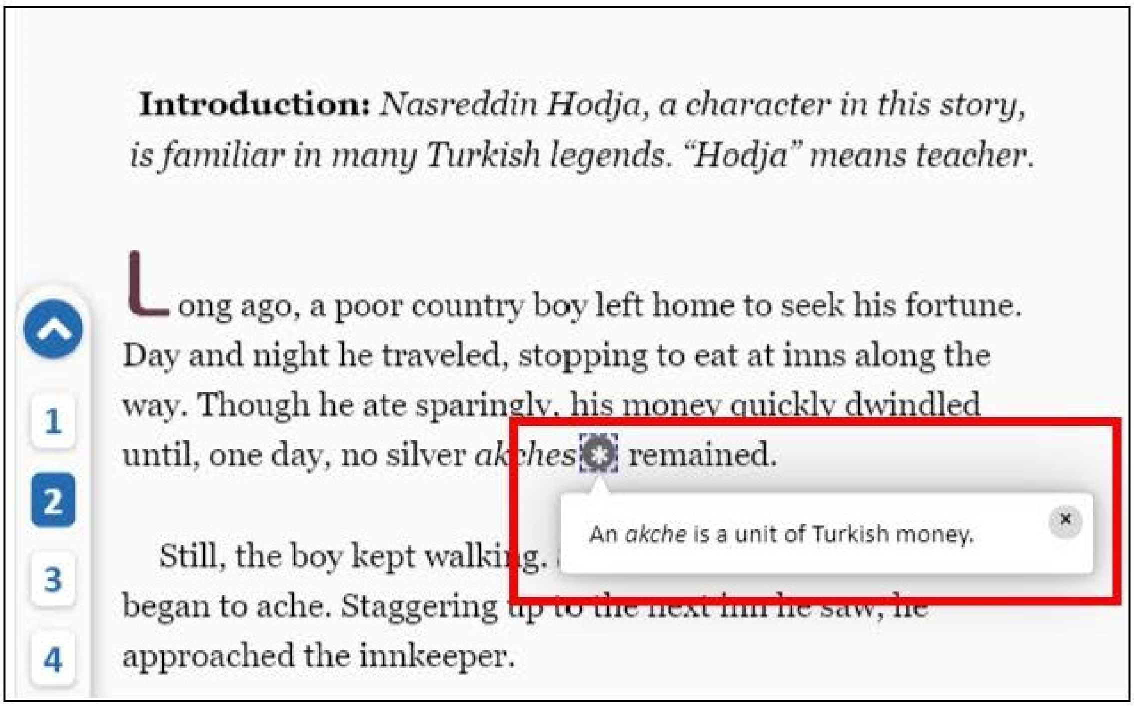
Lots of boys are already on the mats, loosening up. I'm the only girl at the sign-up desk. Some of

Informational UDEs. In the 2026 NAEP Reading Assessment, informational 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 obscure vocabulary unless a word is explicitly tested in a comprehension test item. Obscure vocabulary refers to words of very limited application, such as highly technical terms or non-English referents. In most cases, obscure words will already be defined in the authentic texts, but occasionally the assessment developer may consider whether an additional definition is necessary. 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 when additional context about the author or text is needed to orient students to the passage. A determination must be made by assessment developers about whether a UDE is construct relevant. Finally, as noted in Chapter 2, blocks without UDEs, including those without informational UDEs, are part of the current assessment and will continue to exist in the 2026 NAEP Reading Assessment.

Importantly, informational 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 3.38, from a NAEP Grade 4 block, illustrates two informational UDEs. The first informational 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 informational UDE appears in the form of a vocabulary pop-up box defining the Turkish word "akche."

Exhibit 3.38 Example of Two Informational UDEs from NAEP's "Five Boiled Eggs" Block



Because the meaning or use of the words "Hodja" and "akches" is not directly assessed in this block, this informational 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 introduction defining "Hodja" and the pop-up box defining an "akche" is designed to decrease construct-irrelevant variance. The NAEP Reading Assessment does not assess what students know about obscure topics; that is the job of disciplinary assessments (e.g., the NAEP Science Assessment). Instead, the NAEP Reading Assessment measures how well students can reason about the information provided in texts. Therefore, informational UDEs like these two orient readers to the topic of the text and ensure that all students have an opportunity to make sense of the story and make inferences about the characters.

Exhibit 3.39, from PISA's Reading Literacy test for 15-year-olds, provides an informational 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. This example also 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.

Exhibit 3.39. Example of a Specific Reading Purpose and an Informational UDE from PISA's Rapa Nui Block

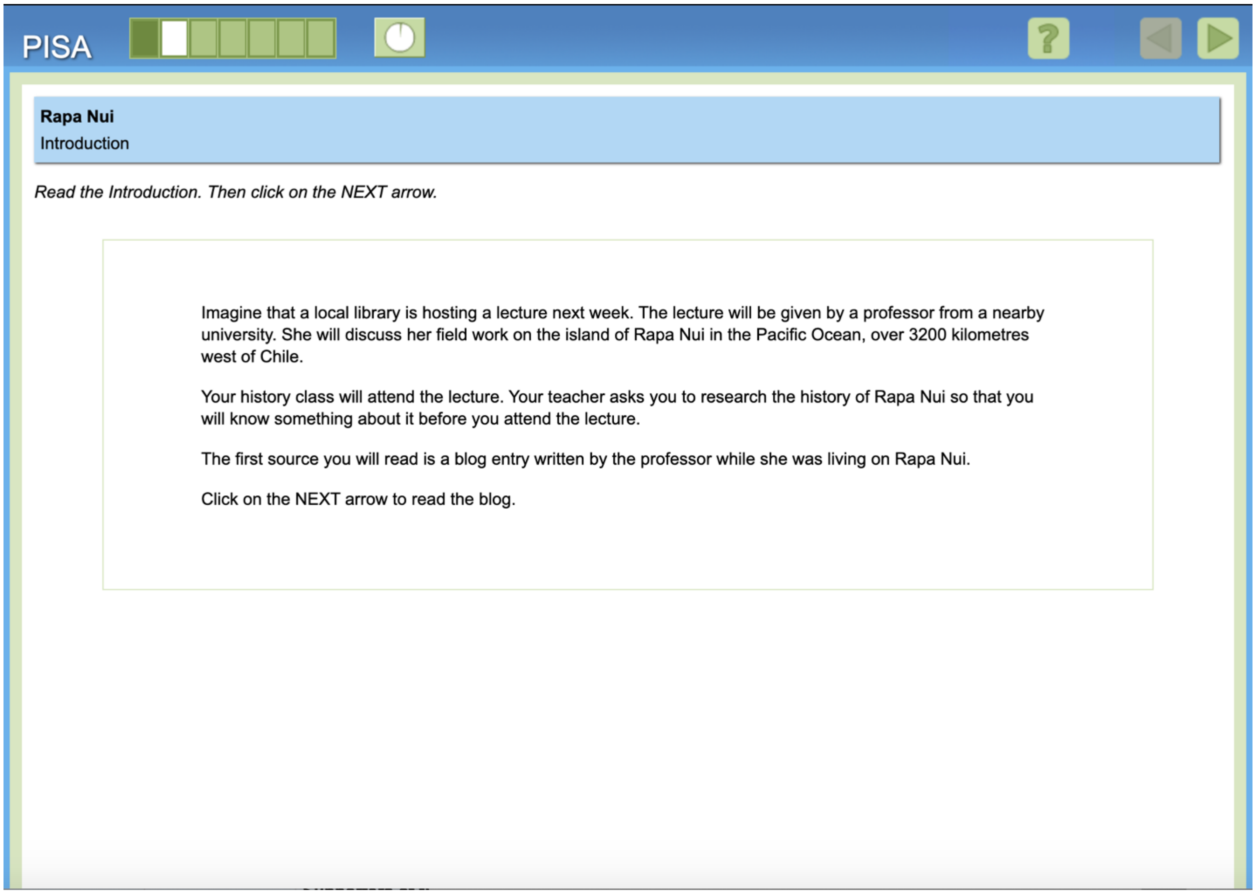


Exhibit 3.40 illustrates two different written introductions, one for each of two texts. In Example 1, an informational UDE appears in the form of an introduction to an article about the writer E. B. White. In Example 2, an informational UDE appears in the form of an introduction to an essay by E. B. White, which explains that the author of the essay is also a children's author.

Exhibit 3.40. Two Examples of Informational 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 3.41, from Michigan's reading assessment for grade 4 students, illustrates three informational 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.

Exhibit 3.41. Example of Three Informational 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.

Selecting appropriate locations for UDEs. Developers decide on appropriate locations in which to insert UDEs into each block of the assessment. Because some 2026 NAEP Reading Assessment tasks will involve complexities in response to handling multiple tasks and texts, readers may be asked to check and reflect on their reading progress in an 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.

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 variables relate to 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);
- 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 3.42 shows that navigation data can be collected via an embedded element (an advertisement as would be seen in a real-world website page context) alongside the task's text.

Exhibit 3.42. 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 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 in complex comprehension tasks while also generating information to further analyze the reading performance of students in grades 4, 8, and 12. Additional research by NCES can inform decisions about the continued use of UDEs.

The purpose of Chapter 4-S 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 Assessments in what has become known as The Nation's Report Card (Governing Board, n.d.). 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.

Reporting Results

Historically, NAEP Reading assessments have 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* (Governing Board, n.d.) allows for examination of results by school characteristics (urban, suburban, rural; public and nonpublic) and other student characteristics (race/ethnicity, gender, English learner status, socioeconomic status, and disability status, i.e., supported by an Individualized Education Program), as required by law. The NAEP Data Explorer is a publicly accessible online tool that allows users to customize reports and investigate specific aspects of student reading achievement, such as performance on different Comprehension Targets or by selected contextual variables. Additionally, reports of the results of survey questionnaires are produced each year on relevant topics, such as instructional emphasis on reading activities, confidence in reading knowledge and skills, teachers' satisfaction with and views on available school resources.

Legislative Provisions for NAEP Reporting

As stated in Chapter 1, under the provisions of the Every Student Succeeds Act (ESSA, 2015) 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 assessment 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. An important development over the last 20 years has been an evolving understanding of how NAEP complements state assessments, which are tightly aligned with state standards.

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. TUDA is administered biennially in odd-numbered years in tandem with NAEP state-level assessments. Sampled students in TUDA districts are assessed in the same subjects and use the same NAEP field materials as students selected as part of national main or state samples. TUDA results are reported separately from the state in which the TUDA is located, but results are not reported for individual students or schools. With student performance results by district, participating TUDA districts can use results for evaluating their achievement trends and for comparative purposes. Here too the complementarity of NAEP with state and local assessments is important to support so as to avoid unnecessary additional testing and to maximize useful information for educators and policymakers to use.

Through ESSA and the NAEP TUDA program, the NAEP Reading Assessment 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

Reporting on achievement levels provides one way for NAEP results to reach the general public and policymakers. 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. Exhibit 4.1 presents the generic policy definitions. See Appendix A for the preliminary achievement level descriptions.

Exhibit 4.1. Generic Achievement Level Policy Definitions for NAEP

Achievement Level	**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 2026 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:

- Revise items included in the reading-specific and the general (i.e., core) parts 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 Assessment databases (including enhancements for the NAEP Data Explorer) in ways that encourage educators, policymakers, and researchers to conduct more nuanced analyses of NAEP Reading Assessment performance.

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

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

Reporting Categories

The reporting system from the 2026 NAEP Reading Framework, and also described below, provides opportunities to interpret results from the 2026 NAEP Reading Assessment 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, whenever feasible, 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, and socioeconomic level. By expanding reporting categories and adding contextual variables, NAEP reporting will provide more robust information on the factors that influence student reading development. 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 NAEP Reading Framework describes two subscales for reporting: reading for literary experience and reading for information. The 2026 NAEP Reading Framework identifies three subscales for reporting 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 reporting the percentage of students whose performance falls within each achievement level (*NAEP Basic, NAEP Proficient, NAEP Advanced*), the 2026 NAEP Reading Assessment will report disciplinary context subscales. This enhancement aligns with increased attention to reading in the content areas in state standards across the nation.

Disaggregating Results Within Demographic Categories

NAEP law (Governing Board, 2017b) requires reporting according to various student populations (see section 303[b][2][G]), including:

- Gender,
- Race/ethnicity,
- Eligibility for free/reduced-price lunch,
- Students with disabilities, and
- English language learners.

Therefore, NAEP will continue to report reading scores by selected student subgroups. In addition, results will be 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 SES within the student subgroup of race/ethnicity. In the NAEP Reading Assessment, as in other large-scale assessments, lower levels of achievement historically are correlated with poverty. Disaggregating results by SES within subgroups will reveal subgroup differences in reading achievement that are associated with SES. At the same time, the success of many schools in supporting high levels of achievement among students from low-SES backgrounds suggests that SES alone does not offer a sufficient explanation for reading performance and that additional contextual variables are crucial to better understand variability in reading (Mullis & Martin, 2019; OECD, 2019). Enhanced reporting can help stakeholders better understand reading performances in context. For example, these data may drive stakeholders 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 the ways described above could be important because the results might suggest that what is, on the surface, presumed to be a cohesive and static category (e.g., a particular subgroup of students) may in fact include significant differences in subgroup members' access to resources. Examining SES and race/ethnicity with a more nuanced lens has the potential to surface factors that are highly amenable to change, such as 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

ELs are defined by NAEP as students "who are in the process of acquiring English language skills and knowledge" (NCES, 2019b). 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; thus, there was no way to disaggregate data to observe or track the successes and increases in achievement of former ELs.

The 2026 NAEP Reading Framework Update expands 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, et al., 2013; 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 outlined in the 2026 NAEP Reading Framework disaggregates scores by three English proficiency categories for which school systems that participate in NAEP already collect data:

- *Current English Learners* – students designated as English learners at the time of the assessment
- *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
- *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. 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

The *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014) recommend that reports of group differences in assessment performance be accompanied by relevant contextual information, where possible, to both discourage erroneous interpretation and enable meaningful analysis of the differences. That standard reads as follows:

> Reports of group differences in test performance should be accompanied by relevant contextual information, where possible, to enable meaningful interpretation of the differences. If appropriate contextual information is not available, users should be cautioned against misinterpretation. (Standard 13.6)

Contextual data about students, teachers, and schools are needed to fulfill the statutory requirement that NAEP include information, whenever feasible, that is disaggregated by race or ethnicity, SES, 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. 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.

Contextual variables are selected to be of topical interest, timely, and directly related to academic achievement and current trends and issues in reading. Data for contextual variables are gathered from student, teacher, and school administrator responses to survey questionnaires and from process data derived from computer monitoring of students' navigation within the assessment tasks completed. Survey questions are intended to be non-intrusive; free from bias; secular, neutral, and non-ideological; and they do not elicit personal values or beliefs. To minimize the burden on those asked to complete the questionnaires, demographic information regarding school and student characteristics are gathered from non-NAEP sources, such as state, district, or school records, when possible.

Current NAEP contextual variables consist of factors that shape students' opportunities to learn, including time, content, instructional strategies, and instructional resources. Resulting data 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 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 Assessment uses an expanded set of research-based contextual variables (Guthrie & Klauda, 2015; Guthrie, Wigfield & Von Secker, 2000) to understand reading achievement (Solano-Flores, 2011; Solano-Flores & Nelson-Barber, 2001). Contextual variables are measurable, and some are also malleable (i.e., they can be influenced). These include *reader characteristics* (e.g., students' self-reports about engagement and motivation, knowledge, 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 by researchers to try 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 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 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, educators, and policymakers to interpret students' differential performance on the NAEP Reading Assessment.

The 2026 NAEP Reading Framework 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 is a complex process shaped by many factors. Factors may include how social and cultural practice influences how readers approach, engage with, and make meaning from texts (Mislevy, 2019; Moje, Afflerbach, Enciso, & Lesaux, 2020; Moje & Luke, 2009; NASEM, 2019; 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; 2020; Phillips Galloway, Brown, & Uccelli, 2020).

The 2026 NAEP Reading Framework envisions an integrated and coherent system of reporting. This reporting system would include research-based contextual variables that form an interrelated network intended to capture reader and environmental characteristics. Taken together, the network of contextual variables is intended to a) correlate with performance on the outcome measure of reading comprehension, b) be malleable (that is, influenced by differences in school and community settings), and c) 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, & Pearson, 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, 2018).
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, & Sabatini, 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

Like reader characteristics, environmental characteristics are also 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, 2020). 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 (e.g., 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 include the following:

Self-Reports 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 the pursuit of a variety of activities (Bowen, Bowen & Ware, 2002).

Self-Reports 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 and reading goals (Wigfield & Wentzel, 2007).
3. **Teacher support for students' background experiences** refers to the students' perceptions that their teacher(s) 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. This may also refer to the extent to which students perceive that the school's reading program and curriculum supports their 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' potential correlations with 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 (Durlak et al., 2015; Guthrie & Klauda, 2016; Katz et al., 2019; Shin et al., 2007; Skerrett, 2020), 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 are 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			
Reports of School and Community Resources			
School social support	√	√	
Belonging in school	√	√	
Participation in out-of-school reading/literacy activities	√		
Reports 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 Reporting Capacity

This chapter describes information that can help contextualize and explain the results NAEP reports and serve as a useful resource for informing educational policy related to teaching reading and learning to read. The evidence collected has the potential 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 at both the national and state levels. 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; NASEM, 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 for either 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.

1. **Reframe the Reporting System Within the Larger Assessment Construct.** The Framework reflects the field's evolving understanding of reading comprehension, cognitive processes, and the changing nature of reading demands in today's society (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). Importantly, it optimizes readers' opportunities to demonstrate reading comprehension that reflect the changing demands of our increasingly complex world (Mislevy, 2016; NASEM, 2018). Reframing and expanding the reporting system is as important as the assessment construct itself in enhancing the appropriateness of inferences based on NAEP results.
2. **Revise Questionnaires**. To increase the capacity to examine the relationships between readers and their environments, NAEP seeks to revise and refresh survey questions. 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 data that reflect the constructs outlined for questionnaires in Exhibit 4.2.
3. **Disaggregate Scores to Achieve More Nuanced 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, which is important 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 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 Nation's Report Card*, and data can be connected 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 that, 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 Analyze 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 Assessment 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 offer increased opportunities to understand the validity, efficacy, and utility of students' assets and needs as readers.

The 2026 NAEP Reading Framework provides opportunities to examine malleable contextual variables that may be correlated with comprehension scores. The identification of malleable factors in the 2026 NAEP Reading Assessment reporting system also provides information that may eventually lead to policies and practices that improve students' reading comprehension instruction and performance. Moreover, the disaggregation of reporting that examines heterogeneity within groups (e.g., race/ethnicity, SES, gender, English learners) along with further disaggregations will provide opportunities for further understanding and greater utility for practice and research and facilitate the avoidance of 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. 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.

The NAEP Reading Framework 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 Framework 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 Specifications presents range ALDs for the NAEP Reading Assessment. 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 2026 NAEP Reading Framework does not provide reporting ALDs; those will be constructed using empirical data during a later stage in the NAEP cycle, i.e., an operational 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.

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) the 2026 Framework has identified three (literature, science, and social studies). In the ALDs below, all three disciplinary contexts are first described within each performance level at each grade, as also seen in the 2026 Framework. Following those general descriptions, for each grade, are listed (in bullet points) some of the possible disciplinary context-specific skills that are associated with the Comprehension Targets and which may appear in an administered NAEP assessment block. **The skills included in these ALDs are illustrative of the range of possible skills that could be addressed in the NAEP Reading Assessment.** A NAEP Reading Assessment block targets a selection of skills appropriate to the reading text(s) in a disciplinary context at a given grade.

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, 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. Items must be developed to address the range of Comprehension Targets with the expectation that there will be a distribution of Comprehension Targets at each achievement level. 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

Reading activities in an assessment block are situated within a disciplinary context as well as a broad reading purpose. 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 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 in the real world occurs. The subsections below describe how specific reading purposes map to disciplinary contexts.

Literature 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 literature 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:

- 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, maps, data, 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

As noted above in the section "Organizational Features and Structures of the Reading Construct: Contexts, Purposes, Comprehension Targets, and Text Complexity," in regards to Comprehension Targets and text complexity, students will engage with texts of various discourse structures and an appropriate grade-level range of text complexity. While reading these texts, 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. Items must be developed to address the range of Comprehension Targets with the expectation that there will be a distribution of Comprehension Targets at each achievement level. **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.

NAEP Basic

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

When engaged in reading literature 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 (e.g., text features) among literature subgenres appearing in specific task texts, and show understanding of vocabulary and simple figurative language. Readers should be able to determine the accuracy of 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 form an evidence-based opinion about a text. 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 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 form an evidence-based opinion about a text. Readers should be able to describe text structures as they pertain to the presentation of content in a specific text, and compare and contrast explicit information found in a firsthand and secondhand account of the same event or topic. Readers should be able to determine the accuracy of a simple summary of a text and integrate information from lower complexity sources to apply to a new context.

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 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 literature 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 information from a multimedia source contributes to understanding of a printed text and show understanding of nuances in word meaning. Readers should be able to apply understanding of a character to an interpretation of another character's point of view.

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 based upon content in the text and to interpret an author's point of view or purpose, including in reference to a procedure and in comparison to another text's author. Readers should be able to determine missing steps in a procedure (e.g., a simple investigation; craft-making related to a scientific concept) based on 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 adopt the persona of a historical figure when applying information learned to a new context.

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 literature 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 common human experiences. Readers should be able to apply knowledge acquired about author's craft to explain how an author achieves an effect.

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 based upon content in the text, 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 based upon content in the text, 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 and apply information from texts in a new context, such as proposing a caption for an illustration or cartoon, or to create a set of recommendations.

To summarize these NAEP Reading Achievement Levels for Grade 4, Exhibit A.1 shows some of the illustrative skills in table format. Additional illustrative skills are located below the table.

Exhibit A.1. NAEP Reading Achievement Levels: Grade 4

<table>
<tr><th></th><th>Reading to Engage in Literature</th><th>Reading to Engage in Science</th><th>Reading to Engage in Social Studies</th></tr>
<tr><td></td><td colspan="3">When reading static, dynamic, and multimodal texts, students will:</td></tr>
<tr><td rowspan="2">NAEP Basic</td><td colspan="3">• Locate, recall, and/or record specific pieces of information.
• Identify relationships between explicitly stated pieces of information.
• Make simple inferences and interpretations in texts.
• Determine the accuracy of summaries.
• Show understanding of vocabulary in the disciplinary contexts.</td></tr>
<tr><td>• Use textual evidence as support to identify or determine literary elements such as character point of view, theme or central message, problem, and setting.
• Explain how a text's illustrations contribute to what is conveyed by the text.
• Explain the differences (e.g., text features) among literature subgenres</td><td>• 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.
• Form an evidence-based opinion about a text.
• Interpret and integrate information presented in a text visually, quantitatively, and orally.</td><td>• Determine the main idea and how it is supported by key details.
• Determine and interpret an author's point of view or purpose.
• Form an evidence-based opinion about a text.
• Describe the text structures as they pertain to the presentation of content in a specific text.</td></tr>
</table>

	appearing in specific task texts. • Show understanding of vocabulary and simple figurative language. • Determine the accuracy of a simple summary of a text. • Continue the narration of an incomplete story to a conclusion of their making.	• Analyze specific results of a simple multistep procedure. • Show understanding of academic and domain-specific vocabulary. • Apply simpler ideas acquired through reading to solve a new problem.	• Compare and contrast explicit information found in a firsthand and secondhand account of the same event or topic. • Determine the accuracy of a simple summary of a text. • Integrate information from lower complexity sources to apply to a new context.
NAEP Proficient	• Make more complex inferences and interpretations. • Reconcile inconsistencies within and across texts. • Explain how an author uses reasons and evidence to support particular points in a text.		
	• 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. • Analyze how information from a multimedia source contributes to understanding of a printed text. • Show understanding of nuances in word meaning. • Apply understanding of a character to an interpretation of another character's point of view.	• Use textual evidence as support to explain events, procedures, ideas, and concepts based on specific information in and across texts. • Make predictions based upon content in the text. • Interpret an author's point of view or purpose, including in reference to a procedure and in comparison to another text's author. • Determine missing steps in a procedure (e.g., a simple investigation; craft-making related to a scientific concept) based on information gained from reading texts.	• Use textual evidence as support to explain events, procedures, ideas, and concepts based on specific information in and across texts. • Explain how information presented in a text visually, quantitatively, and orally contributes to an understanding of a text. • Adopt the persona of a historical figure when applying information learned to a new context.
NAEP Advanced	• Make complex inferences. • Support their interpretations, conclusions, and their judgments based upon evidence within and across texts.		
	• Use textual evidence as support to explain character motivation and behavior and how characters interact with setting and plot. • Evaluate how characters or themes resonate with common human experiences.	• Determine the significance of information and arguments made in a text. • Make predictions based upon content in the text. • Interpret an author's point of view or purpose. • Argue for or against a particular interpretation.	• Determine the significance of information and arguments made in a text. • Make predictions based upon content in the text. • Interpret an author's point of view or purpose. • Argue for or against a particular interpretation.

	• Apply knowledge acquired about author's craft to explain how an author achieves an effect.		• Use and apply information from texts in a new context, such as proposing a caption for an illustration or cartoon, or to create a set of recommendations.

Illustrative Skills Associated with NAEP Reading Comprehension Targets: Grade 4

At each achievement level and with texts at each of the three text complexity levels (low; medium; high), students are expected to demonstrate to varying degrees, per achievement level and text complexity, skills associated with the Comprehension Targets, including but not limited to the skills listed below for each disciplinary context.

Reading to Engage in Literature

- Locate/recall/record specific information or details related to the text
- Determine theme, central message, lesson, moral, or central idea
- Explain how a theme, central message, lesson, or moral is conveyed through details in a text
- Evaluate and form an opinion about a specified aspect of a text or texts and support that opinion with text-based information
- Demonstrate an understanding of the purpose/function of specified text features (e.g., introductions, sidebars, headings, illustrations, charts)
- Determine the accuracy of a summary of a text
- Describe in depth character(s), setting(s), and event(s) in the plot
- Explain how characters' actions contribute to the sequence of events
- Demonstrate an understanding of how the parts of a text are related
- Demonstrate an understanding of differences in point of view across texts
- Determine and interpret the point of view of character(s)
- Explain how specific aspects of a text's illustrations contribute to what is conveyed by the words in a text
- Compare two or more texts in relation to the above skills
- Explain and/or evaluate how information from a multimedia source contributes to understanding of a printed text
- Show understanding of vocabulary, figurative language, word relationships, and nuances in word meanings (e.g., shades of meaning)
- Use information from text(s) in a new situation

Reading to Engage in Science

- Locate/recall/record specific information or details related to the text
- Determine the main idea and explain how the main idea is supported by key details
- Evaluate and form an opinion about a specified aspect of a text or texts and support that opinion with text-based information
- Demonstrate an understanding of the purpose/function of specified text features (e.g., introductions, sidebars, headings, illustrations, charts)
- Determine the accuracy of a summary of a text
- Explain events, procedures, or ideas or concepts, based on specific information in the text
- Show understanding of how to follow precisely a simple multistep procedure

- Analyze the specific results of a simple multistep procedure based on explanations in the text
- Demonstrate an understanding of how an author organizes information in a text or part of a text
- Determine and interpret an author's point of view or purpose
- Explain how specific aspects of a text's illustrations (e.g., maps, photographs) contribute to what is conveyed by the words in a text
- Interpret information presented in a text visually, quantitatively, and orally (e.g., in charts, graphs, diagrams, timelines, animations, interactive elements on web pages) and explain how the information contributes to an understanding of the text
- Evaluate the type and nature of information in a text
- Explain how an author uses reasons and evidence to support particular points in a text
- Explain and/or evaluate how information from a multimedia source contributes to understanding of a printed text
- Compare two or more texts in relation to the above skills
- Show understanding of general academic and domain-specific vocabulary and of figurative language, word relationships, and nuances in word meanings (e.g., shades of meaning)
- Use information from text(s) in a new situation

Reading to Engage in Social Studies

- Locate/recall/record specific information or details related to the text
- Determine the main idea and explain how the main idea is supported by key details
- Evaluate and form an opinion about a specified aspect of a text or texts and support that opinion with text-based information
- Demonstrate an understanding of the purpose/function of specified text features (e.g., introductions, sidebars, headings, illustrations, charts)
- Determine the accuracy of a summary of a text
- Explain events, procedures, or ideas or concepts, based on specific information in the text
- Describe the overall structure (e.g., chronology, comparison, cause/effect, problem/solution) of events, procedures, concepts, or information in a text or a part of a text
- Determine and interpret an author's point of view or purpose
- Compare and contrast a firsthand account and a secondhand account of the same event or topic, including describing the differences in focus and the information presented
- Explain how specific aspects of a text's illustrations (e.g., maps, photographs) contribute to what is conveyed by the words in a text
- Interpret information presented in a text visually, quantitatively, and orally (e.g., in charts, graphs, diagrams, timelines, animations, interactive elements on web pages) and explain how the information contributes to an understanding of the text
- Evaluate the type and nature of information in a text
- Explain how an author uses reasons and evidence to support particular points in a text
- Explain and/or evaluate how information from a multimedia source contributes to understanding of a printed text
- Compare two or more texts in relation to the above skills
- Show understanding of general academic and domain-specific vocabulary and of figurative language, word relationships, and nuances in word meanings (e.g., shades of meaning)
- Use information from text(s) in a new situation

NAEP Reading Achievement Levels: Grade 8

As noted above in the section "Organizational Features and Structures of the Reading Construct: Contexts, Purposes, Comprehension Targets, and Text Complexity," in regards to Comprehension Targets and text complexity, students will engage with texts of various discourse structures and an appropriate grade-level range of text complexity. While reading these texts, 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. Items must be developed to address the range of Comprehension Targets with the expectation that there will be a distribution of Comprehension Targets at each achievement level. **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.

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 across texts, make predictions based upon content in the text, determine the accuracy of summaries, analyze word choice, and show understanding of vocabulary in the disciplinary contexts.

When engaged in reading literature 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 should be able to determine the accuracy of a summary of a text and construct an argument 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, and show understanding of academic and domain-specific vocabulary, key terms, and symbols. Readers 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. They should be able to demonstrate an understanding of the purpose/function of specified text features (e.g., introductions, sidebars, headings, illustrations, charts). 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 should be able to use information from multiple sources to apply to a new context.

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 literature texts such as fiction, drama, film, poetry, and literary nonfiction, eighth-grade readers performing at the *NAEP Proficient* level should be able to 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 information from a multimedia source contributes to understanding of a printed text and how text structure contributes to meaning and style. They should be able to analyze how word choice impacts a text's meaning and tone. Readers should be able to apply analysis of multiple texts to an explanation of 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 multimedia sources with that gained from reading a text on the same topic. Readers should be able to generate an alternative procedure 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. Readers should be able to present an argument 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 texts. Students should be able to evaluate the relevance and strength of evidence to support an author's claims.

When engaged in reading literature 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. They should be able to determine how the text structure contributes to the development of theme, setting, or plot. Readers should be able to describe how a story might change if written from the perspective of another character.

When engaged in reading science texts such as exposition (including literary nonfiction), argumentation, and procedural texts, 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 should be able to construct an argument or explanation 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 should be able to trace and connect various factors (e.g., economic and societal) by incorporating acquired knowledge through reading multiple sources and conducting brief research.

To summarize these NAEP Reading Achievement Levels for Grade 8, Exhibit A.2 shows some of the illustrative skills in table format. Additional illustrative skills are located below the table.

Exhibit A.2. NAEP Reading Achievement Levels: Grade 8

<table>
<tr><th></th><th>Reading to Engage in Literature</th><th>Reading to Engage in Science</th><th>Reading to Engage in Social Studies</th></tr>
<tr><td></td><td colspan="3">When reading static, dynamic, and multimodal texts, students will:</td></tr>
<tr><td rowspan="2">NAEP Basic</td><td colspan="3">• Find information in texts.
• Make simple inferences and interpretations within and across texts.
• Make predictions based upon content in the text.
• Determine the accuracy of summaries.
• Analyze word choice.
• Show understanding of vocabulary in the disciplinary contexts.</td></tr>
<tr><td>• Use textual evidence as support to determine theme or central idea and aspects of character, setting, and plot.
• Compare basic literary attributes of two or more texts and make judgments about how each author presents events.
• Show understanding of vocabulary and figurative language.
• Determine the accuracy of a summary of a text.
• Construct an argument that prosecutes or defends the actions of a character by using evidence from the reading text.</td><td>• Use textual evidence as support to determine the central ideas and conclusions of a text.
• Explain how a text makes connections among and distinctions between individuals, ideas, and/or events.
• 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.
• Show understanding of academic and domain-specific vocabulary, key terms, and symbols.
• Apply simpler ideas acquired through reading to solve a new problem.</td><td>• Determine the central ideas.
• Determine and interpret an author's point of view or purpose.
• Distinguish between fact, opinion, and reasoned judgment in a text.
• Demonstrate an understanding of the purpose/function of specified text features (e.g., introductions, sidebars, headings, illustrations, charts).
• Identify key steps in a text's description of a process related to social studies (e.g., how a bill becomes law).
• Use information from multiple sources to apply to a new context.</td></tr>
<tr><td>NAEP Proficient</td><td colspan="3">• Make more complex inferences and interpretations.
• Form explanations and generalizations.
• Generate alternatives.
• Apply new ideas acquired through reading to a new problem or context.</td></tr>
</table>

	• Use text-based evidence to support arguments and conclusions.		
	• Analyze the development of the theme or central idea over the course of a text. • Analyze how particular lines of dialogue or incidents in a text propel, the action, provoke a decision, or reveal aspects of character. • Analyze how information from a multimedia source contributes to understanding of a printed text. • Analyze how text structure contributes to meaning and style. • Analyze how word choice impacts a text's meaning and tone. • Apply analysis of multiple texts to an explanation of how different authors developed a similar theme or central idea.	• 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. • Analyze how two or more texts provide conflicting information on the same topic, identifying where the texts disagree on matters of fact or interpretation. • Compare and contrast information gained from multimedia sources with that gained from reading a text on the same topic. • Generate an alternative procedure based on knowledge acquired from information gained from reading texts.	• 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). • Analyze the relationship between a primary and secondary source on the same topic. • Analyze how two or more texts provide conflicting information on the same topic, identifying where the texts disagree on matters of fact or interpretation. • Analyze the structure an author uses to organize a text. • Present an argument that proposes a form of social action based on knowledge acquired and opinions formed from the reading texts.
NAEP Advanced	• Make complex inferences • Support their interpretations, conclusions, and their judgments based upon evidence within and across texts. • Evaluate the relevance and strength of evidence to support an author's claims.		
	• 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. • Determine how the text structure contributes to the development of theme, setting, or plot. • Describe how a story might change if written from the perspective of another character.	• Analyze the development of the central idea over the course of the text. • Delineate and evaluate the argument, claims, and reasoning in a text, including whether the evidence is relevant and sufficient to support the claims. • Construct an argument or explanation that synthesizes information from a range of sources to demonstrate a coherent understanding of a	• Analyze the development of the central idea over the course of the text. • Analyze how the author acknowledges and responds to conflicting evidence and/or viewpoints. • Delineate and evaluate the argument, claims, and reasoning in a text, including whether the evidence is relevant and sufficient to support the claims.

		process, phenomenon, or concept.	• Trace and connect various factors (e.g., economic and societal) by incorporating acquired knowledge through reading multiple sources and conducting brief research.

Illustrative Skills Associated with NAEP Reading Comprehension Targets: Grade 8

At each achievement level and with texts at each of the three text complexity levels (low; medium; high), students are expected to demonstrate to varying degrees, per achievement level and text complexity, skills associated with the Comprehension Targets, including but not limited to the skills listed below for each disciplinary context.

Reading to Engage in Literature

- Locate/recall/record specific information or details related to the text
- Determine theme or central idea and aspects of character, setting, and plot
- Analyze the development of the theme or central idea over the course of the text
- Analyze how literary elements relate to each other
- Analyze how particular lines of dialogue or incidents in a story, drama, or narrative poem propel the action, provoke a decision, or reveal aspects of character
- Analyze points of view of and between character(s) and the reader/audience
- Evaluate and form an opinion about a specified aspect of a text or texts and support that opinion with text-based information
- Demonstrate an understanding of the purpose/function of specified text features (e.g., introductions, sidebars, headings, illustrations, charts)
- Determine the accuracy of a summary of a text
- Determine how the text structure contributes to meaning and style, or to the development of theme, setting, or plot
- Compare two or more texts in relation to the above skills
- Explain and/or evaluate how information from a multimedia source contributes to understanding of a printed text
- Show understanding of vocabulary, figurative language, word relationships, and nuances in word meanings (e.g., connotations)
- Analyze how word choice impacts a text's meaning and tone, including how rhymes and other repetitions of sounds (e.g., alliteration) impact a specific section of a text
- Use information from text(s) in a new situation

Reading to Engage in Science

- Locate/recall/record specific information or details related to the text
- Determine the central ideas and conclusions of a text
- Analyze the development of the central idea over the course of the text
- Evaluate and form an opinion about a specified aspect of a text or texts and support that opinion with text-based information
- Demonstrate an understanding of the purpose/function of specified text features (e.g., introductions, sidebars, headings, illustrations, charts)
- Determine the accuracy of a summary of a text

- Explain how a text makes connections among and distinctions between individuals, ideas, and/or events (e.g., through comparisons, analogies, or categories)
- Show understanding of how to follow precisely a multistep procedure
- Analyze the specific results of a multistep procedure based on explanations in the text
- Analyze the structure an author uses to organize a text, including how major sections contribute to the whole and to the development of the ideas, or how a specific paragraph in a text develops and refines a key concept
- Analyze an author's point of view or purpose
- Analyze how the author acknowledges and responds to conflicting evidence and/or viewpoints
- Understand relations among 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)
- Distinguish among fact, opinion, and reasoned judgment within and across multiple sources of information
- Delineate and evaluate the argument, claims, and reasoning in a text, including whether the evidence is relevant and sufficient to support the claims
- Explain and/or evaluate how information from a multimedia source contributes to understanding of a printed text
- Analyze how two or more texts provide conflicting information on the same topic, identifying where the texts disagree on matters of fact or interpretation
- Compare and contrast information gained from multimedia sources with that gained from reading a text on the same topic
- Show understanding of general academic and domain-specific vocabulary, key terms, and symbols
- Show understanding of figurative language, word relationships, and nuances in word meanings (e.g., connotations)
- Analyze how word choice impacts a text's meaning and tone
- Use information from text(s) in a new situation

Reading to Engage in Social Studies

- Locate/recall/record specific information or details related to the text
- Determine the central ideas
- Analyze the development of the central idea over the course of the text
- Evaluate and form an opinion about a specified aspect of a text or texts and support that opinion with text-based information
- Demonstrate an understanding of the purpose/function of specified text features (e.g., introductions, sidebars, headings, illustrations, charts)
- Determine the accuracy of a summary of a text
- Explain how a text makes connections among and distinctions between individuals, ideas, and/or events (e.g., through comparisons, analogies, or categories)
- Identify key steps in a text's description of a process related to history/social studies (e.g., how a bill becomes law; how interest rates are raised or lowered)
- Analyze the relationship between a primary and secondary source on the same topic
- Analyze the structure an author uses to organize a text, including how major sections contribute to the whole and to the development of the ideas, or how a specific paragraph in a text develops and refines a key concept
- Determine an author's point of view or purpose
- Analyze how the author acknowledges and responds to conflicting evidence and/or viewpoints

- Integrate visual information (e.g., in charts, graphs, photographs, videos, or maps) with other information in print and digital texts
- Distinguish among fact, opinion, and reasoned judgment within and across multiple sources of information
- Delineate and evaluate the argument, claims, and reasoning in a text, including whether the evidence is relevant and sufficient to support the claims
- Explain and/or evaluate how information from a multimedia source contributes to understanding of a printed text
- Analyze how two or more texts provide conflicting information on the same topic, identifying where the texts disagree on matters of fact or interpretation
- Show understanding of general academic and domain-specific vocabulary and of figurative language, word relationships, and nuances in word meanings (e.g., connotations)
- Analyze how word choice impacts a text's meaning and tone
- Use information from text(s) in a new situation

NAEP Reading Achievement Levels: Grade 12

As noted above in the section "Organizational Features and Structures of the Reading Construct: Contexts, Purposes, Comprehension Targets, and Text Complexity," in regard to Comprehension Targets and text complexity, students will engage with texts of various discourse structures and an appropriate grade-level range of text complexity. While reading these texts, 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. Items must be developed to address the range of Comprehension Targets with the expectation that there will be a distribution of Comprehension Targets at each achievement level. **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.

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 across texts, make predictions based upon content in the text, determine the accuracy of summaries, analyze word choice, and show understanding of vocabulary in the disciplinary contexts.

When engaged in reading literature 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 should be able to determine the accuracy of a summary of a text and apply a common theme or central idea culled from multiple texts to common human experiences.

When engaged in reading science texts such as exposition (including literary nonfiction), argumentation, and procedural texts, 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 a text structures information to serve an author's purpose and help readers organize their thinking. 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 apply findings described in a text to a new context or situation.

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. They should be able to use information from multiple sources to construct an explanation or argument.

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 literature 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 should be able to analyze how word choice impacts a text's meaning and tone. Readers should be able to present an opinion regarding a universal problem that is elicited from an analysis of the text.

When engaged in reading science texts such as exposition (including literary nonfiction), argumentation, and procedural texts, 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 or describing a procedure, 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 should be able to construct an argument or an explanation 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 historical documents and evaluate the effectiveness of the structure in the text's exposition or argument. Readers should be able to evaluate multiple sources of information presented in different media or formats (visually or in words) in order to construct an argument with evidence to 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 literature 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) compared to the source text. Readers should be able to use or apply information gained from a literary text or a poem to analyze a new text.

When engaged in reading science texts such as exposition (including literary nonfiction), argumentation, and procedural texts, 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 analyze 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 should be able to construct an argument, explanation, or recommendation that requires the application of scientific content from a text.

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 should be able to construct an argument, explanation, or recommendation that utilizes an understanding of legal and ethical principles to address a societal matter of debate (e.g., indigenous peoples' land rights).

To summarize these NAEP Reading Achievement Levels for Grade 12, Exhibit A.3 shows some of the illustrative skills in table format. Additional illustrative skills are located below the table.

Exhibit A.3. NAEP Reading Achievement Levels: Grade 12

<table>
<tr><th></th><th>Reading to Engage in Literature</th><th>Reading to Engage in Science</th><th>Reading to Engage in Social Studies</th></tr>
<tr><td></td><td colspan="3">When reading static, dynamic, and multimodal texts, students will:</td></tr>
<tr><td rowspan="2">NAEP Basic</td><td colspan="3">• Find information in texts.
• Make inferences and interpretations within and across texts.
• Make predictions based upon content in the text.
• Determine the accuracy of summaries.
• Analyze word choice.
• Show understanding of vocabulary in the disciplinary contexts.</td></tr>
<tr><td>• 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.
• Compare literary attributes of two or more texts and make judgments about how each author presents events.
• Show understanding of vocabulary and figurative language.
• Determine the accuracy of a summary of a text.
• Apply a common theme or central idea culled from</td><td>• 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.
• Analyze how a text structures information to serve an author's purpose and help readers organize their thinking.
• Compare and contrast findings presented in a text to those from other sources.</td><td>• 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.
• Distinguish between fact, opinion, and reasoned judgment in a text.
• Show understanding of general academic and domain-specific vocabulary and of figurative language.
• Use information from multiple sources to</td></tr>
</table>

	multiple texts to common human experiences.	• Show understanding of general academic and domain-specific vocabulary, key terms, and symbols. • Apply findings described in a text to a new context or situation.	construct an explanation or argument.
NAEP Proficient	• Make more complex inferences and interpretations. • Form explanations and generalizations. • Generate alternatives. • Apply new ideas acquired through reading to a new problem or context. • Use text-based evidence to support arguments and conclusions.		
	• 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. • Analyze how text structure contributes to meaning and style. • Analyze how word choice impacts a text's meaning and tone. • Present an opinion regarding a universal problem that is elicited from an analysis of the text.	• Use textual evidence as support to analyze an author's point of view or purpose, including in providing an explanation or describing a procedure, identifying important issues that remain unresolved. • 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. • Construct an argument or an explanation that synthesizes information from a range of sources to demonstrate a coherent understanding of a process, phenomenon, or concept.	• Use textual evidence as support to analyze how the central ideas interact and build on one another to produce a complex account. • Analyze the themes, purposes, and rhetorical features of historical documents. • Evaluate the effectiveness of the structure in the text's exposition or argument. • Evaluate multiple sources of information presented in different media or formats (visually or in words) in order to construct an argument with evidence to support a judgment.
NAEP Advanced	• Make complex inferences. • Support their interpretations, conclusions, and their judgments based upon evidence within and across texts. • Use an understanding of legal and ethical principles to develop a text or presentation on a matter of social debate.		
	• Use textual evidence as support to analyze and evaluate multiple interpretations of text (e.g., multimedia versions of a text) to the source text.	• Delineate and evaluate the argument, claims, and reasoning in a text. • Analyze and evaluate the hypotheses, data, analysis, and conclusions in a text.	• Delineate and evaluate argument, claims, and reasoning in a text. • Explain how style and content contribute to the power, persuasiveness, or beauty of the text.

	• Use or apply information gained from a literary text or a poem to analyze a new text.	• Explain how style and content contribute to the power, persuasiveness, or beauty of the text. • Construct an argument, explanation, or recommendation that requires the application of scientific content from a text.	• Construct an argument or explanation that utilizes an understanding of legal and ethical principles to address a societal matter of debate (e.g., indigenous peoples' land rights).

Illustrative Skills Associated with NAEP Reading Comprehension Targets: Grade 12

At each achievement level and with texts at each of the three text complexity levels (low; medium; high), students are expected to demonstrate to varying degrees, per achievement level and text complexity, skills associated with the Comprehension Targets, including but not limited to the skills listed below for each disciplinary context.

Reading to Engage in Literature

- Locate/recall/record specific information or details related to the text
- Determine theme or central idea and aspects of character, setting, and plot
- 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
- Analyze how literary elements relate to each other
- Analyze points of view of and between character(s) and the reader/audience
- Evaluate and form an opinion about a specified aspect of a text or texts and support that opinion with text-based information
- Demonstrate an understanding of the purpose/function of specified text features (e.g., introductions, sidebars, headings, illustrations, charts)
- Determine the accuracy of a summary of a text
- Determine how the text structure contributes to meaning and style
- Compare two or more texts in relation to the above skills
- Analyze and evaluate multiple interpretations of text (e.g., multimedia versions of a text) to the source text
- Show understanding of vocabulary, figurative language, word relationships, and nuances in word meanings (e.g., connotations)
- Analyze how word choice impacts a text's meaning and tone
- Use information from text(s) in a new situation

Reading to Engage in Science

- Locate/recall/record specific information or details related to the text
- Determine the central ideas and conclusions of a text
- Evaluate and form an opinion about a specified aspect of a text or texts and support that opinion with text-based information
- Demonstrate an understanding of the purpose/function of specified text features (e.g., introductions, sidebars, headings, illustrations, charts)
- Determine the accuracy of a summary of a text
- Explain how specific individuals, ideas, and/or events interact and develop over the course of a text
- Show understanding of how to follow precisely a complex multistep procedure

- Analyze the specific results of a multistep procedure based on explanations in the text
- Analyze how a text structures information to serve an author's purpose and help readers organize their thinking
- 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
- Explain how style and content contribute to the power, persuasiveness, or beauty of the text
- 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
- Delineate and evaluate the argument, claims, and reasoning in a text
- Compare and contrast findings presented in a text to those from other sources
- Show understanding of general academic and domain-specific vocabulary, key terms, and symbols
- Analyze how word choice impacts a text's meaning and tone
- Analyze the themes, purposes, and rhetorical features of primary sources
- Use information from text(s) in a new situation

Reading to Engage in Social Studies

- Locate/recall/record specific information or details related to the text
- Determine the central ideas and how the central ideas interact and build on one another to produce a complex account
- Evaluate and form an opinion about a specified aspect of a text or texts and support that opinion with text-based information
- Demonstrate an understanding of the purpose/function of specified text features (e.g., introductions, sidebars, headings, illustrations, charts)
- Determine the accuracy of a summary of a text
- Explain how specific individuals, ideas, and/or events interact and develop over the course of a text
- Analyze the themes, purposes, and rhetorical features of primary sources
- Analyze how a text structures information to serve an author's purpose and help readers organize their thinking
- Evaluate the effectiveness of the structure in the text's exposition or argument
- Determine an author's point of view or purpose
- Explain how style and content contribute to the power, persuasiveness, or beauty of the text
- Evaluate multiple sources of information presented in different media or formats (visually or in words)
- Delineate and evaluate the argument, claims, and reasoning in a text
- Show understanding of general academic and domain-specific vocabulary and of figurative language, word relationships, and nuances in word meanings (e.g., connotations)
- Analyze how word choice impacts a text's meaning and tone
- Use information from text(s) in a new situation

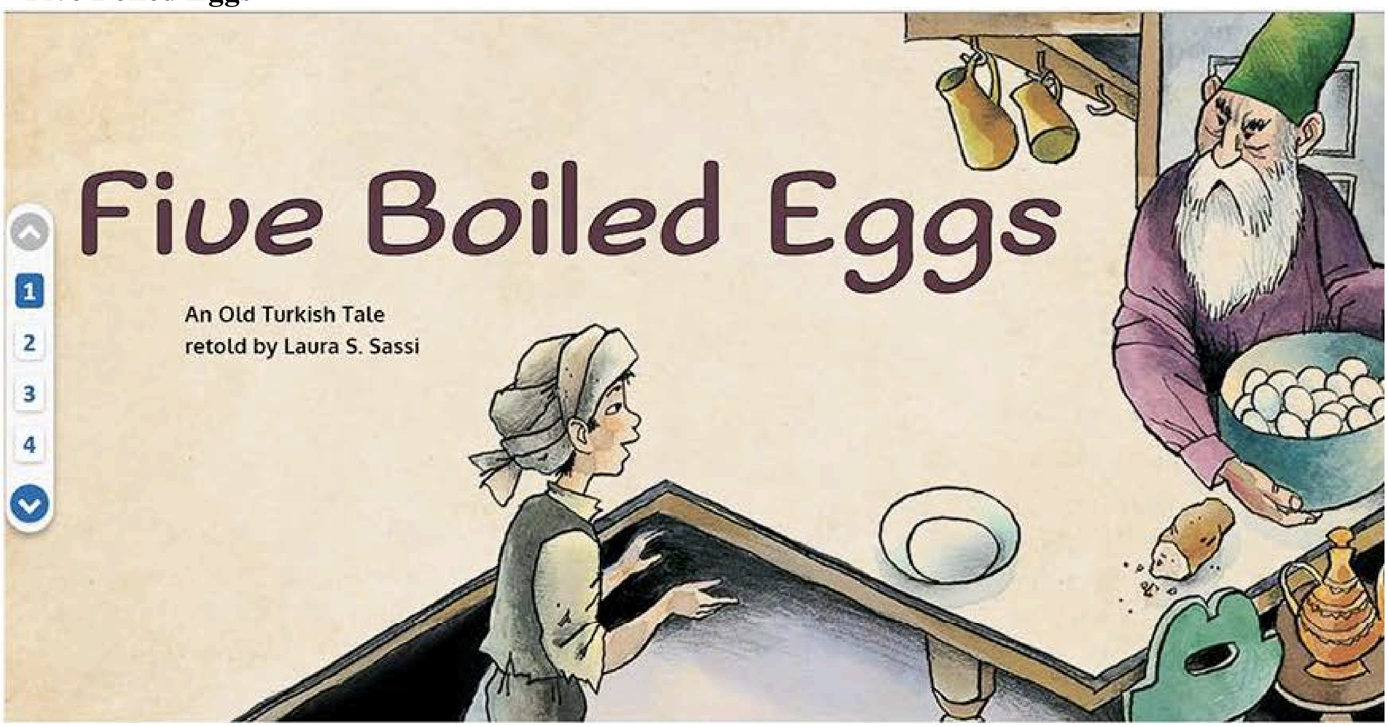
The reading items illustrating NAEP ALDs are organized by disciplinary context (literature; science; social studies), then by each grade level (4; 8; 12) and the achievement levels (Basic; Proficient; Advanced).

Disciplinary Context: Literature

Grade 4

The items illustrating NAEP ALDs at the Basic, Proficient, and Advanced levels in the Grade 4 literature disciplinary context are associated with the text “Five Boiled Eggs.”

“Five Boiled Eggs”



1 2 3 4

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 approached the innkeeper.

"Please feed me!" he said. "I don't have any money now, but I promise to pay you as soon as I can."

"I'll see what I can spare," the innkeeper grumbled. He took five boiled eggs out of a large bowl and put them on a plate with some stale bread. "Here," he said, plopping the platter in front of the boy.

The famished lad gratefully gobbled every morsel. Then, repeating his promise to pay back the innkeeper, he journeyed on.

Revived by his five-egg breakfast, the boy soon reached a bustling seaport. Intent on finding his fortune, he set sail on the first ship that was leaving the harbor.

Years passed, and the lad prospered. As a sea merchant, he sailed far, stopping in many exotic ports. However, he never forgot his humble beginnings or the money he owed the innkeeper.

When he finally returned home, he stopped by the old roadside inn.

1 2 3 4

"Kind sir," he respectfully inquired, "how much for the five boiled eggs that you served me so long ago?"

In truth, the innkeeper did not remember him, for this fine-looking fellow looked nothing like the scrawny lad who had begged for food some ten years before. Still, eager to make a profit, he readily added up the charges. "That'll be ten thousand *akches*," he declared.

"For five eggs?" The rich stranger gasped. He had thought that he would have to pay no more than ten or twenty *akches*.

"Ah, but you must consider their lost worth," the greedy innkeeper replied. "Had you not eaten those eggs, they would have hatched into hens. Those hens, in turn, would have laid eggs that would have hatched into hens. . . ." On and on he ranted until at last he reached his grand total.

When the stunned merchant refused to pay, the innkeeper declared that he would take him to court.

A trial was set for the following week. Alas, rumor had it that the judge was a close friend of the innkeeper.

"I'm ruined!" the merchant muttered as he sat in the village square. "What will I do?"

At that moment, he was approached by a sturdy little man wearing a white turban and riding a donkey. "Nasreddin Hodja, at your service," the man said with a friendly nod. "What seems to be the problem?"

After hearing the merchant's story, Hodja announced, "This is your lucky day! It would be my honor to defend you. I have great experience in these matters."

"Thank you," the merchant said, amazed at his good fortune. But when the court date finally arrived, Nasreddin Hodja was nowhere in sight.

"Woe is me," mumbled the merchant.

"I'll soon be rich!" cried the innkeeper.

"Where is Hodja?" demanded the judge, growing angrier by the minute. He was about to render judgment in the innkeeper's favor when Hodja boldly barged in.

"Pardon me," he said, panting, as he hastily took the witness stand. "I would have been here sooner, but this morning I had the cleverest plan. Instead of eating my boiled corn for breakfast, I planted it. Think of the rich harvest I'll reap!"

"That's absurd," the innkeeper scoffed. "You can't grow corn from cooked kernels!" "Indeed?" Hodja said with mock wonder. "Then, sir, how is it that you would have been able to hatch chickens from boiled eggs?"

At that, the whole room reeled with laughter.

"Order in the court!" shouted the judge, pounding his gavel and scowling at the innkeeper.

The judge then ruled that the merchant would not have to pay even one *akche* for the eggs. Instead, the innkeeper would have to pay a fine for wasting the court's time with such foolishness.

VH151163

This is the end of the story. Tap the Show Questions button to answer the questions.

This text appeared in the 2017 NAEP grade 4 Reading administration in Block 4R5.

NAEP Basic

In this item, students are given options of how a main character in the story became successful, addressing the NAEP Basic level language of "identify relationships between explicitly stated pieces of information."

Comprehension Target: Locate and Recall; **Item Format:** SR-SSMC; **Key:** B
Skill: Recognize paraphrase of explicit details about main character in a story

What did the boy do to become successful?

A He raised hens from the eggs the innkeeper gave him.

B He became a sea merchant and traveled to many places.

C He learned from the innkeeper how to make his fortune.

D He borrowed money to buy a new sailing ship.

This item appeared in the 2017 NAEP grade 4 Reading administration with NAEP Item ID 2017-4R5 #1.

NAEP Proficient
In this item, students are given options regarding how the plot of a story is resolved based upon analysis of the story's plot and character interactions, addressing the NAEP Proficient level language of making "more complex inferences and interpretations."

Comprehension Target: Integrate and Interpret; **Item Format:** SR-SSMC; **Key:** A
Skill: Recognize reason for plot resolution in a story

Why does the judge decide that the merchant does not have to pay?

A ○ Nasreddin Hodja shows that the innkeeper's demand is silly.

B ○ The innkeeper finally agrees that the merchant is right.

C ○ The amount of money the innkeeper wants is much too high.

D ○ Nasreddin Hodja proves that he is a good friend of the judge.

This item appeared in the 2017 NAEP grade 4 Reading administration with NAEP Item ID 2017-4R5 #10.

NAEP Advanced
In this item, students are asked to evaluate how a character does or does not change over the course of a story, addressing the NAEP Advanced level language of making "complex inferences and to support their interpretations, conclusions, and their judgments based upon evidence."

Comprehension Target: Analyze and Evaluate; **Item Format:** ECR; **Key:** N/A
Skill: Evaluate character development using text support from beginning and end of a story

Do you think that the innkeeper changes in the story? Use specific information from the beginning and end of the story to support your opinion.

This item appeared in the 2017 NAEP grade 4 Reading administration with NAEP Item ID 2017-4R5 #6.

Grade 8

The items illustrating NAEP ALDs at the Basic, Proficient, and Advanced levels in the Grade 8 literature disciplinary context are associated with the task-based texts of an excerpt from the novel *The Black Pearl* and the poem "The Last Bargain."

Read the passage from *The Black Pearl*, in which pearl dealers visit the Salazar family. Then answer the questions.

From *The Black Pearl* by
Scott O'Dell

They came early in the afternoon, dressed in their best black suits and carrying a scale and calipers and their money in a crocodile bag. The excitement in the town had died after a couple of days, but when word got around that the dealers were going to the Salazars to buy the great black pearl a crowd followed them and stood outside our gate.

My mother and my two sisters had come back from Loreto, for they too had heard the news of the pearl, and so the fountain in the patio was turned on and the parlor was fixed up with flowers and all the furniture shone.

The four men wore serious faces and they put their calipers and scales on the parlor table and their brown crocodile bag. They sat down and folded their hands and said nothing.

Then my father said, "The bag is very small, gentlemen. I doubt that it holds enough money to buy the great Pearl of Heaven."

The four dealers did not like this. One of them, named Arturo Martín, was big and shaped like a barrel and had small white hands.

“I have heard that the pearl is the size of a grapefruit,” he said. “In which case we have more money than we need. For as you know the large ones are of little value.”

“They do not live long, these monsters,” said Miguel Palomares, who was as fat as Martín and had a bald head that glistened. “They often die or become dull before a year passes.

“And so do many of the small ones,” my father said. “Like the pink one Señor Palomares sold us last month.”

Señor Palomares shrugged his shoulders.

“Before I show the Pearl of Heaven,” my father said, “I will tell you the price. It is twenty thousand pesos, no more and no less.”

The four men looked at each other and smiled thin smiles, as if to say that they had made up their minds already about what they would pay.

My father went out of the room and came back with the pearl wrapped in a piece of white velvet. He laid it on the table in front of the four dealers.

“Now, gentlemen.” With a flourish he unwrapped the pearl and stepped back so all of them could see it. “The Pearl of Heaven!”

The great pearl caught the light, gathered it and softened it into a moon of dark fire. None of the dealers spoke for a moment or two.

Then Señor Martín said, “It is as I feared, more like a grapefruit than a pearl.”

“It is a monster all right,” Señor Palomares said. “The kind that often has a brief life and is very hard to sell.”

One of the dealers who had not spoken cleared his throat and said, “But still we will make an offer.”

The other dealers nodded solemnly.

“Ten thousand pesos,” said Martín.

Señor Palomares grasped the pearl in a small, white hand and studied it.

“I think that I see a flaw,” he said after a long time. “Ten thousand is too much.”

“There is no flaw,” my father said. “And the price, gentlemen, remains twenty thousand pesos.”

The great pearl was passed around to the other dealers and they all turned it in their hands and squinted at it. At last Señor Martín used the calipers and placed the pearl on the scales. His readings were the same as I had made, almost.

“Eleven thousand pesos,” he said.

“Nine thousand more is required,” my father answered. “In your lives you have never seen a pearl like this one nor will you.”

“Twelve thousand,” said Señor Palomares.

After that and for most of an hour the price the dealers offered went up two hundred and fifty pesos at a time until the figure reached the sum of fifteen thousand pesos. And then tempers began to rise and my mother brought in a pitcher of cold juice and a platter of buñuelos. I knew that she wanted to take the dealers’ offer, for I stood where I could see her in the hall making gestures to my father. She had set her mind on a beautiful red carriage and four white horses she had seen in Loreto and was fearful of losing her wish if my father did not lower the price.

Señor Martín wiped his mouth and said, “Fifteen thousand pesos is our last offer.”

“Then,” said my father, “I shall take the great pearl to Mexico City and ask twice that amount and sell it without haggling to dealers who know its true worth.”

Señor Palomares picked up the pearl and put it down. His small head was sunk deep in the folds of his fat neck. Suddenly his head came forth like the head of a turtle and he looked at my father who was pacing back and forth.

“If you remember,” he said, “you made the long journey to the City of México once before. And what did you find there? You found that the dealers are not so generous with their money as we are here in La Paz. And you came home after the long journey with your tail between your legs.”

Señor Palomares got to his feet and the others followed him.

“Fifteen thousand, two hundred and fifty pesos,” he said. “This is final offer.”

Read the poem “The Last Bargain.” Then answer the questions.

The Last Bargain

by Rabindranath Tagore

“Come and hire me,” I cried, while in the morning I was walking on the stone-paved road.

Sword in hand, the King came in his chariot.

He held my hand and said, “I will hire you with my power.”

But his power counted for nought, and he went away in his chariot.

In the heat of the midday the houses stood with shut doors.

I wandered along the crooked lane.

An old man came out with his bag of gold.

He pondered and said, “I will hire you with my money.” He weighed his coins one by one, but I turned away.

It was evening. The garden hedge was all aflower.

The fair maid came out and said, “I will hire you with a smile.”

Her smile paled and melted into tears, and she went back alone into the dark.

The sun glistened on the sand, and the sea waves broke waywardly. A child sat playing with shells.

He raised his head and seemed to know me, and said, “I hire you with nothing.”

From thenceforward that bargain struck in child’s play made me a free man.

These texts appeared in the 2019 Grade 8 Released Items published by New Meridian.

NAEP Basic

In this item, students are asked to determine the tone of a paragraph in a fictional text, addressing the NAEP Basic level language of “analyze word choice, and show understanding of vocabulary.”

Comprehension Target: Integrate and Interpret; **Item Format:** SR-SSMC; **Key:** A
Skill: Analyze how word choice impacts a text's meaning and tone.

Which word **best** describes the tone of the excerpt from *The Black Pearl*?

A. tense
B. eager
C. consoling
D. desperate

This item is adapted from a New Meridian item. The original item appeared in the 2019 Grade 8 Released Items published by New Meridian with Item ID FF429340799.

NAEP Proficient
In this item, students are asked to compare and contrast structure of two texts, addressing the NAEP Proficient level language of "make more complex inferences and interpretations."

Comprehension Target: Analyze and Evaluate; **Item Format:** SR-Match; **Key:** See below the item.
Skill: Compare two or more texts in relation to text structure and literary elements.

Compare and contrast the structure of the passage from *The Black Pearl* and the structure of the poem "The Last Bargain." Drag **each** description into the appropriate box. All descriptions will be used.

The text is a series of different interactions.	The setting changes as the text progresses.	The text is one interaction among characters.
The setting remains the same throughout the text.	The repetition of a key idea fails to resolve the conflict.	The repetition of a key idea leads to a final understanding.

from *The Black Pearl*

"The Last Bargain"

KEY:

from *The Black Pearl*

- The text is one interaction among characters.
- The setting remains the same throughout the text.
- The repetition of a key idea fails to resolve the conflict.

"The Last Bargain"

- The text is a series of different interactions.
- The setting changes as the text progresses.
- The repetition of a key idea leads to a final understanding.

This item appeared in the 2019 Grade 8 Released Items published by New Meridian with Item ID FF429350528.

NAEP Advanced

In this item, students are asked to analyze how events contribute to the development of theme in each text of a paired text task, addressing the NAEP Advanced level language of "make complex inferences and to support their interpretations, conclusions, and their judgments based upon evidence."

Comprehension Target: Analyze and Evaluate; **Item Format:** ECR; **Key:** N/A
Skill: Analyze how events in a text contribute to the text's theme.

You have read a passage from *The Black Pearl* and the poem "The Last Bargain."

Analyze how the events in each text contribute to the development of each text's theme. Be sure to use evidence from **both** texts in your analysis.

This item is adapted from a New Meridian item. The original item appeared in the 2019 Grade 8 Released Items published by New Meridian with Item ID FF429354786.

Grade 12

The items illustrating NAEP ALDs at the Basic, Proficient, and Advanced levels in the Grade 12 literature disciplinary context are associated with the task-based text from *The Odyssey* and the painting *Penelope and the Suitors*.

Today you will read a passage from the epic poem *The Odyssey* and view the painting *Penelope and the Suitors*. *The Odyssey* tells the story of Ulysses, who, after the decade-long war between the Greeks (the Achaeans) and the Trojans, spends an additional ten years journeying back to his home in Greece, where his son Telemachus and wife Penelope await his return. As you read, you will gather information about the passage and the painting and answer questions about them so you can write a narrative story.

Read the passage from *The Odyssey*. Then answer the questions.

from *The Odyssey*

by Homer

"The sons of all the chief men among you are pestering my mother to marry them against her will. They are afraid to go to her father Icarius, asking him to choose the one he likes best, and to provide marriage gifts for his daughter, but day by day they keep hanging about my father's house, sacrificing our oxen, sheep, and fat goats for their banquets, and never giving so much as a thought to the quantity of wine they drink. No estate can stand such recklessness; we have now no Ulysses to ward off harm from our doors, and I cannot hold my own against them. I shall never all my days be as good a man as he was, still I would indeed defend myself if I had power to do so, for I cannot stand such treatment any longer; my house is being disgraced and ruined. Have respect, therefore, to your own consciences and to public opinion. Fear, too, the wrath of heaven, lest the gods should be displeased and turn upon you. I pray you by Jove and Themis, who is the beginning and the end of councils, [do not] hold back, my friends, and leave me singlehanded—unless it be that my brave father Ulysses did some wrong to the Achaeans which you would now avenge on me, by aiding and abetting these suitors. Moreover, if I am to be eaten out of house and home at all, I had rather you did the eating yourselves, for I could then take action against you to some purpose, and serve you with notices from house to house till I got paid in full, whereas now I have no remedy."

With this Telemachus dashed his staff to the ground and burst into tears. Every one was very sorry for him, but they all sat still and no one ventured to make him an angry answer, save only Antinous, who spoke thus:—

"Telemachus, insolent braggart that you are, how dare you try to throw the blame upon us suitors? It is your mother's fault not ours, for she is a very artful woman. This three years past, and close on four, she has been driving us out of our minds, by encouraging each one of us, and sending him messages without meaning one word of what she says. And then there was that other trick she played us. She set up a great tambour frame in her room, and began to work on an enormous piece of fine needlework. 'Sweet hearts,' said she, 'Ulysses is indeed dead, still do not press me to marry again immediately, wait—for I would not have my skill in needlework perish unrecorded—till I have completed a pall for the hero Laertes, to be in readiness against the time when death shall take him. He is very rich, and the women of the place will talk if he is laid out without a pall.'

"This was what she said, and we assented; whereon we could see her working on her great web all day long, but at night she would unpick the stitches again by torchlight. She fooled us in this way for three years and we never found her out, but as time wore on and she was now in her fourth year, one of her maids who knew what she was doing told us, and we caught her in the act of undoing her work, so she had to finish it whether she would or no. The suitors, therefore, make you this answer, that both you and the Achaeans may understand—'send your mother away, and bid her marry the man of her own and of her father's choice'; for I do not know what will happen if she goes on plaguing us much longer with the airs she gives herself on the score of the accomplishments Minerva has taught her, and because she is so clever. We never yet heard of such a woman; we know all about Tyro, Alcmena, Mycene, and the famous women of old, but they were nothing to your mother any one of them. It was not fair of her to treat us in that way, and as long as she continues in the mind with which heaven has now endowed her, so long shall we go on eating up your estate, and I do not see why she should change, for she gets all the honour and glory, and it is you who pay for it, not she. Understand, then, that we will not go back to our lands neither here nor elsewhere, till she has made her choice and married some one or other of us."

This text and art appeared in the 2017 Grade 11 Released Items published by Partnership for Assessment of Readiness for College and Careers (PARCC).

NAEP Basic

In this item, students are asked to distinguish the key events in a text and order them chronologically into an objective summary of the text, addressing the NAEP Basic level language of "create objective summaries."

Comprehension Target: Integrate and Interpret; **Item Format:** SR-Match; **Key:** See below the item.
Skill: Determine an accurate summary of a text.

Provide an objective summary of the passage from *The Odyssey*. Drag and drop key events from the list of sentences into the boxes in chronological order.

- Penelope confesses her love for Ulysses in his absence.
- Telemachus is told that until his mother chooses a new husband, the suitors will not leave his estate.
- The leaders acknowledge the artistry of Penelope's needlework.
- One of the suitors challenges Telemachus's claim that the situation is their fault.
- Telemachus details his financial and emotional challenges and asks for understanding.
- Antinous reveals how Penelope has delayed making a decision about a new husband.
- Telemachus gets into a physical confrontation with one of the suitors.

1.

2.

3.

4.

KEY:

1. Telemachus details his financial and emotional challenges and asks for understanding.
2. One of the suitors challenges Telemachus's claim that the situation is their fault.
3. Antinous reveals how Penelope has delayed making a decision about a new husband.
4. Telemachus is told that until his mother chooses a new husband, the suitors will not leave his estate.

This item is adapted from a PARCC item. The original item appeared in the 2017 Grade 11 Released Items published by Partnership for Assessment of Readiness for College and Careers (PARCC) with Item ID VH130123.

NAEP Proficient

In this item, students are asked to determine which central ideas apply to the text, the painting, or to both stimuli, addressing the NAEP Proficient level language of "make more complex inferences and interpretations."

Comprehension Target: Analyze and Evaluate; **Item Format:** SR-Grid; **Key:** See below the item.
Skill: Determine the central ideas of a text.

Complete the table by selecting the boxes in the appropriate columns to indicate whether the central ideas listed apply to either the passage from *The Odyssey*, the painting *Penelope and the Suitors*, or both the passage and the painting.

Central Idea	Penelope focuses on her project and ignores the suitors.	Penelope's weaving is revealed as a way for her to challenge the suitors.	The suitors are destroying Ulysses's estate.	The suitors are determined to achieve their objective.
The Odyssey				
Penelope and the Suitors				
Both				

KEY:

Central Idea	Penelope focuses on her project and ignores the suitors.	Penelope's weaving is revealed as a way for her to challenge the suitors.	The suitors are destroying Ulysses's estate.	The suitors are determined to achieve their objective.
The Odyssey	☐	☑	☑	☐
Penelope and the Suitors	☑	☐	☐	☐
Both	☐	☐	☐	☑

This item appeared in the 2017 Grade 11 Released Items published by Partnership for Assessment of Readiness for College and Careers (PARCC) with Item ID VH130226.

NAEP Advanced

In this item, students are asked to write an original narrative, using the point of view of a character, based on what they have learned from the text and the painting, addressing the NAEP Advanced level language of "make complex inferences and to support interpretations, conclusions, and judgments based upon evidence."

Comprehension Target: Use and Apply; **Item Format:** ECR; **Key:** N/A
Skill: Analyze the point of view of characters and of multiple interpretations of a text (e.g., multimedia versions of a text) to the source text.

You have read a passage from *The Odyssey* and viewed the painting *Penelope and the Suitors.* Using what you have learned from these sources, write a journal entry from Penelope's point of view, describing what happens after she finishes weaving the cloth. Using details from the passage, your journal entry should offer insight into Penelope's thoughts and interactions with other characters.

This item is adapted from a PARCC item. The original item appeared in the 2017 Grade 11 Released Items published by Partnership for Assessment of Readiness for College and Careers (PARCC) with Item ID VH130242.

Disciplinary Context: Science

Grade 4

The items illustrating NAEP ALDs at the Basic, Proficient, and Advanced levels in the Grade 4 science disciplinary context are associated with the task-based texts "An Eye for Ants" and "Life in an Ant Colony."

Source #1

This article from *Appleseeds* magazine is about a scientist who studies ants.

An Eye for Ants
by Gretchen Noyes-Hull

Dr. Edward O. Wilson, scientist and teacher, has spent his life peeking into the nests of ants. He's curious about the job of each ant in the colony. He wants to uncover the secrets of any colonies' success.

As a child, Edward was often alone. . . . Wherever he lived, snakes, fish, and insects became his friends. For a time, he even kept a colony of harvester ants in a jar under his bed.

The summer he was 7, Edward hurt his right eye in a fishing accident. As he says: "The attention of my surviving eye turned to the ground." It wasn't long before Edward decided to become an entomologist—a scientist who studies insects.

Ants live almost everywhere—from tropical climates to beyond the Arctic Circle, from dry deserts to shady rain forests, from city sidewalks to wild woodlands, and from deep in the ground to the tops of the tallest trees. They live in colonies. An ant colony can have as many as 20 million members.

There is only one queen ant in a colony. It's the queen's task to lay the eggs. Out of the eggs grow worker ants and sometimes a new queen. Every ant in a colony has a job. The main goal of all the worker ants is to take care of the queen and her offspring. This they do in some amazing ways.

For 50 years, Dr. Wilson has traveled around the world looking for new kinds of ants. Sometimes he brings entire colonies back to his laboratory to observe them more closely. He wants to learn about each ant's job within its colony. He wants to know how each ant's job contributes to the future survival of its species.

Dr. Wilson's discoveries help us understand why many animal species develop social organization. In a social organization, each member of the group has a specific job. Each job is important to the entire species' success.

Whenever possible, Dr. Wilson still returns to the place where he first watched ants. He notes the changes in ant species that have occurred over the past 60 years. And today he still relies on the observations and collections of the specimens that he made when he was a young boy.

An Ant Experiment to Try

Worker ants must build, feed, and guard their colony. To do this, they need to communicate with each other. Like most living things, ants depend on chemical odors (known as pheromones) to send messages, such as, "I found food over here . . . alert! there's a stranger in here." Over the years, Dr. Wilson has carried out hundreds of experiments to find the meanings of these odor signals. Although he's made important discoveries, many mysteries remain.

You can do an experiment to test the odor signals of ants. Put several drops of sugar water on a piece of paper. Place the paper near some ants. Watch as one ant discovers the food. Other ants will soon follow the first ant's odor trail. Turn the paper sideways. The ants will still follow the scent of the odor trail, although the sugar water is now in a different place.

Amazing Ant Facts

There are almost 10,000 known species of ants and many more remaining to be discovered. At any one time, 10 million billion (that's 10,000,000,000,000,000) ants are alive. (The world's population of humans is only about 6.6 billion!)

Most ants are scavengers. They find food outside the nest. But some kinds of ants actually "farm" their food. Some "farming" ants grow fungus on underground leaf farms. . . .

Some ants drop pebbles down other colonies' holes. The pebbles block the other ants and keep them from going after the same food.

Some worker ants act like storage containers. They fill themselves up with food like a balloon. If food becomes scarce, they regurgitate it for the rest of the colony. (Regurgitate is the scientific way to say "throw up.")

Source #2

This source is about what happens inside an ant colony.

Life in an Ant Colony

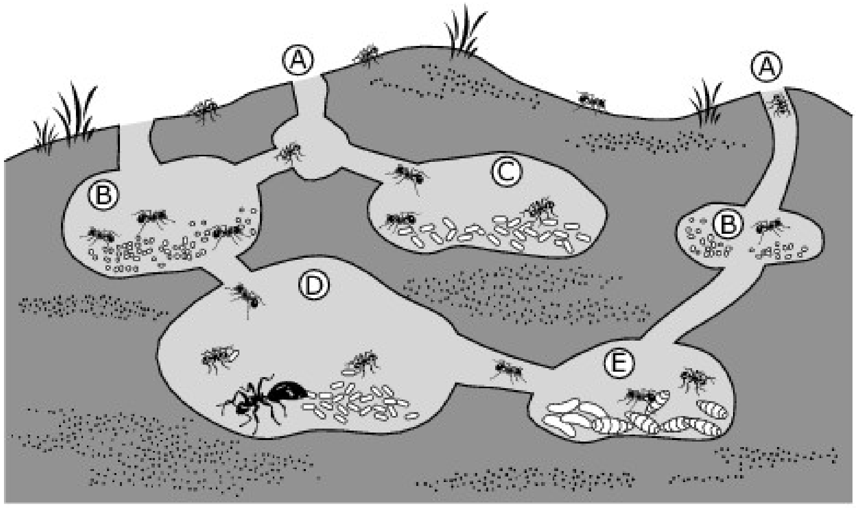
You can find an ant almost anywhere in the world. You will know it is an ant by the six tiny legs, the skinny body in three parts, the bent antennae, and the strong chewing parts. It will be brown, red, yellow, or black.

Ants like to live in the tropics best, but they live on every continent except Antarctica. There are so many ants in the world that if you piled them together, they would weigh about as much as all the people on Earth.

As small as ants are, they are very tough. Many ants bite, and some sting. An ant can carry up to 100 times its own body weight. That is like a child picking up a car! But ants are also strong in another way, and that is in working together in big groups in order to survive. Ant nests are a great example.

Many ants build nests. These ants build nests on the ground, inside logs, under stones, and in trees. They often use wood, leaves, or soil to build nests. Some ant colonies are small enough to nest inside an acorn. Other nests rise above the earth in large mounds. Some colonies extend for a mile or more underground.

Most ant nests have layers of chambers with tunnels to connect them. Most nests also have nurseries where eggs hatch and workers care for young ants.



An ant colony usually has many entrances (A). There are many chambers located underground. Some areas are used to store food (B). One chamber is just for the queen (D). Workers look after unhatched eggs in another chamber (C). One room deep below the surface is used as a nursery for larvae and cocoons (E).

Within the nest there are storerooms for the food that the ants collect. If stored food gets damp during heavy rains, workers bring the wet food up to the surface on the first sunny day. When the food dries out, they return it to the nest.

There are even "stables" within the nest where workers hold and care for other insects. An aphid is a sap-sucking insect that gives off a sugary substance called honeydew. Some ants love honeydew, so they keep a group of aphids to make it for them. This is similar to a farmer having cows that produce milk.

Other ant nests include fungus gardens. Farmer ants grow and take care of this food made from leaves and bits of vegetable matter.

Scout ants go out looking for food. They may wander as far as 700 feet from the nest. If they find food—seeds, grains, or animal matter—they eat it. The food to take home goes into a separate stomach. When the scout ants return to the colony, they regurgitate this food to feed the other ants.

The scouts leave a special chemical called pheromone along the way to the nest. Their nest mates will pick up the scent and follow it back to get more food.

Ants also communicate to protect the colony. When there is danger, the ants release alarm chemicals from their bodies to warn the other ants.

In some ant colonies, a soldier ant sits inside the nest, facing outward. The soldier's head matches the size of the nest entrance. When a worker ant wants to come back inside the nest, it touches the soldier ant's head or antennae to let the soldier know it belongs to the colony.

More than 12,000 species of ants have been classified. There are many differences between them. But they have one important thing in common. Each ant colony thrives on working together for the good of all.

Sources Used:

Ant facts for kids/Ants habitat/Ants diet. (2013). *Animals Time*. Retrieved from http://animalstime.com/ant-facts-kids-ant-habitat-ant-diet/

Ants.(2002-2013). *BioKIDS University of Michigan.* Retrieved from http://www.biokids.umich.edu/critters/Formicidae/

Hadley, D.(2013). 10 fascinating facts about ants. *About.com.Insects.* Retrieved from http://insects.about.com/od/antsbeeswasps/a/10-cool-facts-about- ants.htm

These texts appeared in the 2019–20 Smarter Balanced Grade 4 Sample Items published by The Regents of the University of California.

NAEP Basic

In this item, students are asked to match main ideas from multiple source texts to its appropriate source text, addressing the NAEP Basic level language of "locate specific pieces of information."

Comprehension Target: Locate and Recall; **Item Format:** SR-Grid; **Key:** People know so much about ants because of entomologists.: Source #1: An Eye for Ants; Ant colonies are able to survive because of the different jobs that the ants have.: Source #1: An Eye for Ants, Source #2: Life in an Ant Colony
Skill: Determine key ideas in a text.

Click on the boxes to match each source with the idea or ideas that it supports.

	Source #1: An Eye for Ants	Source #2: Life in an Ant Colony
People know so much about ants because of entomologists.	☐	☐
Ant colonies are able to survive because of the different jobs that the ants have.	☐	☐

This item appeared in the 2019–20 Grade 4 Smarter Balanced Sample Items published by The Regents of the University of California with Item ID 61164.

NAEP Proficient

In this item, students are asked to briefly explain, using textual evidence, key details that support central ideas by locating the details in multiple source texts, addressing the NAEP Proficient level language of "uses reasons and evidence to support particular points in a text."

Comprehension Target: Locate and Recall; **Item Format:** SCR; **Key:** N/A
Skill: Locate information to support central ideas.

The sources describe jobs ants do in a colony. Explain some of the specific jobs ants do that help the colony. Give **one** detail from Source #1 and **one** detail from Source #2 to support your answer. For each detail, give the source title or number.

This item appeared in the 2019–20 Grade 4 Smarter Balanced Sample Items published by The Regents of the University of California with Item ID 61160.

NAEP Advanced

In this item, students are asked to produce a narrative, synthesizing and incorporating information learned from multiple source texts, addressing the NAEP Advanced level language of "make complex inferences and to support their interpretations, conclusions, and their judgments based upon evidence."

Comprehension Target: Use and Apply; **Item Format:** ECR; **Key:** N/A
Skill: Use information learned from sources and apply to a new context.

A book author comes to your class and talks about his book of short stories. After his talk, he asks your class to write their own short stories and says he will come back to the class and listen to all of the stories.

After your research on ants, you decide to write a story about what happens when you shrink, fall into a hole in the ground, and find yourself part of an ant colony.

Write a scene for your story about how you help the ants with their food on a rainy day. Use information and details from the two sources in your story. Make sure you include one or more characters, a setting, and a plot.

This item is adapted from a Smarter Balanced Item. The original item appeared in the 2019–20 Grade 4 Smarter Balanced Sample Items published by The Regents of the University of California with Item ID 54697.

Grade 8

The items illustrating NAEP ALDs at the Basic, Proficient, and Advanced levels in the Grade 8 science disciplinary context are associated with the task-based texts "How Do We Remember," "Memory Masters," and "Interpreters: Silver-Tongued Masters of Memory."

Task:

You are learning about the brain in science class. You are curious. Why are some people able to memorize their student number with ease, but you can't remember the four items you are supposed to pick up at the store? You decide to do some research on memory and the brain. As part of your initial research, you have uncovered three articles about memory.

After you have reviewed these sources, you will answer some questions about them. Briefly scan the sources and the questions that follow. Then, go back and read the sources carefully to gain the information you will need to answer the questions and finalize your research. You may click on the NOTES box to take notes on the information you find in the sources as you read. Your notes will be available to you as you answer the questions.

In Part 2, you will write an explanatory article on a topic related to the sources.

Directions for Beginning:
You will now examine several sources. You can reexamine any of the sources as often as you like.

Research Questions:
After examining the research sources, use the remaining time in Part 1 to answer question(s) about them. Your answers will be scored. Also, your answers will help you think about the research sources you have read and viewed, which should help you write your explanatory article.

You may click on the NOTES button above the sources to look at your notes when you think it would be helpful. Answer the questions in the spaces provided below them.

Part 1

Sources for Performance Task:

Source #1
Read the article about memory from a popular science website for kids.

How Do We Remember?

You need to go to the store and pick up milk, eggs, butter, and bread. You repeat the list of foods over and over on the way to the store. When you arrive at the store, you collect the milk, eggs, bread, and . . . What was the other thing? How did you already forget the other item that was on your mental list? How does your memory work, and why does it let you down sometimes?

When most people refer to memory, they think of it as one part of the brain. The truth is your memory isn't one particular part of your brain. Memory involves several parts of your brain working together. It is a concept. It is the idea of remembering.

Formerly, scientists used to describe memory as a miniature filing cabinet full of many files that contained memories. Others described memory as a tiny supercomputer located in the brain. Today, scientists believe that memory is much more complicated than that.

How Memory Works
Memories begin as a result of the senses. The memory is then encoded, or stored, in your brain with electrical impulses and chemicals. Your brain is full of nerve cells. There are electrical pulses carrying messages from one cell to another. The electrical pulses trigger chemical messengers to be released. The chemical messengers are called neurotransmitters. The connection that is made between the cells isn't necessarily permanent. It is changing all of the time. Brain cells work together as a team, organizing themselves into groups. The groups specialize in different kinds of information processing. Each time one cell sends a message to another, the connection between those two cells gets stronger. With each new experience your brain changes a little. If you keep using your brain the same way over and over again, it shapes how your brain will be organized.

Types of Memory
There are three types of memory: sensory memory, short-term memory, and long-term memory.

Sensory memory hangs on to information for a very short period of time, only a second or two. When you look at a picture of a beautiful landscape, an almost exact image of that landscape is stored momentarily in your visual sensory memory. Your visual sensory memory requires your eyes and parts of your brain to work together. Unless you make an active effort to think about the landscape the image will quickly fade.

Short-term memory stores what you are actively thinking about at any given moment. Your short-term memory is able to hold on to information for as long as you are thinking about it. You use your short-term memory to remember the list of things your mom wants you to pick up at the store. If you continually repeat this information to yourself, you can remember it, but the moment you start thinking about something else, like where in the store the milk is located, the list of groceries will only stick around for about 20 or 30 seconds.

Long-term memory stores information, experiences, and ideas long after you stop thinking about them. When you consciously process information, short-term and long-term memory work together. For example, when you think or solve problems, the short-term and long-term memory systems are working together. Long-term memory includes an enormous amount of information. Some of this information is there for a lifetime. Scientists believe that over the course of a lifetime, the long-term memory has stored vast amounts of information. Much more than an encyclopedia!

Forgetting
As time passes, memory fades or we forget all of the specific details. An hour after you read a book, you can remember most of what it was about. Two days later, you might recall only a bit of the information that was in the book. After a month has passed, you probably remember even less.

There are several explanations as to why we may forget things. Maybe the information was not encoded in our memory properly. For instance, while reading over your notes for the test you were trying to watch your favorite show on television. This type of distraction can really interfere in encoding memories and the information is not successfully saved in your memory.

Alternatively, another reason that you may not be able to remember something is not because you actually have forgotten the information. The problem could be that you are having trouble retrieving it from your memory. You can't remember the answer to write it down on the test. It is right there, you know the answer, but it just won't come to you. As soon as the test is over and you walk out of the classroom, there it is—that answer you were trying so hard to come up with. This is a problem with retrieval. Your brain is having trouble locating that information again. It is similar to looking for a small object inside a room that is full of stuff. It can be very frustrating!

References

Loftus, E.F. (2013). Memory. *World Book Advanced*. Retrieved from http://www.worldbookonline.com/advanced/printarticle?id=ar354840

Mohs, R.C. (2007, May 8). How human memory works. *HowStuffWorks.com*. Retrieved from http://science.howstuffworks.com/life/inside-the-mind/human-brain/human-memory.htm

Source #2
Read the article about people who participate in memory championships from a 2012 issue of *Appleseeds* magazine.

Memory Masters
by Alice Andre-Clark

Nelson Dellis can look at a deck of cards for 5 minutes and then tell you the order of every single card in 63 seconds.

If you give teenager Sophia Hu a list of random words and let her study it for just 15 minutes, she might remember as many as 120 words.

Dellis and Hu were contestants in the USA Memory Championship, which has been crowning our top "mental athletes" since 1997. At the Memory Championship you start by studying the pictures of 117 strangers for 15 minutes, then try to remember all their names. In 2010, Hannan Khan listed 159 first and last names. Later, try meeting five guests at a pretend tea party and see if you can later recall their names, addresses, pets' names, hobbies, favorite foods, and more.

Think you have a knack for numbers? Try memorizing a sheet of 500 digits. It'll be tough to beat Dellis, who once remembered 248 numbers after only 5 minutes of studying.

Most of our top mental athletes say they weren't born with amazing memories. Brain scientists agree that there's probably nothing physically unusual about the brains of memory champions. They just happen to know a few tricks for keeping a lot of facts in their minds at once. . . .

Building a Memory Palace
Memories get stronger if you associate them with a place. To remember your shopping list, build it a "memory palace." Picture a building you know well, perhaps your own house. Now imagine each item in a different part of the house. Marshmallows strung like pearls, dangling from your mom's jewelry drawer. A graham-cracker fan on the coffee table. Chocolate bars popping out of the toaster.

Person + Action + Object = ?
Need to memorize a long string of numbers? Start by thinking of a person, an action, and an object for each number from 00 to 99. . . .

Now you're ready to learn a bigger number. For 872,936, combine the person from 87 with the action from 29 and the object from 36. . . .

What's in a Name? A Picture
Names can be hard to recall. Words like "mirror" and "table" may bring up lots of memories, but the first time you meet a Peyton or a Mrs. Cohen, you might not associate those words with anything. Change names just a little, and Cohen becomes "cold hen," an unhappy chicken sitting on a nest filled with ice cubes.

Use pictures to match faces with names too. If Mrs. Cohen has curly red hair, give the hen some fluffy red feathers. Long-necked Peyton ("pay ten") could become a stretched-out ten-dollar bill. Soon you'll rarely forget a name.

Andre-Clark, Alice. (2012, July/August). Memory Masters. *Appleseeds*, pp. 8-11.

Source #3
Read the article about interpreters and memory from a 2008 issue of *Odyssey* magazine.

Interpreters: Silver-Tongued Masters of Memory
by Charles Capaldi

Today, Murielle Pérégovoy sits in a glass-enclosed booth. An ultra-light headset rests on her ears. A microphone hovers in mid-air, inches from her mouth. Pérégovoy doesn't see any of it. Her attention is riveted on the space between her ears, which is currently filled with short bursts of angry Russian from a participant who has the floor on the other side of the conference room. Her voice rises and falls to match that of the speaker, filling the booth and the headsets of everyone tuned to the French channel. The participant finishes speaking and sits down. Murielle finishes one sentence behind him and reaches out to turn off her microphone. On any given day, she could be the voice of an ambassador, a distraught mother in war-torn Iraq, or an orthopedic surgeon. Pérégovoy is a simultaneous interpreter, and her workday has just ended.

In addition to knowing their native languages, professional interpreters are expected to understand two or more languages as well as any educated native speaker. More than 50 percent of the world's population is *bilingual* (speaks a second language from early infancy), and many bilingual people are drawn to the field.

By the age of two, most children have a vocabulary of about 2,000 words. The average American high school graduate has a vocabulary of about 50,000 words. A bilingual high school graduate can possess a vocabulary twice that size, split across two languages. Imagine the vast vocabulary stored in the long-term memory of an interpreter. Interpreters, then, seem to have amazing memories. But do they really? Questions like this one keep neuroscientists up late at night.

One of these neuroscientists is Dr. Michel Paradis, who teaches at McGill University in Montreal, Canada, and researches aphasia in bilingual people. *Aphasia* (the inability to understand or use language) usually results from a traumatic brain injury such as a stroke or accident. In the course of his research, Paradis has learned a lot about memory and language in people who are not aphasic. So, when asked whether interpreters have better memories than average, he says, "In the same way that the term intelligence covers many different types of capabilities, memory is an umbrella term that refers to many different kinds of capacity."

"Much of an interpreter's brain power is devoted to keeping information in short-term memory," says Paradis. "Simultaneously listening in one language and speaking in another makes the task much more challenging." How then does Murielle's brain undertake this seemingly impossible task?

As the message flows through her headphones, Pérégovoy must decode it. Decoding does not mean knowing what each word means. Interpretation focuses on the message being conveyed, rather than the words used to convey it. Understanding the speech flowing through her headset requires the use of procedural (a type of non-declarative) memory—the kind of memory we use for automated tasks, skills, and habits. The interpreter knows the language of the speaker well enough to understand it effortlessly. Similarly, when you hear an utterance in English you probably aren't even aware of trying to understand it. The fact that you comprehend it subconsciously is the hallmark of procedural memory.

Once Pérégovoy's brain has decoded the message, it identifies blocks of information that should be stored for later use. This identification process is a conscious activity. Murielle's memory clings to facts, events, people, and objects, relying on what neuroscientists call declarative memory. Where procedural memory is subconscious, declarative memory requires effort and focused attention.

Murielle stores the decoded message in her short-term memory and holds it there until it has been correctly translated. She must retrieve the information and compare it to her translation before uttering a single word into the microphone. This step involves working memory. Think of working memory as a tub being filled with water and drained simultaneously. Water cannot flow into the tub at a faster rate than water drains from the tub or else the tub will overflow.

While all this is happening in Pérégovoy's brain, the speaker continues talking. The average person speaks at 120 words per minute, with bursts that reach 180 words per minute. Neuroscientists have identified that working memory has about 10 seconds (or 20 words) of storage capacity. As new information is continually added to the tub, previously stored information is constantly being compared to the memory store, putting an extra burden on working memory.

For instance, Dr. Franco Fabbro at the University of Udine in Italy found that advanced interpreting students remembered fewer details of a story when they were asked to interpret it than when they just listened to it. Other studies show that sign language interpreters have better recall than interpreters of spoken languages. Sign language interpreters undertake the same process of decoding and encoding the message in another language, but sign language does not require them to speak their translation. Instead, they deliver the message through their hands and upper bodies. Dr. Fabbro and his colleagues reasoned that the demand on interpreters to speak and listen simultaneously might be at the root of the memory interference. To test this hypothesis, he asked the students to listen to another set of stories and told them not to interpret, but to keep

repeating "the . . . the . . . the. . ." while they listened. He found that these students remembered fewer details than when just listening to the stories. Working memory is taxed by the need to listen and speak at the same time, and when working memory is burdened, memorizing information becomes more difficult.

Interpreters may start out with the same three pounds of gray matter that everyone else has, but they have trained their short-term memory to help perform a particular task. Not everyone with a three-pound brain will have what it takes to become an interpreter, in much the same way that not everyone with a good pair of lungs and a love of music will grow up to become an opera singer. A lot depends on how you train, how committed you are, and your natural inclinations. "You can be good at one type of memory and poor at another," Paradis explains. "But you can improve each type of memory with practice. If you want to increase your memory, EXERCISE IT!" Do interpreters have better memories than the average person? Probably not better—just more buff.

Sweating to the Oldies? A Short-Term Memory Workout
Student interpreters often begin their studies with short-term memory workouts, called "lag exercises," which also teach them to listen and speak at the same time.

Record the following list of words into a tape recorder, or have a friend read them to you at a slow, steady pace. Leave a gap between one word and the next by reading one word every two seconds (approximately 30 words per minute).

tree	memory
car	tin
baby	ocean
tool	house
burp	computer
box	scratch
smooth	look
letter	lunch
pretty	pet
write	type
hello	table
lady	game
groove	bowl
tongue	dream
talk	breakfast

Play the tape, or have your friend start reading. Listen to the first word. When you hear the second word, cover it up by saying the first word. You'll be saying "tree" as you hear the word "car." Be careful not to speak in the gap between words—it's important to be speaking and listening at the same time. Student interpreters often practice this exercise in the same language until they can maintain a seven-word lag.

Capaldi, C. (2008, May/June). Silver-Tongued Masters of Memory. *Odyssey Adventures in Science*, pp. 30-33.

These texts appeared in the 2019–20 Smarter Balanced Grade 8 Sample Items published by The Regents of the University of California.

NAEP Basic

In this item, students are asked to determine the whether the textual evidence of each source supports the stated claims provided in the item, addressing the NAEP Basic level language of "make simple inferences and interpretations."

Comprehension Target: Integrate and Interpret; **Item Format:** SR-Grid; **Key:** See below the item.
Skill: Determine whether textual evidence supports claims.

Click on the boxes to show the claim(s) that each source supports. Some sources will have more than one box selected.

	Source #1: How Do We Remember?	Source #2: Memory Masters	Source #3: Interpreters: Silver-Tongued Masters of Memory
Find out how your memory systems process information.	☐	☐	☐
Learn how to improve your memory skills.	☐	☐	☐
Learn about the kinds of challenges presented at a memory competition.	☐	☐	☐

KEY:

Find out how your memory systems process information.: Source #1: How Do We Remember?, Source #3: Interpreters: Silver-Tongued Masters of Memory

Learn how to improve your memory skills.: Source #2: Memory Masters

Learn about the kinds of challenges presented at a memory competition.: Source #2: Memory Masters

This item appeared in the 2019–20 Grade 8 Smarter Balanced Sample Items published by The Regents of the University of California with Item ID 61235.

NAEP Proficient

In this item, students are asked to determine which source is most relevant to a specified topic and provide written justification with evidence in their response, addressing the NAEP Proficient level language of "form explanations and generalizations."

Comprehension Target: Analyze and Evaluate; **Item Format:** SCR; **Key:** N/A
Skill: Gather relevant information from multiple print and digital sources.

All of the sources provide information about memory. Which source would be **most** relevant to students researching ways to help remember information? Justify your answer and support it with details from the source.

This item appeared in the 2019–20 Grade 8 Smarter Balanced Sample Items published by The Regents of the University of California with Item ID 55409.

NAEP Advanced
In this item, students are tasked to write a multi-paragraph article using relevant information from the source texts, addressing the NAEP Advanced level language of "make complex inferences and to support their interpretations, conclusions, and judgments based upon evidence" and to "evaluate the relevance and strength of evidence to support an author's claims."

Comprehension Target: Use and Apply; **Item Format:** ECR; **Key:** N/A
Skill: Synthesize information from a range of sources into a coherent understanding of a process, phenomenon, or concept.

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 a brief explanatory article about improving memory.

Using more than one source, explain how to improve memory. Be sure to include information from the sources you choose to use, and to reference any quotations or paraphrasing of details or facts from the sources.

This item is adapted from a Smarter Balanced item. The original item appeared in the Grade 8 Smarter Balanced Sample Items published by The Regents of the University of California with Item ID 55074.

Grade 12

The items illustrating NAEP ALDs at the Basic, Proficient, and Advanced levels in the Grade 12 science disciplinary context are associated with the text "Blue Crabs Provide Evidence of Oil Tainting Gulf Food Web."

Blue Crabs Provide Evidence of Oil Tainting Gulf Food Web
by John Flesher, *Staff Writer*

Weeks ago, before engineers pumped in mud and cement to plug the gusher, scientists began finding specks of oil in crab larvae plucked from waters across the Gulf coast.

The government said last week that three-quarters of the spilled oil has been removed or naturally dissipated from the water. But the crab larvae discovery was an ominous sign that crude had already infiltrated the Gulf's vast food web—and could affect it for years to come.

"It would suggest the oil has reached a position where it can start moving up the food chain instead of just hanging in the water," said Bob Thomas, a biologist at Loyola University in New Orleans. "Something likely will eat those oiled larvae . . . and then that animal will be eaten by something bigger and so on."

Tiny creatures might take in such low amounts of oil that they could survive, Thomas said. But those at the top of the chain, such as dolphins and tuna, could get fatal "megadoses."

Marine biologists routinely gather shellfish for study. Since the spill began, many of the crab larvae collected have had the distinctive orange oil droplets, said Harriet Perry, a biologist with the University of Southern Mississippi's Gulf Coast Research Laboratory.

"In my 42 years of studying crabs I've never seen this," Perry said.

She wouldn't estimate how much of the crab larvae are contaminated overall, but said about 40 percent of the area they are known to inhabit has been affected by oil from the spill.

While fish can metabolize dispersant and oil, crabs may accumulate the hydrocarbons, which could harm their ability to reproduce, Perry said in an earlier interview with *Science* magazine.

She told the magazine there are two encouraging signs for the wild larvae—they are alive when collected and may lose oil droplets when they molt.

Tulane University researchers are investigating whether the splotches also contain toxic chemical dispersants that were spread to break up the oil but have reached no conclusions, biologist Caz Taylor said.

If large numbers of blue crab larvae are tainted, their population is virtually certain to take a hit over the next year and perhaps longer, scientists say. The spawning season occurs between April and October, but the peak months are in July and August.

How large the die-off would be is unclear, Perry said. An estimated 207 million gallons of oil have spewed into the Gulf since an April 20 drilling rig explosion triggered the spill, and thousands of gallons of dispersant chemicals have been dumped.

Scientists will be focusing on crabs because they're a "keystone species" that play a crucial role in the food web as both predator and prey, Perry said.

Richard Condrey, a Louisiana State University oceanographer, said the crabs are "a living repository of information on the health of the environment."

Named for the light-blue tint of their claws, the crabs have thick shells and 10 legs, allowing them to swim and scuttle across bottomlands. As adults, they live in the Gulf's bays and estuaries amid marshes that offer protection and abundant food, including snails, tiny shellfish, plants and even smaller crabs. In turn, they provide sustenance for a variety of wildlife, from redfish to raccoons and whooping cranes.

Adults could be harmed by direct contact with oil and from eating polluted food. But scientists are particularly worried about the vulnerable larvae.

That's because females don't lay their eggs in sheltered places, but in areas where estuaries meet the open sea. Condrey discovered several years ago that some even deposit offspring on shoals miles offshore in the Gulf.

The larvae grow as they drift with the currents back toward the estuaries for a month or longer. Many are eaten by predators, and only a handful of the 3 million or so eggs from a single female live to adulthood.

But their survival could drop even lower if the larvae run into oil and dispersants.

"Crabs are very abundant. I don't think we're looking at extinction or anything close to it," said Taylor, one of the researchers who discovered the orange spots.

Still, crabs and other estuary-dependent species such as shrimp and red snapper could feel the effects of remnants of the spill for years, Perry said.

"There could be some mortality, but how much is impossible to say at this point," said Vince Guillory, biologist manager with the Louisiana Department of Wildlife and Fisheries.

Perry, Taylor and Condrey will be among scientists monitoring crabs for negative effects such as population drop-offs and damage to reproductive capabilities and growth rates.

Crabs are big business in the region. In Louisiana alone, some 33 million pounds are harvested annually, generating nearly $300 million in economic activity, Guillory said.

Blue crabs are harvested year-round, but summer and early fall are peak months for harvesting, Guillory said.

Prices for live blue crab generally have gone up, partly because of the Louisiana catch scaling back due to fishing closures, said Steve Hedlund, editor of SeafoodSource.com, a website that covers the global seafood industry.

Fishers who can make a six-figure income off crabs in a good year now are now idled—and worried about the future.

"If they'd let us go out and fish today, we'd probably catch crabs," said Glen Despaux, 37, who sets his traps in Louisiana's Barataria Bay. "But what's going to happen next year, if this water is polluted and it's killing the eggs and the larvae? I think it's going to be a long-term problem."

This text appeared in the 2019–20 Smarter Balanced Grades 11–12 Sample Items published by The Regents of the University of California.

NAEP Basic

In this item, students are asked to determine which two pieces of textual evidence support an inference provided in the stem, addressing the NAEP Basic level language of "find information" and "make inferences and interpretations."

Comprehension Target: Locate and Recall; **Item Format:** SR-MSMC; **Key:** 1) She told the magazine there are two encouraging signs . . .; 2) "Crabs are very abundant. I don't think we're looking at extinction . . ."
Skill: Cite explicit text evidence to support inferences made or conclusions drawn from a text.

Select the **two** sentences from the text that **best** support the inference that blue crabs may be less impacted by the oil spill than some scientists predict.

- ☐ Tiny creatures might take in such low amounts of oil that they could survive, Thomas said.
- ☐ "In my 42 years of studying crabs I've never seen this," Perry said.
- ☐ She told the magazine there are two encouraging signs for the wild larvae—they are alive when collected and may lose oil droplets when they molt.
- ☐ "Crabs are very abundant. I don't think we're looking at extinction or anything close to it," said Taylor, one of the researchers who discovered the orange spots.
- ☐ Still, crabs and other estuary-dependent species such as shrimp and red snapper could feel the effects of remnants of the spill for years, Perry said.

This item appeared in the 2019–20 Grades 11–12 Smarter Balanced Sample Items published by The Regents of the University of California with Item ID 183102.

NAEP Proficient

In this item, students are asked to determine an author's point of view based upon the author's inclusion of conflicting information in the text, addressing the NAEP Proficient level language of "use text-based evidence to support arguments and conclusions."

Comprehension Target: Analyze and Evaluate; **Item Format:** SR-SSMC; **Key:** A
Skill: Determine an author's point of view or purpose in a text by analyzing style and content.

What does the conflicting information about the effects of oil on blue crab larvae reveal about the author's point of view?

Ⓐ It reinforces the author's belief that scientists do not yet know how the oil will affect the blue crab population.

Ⓑ It suggests that the author disagrees with scientists who predict long-term damage to the blue crab population.

Ⓒ It reinforces the author's feeling that scientists may never know the true effects of oil on the blue crab population.

Ⓓ It suggests that the author feels scientists have not devoted enough attention to the effects of oil on blue crab larvae.

This item appeared in the 2019–20 Grades 11–12 Smarter Balanced Sample Items published by The Regents of the University of California with Item ID 183143.

NAEP Advanced

In this item, students are asked to write a short response based upon an analysis of the evidence an author uses to support claims in a text, addressing the NAEP Advanced level language of "make complex inferences and to support their interpretations, conclusions, and their judgments based upon evidence."

Comprehension Target: Analyze and Evaluate; **Item Format:** SCR; **Key:** N/A
Skill: Delineate and evaluate the argument, claims, and reasoning in a text.

What inference can be made about the evidence the author uses to support claims in the text? Support your answer with evidence from the text.

This item appeared in the 2019–20 Grades 11–12 Smarter Balanced Sample Items published by The Regents of the University of California with Item ID 183109.

Disciplinary Context: Social Studies

Grade 4

The items illustrating NAEP ALDs at the Basic, Proficient, and Advanced levels in the Grade 4 social studies disciplinary context are associated with the text "Marian's Revolution."

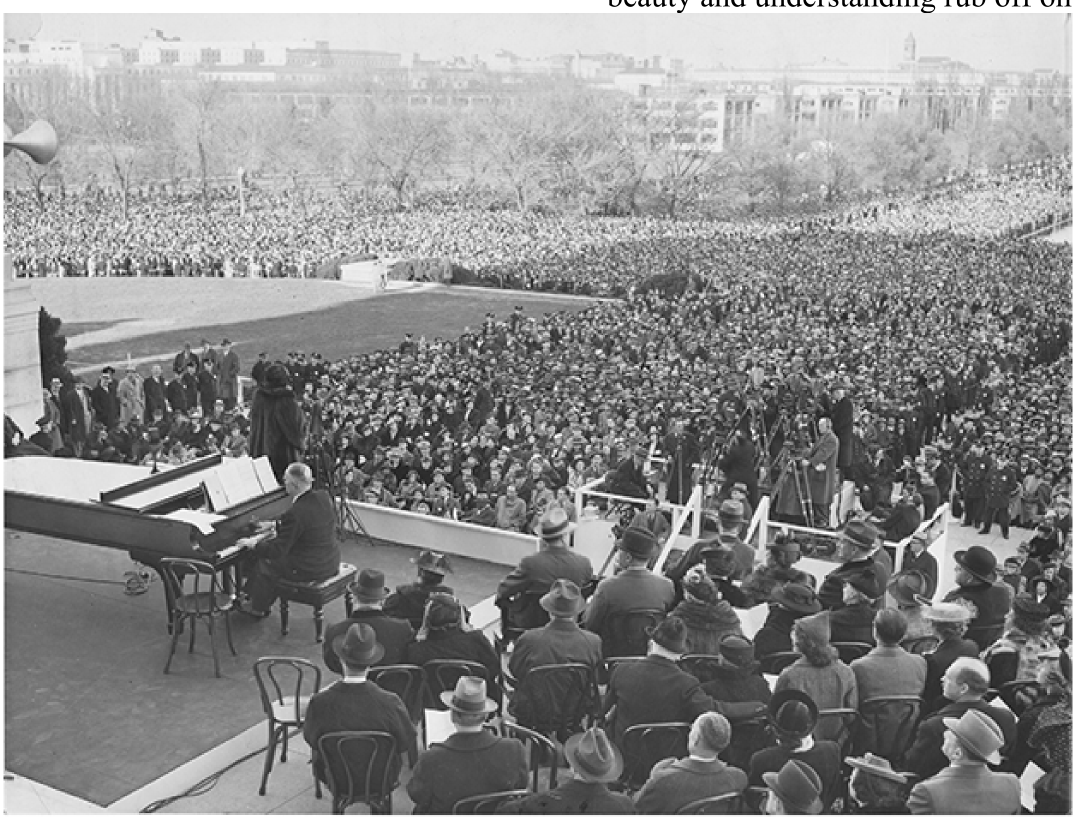
Marian's Revolution

by Sudipta Bardhan-Quallen

By 1939, Marian Anderson had performed for presidents and kings. She had been praised for having "a voice ... one hears once in a hundred years." Despite her success, when Marian wanted to sing at Constitution Hall that year, she was banned from doing so. The owner of the hall, an organization called the Daughters of the American Revolution (DAR), felt that Marian couldn't be allowed to sing there because she was African American.

Chosen by Music

That wasn't the first time Marian had been turned away because she was black. When she was 18 years old, she applied to music school. The clerk at the desk rudely sent her home because of her race. Marian was shocked by the clerk's words. "I could not conceive of a person," Marian said, "surrounded as she was with the joy that is music without having some sense of its beauty and understanding rub off on her."



Marian Anderson sings to a crowd of 75,000 people at the Lincoln Memorial on April 10, 1939.

Page 2

“I don’t think I had much
to say in choosing it.
I think music chose me.”
—Marian Anderson

Because of segregation—the practice of keeping blacks and whites separate—the early 1900s were a difficult time for a young black woman to begin a professional singing career. But Marian was determined to sing. "It was something that just had to be done," she remembered. "I don't think I had much to say in choosing it. I think music chose me."

In 1925, Marian won a voice contest in New York, and sang with the New York Philharmonic. Still, her chances to perform in the United States were limited. To build her career, Marian traveled to Europe in 1928, where she became very successful.

A World-Class Singer Faces Racism

By 1939, Marian was a world-class singer. She returned to the United States to continue her career. But back at home, she faced racism in many ways. Segregation was still common on trains and in hotels and restaurants. No amount of vocal talent could spare Marian from that.

Even concert halls were segregated, although usually that was limited to the audience. Because black performers often appeared on stage in segregated halls, Marian had no reason to think she would be turned away from Constitution Hall. She believed that musical skill would be the only factor that the DAR would consider.

At first, the DAR told Marian that the date she requested was not available. Then they told her that all of her alternate dates were booked. Eventually, the DAR upheld their policy that only white performers could appear in Constitution Hall.

A Voice for Civil Rights

When news of the DAR's policy got out, many people were outraged. First Lady Eleanor Roosevelt resigned from the DAR. In a letter, she wrote: "I am in complete disagreement with the attitude taken in refusing Constitution Hall to a great artist .... You had an opportunity to lead in an enlightened way, and it seems to me your organization has failed."

Marian believed strongly in the civil rights movement. She knew firsthand the pain that racism caused. She understood that the way the controversy with the DAR was resolved would be a milestone for civil rights.

Despite public outcry, the DAR would not back down and let Marian sing. With Mrs. Roosevelt's support, the Secretary of the Interior arranged a special concert for Marian, to be held at the Lincoln Memorial. Seventy-five thousand people attended. In many ways, Marian's concert was considered to be America's first civil rights rally. That night, she took a stand against discrimination and for equality. The first words she sang were: "My country, 'tis of thee, sweet land of liberty, of thee I sing."

The Open-Hearted Way

Marian realized that equality in the United States would be achieved when every person was willing to stand up for what is right. As a public figure, she felt a responsibility to set an example. After the 1939 incident, she did her part by turning down concerts for segregated audiences.

"The minute a person whose word means a great deal dares to take the openhearted and courageous way," she said, 'many others follow."

As Marian's career progressed, America changed. She performed in many prestigious locations, including Constitution Hall, where she sang after the DAR changed its policies. By 1954, segregation was declared unconstitutional. The Civil Rights Act was signed into law in 1964, the year Marian retired from performing. By then, many of the barriers she'd had to fight through were disappearing. Marian's farewell tour began in front of an admiring crowd at Constitution Hall.

Eleanor Roosevelt honors singer Marian Anderson.

Photo credits for "Marian Anderson": Marian Anderson Collection, Rare Book and Manuscript Library, University of Pennsylvania, Philadelphia, PA.

This text appeared in the 2011 NAEP grade 4 Reading administration in Block 4R10.

NAEP Basic
In this item, students are asked to identify the main idea of an article, addressing the NAEP Basic level language of "make simple inferences and interpretations."

Comprehension Target: Integrate and Interpret; **Item Format:** SR-SSMC; **Key:** D
Skill: Identify the main idea of an article.

What is the article mainly about?

A. Civil rights songs that Marian Anderson liked to sing

B. Marian Anderson's friendship with Eleanor Roosevelt

C. How Marian Anderson learned to sing

D. How segregation affected Marian Anderson's career

This item appeared in the 2011 NAEP grade 4 Reading administration with NAEP Item ID 2011-4R10 #1.

NAEP Proficient
In this item, students are asked to determine the meaning of a vocabulary word via surrounding context, addressing the NAEP Proficient level language of "make more complex inferences and interpretations."

Comprehension Target: Integrate and Interpret; **Item Format:** SR-SSMC; **Key:** B
Skill: Show understanding of vocabulary, figurative language, word relationships, and nuances in word meanings (e.g., shades of meaning).

On page 4, the article says that Marian Anderson performed in many **prestigious** locations. This means that she sang in places that were

A. far away from each other

B. famous and important

C. open to people of all races

D. large and crowded

This item appeared in the 2011 NAEP grade 4 Reading administration with NAEP Item ID 2011-4R10 #9.

NAEP Advanced

In this item, students are asked to explain how a key detail supports the main idea of an article, addressing the NAEP Advanced level language of "make complex inferences and to support their interpretations, conclusions, and judgments based upon evidence."

Comprehension Target: Analyze and Evaluate; **Item Format:** SCR; **Key:** N/A
Skill: Analyze how key details support the main idea.

Why do you think Marian Anderson began her concert by singing the words, "My country, 'tis of thee, sweet land of liberty, of thee I sing"? Use information from the article to support your answer.

This item appeared in the 2011 NAEP grade 4 Reading administration with NAEP Item ID 2011-4R10 #8.

Grade 8

The items illustrating NAEP ALDs at the Basic, Proficient, and Advanced levels in the Grade 8 social studies disciplinary context are associated with the text "1920: Women Get the Vote."

1920: Women Get the Vote

by Sam Roberts

The 19th Amendment was ratified in 1920, after decades of campaigning by the women's suffrage movement.

When John Adams and his fellow patriots were mulling independence from England in the spring of 1776, Abigail Adams famously urged her husband to "remember the ladies and be more generous and favorable to them than your ancestors." Otherwise, she warned, "we are determined to foment a rebellion, and will not hold ourselves bound by any laws in which we have no voice or representation."

That summer, the Declaration of Independence proclaimed that all men are created equal but said nothing of women's equality. It would take another 144 years before the U.S. Constitution was amended, giving women the right to vote in every state.

That 19th Amendment says simply: "The right of citizens of the United States to vote shall not be denied or abridged by the United States or by any State on account of sex." It took effect after a dramatic ratification battle in Tennessee in which a 24-year-old legislator cast the deciding vote.

The amendment was a long time coming. At various times, women could run for public office in some places, but

More than 20,000 marchers took part in this 1915 parade in New York City in support of women's suffrage.

Courtesy of Library of Congress #LC-USZ62-50393

Page 2

ELIZABETH CADY STANTON
Courtesy of Library of Congress
#LC-USZ62-28195

SUSAN B. ANTHONY
Courtesy of Library of Congress
#LC-USZ62-111423

could rarely vote. (As far back as 1776, New Jersey allowed women property owners to vote, but rescinded that right three decades later.)

"WOMANIFESTO"

The campaign for women's rights began in earnest in 1848 at a Women's Rights convention in Seneca Falls, N.Y., organized by 32-year-old Elizabeth Cady Stanton and other advocates. Stanton had drafted a "Womanifesto" patterned on the Declaration of Independence, but the one resolution that shocked even some of her supporters was a demand for equal voting rights, also known as universal suffrage. "I saw clearly," Stanton later recalled, "that the power to make the laws was the right through which all other rights could be secured."

Stanton was joined in her campaign by Susan B. Anthony, Sojourner Truth, Lucretia Mott, and other crusaders who would become icons of the women's movement. Some were militant. Many were met with verbal abuse and even violence. Already active in the antislavery movement and temperance campaigns (which urged abstinence from alcohol), women often enlisted in the fight for voting rights too.

WYOMING IS FIRST

They staged demonstrations, engaged in civil disobedience, began legal challenges, and pressed their case state by state. In 1869, the Wyoming Territory gave women the vote, with the first permanent suffrage law in the nation. ("It made sense that a place like Wyoming would embrace women's rights," Gail Collins of *The New York Times* wrote in her book *America's Women*. "With very few women around, there was no danger that they could impose their will on the male majority.")

In 1878, a constitutional amendment was introduced in Congress. The legislation languished for nine years. In 1887, the full Senate considered the amendment for the first time and defeated it by about 2-to-1.

But the suffrage movement was slowly gaining support. With more and more women graduating from high school, going to college, and working outside the home, many Americans began asking: Why couldn't women vote too?

Plenty of opposition existed, according to Collins: Democrats feared women would vote for more socially progressive Republicans. The liquor industry, afraid of prohibition, also opposed women's suffrage, as did many people in the South, where blacks had been largely disenfranchised since Reconstruction.

In 1918, after much cajoling and picketing by suffragists, President Woodrow Wilson changed his mind and backed the amendment. The next year, both houses of Congress voted to amend the Constitution. Suffrage advocates predicted quick ratification by the states. (By 1919, 28 states permitted women to vote, at least for President.) Within a little more than a year, 35 of the required 36 states had voted for ratification.

The last stand for anti-suffragists was in Tennessee in the summer of 1920. Their showdown in the State Legislature became known as the "War of the Roses." (Pro-amendment forces sported yellow roses; the antis wore red.)

After two roll calls, the vote was still tied, 48-48. On the third, Harry T. Burn, a Republican and, at 24, the youngest member of the legislature. switched sides. He was wearing a red rose but voted for ratification because he had received a letter from his mother that read, in part: "Hurrah and vote for suffrage! Don't keep them in doubt!"

Burn said later: "I know that a mother's advice is always safest for her boy to follow and my mother wanted me to vote for ratification. I appreciated the fact that an opportunity such as seldom comes to mortal man-to free 17,000,000 women from political slavery-was mine."

GRADUAL CHANGE

In 1920, women across America had the right to vote in a presidential election. (In the South, black women and men would be kept off voter rolls in large numbers until 1965, after passage of the Voting Rights Act.)

But newly enfranchised women voted in much smaller numbers than men. "Women who were adults at that time had been socialized to believe that voting was socially inappropriate for women," says Susan J. Carroll, senior scholar at the Center for American Women and Politics.

The political and social change sought by suffragists came gradually and not without fits and starts. An Equal Rights Amendment, stipulating equal treatment of the sexes under the law, was passed by Congress and sent to the states in 1972, but later failed after being ratified by only 35 of the necessary 38 states.

In 1980, however, women surpassed men for the first time in turnout for a presidential election. Since then, there has also been a substantial rise in the number of women running for and holding political office.

VC176436

From *THE NEW YORK TIMES UPFRONT* magazine, September 5, 2005 issue.

Page 4

This text appeared in the 2011 NAEP grade 12 Reading administration in Block 8R11.

NAEP Basic
In this item, students are asked to determine the meaning of a word based on context, addressing the NAEP Basic level language of "show understanding of vocabulary."

Comprehension Target: Integrate and Interpret; **Item Format:** SR-SSMC; **Key:** C
Skill: Show understanding of general academic and domain-specific vocabulary.

On page 3, the article says that Elizabeth Cady Stanton and Susan B. Anthony would beomce **icons** of the women's movement. This means that the two women would

A. beome religious leaders

B. be pictured on the "Womanifesto" document

C. become important symbols of the movement

D. be ready to sacrifice everything for the movement

This item appeared in the 2011 NAEP grade 8 Reading administration with NAEP Item ID 2011-8R11 #5.

NAEP Proficient
In this item, students are asked to explain the how events described in the text affect the central idea of the text, addressing the NAEP Proficient level language of "make more complex inferences and interpretations."

Comprehension Target: Integrate and Interpret; **Item Format:** SCR; **Key:** NA
Skill: Explain how a text makes connections between individuals, ideas, and/or events.

The section "Wyoming Is First" describes changes in United States society in the late 1800s and early 1900s. Choose one of these changes and explain its effect on women's progress in getting the vote.

This item appeared in the 2011 NAEP grade 8 Reading administration with NAEP Item ID 2011-8R11 #10.

NAEP Advanced
In this item, students are asked to XXX, addressing the NAEP Advanced level language of "make complex inferences and to support their interpretations, conclusions, and their judgments based upon evidence."

Comprehension Target: Analyze and Evaluate; **Item Format:** ECR; **Key:** NA
Skill: Analyze how word choice impacts a text's meaning and tone.

In describing the women's suffrage movement, the author uses such words as "battle," "militant," and "showdown." Do you think this is an effective way to describe the women's suffrage movement? Support your answer with two references to the article.

This item appeared in the 2011 NAEP grade 8 Reading administration with NAEP Item ID 2011-8R11 #7.

Grade 12

The items illustrating NAEP ALDs at the Basic, Proficient, and Advanced levels in the Grade 12 social studies disciplinary context are associated with the text "Inaugural Address of Theodore Roosevelt."

Theodore Roosevelt

United States President (1901-1909)

Inaugural Address

Saturday, March 4, 1905

The energetic Republican President had taken his first oath of office upon the death of President McKinley, who died of an assassin's gunshot wounds on September 14, 1901. Mr. Roosevelt had been President himself for three years at the election of 1904. The inaugural celebration was the largest and most diverse of any in memory—cowboys, American Indians (including the Apache Chief Geronimo), coal miners, soldiers, and students were some of the groups represented. The oath of office was administered on the East Portico of the Capitol by Chief Justice Melville Fuller.


Theodore Roosevelt

MY FELLOW CITIZENS, no people on Earth have more cause to be thankful than ours, and this is said reverently, in no spirit of boastfulness in our own strength, but with gratitude to the Giver of Good who has blessed us with the conditions which have enabled us to achieve so large a measure of well-being and of happiness.

To us as a people it has been granted to lay the foundations of our national life in a new continent. We are the heirs of the ages, and yet we have had to pay few of the penalties which in old countries are exacted by the dead hand of a bygone civilization.... Our life has called for the vigor and effort without which the manlier and hardier virtues wither away.

Under such conditions it would be our own fault if we failed; and the success which we have had in the past, the success which we confidently believe the future will bring, should cause in us no feeling of vainglory, but rather a deep and abiding realization of all which life has offered us; a full acknowledgment of the responsibility which is ours; and a fixed determination to show that under a free government a mighty people can thrive best, alike as regards the things of the body and the things of the soul.

Much has been given us, and much will rightfully be expected from us. We have duties to others and duties to ourselves; and we can shirk neither. We have become a great nation, forced by the fact of its greatness into relations with the other nations of the Earth, and we must behave as beseems a people with such responsibilities.

Toward all other nations, large and small, our attitude must be one of cordial and sincere friendship. We must show not only in our words, but in our deeds, that we are earnestly desirous of securing their goodwill by acting toward them in a spirit of just and generous recognition of all their rights.

But justice and generosity in a nation, as in an individual, count most when shown not by the weak but by the strong. While ever careful to refrain from wrongdoing others, we must be no less insistent that we are not wronged ourselves. We wish peace, but we wish the peace of justice, the peace of righteousness. We wish it because we think it is right and not because we are afraid. No weak nation that acts manfully and justly should ever have cause to fear us, and no strong power should ever be able to single us out as a subject for insolent aggression.

Our relations with the other powers of the world are important; but still more important are our relations among ourselves. Such growth in wealth, in population, and in power as this nation has seen during the century and a quarter of its national life is inevitably accompanied by a like growth in the problems which are ever before every nation that rises to greatness. Power invariably means both responsibility and danger. Our forefathers faced certain perils which we have outgrown. We now face other perils, the very existence of which it was impossible that they should foresee.

Modern life is both complex and intense, and the tremendous changes wrought by the extraordinary industrial development of the last half century are felt in every fiber of our social and political being. Never before have men tried so vast and formidable an experiment as that of administering the affairs of a continent under the forms of a democratic republic. The conditions which have told for our marvelous material well-being—which have developed to a very high degree our energy, self-reliance, and individual initiative—have also brought the care and anxiety inseparable from the accumulation of great wealth in industrial centers.

To do so we must show, not merely in great crises, but in the everyday affairs of life, the qualities of practical intelligence, of courage, of hardihood, and endurance, and above all the power of devotion to a lofty ideal, which made great the men who founded this republic in the days of Washington, which made great the men who preserved this republic in the days of Abraham Lincoln.
Text courtesy of Bartleby Library.
Photograph courtesy of Library of Congress #LC-D429-29129.

This text appeared in the 2013 NAEP grade 12 Reading administration in Block 12R11.

NAEP Basic
In this item, students are asked to evaluate why the author uses references to two historical figures in a speech, addressing the NAEP Basic level language of "make inferences and interpretations."

Comprehension Target: Analyze and Evaluate; **Item Format:** SR - SSMC; **Key:** C
Skill: Evaluate author's technique.

Roosevelt most likely refers to Washington and Lincoln at the end of the address in order to

A. praise the speaking styles of previous presidents

B. encourage listeners to study history

C. recall accomplishments from the past

D. suggest that government was more powerful in the past

This item appeared in the 2013 NAEP grade 12 Reading administration with NAEP Item ID 2013-12R11 #10.

NAEP Proficient
In this item, students are asked to determine the relationship between two ideas over the course of the text, addressing the NAEP Proficient level language of "make more complex inferences and interpretations."

Comprehension Target: Integrate and Interpret; **Item Format:** MC; **Key:** C
Skill: Explain how specific individuals, ideas, and/or events interact and develop over the course of a text.

Which of the following best describes Roosevelt's ideas about the relationship between progress and problems?

A. He believes that in the future progress will not lead to problems.

B. He believes progress solves most problems once thought unsolvable.

C. He believes a nation cannot have progress without also having problems.

D. He believes progress can solve only certain types of problems.

This item appeared in the 2013 NAEP grade 12 Reading administration with NAEP Item ID 2013-12R11 #8.

NAEP Advanced

In this item, students are asked to delineate the author's argument and explain why the author makes an argument, addressing the NAEP Advanced level language of "make complex inferences and to support their interpretations, conclusions, and their judgments based upon evidence."

Comprehension Target: Integrate and Interpret; **Item Format:** ECR; **Key:** N/A
Skill: Delineate and evaluate the argument, claims, and reasoning in a text.

Roosevelt emphasizes "responsibility" and "duty" throughout his address. According to Roosevelt, why should the nation take responsibility? What are two responsibilities or duties that Roosevelt believed were important?

This item appeared in the 2013 NAEP grade 12 Reading administration with NAEP Item ID 2013-12R11 #6.

This appendix is provided to describe design considerations, based on the principles outlined in the Framework, that assessment developers might weigh as they develop blocks. (This appendix is also included as Appendix C in the Framework.) 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 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 C.1 for an illustration of how these areas relate to one another.

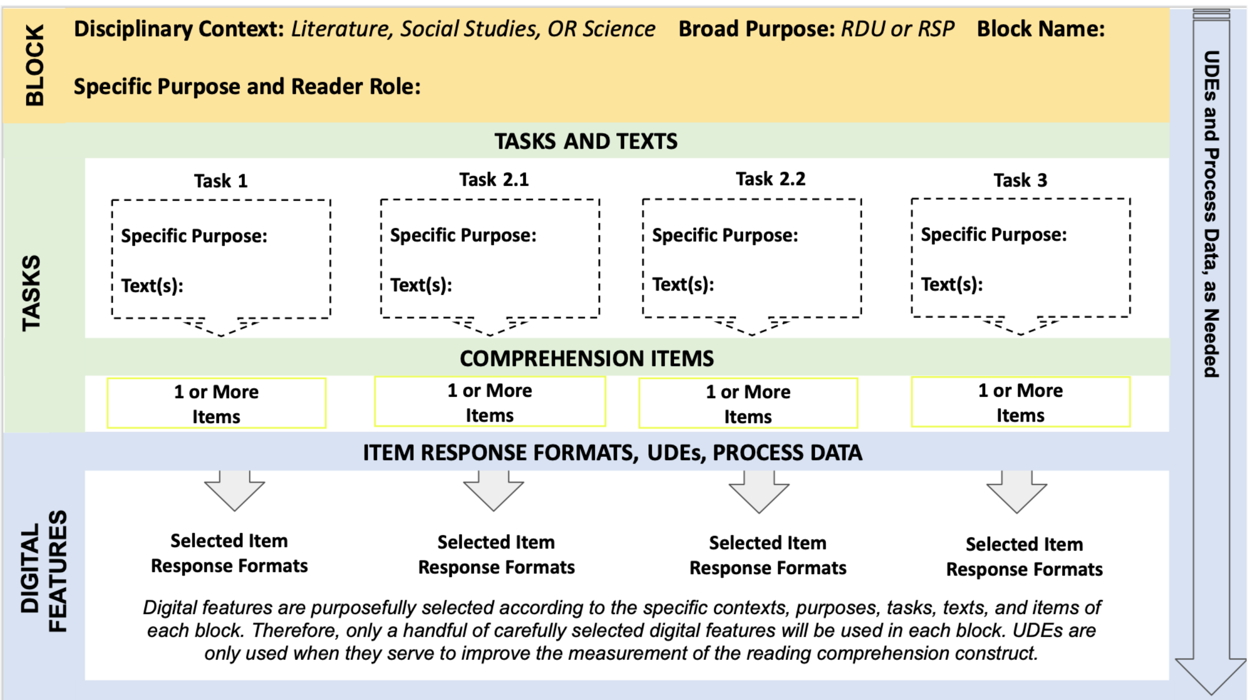
It is important to note that developers do not necessarily make decisions about these three areas in this order; rather, some of these decisions might be iterative and mutually informative. For example, in developing a literature block for a certain grade level, the developer might first choose a text 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 C.1. Third, design features such as item formats, UDEs, and process data are used to leverage the digital assessment environment to measure 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 C.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 C.1. This exhibit provides one sample approach to an assessment block; other approaches are possible that would have variations in the components (e.g., the number of tasks and texts). 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 C.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 C.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 C.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 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 C.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 with less probability of	↔	Purpose-driven tasks are situated in line with disciplinary context but tasks are more tightly structured so that one task builds on the previous; more probability that tasks are

	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.		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 lesser need for task-based UDEs.	⟺	More cumulative reading purposes that may require UDEs for knowledge or motivation and potentially greater 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 block-specific purpose and reader role need to be made apparent at the outset of a block.
    b. Students should be reminded of purpose and role as appropriate 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 (Sabatini, O'Reilly, Weeks & Wang, 2019) 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.

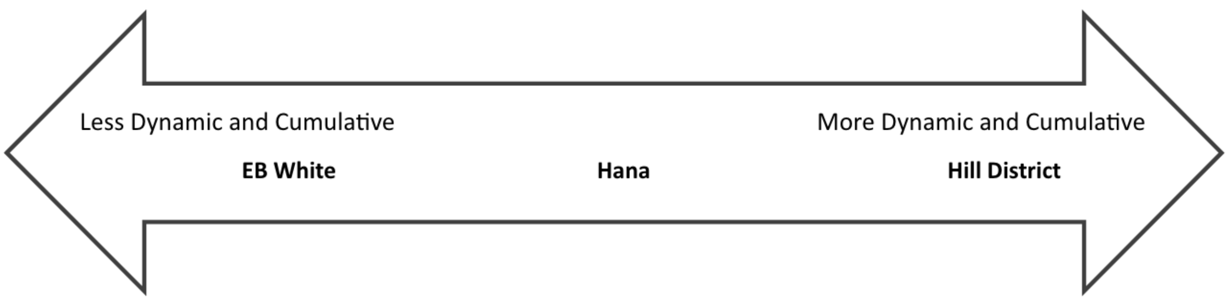
Block Sketches

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 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). 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). And the third (labeled *E. B. 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. Referring to the underlying continuum of variation for assessment components within blocks as detailed in Exhibit C.2 above, these three block sketches are situated on three hypothetical points along that continuum, as illustrated in Exhibit C.3.

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



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 read and interpret story excerpts from the short story, Hana Hashimoto, by Chieri Uegaki in preparation for a book discussion with three peers. First, students are asked to read to develop an understanding of the characters, key events, and author's craft. Second, they apply their insights to describe what Hana is like as a person. so that they are ready to contribute to the discussion.

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 RSP block with texts situated in a social studies context. In this block, students engage in more 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 multiple texts to solve a problem.

E. B. 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.

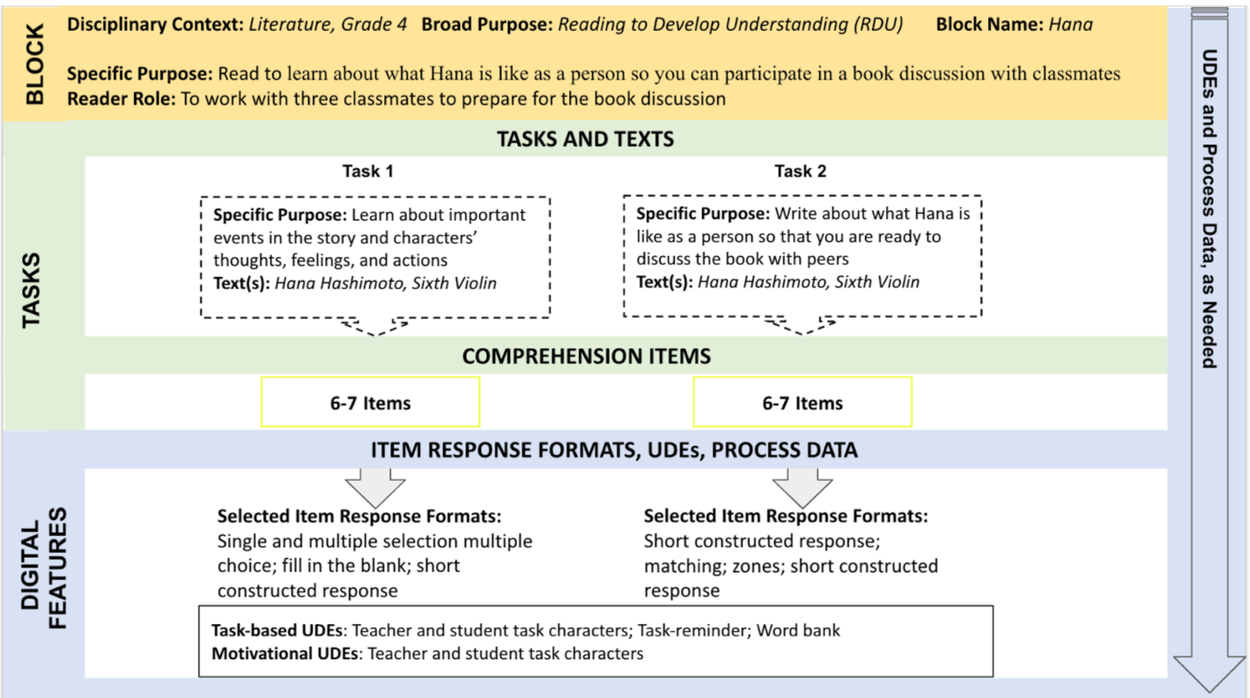
When viewing these examples, it is important to keep in mind the following points:

- 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 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 A.7), 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 *E. B. 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 *E. B. 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 C.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 C.4. Block Design for Hana

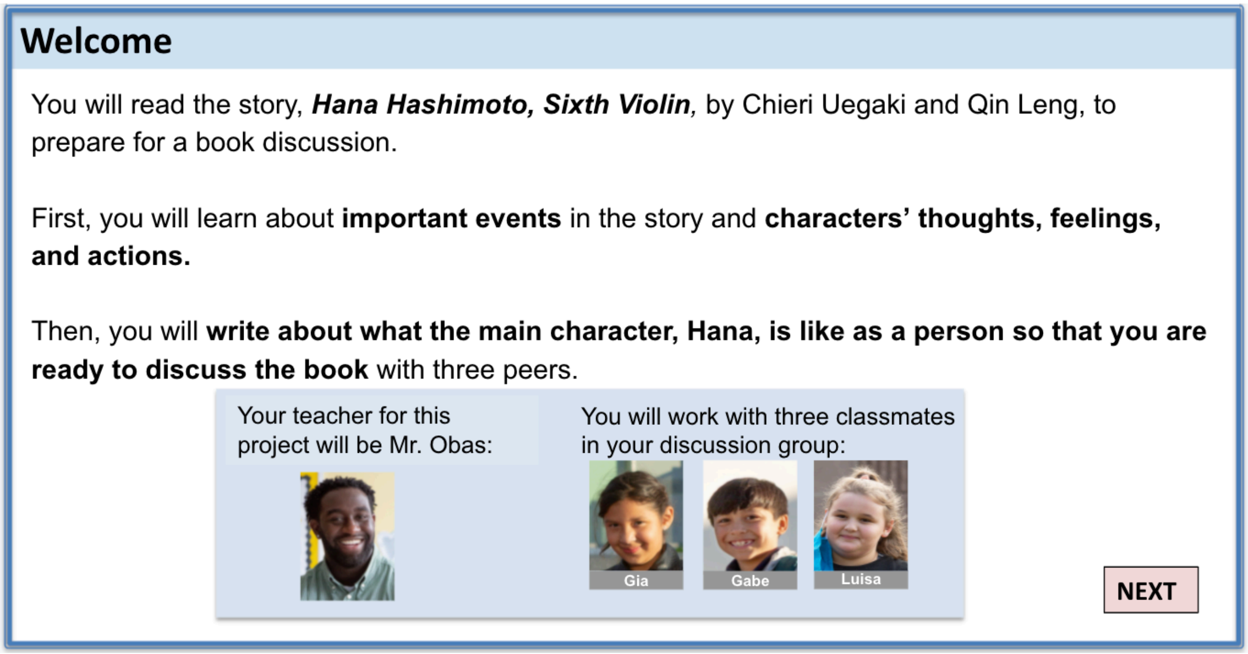


Block Components (Disciplinary Context, Purposes, and Reader Role). This block is designed to assess how Grade 4 readers develop understanding within a single, print text in a literature context. In this block, readers identify important events in the story and analyze how characters' thoughts, feelings, and actions describe the kind of people they are. Then, readers use and apply what they have learned to form an overall interpretation of the main character, Hana. They choose a character trait from a word bank and then 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 C.5), readers are 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 with three fourth grade classmates (represented in the assessment by task characters Gia, Gabe, and Luisa). They are also introduced to their teacher for the project (represented by the task character Mr. Obas).

Then, a **task-based UDE** in the form of two statements informs students what tasks will be expected of them. Here, students are told that, to prepare for the book discussion, they will read the story and 1) learn about important events in the story and characters' thoughts, feelings, and actions; and, 2) use what they have learned about Hana to describe what she is like as a person. **Motivational UDEs** (here, student and teacher avatars) serve to motivate readers to engage with the block.

Exhibit C.5. Specific purpose, reader role, and 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



Throughout Appendix C, the photograph of Mr. Obas is sourced from https://images.all4ed.org/male-sixth-grade-math-teacher-with-protractor (photographer Allison Shelley for EDUimages). The photograph of Gia is sourced from https://images.all4ed.org/elementary-boy-with-backpack-and-girl-with-notebook/ (photographer Allison Shelley for EDUimages). The photograph of Gabe is sourced from https://images.all4ed.org/third-grade-boy-with-backpack-outside/. The photograph of Luisa is sourced from https://images.all4ed.org/fifth-grade-girl-mask-break (photographer Allison Shelley for EDUimages).

Task Components (Tasks, Text(s), and Items).

Tasks*.* 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 C.6). 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 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 C.6).

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 C.6-C.11 illustrate items that help students accomplish the first task of learning about the events and characters. Exhibits C.12-C.14 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.

Item response types 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 might, for example, include single-selection multiple choice items to assess readers’ ability to locate and recall important events and other details (see Exhibit C.6), short constructed-response items that include fill in the blank options (see Exhibit C.7), multiple-selection multiple choice items (see Exhibit C.8), and longer short constructed response items that ask readers to interpret and integrate details about the character’s thoughts, feelings, and actions into their understanding of the story (see Exhibit C.10).

Exhibit C.6. A Grade 4 RDU block illustrating a Locate and Recall multiple choice item. The teacher reminds the reader of the specific purpose (to prepare for a discussion) and the first task (to learn about events and characters)

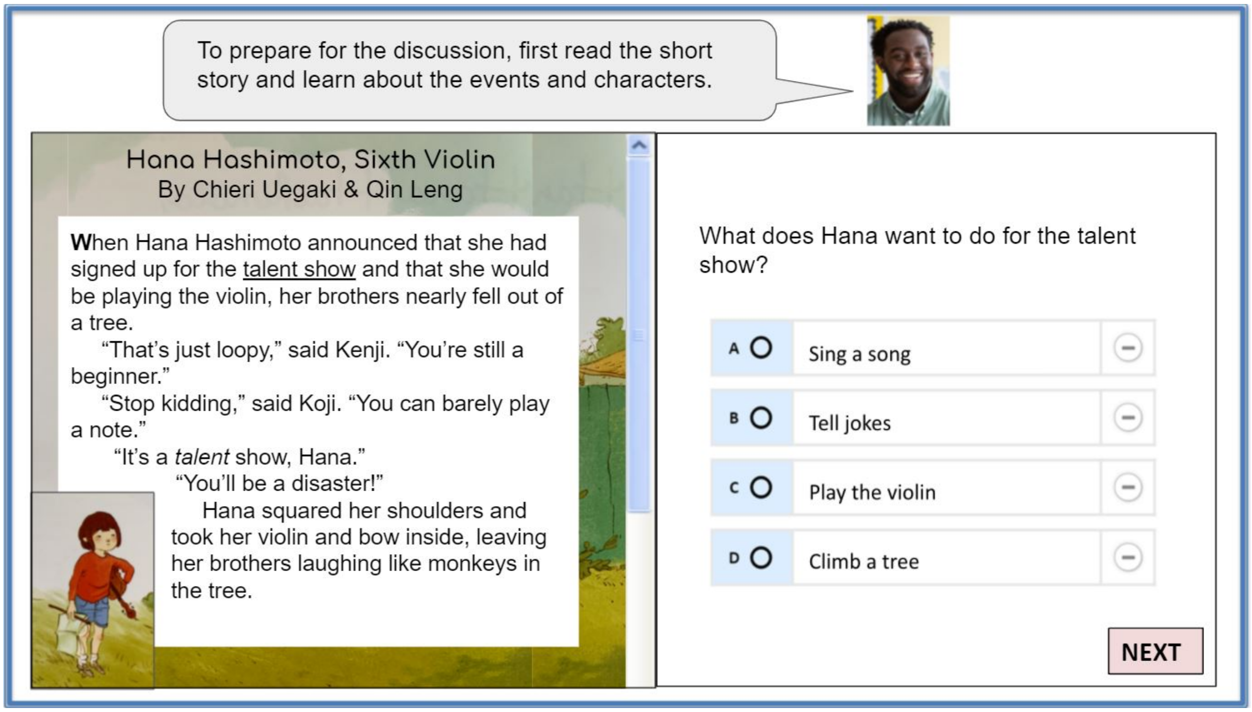


Exhibit C.7. A Grade 4 Locate and Recall item illustrating a fill in the blank short constructed response item

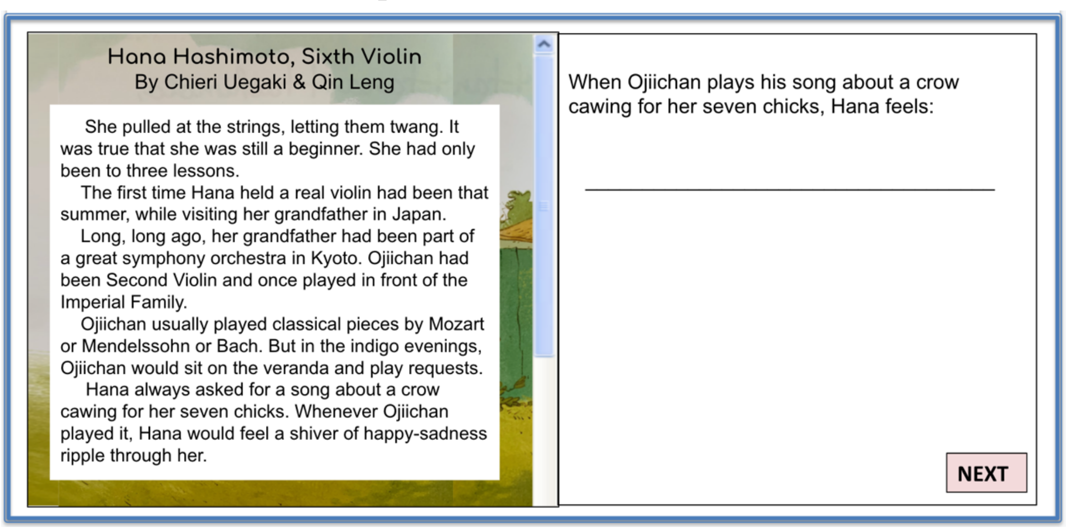
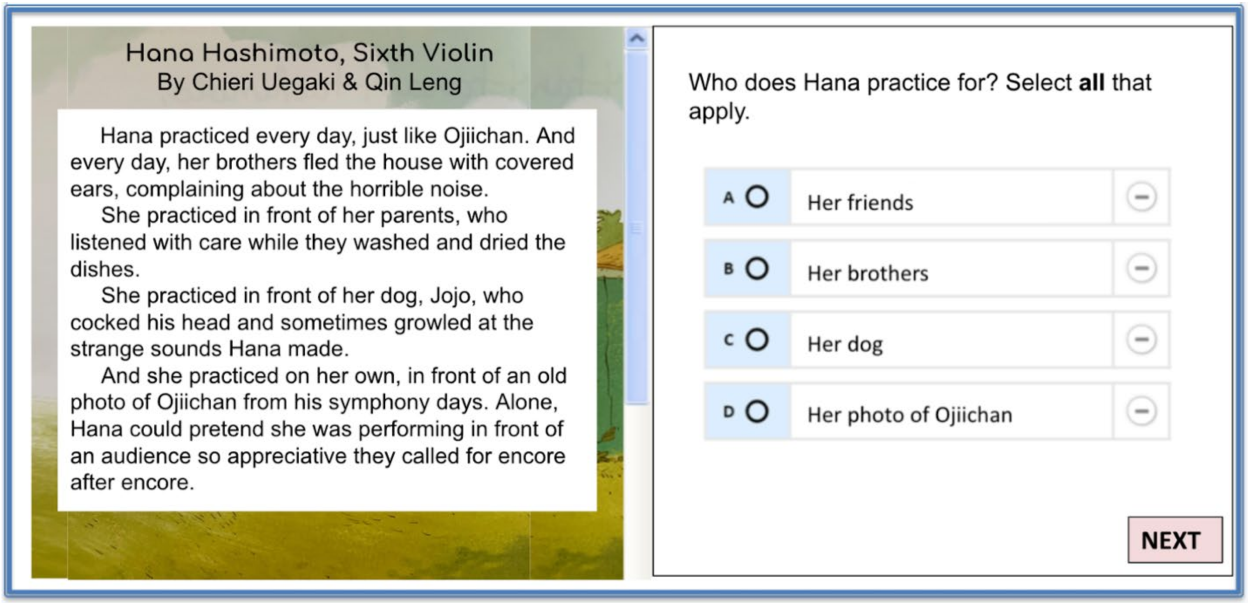
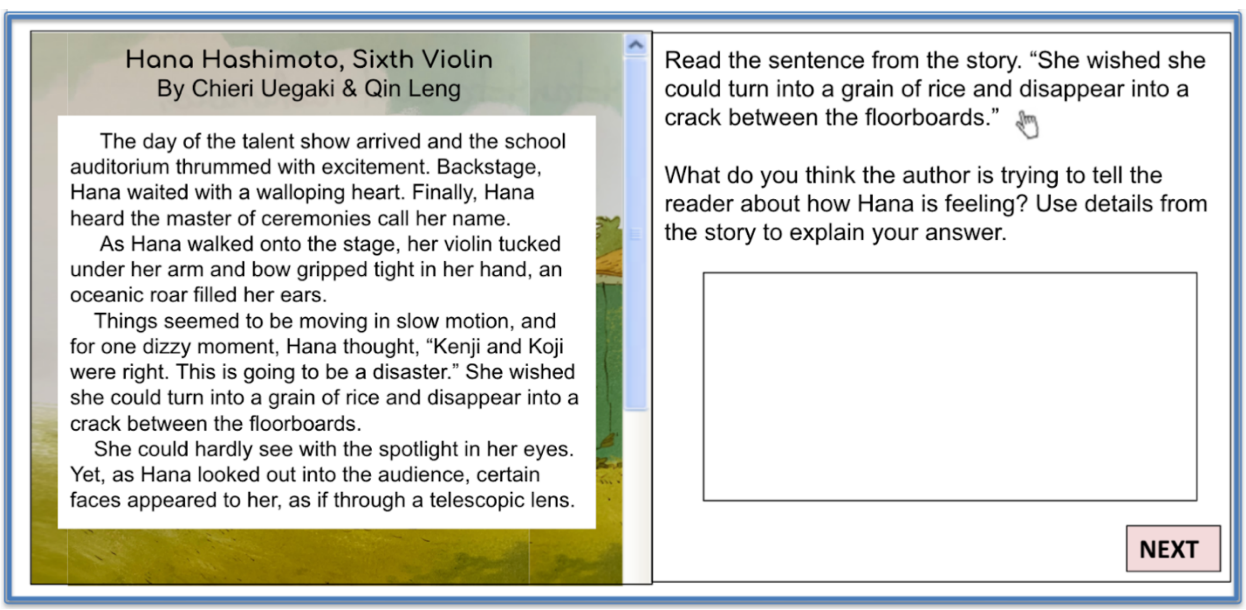


Exhibit C.8. A Grade 4 Locate and Recall item illustrating a multiple-selection multiple choice response format



In addition, **a look-back button (a task-based UDE)** is embedded into items with excerpted text (see Exhibits C.9 and C.10). If readers wish, they can click to see exactly where the excerpted text is located in the context of the original story in the assessment space. Multiple choice and constructed response item formats are interspersed throughout the assessment.

Exhibit C.9. A Grade 4 Analyze and Evaluate short constructed-response item illustrating a task-based UDE in the form of a look-back button that refers readers to the relevant section of text



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. Her family loves her performance so much that later that evening, they ask her to play them more musical notes around the dinner table.

Exhibit C.10. 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

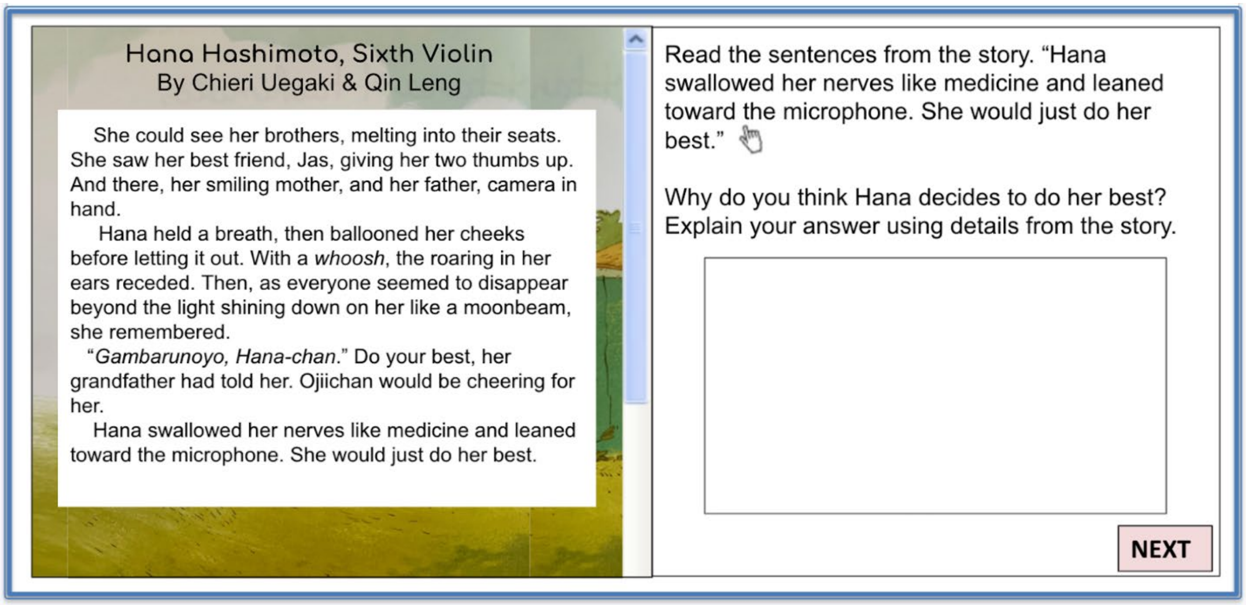
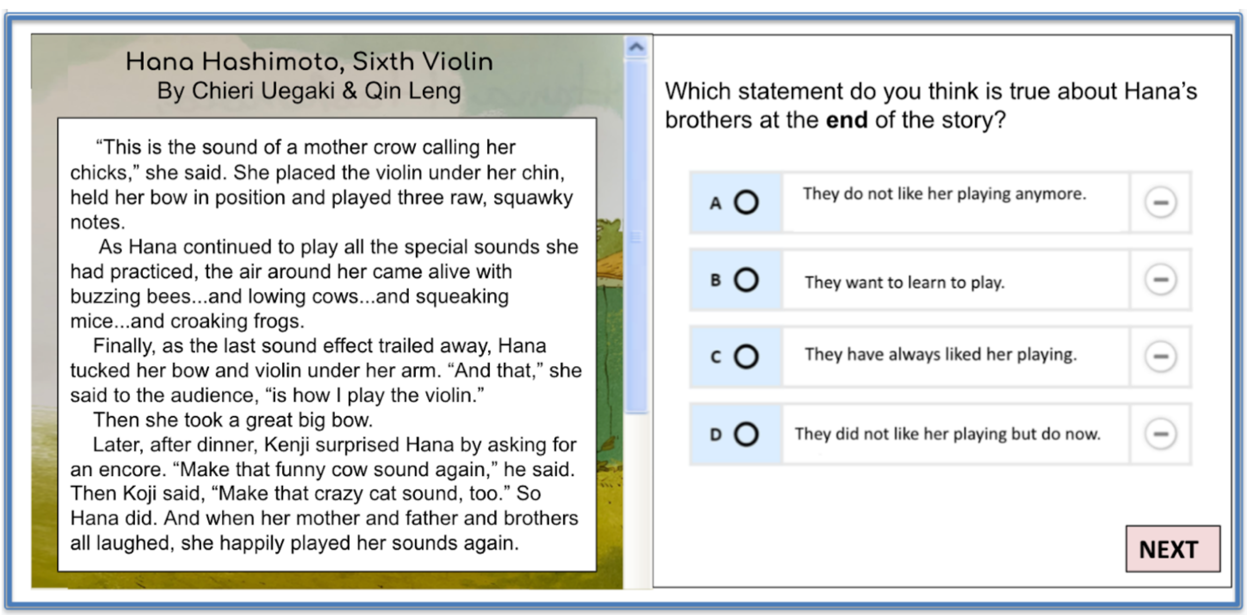


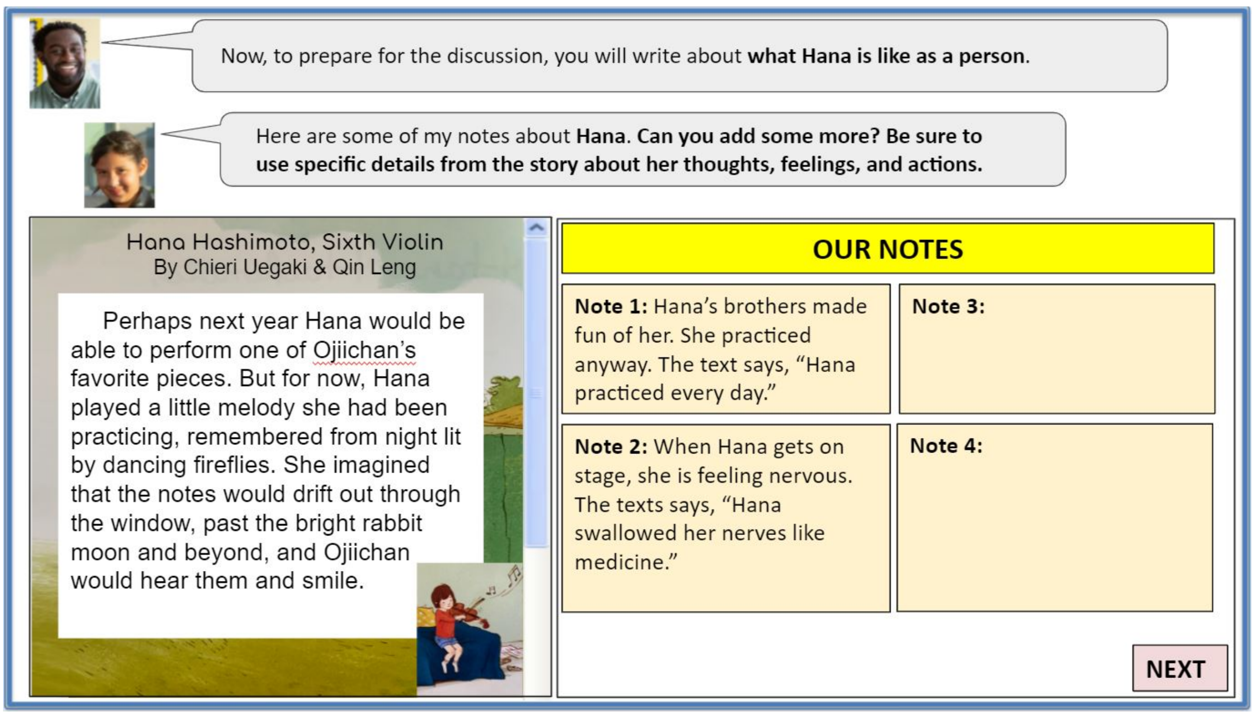
Exhibit C.11. A Grade 4 Integrate and Interpret Item for the first task using a single-selection multiple choice format



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 C.12).

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 C.13).

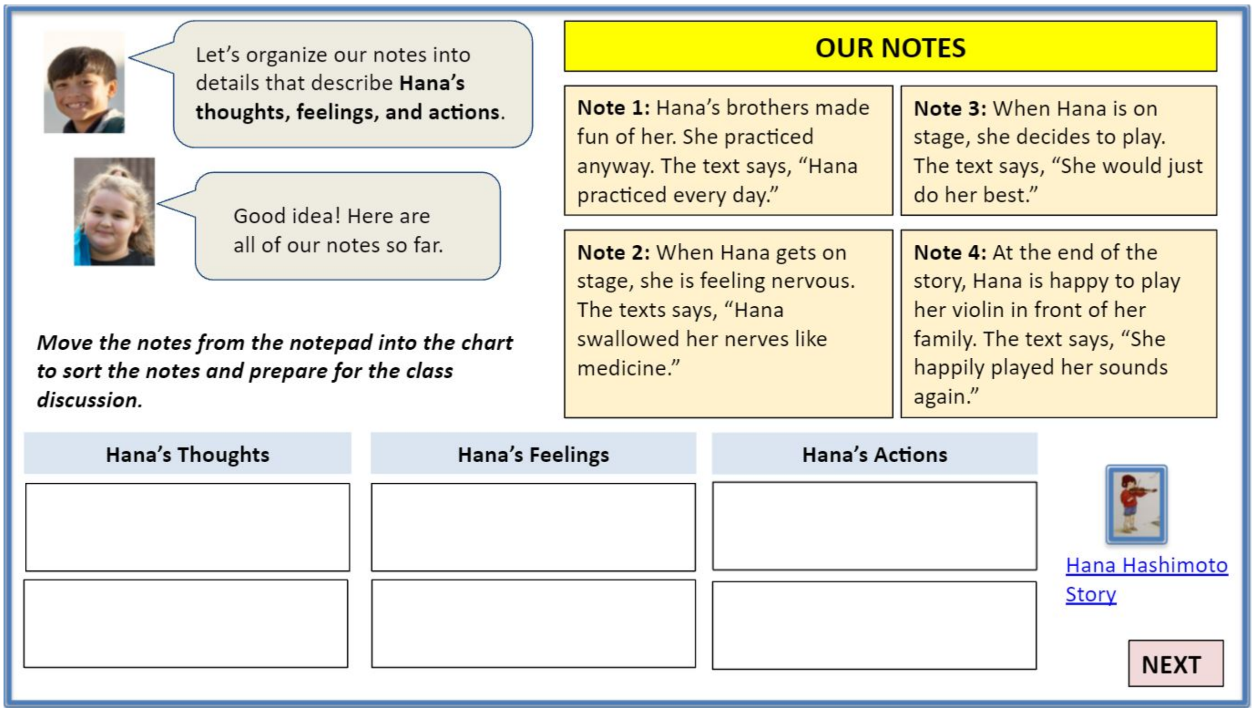
Exhibit C.12. Teacher and student task characters remind readers of the second task goal in this Integrate and Interpret item



In Exhibit C.13, 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 C.13, 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 C.14), 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.

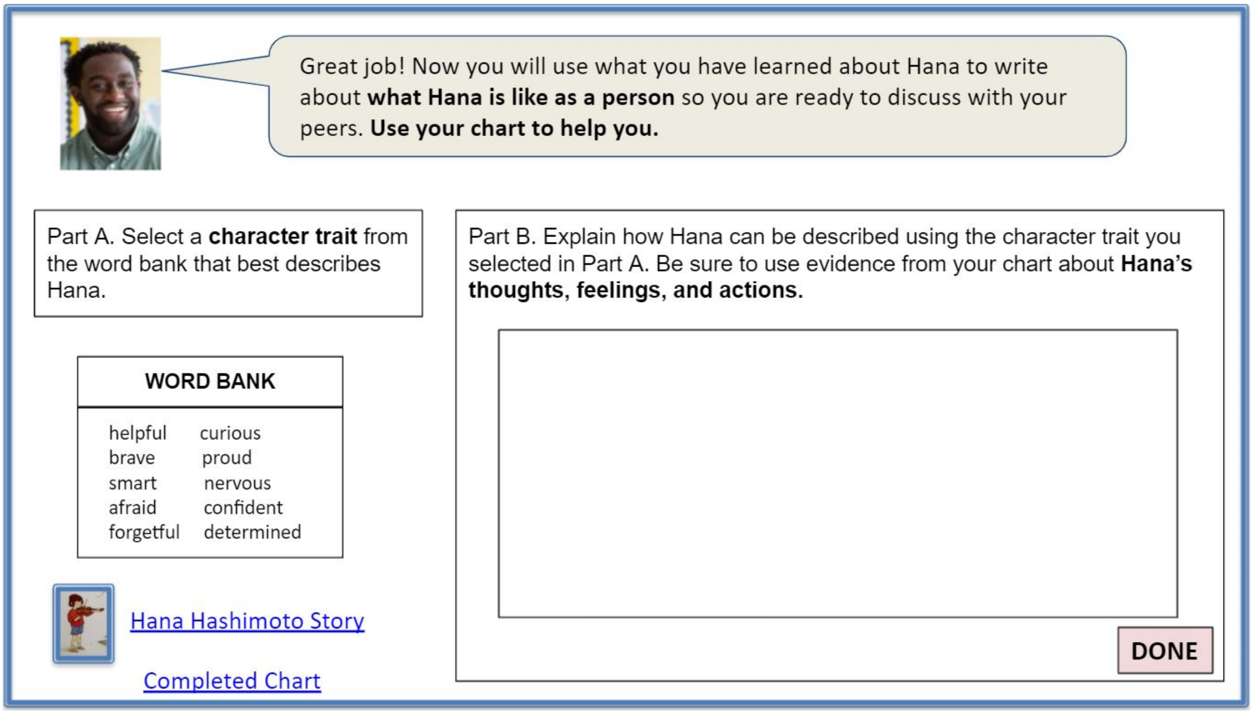
Exhibit C.13. The student's responses from their completion of the previous item are carried over to the next item as the completed notes. A graphic organizer with drag and drop features offers students an efficient way to demonstrate their understanding of how the text conveys the character's thoughts, feelings, and actions in this Grade 4 Integrate and Interpret item



A longer constructed response item such as the example shown in Exhibit C.14 is designed to assess readers' ability to Use and Apply understandings learned from the story to form a characterization of Hana. As readers engage with this final part of the block, 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 C.14, 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; when readers click on a word, the word is highlighted and is recorded as the student's answer to Part A) when asked to describe 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 C.14. 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 an extended constructed-response item format. 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.



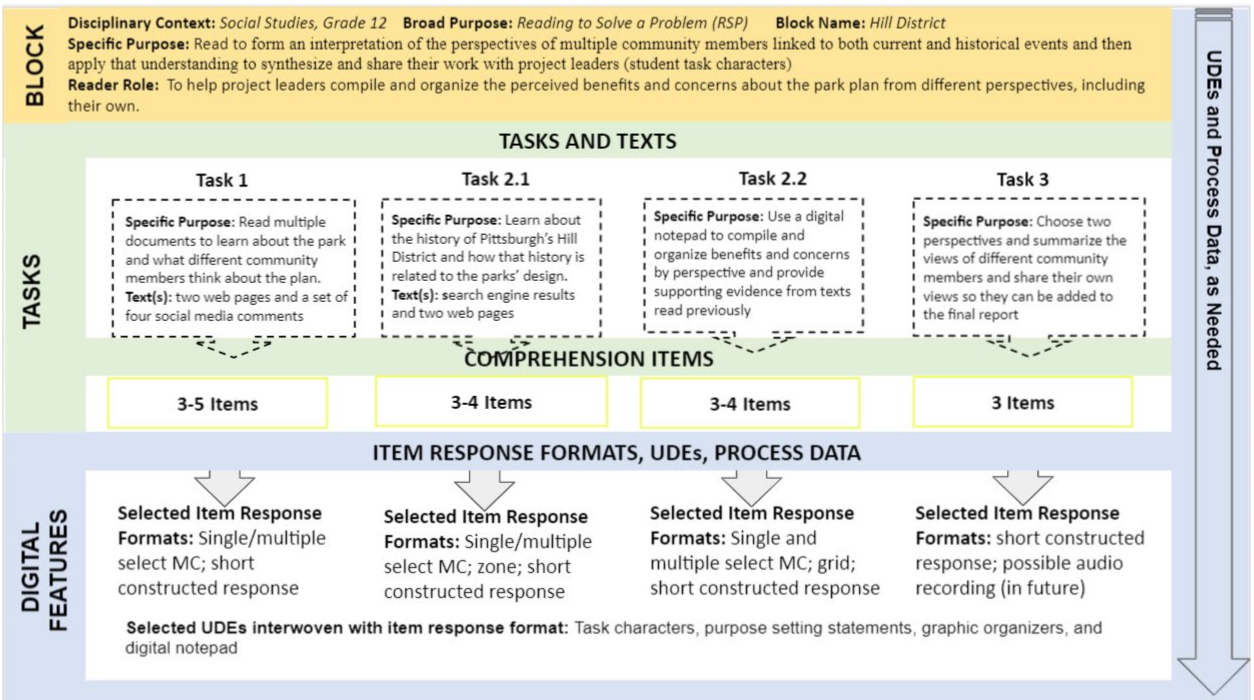
Performance Evidence and Indicators*. When interpreting reading achievement from performance on the 2026 NAEP Reading Assessment, multiple indicators can be used to explain what students are able to do. As indicated earlier in this chapter, each block would be classified with a primary disciplinary context, grade level, and broad purpose. Scores from the Hana Hashimoto, Sixth Violin block, then, 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 students' ability to develop their understanding of a single text and then apply that understanding in a simple culminating event (in this case, describing the kind of person Hana is based on her thoughts, feelings, and actions in the story).*

Test developers keep 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 12

Block Components (Context, Purposes, and Reader Role). This block is designed to assess how 12th 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 C.15 for the block design and Exhibit C.16 for the introduction to the block).

Exhibit C.15. Block Design for Hill District Sketch



More specifically, readers are invited to engage with three students (represented by task characters in the assessment) who have been asked by the Mayor to compile and organize public reactions to an ambitious plan 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.

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 C.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.

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 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 C.16), 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 C.17). 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 C.16. 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

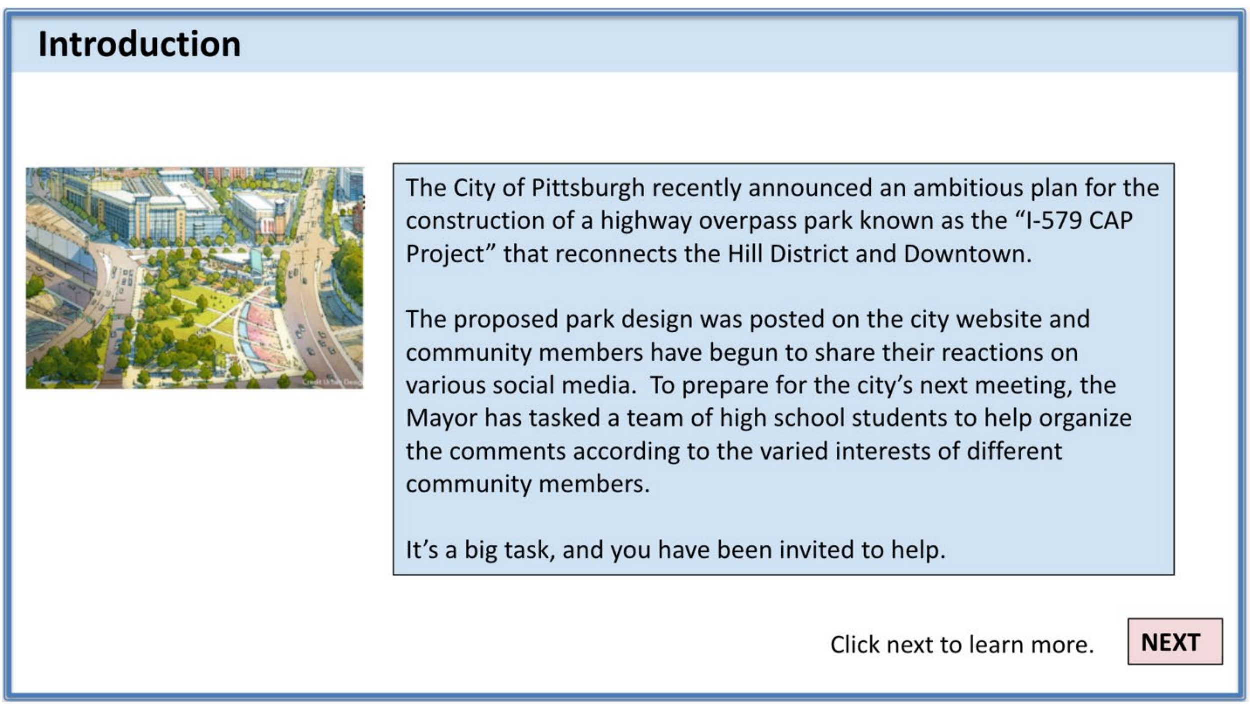
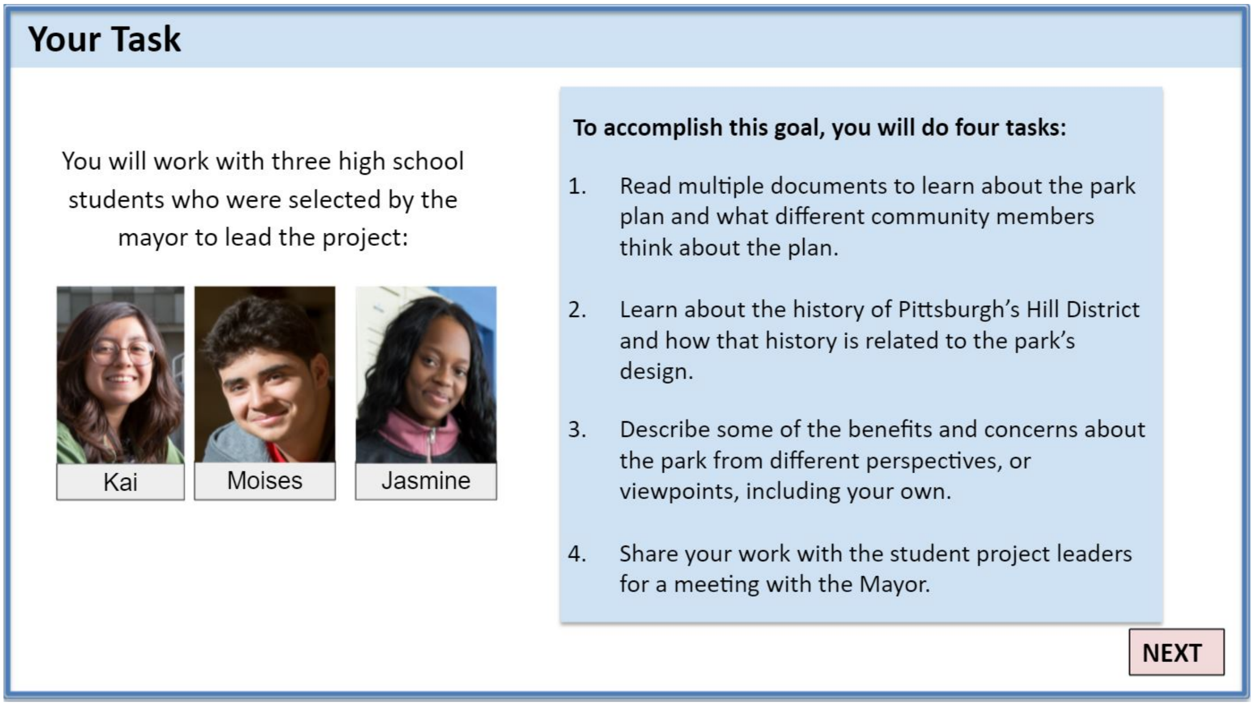


Exhibit C.17. Same-aged task characters and a task-based UDE in the form of four task-specific purposes serve to guide and motivate readers in the RSP block



Throughout Appendix C, the photograph of Kai is sourced from https://images.all4ed.org/high-school-boy-and-girl-near-playground (photographer Allison Shelley for EDUimages). The photograph of Moises is sourced from https://images.all4ed.org/high-school-boy-in-hallway (photographer Allison Shelley/The Verbatim Agency for EDUimages). The photograph of Jasmine is sourced from https://images.all4ed.org/high-school-boy-and-girl-drive-robots (photographer Allison Shelley/The Verbatim Agency for EDUimages).

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 opportunity to express their own opinions about the project) serve to guide and motivate readers to engage with the 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 plan. These include social media comments from community members; 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 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 readers different kinds of opportunities to demonstrate their understanding in the block and apply that understanding to 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 navigate and make meaning across sources representing multiple and diverse perspectives.

After being asked to read text and watch a short video on a website about the park project (Exhibit C.18), 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 C.19 and C.20). 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 C.20). **Task-based UDEs** (e.g., one of three task characters) provide short prompts (shown at the top of Exhibits C.18 and C.21) designed to cue the reader about the steps they are completing as they read across different sources to solve the problem.

Exhibit C.18. 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

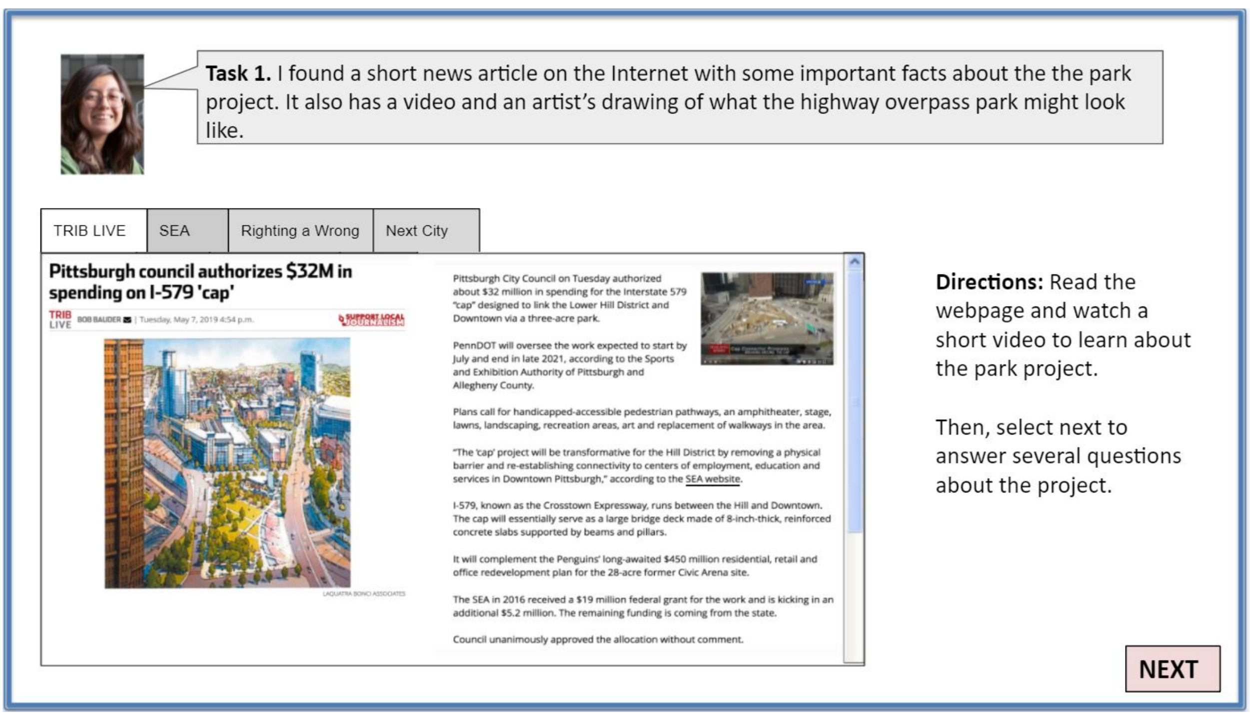


Exhibit C.19. A Grade 12 Locate and Recall item illustrating a multiple-selection multiple choice response format

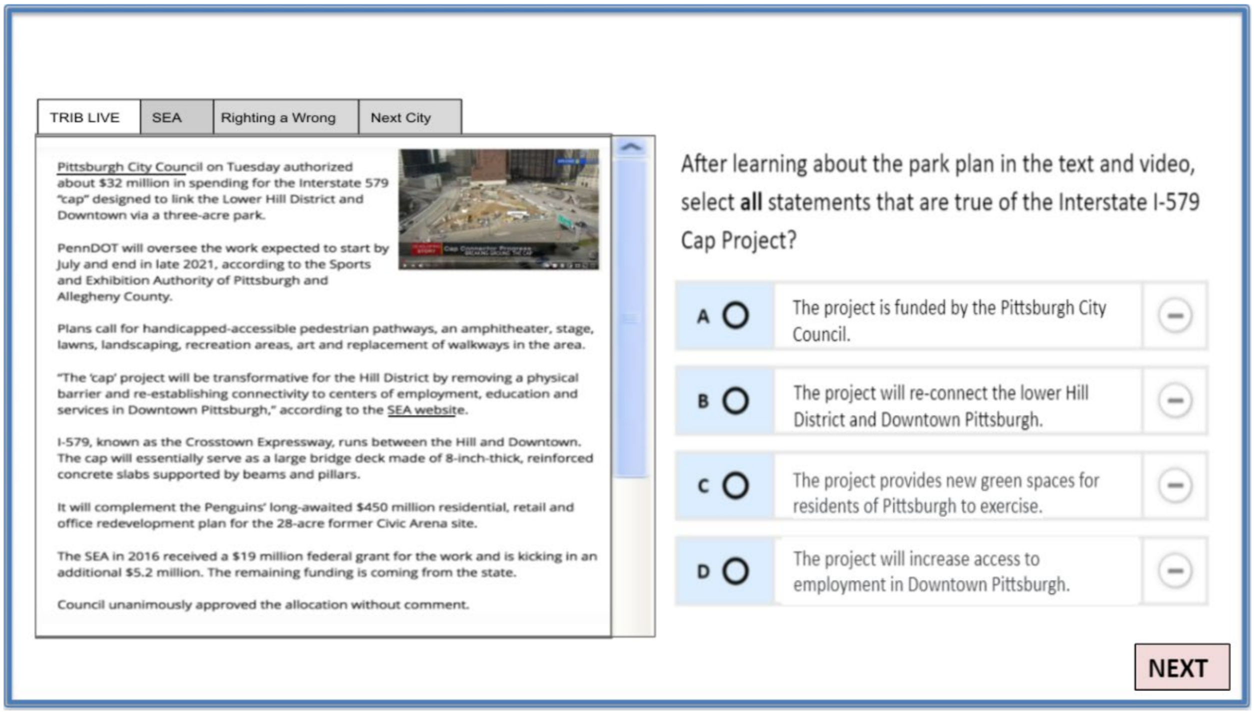


Exhibit C.20. A Grade 12 Locate and Recall item illustrating a single-select multiple choice item response format

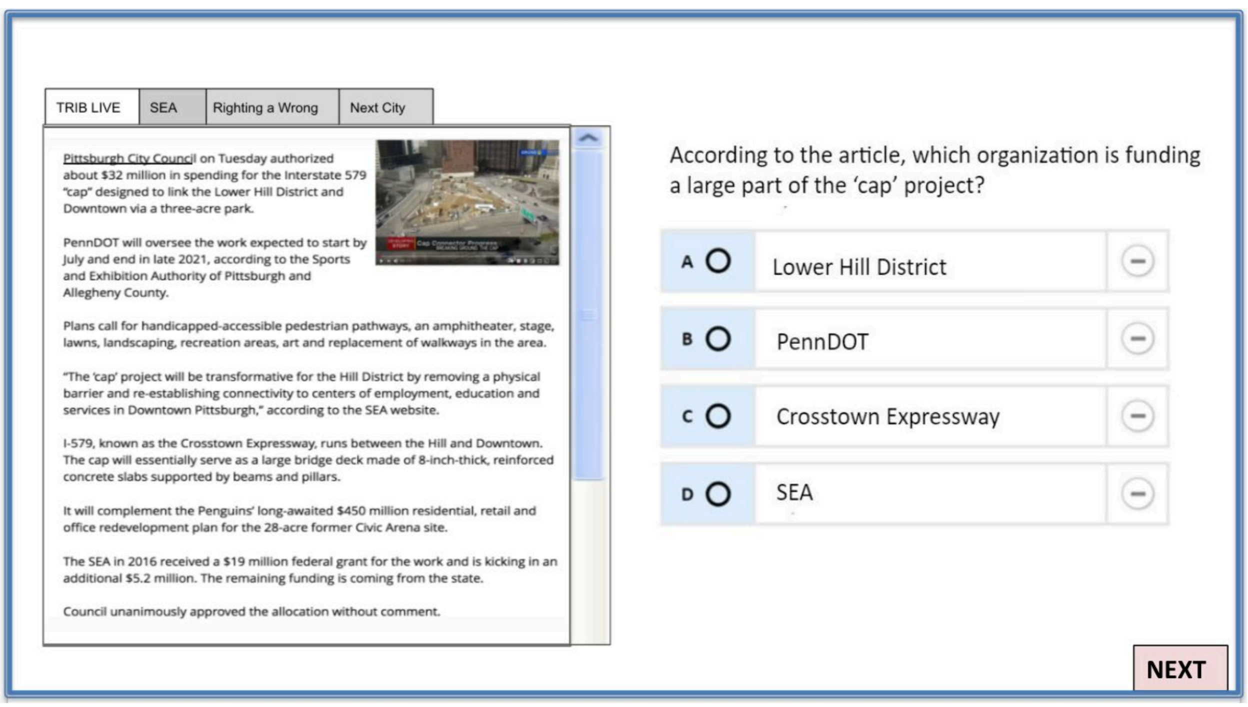
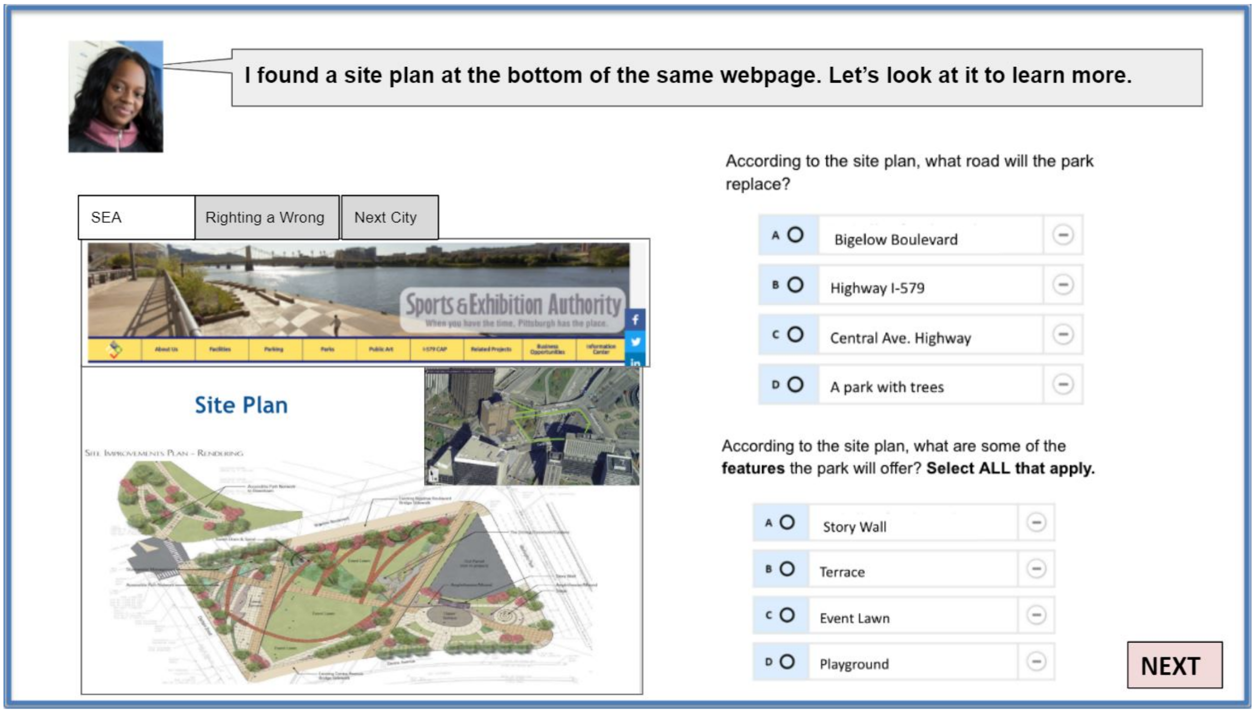


Exhibit C.21. Two Grade 12 items that ask readers to Integrate and Interpret (item 1) and Locate and Recall (item 2) textual and visual information from an artist's rendering of the site improvement plan published on a website



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 C.22). 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 C.23) or to analyze and evaluate the requests of some community members to include park features that honor the history of their neighborhood (Exhibit C.24). Also depicted in Exhibit C.24 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 C.22. 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

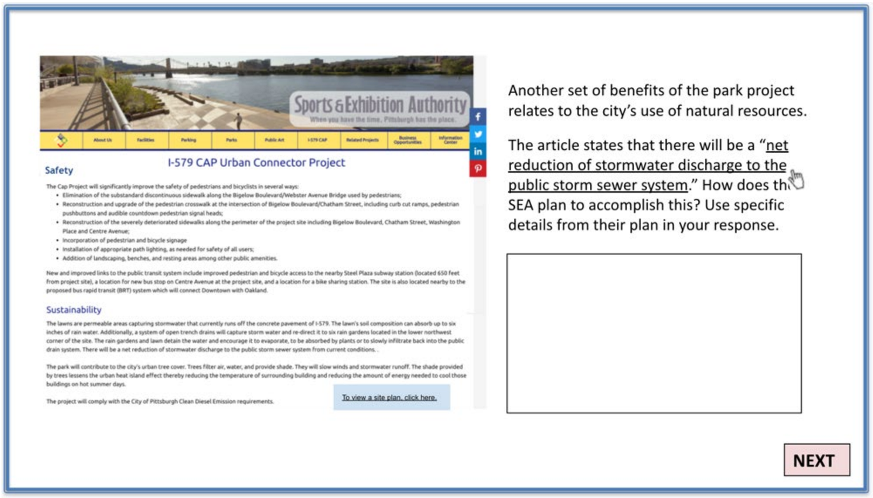


Exhibit C.23. A Grade 12 short constructed-response item with a look-back button (task-based UDE) that asks readers to integrate and interpret information in an online newspaper article about the historical significance of the park's design

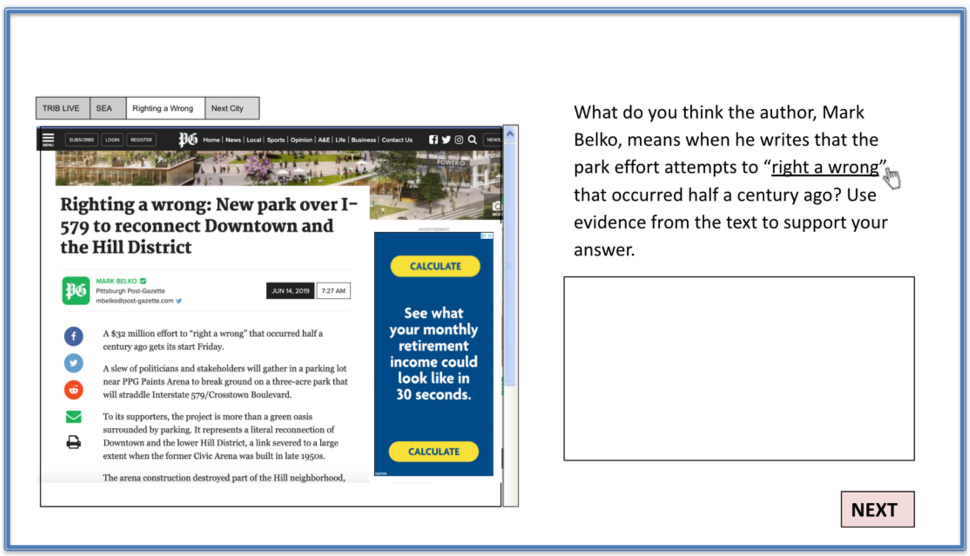
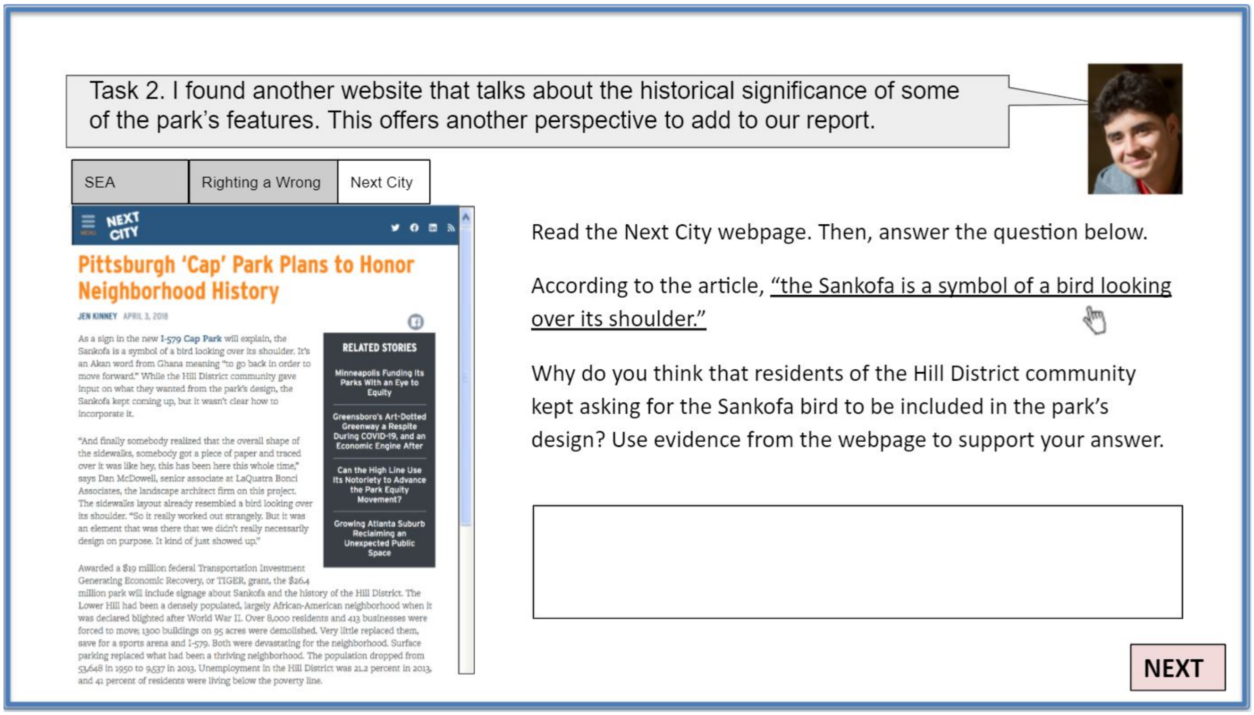


Exhibit C.24. A Grade 12 short constructed-response item that asks readers to integrate and interpret information on a web page with a look-back button (task-based UDE). The task character reminds readers of the specific purpose of the second task



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 C.25) 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 C.26 and C.27 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 C.25. A Grade 12 selected response zone item designed to capture process data about which link is selected and paired with a short constructed response scored item that asks readers to analyze and evaluate the relevance of their search engine choice

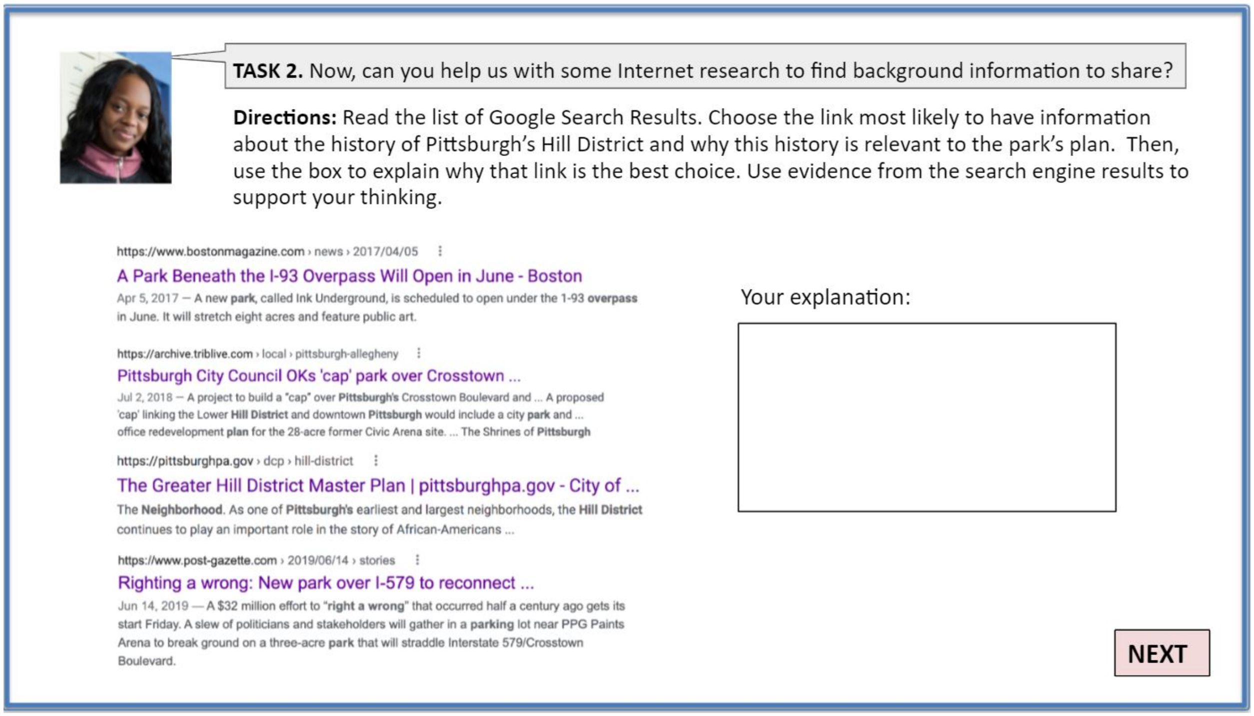


Exhibit C.26. A Grade 12 item selected response zone item designed to capture process data about how readers navigate through hyperlinked web pages

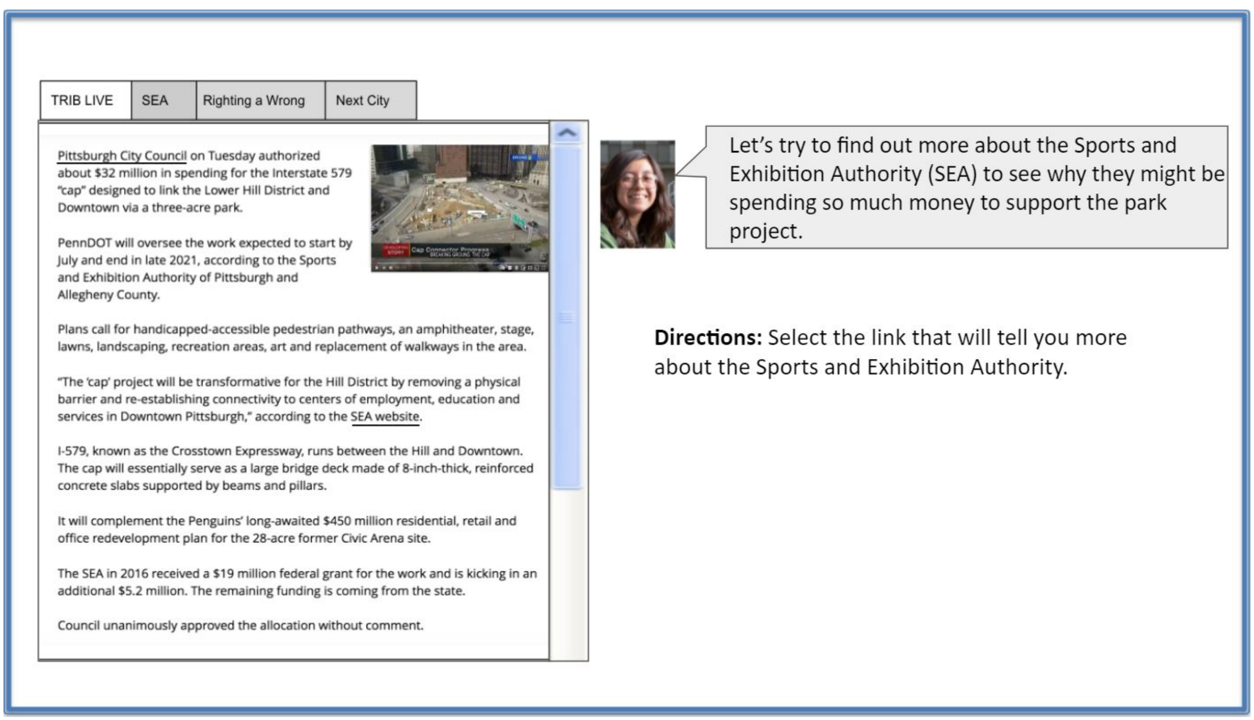
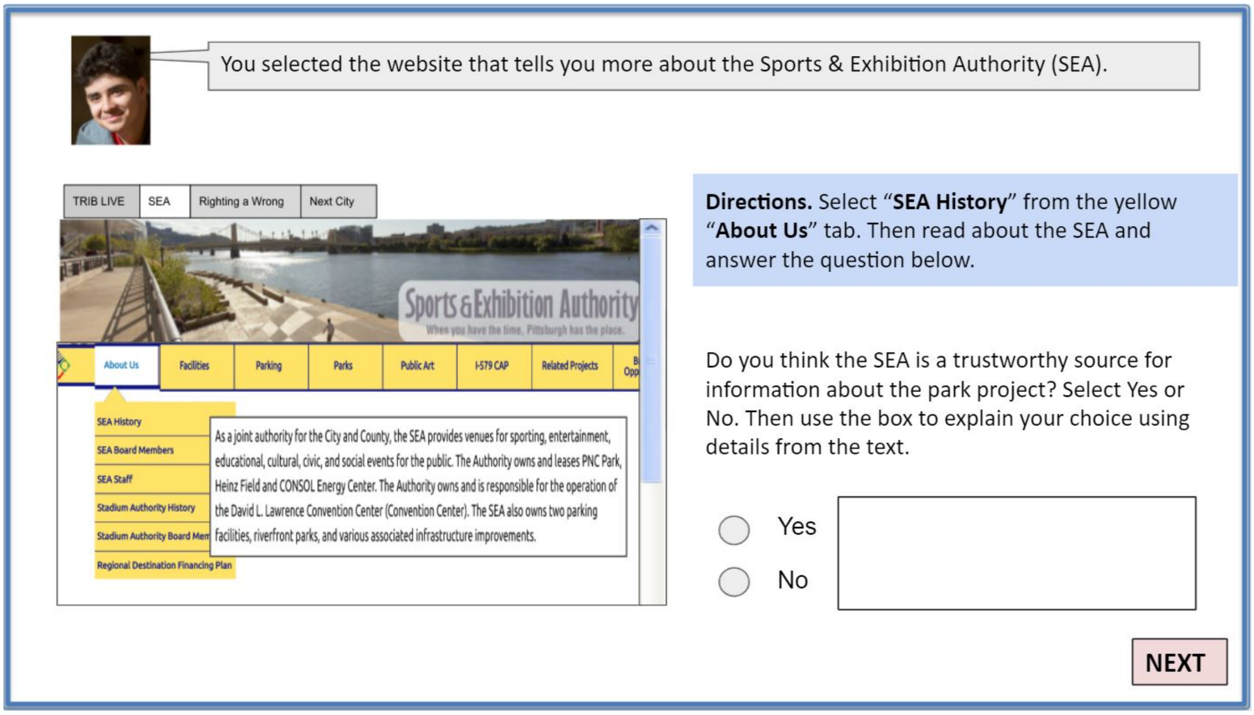
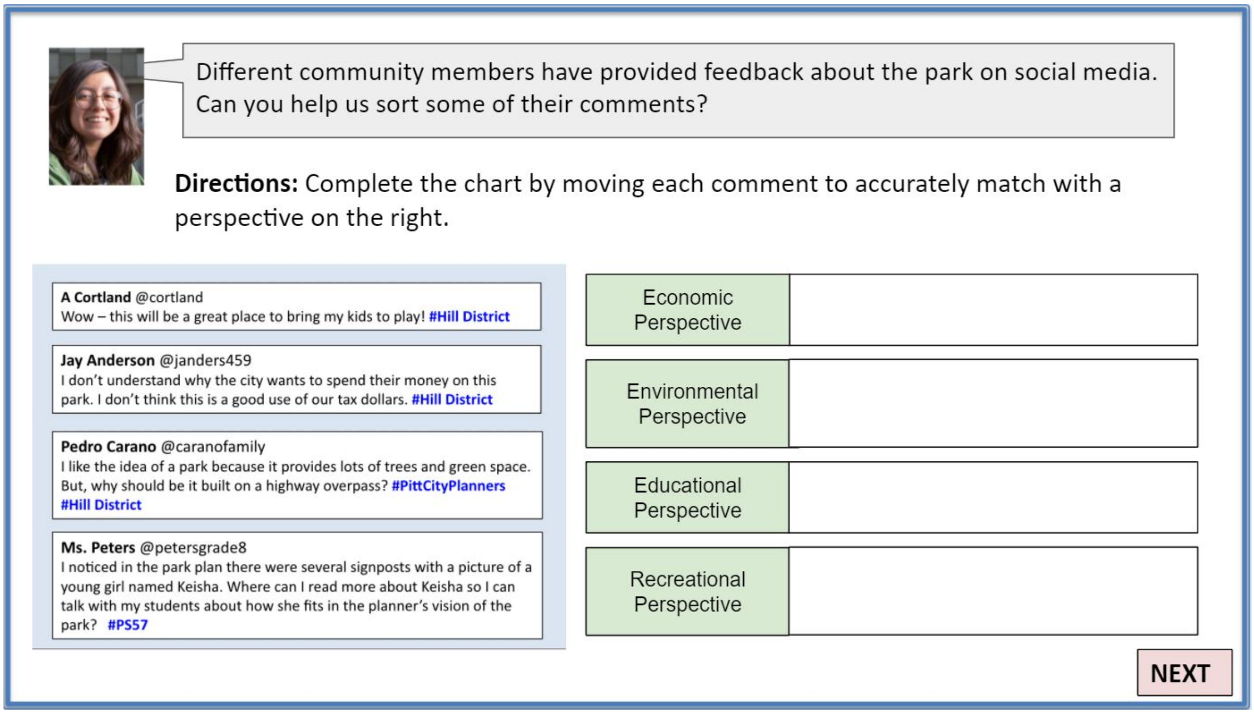


Exhibit C.27. A Grade 12 critical online resource evaluation item that asks readers to analyze and evaluate the extent to which an organization has the appropriate qualifications to publish details about the proposed park plan on their website using a hybrid constructed response



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 C.28, 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 functioning 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 C.28. 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 C.29 and C.30) 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 C.29 applies a multiple-select 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, match each comment with one or more specific benefits on the right.

In contrast, Exhibit C.30 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. 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 C.31 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 C.29. A Grade 12 multiple-select 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

I noticed that there are a lot of different opinions and perspectives on the park in this website. I was thinking we could organize these by topic and add these to our summary report for the Mayor.

Directions. The table below lists comments from two community members and columns with three benefits of the proposed plan. **Select one or more benefit that applies to each person's comment.**

Comments from Community Members as Quoted in Website #1 ("Righting a Wrong")	Connects Hill District to Downtown	Offers Green Space	Rights A Wrong
Longtime Hill District Resident Brenda Tate: For Brenda Tate, who has lived on the same block of Webster Avenue in the Hill for all of her 70 years, the park once again will give her the chance to traverse Wylie Avenue to the park then into Downtown and back. "There won't be separation. There will be a clear avenue to come back and forth. It's symbolic," she said. Ms. Tate, who with her 98-year-old aunt will be attending Friday's groundbreaking, sees positives in the park's construction. "It will be a nice green space, a welcoming space, for people who want to come into the community," she said. (*supportive member of the Hill District*)	○	○	○
City Councilman R. Daniel Lavelle: "What we're going to begin doing [Friday] is finally righting those wrongs of 50 or 60 years ago," added Mr. Lavelle, who represents the Hill. While the park is important, Mr. Lavelle said the greater value lies in providing business and job opportunities within the arena redevelopment for Hill residents and minorities. (*city councilman who represents the Hill district*)	○	○	○

NEXT

Exhibit C.30. 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

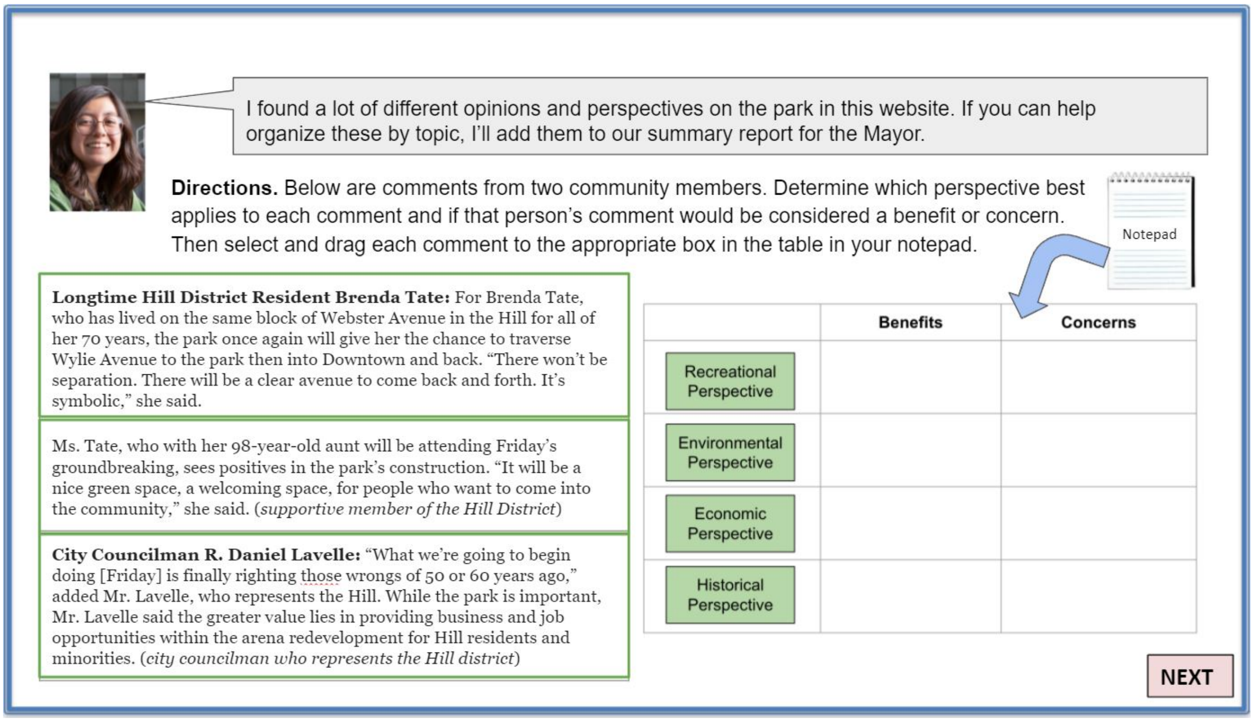
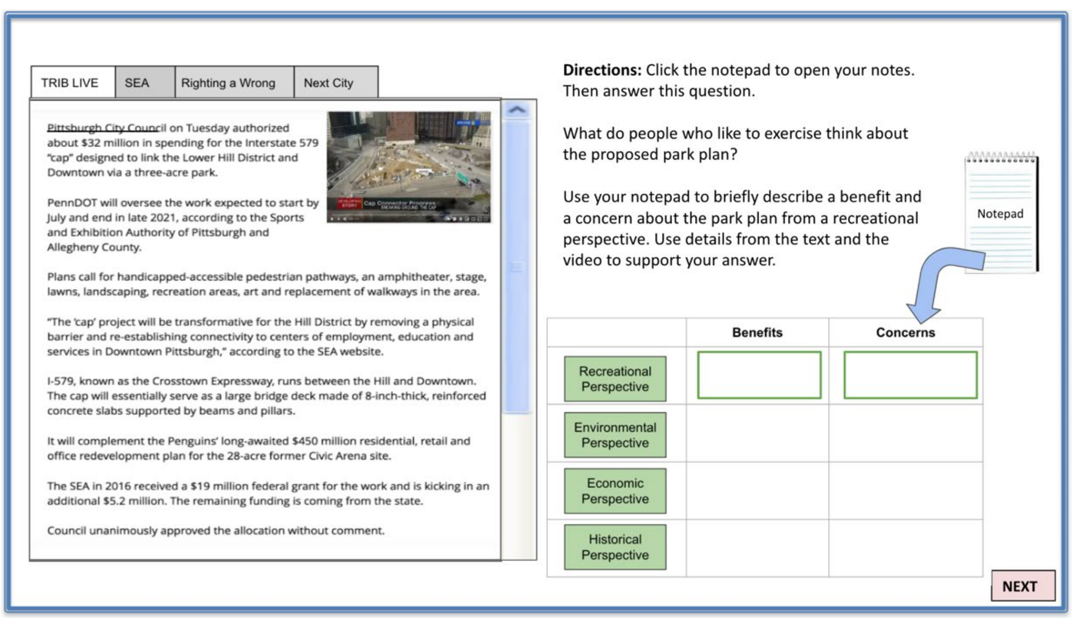


Exhibit C.31. A Grade 12 dynamic matching response grid item with a task-based UDE (retractable digital notepad) that serves to support readers at another point in the RSP block as they read across multiple sources to solve the problem at hand



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 C.32), 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 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 have 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 C.33 illustrates how this item might look using a short-constructed response format, similar to those in existing NAEP assessment blocks, and Exhibit C.34 is included to depict what an item might look like in 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 C.32. 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.

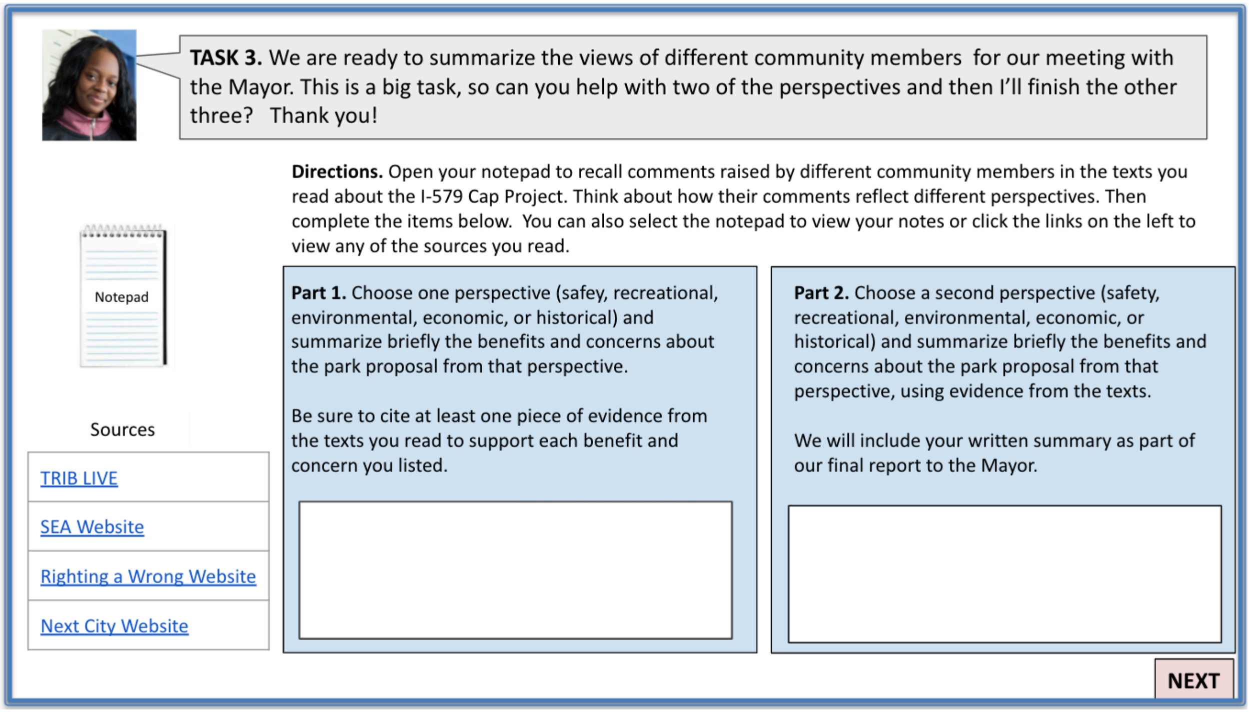


Exhibit C.33. This final Use and Apply item with open-constructed response format illustrates the use of a task character (motivational UDE) who reminds students they have accomplished their goal and validates the test-taker's role by inviting them to use what they learned and apply that understanding by sharing their own opinion.

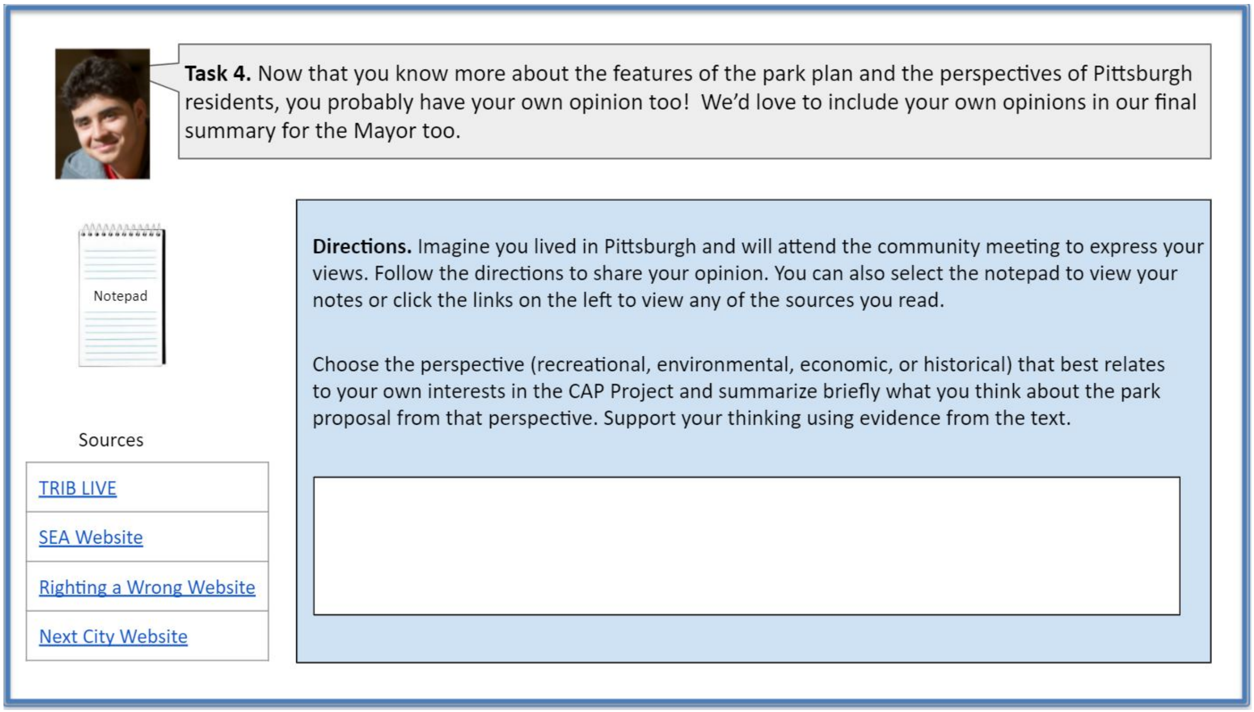
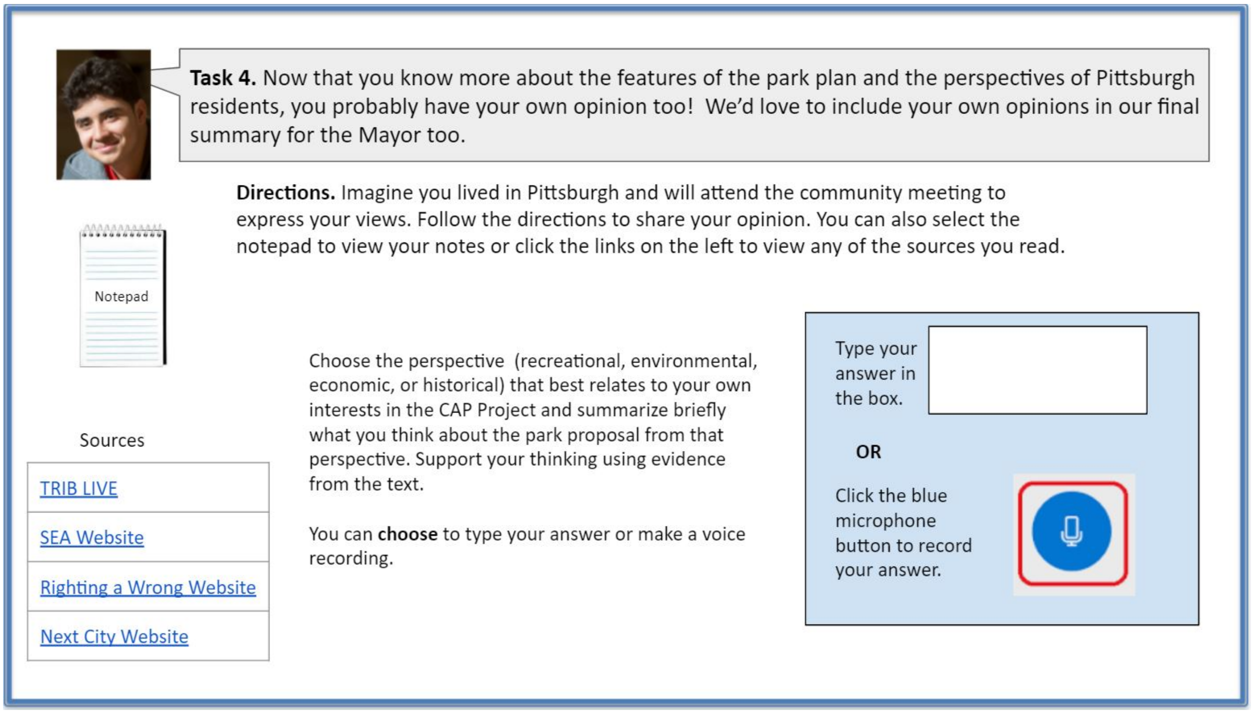


Exhibit C.34. This alternative format for the final Use and Apply item with open-constructed response format illustrates the use of motivational UDEs for two purposes: a task character who invites students' own opinion paired with an opportunity to choose their preferred format (text or audio) for expressing their opinion.



Performance Evidence and Indicators. *Scores from the Hill District block reveals what Grade 12 students can do when Reading to Solve a Problem in a social studies context. Ultimately, NAEP produces descriptions of what 12th graders (or sub-groups of 12th 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 Grade 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.*

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 C.2. Here, students have more time to develop deep understanding of the texts. Tasks are relatively simple, 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 literature 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 C.44 and C.45), 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 literature genres—a biographical sketch and a human-interest essay (see Exhibit C.35).

Exhibit C.35. Introduction to E. B. White

Introduction

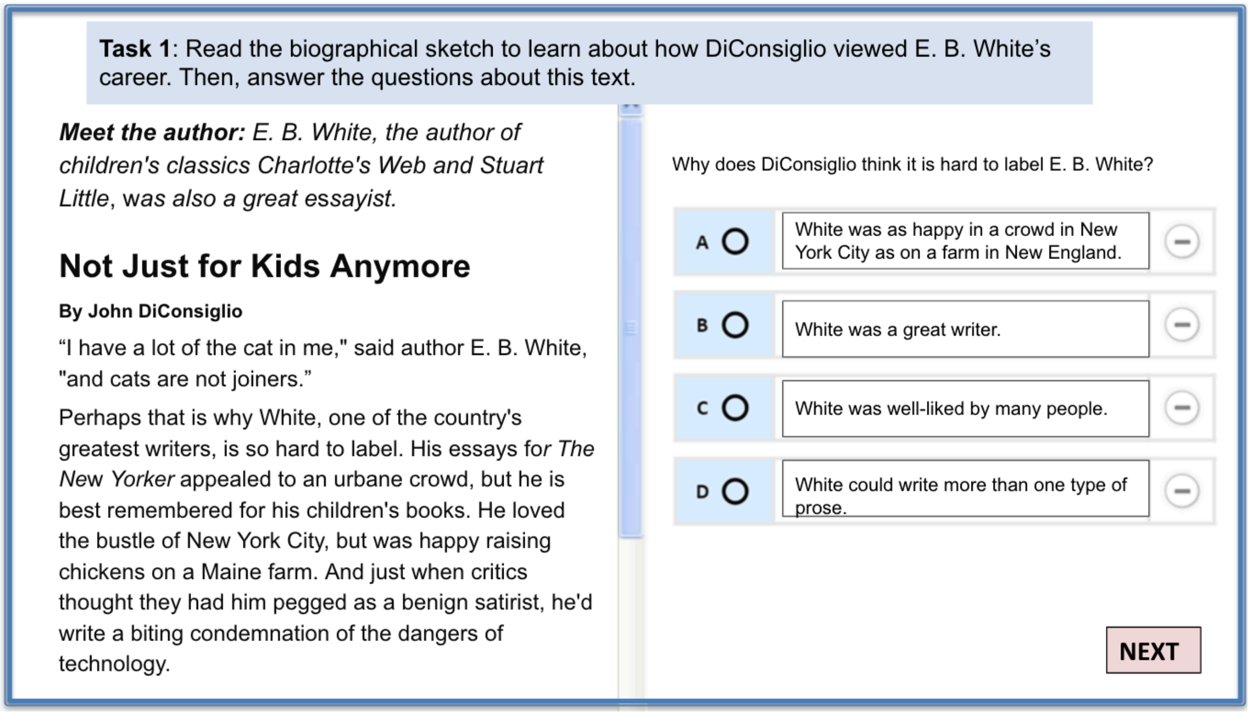
You will read two texts: (1) a biographical sketch about the author E. B. White, most famous for writing *Charlotte's Web*, and (2) an essay that White wrote for *The New Yorker* magazine.

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 Exhibit C.36, which shows task 1.

Exhibit C.36. Introduction to the grade 8 E. B. White literature block

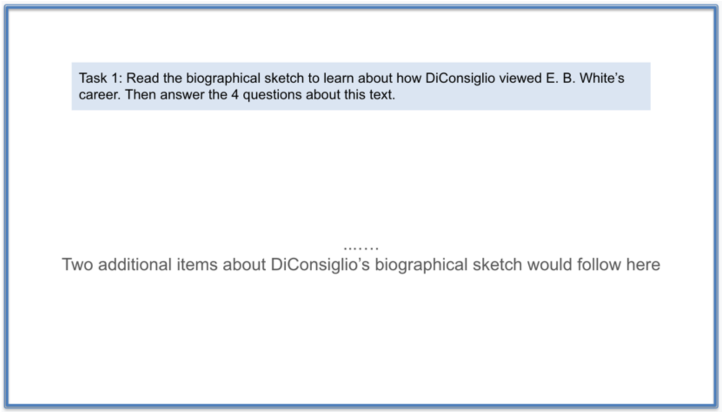


The comprehension items for Task 1 could help the reader develop understanding on segments of the biographical sketch that focus on characteristics of White that might be useful in Task 3 (see Exhibit C.37). Plausible segments for focus could be…

- 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 an informational 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 C.37. Task 1 would involve additional items

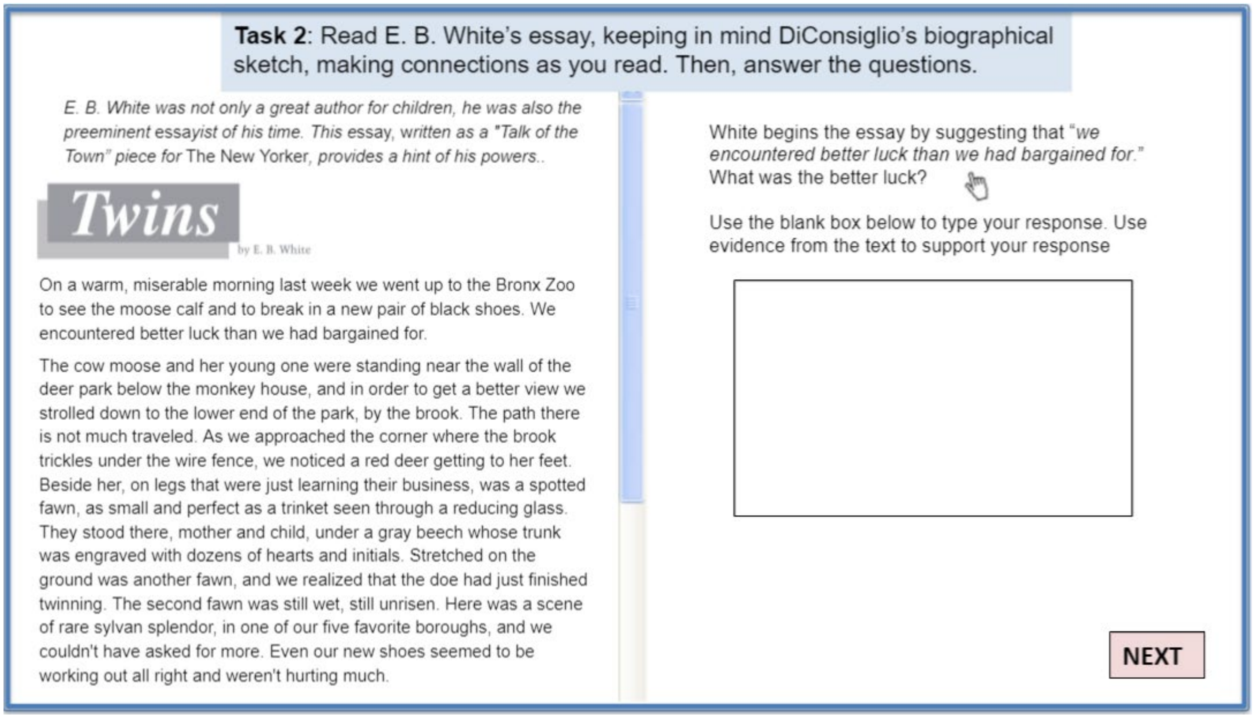


For Task 2, comprehension items should focus on the narrator White's statements that say something about his personality and attitudes toward the world around him (see Exhibits C.38-C.40). Candidates for items include:

- Getting more than we bargained for and the sighting of the doe and her twins.
- White's characterization of the doe being 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 an informational 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 C.38. Task 2 for the grade 8 E. B. White block illustrating an Integrate and Interpret item with a short constructed response item format



Task 2: Read E. B. White's essay, keeping in mind 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

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.

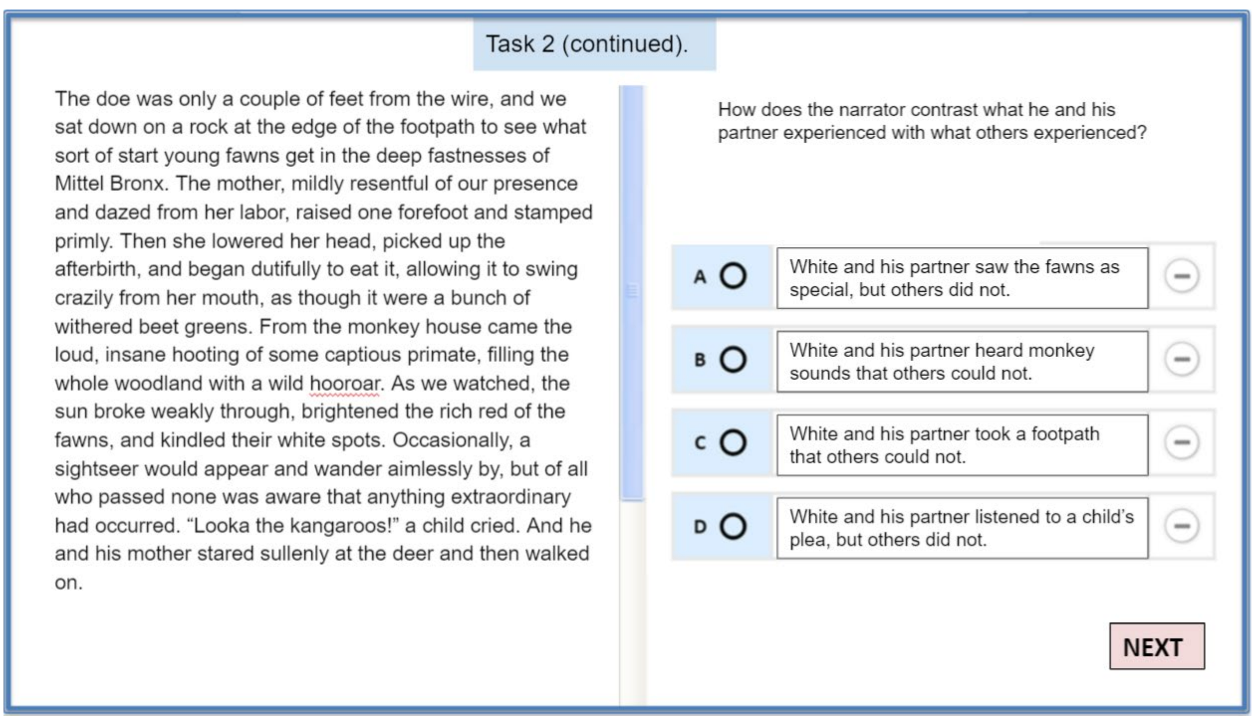
The cow moose and her young one were standing near the wall of the deer park below the monkey house, and in order to get a better view we strolled down to the lower end of the park, by the brook. The path there is not much traveled. As we approached the corner where the brook trickles under the wire 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.

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 support your response

NEXT

Exhibit C.39. Task 2 continues for the grade 8 E. B. 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.

How does the narrator contrast what he and his partner experienced with what others experienced?

A White and his partner saw the fawns as special, but others did not.

B White and his partner heard monkey sounds that others could not.

C White and his partner took a footpath that others could not.

D White and his partner listened to a child's plea, but others did not.

NEXT

Exhibit C.40. 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 C.41 and C.42), might be provided to assist students in organizing their response to a final Use and Apply extended constructed response item (see Exhibit C.43).

Exhibit C.41. 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 blank spaces in the chart.

1. Idea from *Not Just for Kids Anymore*	2. How the idea applies to the narrator of *Twins*
Cats are not joiners.	**White and his companion stayed 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

Exhibit C.42. Integrate and Interpret item illustrating resetting of item responses from prior item

Task 3: Completed Chart: Comparing ideas across the two passages

No questions to answer on this screen. Below is the chart from the previous page with the phrases from the Idea Box dragged into the correct spaces in the chart. You can refer back to this chart when you complete the next (and last) item in this block.

1. Idea from *No Longer Just for Kids*	2. How the idea applies to the narrator of *Twins*
Cats are not joiners.	**White and his companion stayed 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 C.41), students receive feedback about how the chart might best have been completed in Exhibit C.42. The task-based UDE, called resetting, is provided so that students do not carry misconceptions into the final item in Exhibit C.43.

Exhibit C.43. A Final Use and Apply item asks students to use ideas from the first text to develop ideas about the second text

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 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 E. B. White block sketch provides an example of how blocks might look under the auspices of the 2026 assessment 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 C.44 and C.45.

Exhibit C.44. 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 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.

© The New York Times/Redux

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 give *The New 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."

-By John DiConsiglio

Exhibit C.45. The Second Text for the E. B. White Task: An Essay from the *New Yorker*

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 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.

[PLACEHOLDER]

Appendix B: Achievement Level Descriptions

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.

Commented [A1]: This is Appendix B from the Reading Assessment Framework approved by the Board on August 5. The tracked changes reflect edits that would need to be made to be consistent with the Reading Assessment and Item Specifications. These edits to the ALDs were made in response to concerns raised by Board staff, NCES staff, and Technical Advisory Committee members to eliminate references to knowledge and skills that NAEP cannot measure.

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 2026 NAEP Reading Framework does not provide reporting ALDs; those ~~are~~ will be constructed using empirical data during a later stage in the NAEP cycle, i.e., ~~a live~~an operational 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.

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 2026 Framework has identified three (literature, science, and social studies~~, and literature~~). In the ALDs below, all three disciplinary contexts are described within each performance level.

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 *C~~c~~omprehension T~~t~~argets:* (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 C~~c~~omprehension T~~t~~arget. 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 C~~c~~omprehension T~~t~~argets, such as Use and Apply.

Broad and Specific Reading Purposes

Reading activities in an assessment block are situated within ~~not only~~ a disciplinary context ~~but~~ as well as~~also~~ a broad reading purpose. 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 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 in the real world occurs. The subsections below describe how specific reading purposes map to disciplinary contexts.

Literature 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 literature 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:

- 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, maps, data, 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, recall, and/or record specific pieces of information, identify relationships between explicitly stated pieces of information, make simple inferences and interpretations in static, dynamic, and multimodal texts, ~~create~~ determine the accuracy of summaries, and show understanding of vocabulary in the disciplinary contexts.

When engaged in reading literature 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 prosea~~(e.g., text features) among literature subgenres appearing in ~~a~~ specific task texts, and show understanding of vocabulary and simple figurative language. Readers should be able to ~~produce~~ determine the accuracy of 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~~form an evidence-based opinion about a text. 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 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~~form an evidence-based opinion about a text. Readers should be able to describe ~~the overall structure of a text~~text structures as they pertain to the presentation of content in a specific text, and compare and contrast explicit information found in a firsthand and secondhand account of the same event or topic. Readers should be able to ~~produce~~ determine the accuracy of a simple summary of a text and integrate information from lower complexity sources to ~~produce a new text of informational or argumentative purpose~~apply to a new context.

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 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 literature 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~~information from a multimedia source contributes to understanding of a printed text and show understanding of nuances in word meaning. Readers should be able to ~~produce a detailed summary of a text and rewrite a story~~

~~from a different character's perspective~~apply understanding of a character to an interpretation of another character's point of view.

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 based upon content in the text 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~~determine missing steps in a procedure ~~or experiment~~(e.g., a simple investigation; craft-making related to a scientific concept) 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~~applying information learned to a new context.

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 literature 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 common human experiences~~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 based upon content in the text, ~~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 based upon content in the text, ~~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 caption to a political cartoon or a personal bill of rights~~and apply information from texts in a new context, such as proposing a caption for an illustration or cartoon, or to create a set of recommendations.

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 based upon content in the text, ~~create objective~~determine the accuracy of summaries, analyze word choice, and show understanding of vocabulary in the disciplinary contexts.

When engaged in reading literature 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 should be able to ~~develop a simple objective~~determine the accuracy of a summary of a text and ~~produce an argumentative text~~construct an argument 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 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. They should be able to demonstrate an understanding of the purpose/function of a specified text features (e.g., introductions, sidebars, headings, illustrations, charts). 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 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~~use information from multiple sources to apply to a new context.

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 literature texts such as fiction, drama, film, poetry, and literary nonfiction, eighth-grade readers performing at the *NAEP Proficient* level should be able to 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~~information from a multimedia source contributes to understanding of a printed text and how text structure contributes to meaning and style. They should be able to analyze how word choice impacts a text's meaning and tone. Readers 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~~apply analysis of multiple texts to an explanation of 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 should be able to ~~produce~~ present an argument~~ative 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 literature 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~~They should be able to determine how the text structure contributes to the development of theme, setting, or plot. Reachers should be able to ~~rewrite a section of a story from another character's perspective~~describe how a story might change if written from the perspective of another character.

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 should be able to ~~produce a new argumentative or informative text~~construct an argument or explanation 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 should be able to ~~produce an informative text that~~ trace~~s~~ and connect~~s~~ 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 based upon content in the text, ~~create objective~~determine the accuracy of summaries, analyze word choice, and show understanding of vocabulary in the disciplinary contexts.

When engaged in reading literature 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 should be able to ~~develop an objective~~determine the accuracy of a summary of a text and ~~produce an informational text that applies~~apply a common theme or central idea culled from multiple texts to ~~a current societal issue~~ncommon human experiences~~ew context or situation~~.

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~~ a text structures information ~~or ideas into categories or hierarchies~~to serve an author's purpose and help readers organize their thinking. 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~~apply findings described in a text to a new context or situation.

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 should be able to ~~integrate~~ use information from multiple sources to ~~produce a new text of informational or argumentative purpose~~construct an explanation or argument.

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 literature 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 should be able to analyze how word choice impacts a text's meaning and tone. Readers 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~~present an opinion regarding a universal problem that is elicited from an analysis of the text.

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 or~~,~~ 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 should be able to ~~produce a new argumentative or informative text~~construct an argument or an explanation 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.~~historical 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 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~~construct an argument 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 literature 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) compared to the source text. Readers 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~~or apply information gained from a literary text or a poem to analyze a new text.

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~~ analyze 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 should be able to ~~produce a new argumentative or informative text~~construct an argument, ~~or~~ explanation, or recommendation that ~~utilizes an understanding of legal and ethical principles to address a scientific matter of debate (e.g., uses of genetic databases)~~requires the application of scientific content from a text.

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 should be able to ~~produce a new argumentative or informative text~~construct an argument, ~~or~~ explanation, or recommendation that utilizes an understanding of legal and ethical principles to address a societal matter of debate (e.g., indigenous peoples' land rights).

Potential Changes to NAEP 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. In addition, the revised policy focuses on high level guidance rather than procedural details; the latter is intended to be captured by an accompanying Framework Development Procedures Manual. A procedures manual has not yet been created but this task is included in a Technical and Logistical Services contract that has recently been awarded to the Manhattan Strategies Group (MSG) with a subcontract to the Human Resources Research Organization (HumRRO).

The current policy has now been in place since March 2018 and has guided the updates of the NAEP Mathematics Framework (adopted by the Board in November 2019) and the NAEP Reading Framework (adopted by the Board in August 2021). There is a need to evaluate the extent to which the current policy and procedures should be revised, and to determine how procedures should be documented in a Framework Development Procedures Manual that will be developed in advance of the next framework update.

To provide additional background and inform potential recommendations, Board staff commissioned two papers. As a consultant, former Governing Board Executive Director Cornelia Orr synthesized historical information on NAEP framework development. As part of the Board's previous 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 a technical memo on how NAEP framework development relates to procedures for developing other assessments. The two papers were included in previous ADC and Committee on Standards, Design and Methodology (COSDAM) materials and discussions on framework processes.

Board staff worked with the ADC leadership and members to develop preliminary recommendations for revising the Board policy and processes for framework development. The initial recommendations were discussed in a joint planning meeting of ADC and COSDAM in September and revised in preparation for discussions with the full Board at the upcoming quarterly meeting. The goal is to adopt a revised policy and create a Framework Development Procedures Manual prior to convening panels for the NAEP Science Framework update in late spring of 2022, if the Board decides that an update to the NAEP Science Framework is necessary. The Achievement Levels Procedures Manual provides an example of how a Board policy can be further elaborated for implementation.

Several of the recommended changes to the framework update process are related to how the Board can surface and provide direction on important policy and controversial issues upfront and at key points throughout the process rather than waiting until after seeking public comment on a draft framework. The timing of Board input is not specified in the policy but traditionally the Board has prioritized receiving recommendations from content experts without regard to most other factors and releasing those directly for public comment; Board policy discussions have taken place following a public comment period on a draft framework. By this point, framework development panels have worked for a year or more without any policy considerations placed on the process. Based on experience with both the NAEP Mathematics and Reading Framework updates in recent years, it is not ideal for the process if public comment on a draft framework surfaces controversial issues that were not previously considered by either the Board or the panel. More broadly, in the two decades since most NAEP assessment frameworks were last updated, the context surrounding education and assessment has changed significantly – e.g., greater alignment in states' standards in some content areas, transitions to digital assessment, new opportunities to engage with stakeholders through virtual meetings and digital media – necessitating consideration of other factors earlier in the process as well as new approaches.

In the attached policy document, proposed edits are indicated in "tracked changes," rationales for substantial changes are noted in comment boxes, and proposed changes that apply only to the procedures manual are indicated in comment boxes.

The Strategic Vision includes a goal to: *Optimize the utility, relevance, and timing of NAEP subject-area frameworks and assessment updates to measure expectations valued by the public.* To address this goal, staff have also suggested undertaking additional research to inform the implementation of frameworks, including the feasibility of smaller, more frequent updates. It currently takes approximately 5-6 years from the time that the Board adopts a framework until NCES implements the changes in the operational assessment. Staff have suggested additional research and expert consultation in conjunction with NCES; the outcomes of such work may warrant additional updates to the policy statement and procedures manual in the future.

During the November plenary discussion, ADC Vice Chair Mark Miller and ADC member Patrick Kelly will present key highlights of the proposed changes to the Board's processes for framework updates and facilitate Board member discussion. No action is anticipated at this meeting; additional changes to the policy and procedures are intended as an outcome from the November meeting discussion for potential action at the March 2022 Board meeting.

DRAFT FOR BOARD DISCUSSION AT THE NOVEMBER QUARTERLY MEETING

Adopted: TBD

National Assessment Governing Board

Assessment Framework Development

Policy Statement

It is the policy of the National Assessment Governing Board to conduct a comprehensive, inclusive, and deliberative process to determine and update the content and format of all assessments under the National Assessment of Educational Progress (NAEP). 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.

Commented [A1]: RATIONALE: "Assessment" has been added to the name of the policy to ensure it is clear that NAEP frameworks are assessment frameworks; this text from the current policy already indicates that other references to "framework" are shorthand for "assessment framework."

The Governing Board, through its Assessment Development Committee (ADC), shall monitor the framework development and update processes to ensure that the final Governing Board-adopted framework and~~,~~ specifications~~, contextual variables documents,~~ and their development processes comply with all principles and guidelines of the Governing Board Assessment Framework Development Policy.

Introduction

Since its creation by Congress in 1988, the Governing Board has been responsible for determining the content and format of all NAEP assessments. The Governing Board has carried out this important statutory responsibility by engaging a broad spectrum of stakeholders in developing recommendations for the knowledge and skills NAEP should assess in various grades and subject areas. From this comprehensive process, the Governing Board develops a framework to outline the content and format for each NAEP assessment at grades 4, 8, and 12. Development of a framework for a new assessment is guided by the schedule of NAEP assessments adopted by the Governing Board.

Under provisions of the National Assessment of Educational Progress Authorization Act of 2002 (P.L. 107-279), Congress authorized the Governing Board to continue its mandate for determining the content and format of valid and reliable assessments based on widely accepted technical and professional ~~testing~~ standards for test development and active participation of stakeholders. This mandate aligns with the purpose of NAEP, which is to provide fair and accurate measurement of student academic achievement.

Given this mandate, the Governing Board must ensure that the highest standards of test development are employed in framework development to support the validity of educational inferences made using NAEP data. The Governing Board Item Development Policy separately details principles and guidelines for NAEP assessment items, and the Governing Board has final authority on the appropriateness of all assessment items.

By law, NAEP assessments shall not evaluate personal beliefs or publicly disclose personally identifiable information, and NAEP assessment items shall be secular, neutral, and non-ideological and free from racial, cultural, gender, or regional bias.

~~The~~ NAEP framework development ~~and update processes~~ shall be informed by a broad, balanced, and inclusive set of factors. F~~The f~~rameworks 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.

Commented [A2]: RATIONALE: This text was moved from the previous Principle 4

To develop the recommended framework for Board adoption, the Governing Board convenes stakeholders (via panels and broad outreach) to identify and/or provide feedback on the content and design for each NAEP assessment.

In this process, involved stakeholders shall include:

Teachers	Policymakers
Curriculum Specialists	Business Representatives
Content Experts	Parents
Assessment Specialists	Users of Assessment Data
State Administrators	Researchers and Technical Experts
Local School Administrators	Members of the public

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. A procedures manual shall provide additional detail about how this Policy is implemented.

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.

National Center for Education Statistics (NCES) Statistical Standards. (2012).

Principles for Framework Development

Principle 1: Elements of Frameworks

Principle 2: Development and Update Process

Principle 3: Periodic Framework Review

~~**Principle 4: Resources for the Process**~~

Principle 4~~5~~: Elements of Specifications

Principle 5~~6~~: Role of the Governing Board

Guidelines for the Principles

Principle 1: Elements of Frameworks

The Governing Board is responsible for developing a framework for each NAEP assessment. The framework shall define the scope of the domain to be measured by delineating the knowledge and skills to be tested at each grade, the format of the NAEP assessment, ~~and~~ the achievement level descriptions, and recommendations for subject-specific contextual variables~~s~~.

Guidelines

a) The framework shall determine the extent of the domain and the scope of the construct to be measured for each grade level in a NAEP assessment. The framework shall provide information to the public and test developers on three key aspects of the assessment:

- *What* is to be measured, including definitions of the constructs being assessed and reported upon and descriptions of the purpose(s) of the assessment;
- *How* that domain of content is most appropriately measured in a large-scale assessment, including the format requirements of the items and the assessment, the content and skills to be tested at each grade, sample items for each grade to be tested, the weighting of the item pool in terms of content and cognitive process dimensions, and any additional requirements for the assessment administration unique to a given subject area, such as provision of ancillary materials and uses of technology; and
- *How much* of the content domain, in terms of knowledge and skills, should students know and be able to do at the ~~*basic*~~*NAEP Basic*, ~~*proficient*~~*NAEP Proficient*, and ~~advanced~~ *NAEP Advanced* levels in achievement level descriptions for each grade to be tested. The achievement level descriptions shall be based on the Governing Board's policy definitions for ~~basic~~*NAEP Basic*, ~~proficient~~*NAEP Proficient*, and ~~advanced~~ *NAEP Advanced* achievement and shall incorporate the content and process dimensions of the assessment at each grade.

- ~~The framework shall determine the construction of items for each NAEP assessment. The achievement level descriptions in each framework shall also be used in the level-setting process.~~

b) The framework shall inform the development of subject-specific contextual questionnaires for students, teachers, and school administrators by identifying variables that may help contextualize the assessment results (See the Governing Board Policy on Collecting and Reporting Contextual Data).

c) 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.

d) Content coverage in each subject and grade shall be broad, inclusive of content valued by

Commented [A3]: RATIONALE: This sentence is unclear and redundant with the previous guideline.

Commented [A4]: RATIONALE: The Board policy on achievement level setting already describes (with more nuance) how the achievement level descriptions created with the framework will be used in the achievement level setting process.

Commented [A5]: RATIONALE: To make clear that contextual variable recommendations should be incorporated into the framework and specifications documents and do not need to be an additional document.

the public as important to measure, and reflect high aspirations for student achievement. ~~(See *Principle 4* for more detail on the factors balanced in content coverage.)~~

e) Frameworks shall be written to be clear and accessible to educators and the general public. The framework shall use clear language, accessible to educators and to the general public, and contain ~~sufficient~~ information ~~to inform all stakeholders~~ about the nature and scope of the given assessment. Following Governing Board adoption, the framework shall be widely disseminated.

Commented [A6]: PROCECURES MANUAL will describe a role for communication experts in the framework development process. In addition to ensuring that the language is clear and accessible, communications experts can help surface controversial issues in a subject area, provide advice for navigating potential challenges throughout the process.

Principle 2: Development and Update Process

Commented [A7]: RATIONALE: Principle 2 has been revised to reflect the general sequence and steps of the framework development process and to be more parallel with the achievement levels policy. It is intended to address a concern raised by some stakeholders that the steps and sequence of the current process is unclear.

The Governing Board shall develop and update frameworks through a comprehensive, inclusive, and deliberative process that involves active participation of stakeholders listed in the Introduction section.

Guidelines

~~a) 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.~~
- ~~*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.~~

a) When the Board reviews a framework for potential updates (see *Principle 3*), Board deliberations shall begin by discussing major policy and assessment issues in the content area. Such issues may be identified through seeking and collecting public comment, as well as through engaging relevant content experts.

Commented [A8]: RATIONALE: It is important to surface a broad range of views and potential controversies for a given subject at the outset of the framework development process so that it can inform initial Board direction and the selection of panelists to represent diverse perspectives on the issues that are of most importance to the Board.

b) After considering the policy and assessment issues in the content area, the Board shall develop a charge to articulate the need for an update to the framework and to specify policy guidance, constraints (including but not limited to those imposed by the NAEP legislation), and any specific tensions to resolve in the development of framework recommendations. The Board charge shall explicitly address whether maintaining trends with assessment results from the previous framework should be prioritized above other factors.

Commented [A9]: RATIONALE: Providing more direction from the Board in the initial charge will help prevent a situation where the Board provides policy guidance and imposes constraints only after framework panels have worked for a year or more on a draft document.

~~b)~~c) Framework development and update processes shall be executed primarily via two panels: a ~~Visioning~~ Steering Panel with a subset of members continuing as the Development Panel. This process shall result in ~~three~~ two documents for Board consideration: a recommended framework and~~,~~ recommended assessment and item specifications~~, and recommendations for contextual variables that relate to the subject being assessed~~. For each framework,

Commented [A10]: RATIONALE: The name "Visioning Panel" may contribute to a perception that the group's charge is very aspirational; "Steering Panel" may better communicate the necessity of being bound by constraints and parameters. (Prior to the 2018 Board policy, the initial panel was called the "Steering Committee"; "Steering Panel" is proposed until/unless a better alternative can be identified).

- *The Framework ~~Visioning~~Steering Panel* shall formulate high-level guidance about the state of the field and how to implement the Board charge to inform the process~~, providing these in the form of guidelines~~. The major part of the ~~Visioning~~Steering Panel work will be at the beginning to provide initial guidance for developing a recommended framework. The ~~Visioning~~Steering Panel shall be comprised of the stakeholders referenced in the I~~i~~ntroduction ~~above~~section. At least 20 percent of this panel shall ~~have classroom teaching experience~~be current classroom teachers in the subject areas under consideration. This panel may include up to 30 members with additional members as needed.
- *The Framework Development Panel* shall develop drafts of the ~~three~~ two project documents and engage in the detailed deliberations about how issues outlined in the Board charge and ~~Visioning~~Steering Panel discussion should be reflected in a recommended framework. As a subset of the ~~Visioning~~Steering Panel, the Development Panel shall have a proportionally higher representation of content experts and educators, whose expertise collectively addresses all grade levels designated for the assessment under development. At least 30 percent of this panel shall be current classroom teachers in the subject areas under consideration. Educators shall be drawn from schools across the nation, including individuals who work with students from high-poverty and low-performing schools, as well as public and private schools. This panel may include up to ~~15~~ 20 members, with additional members as needed.

~~c) In addition to a recommended framework, the framework development or update process shall result in assessment and item specifications (see *Principle 5*) and recommendations on related contextual variables to be collected from students, teachers, and school administrators. Recommendations shall take into account burden, cost, quality of the data to be obtained, and other factors. (See the Governing Board Policy on Collecting and Reporting Contextual Data.)~~

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. (See *4.b*).~~

e) A nominations process shall be used to seek broad input on recommendations for well-qualified individuals who represent diverse demographic characteristics, stakeholder groups, and perspectives on the key issues identified in the Board charge to the panels.

~~d)~~f) From the pool of nominees, the Board will select those with the most outstanding content and education credentials to represent multiple perspectives on the key issues identified in the Board charge to the panels. The ADC shall review panelist nomination materials and recommend a slate of panelists, which shall be subject to Executive Committee approval.

Commented [A11]: PROCEDURES MANUAL will contain additional detail about the qualifications of panelists, including diversity of demographic characteristics, definition of classroom teaching experience, need for individuals with previous assessment experience which may include a state testing director, etc.

Commented [A12]: RATIONALE: This is redundant with information included in Principle 1.

Commented [A13]: RATIONALE: This is redundant with the Principles 2a and 2b.

Commented [A14]: RATIONALE: An open call for nominations enables all interested stakeholders to recommend potential panelists.

Commented [A15]: RATIONALE: To allow for more Board member involvement in the selection of panelists without unnecessarily adding significant time to the framework development process.

~~e)~~g) The process that panels employ to develop recommendations for new or updated frameworks ~~Framework development and updating~~ shall be comprehensive in approach and conducted in an environment that is open, balanced, and even-handed. Panels shall consider all viewpoints and debate all pertinent issues in formulating consensus recommendations on the content and design of a NAEP assessment, including findings from research. Reference materials shall represent multiple views.

h) For each ~~project~~new or updated framework, 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 Board deliberations.

i) The framework panels shall consider a wide variety of resources during deliberations, including but not limited to relevant research, trends in state and local standards and assessments, use of previous NAEP results, curriculum guides, 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, if available.

Commented [A16]: RATIONALE: Moved from Principle 4 and edited for clarity.

j) A Technical ~~experts shall be involved~~Advisory Committee shall be convened to uphold the highest technical standards for development of the NAEP framework and specifications. As a resource to the framework panels, these ~~-~~ experts shall respond to technical issues raised during panel deliberations.

Commented [A17]: RATIONALE: Moved from Principle 4 and the convening of a TAC was made more explicit.

k) An Educator Advisory Committee shall be convened to include additional practitioners in the framework development process. As a resource to the framework panels, these practitioners shall provide meaningful consultation on issues raised during panel deliberations that need input from those in the field teaching the subjects being assessed.

Commented [A18]: RATIONALE: In addition to teachers now comprising 6 out of 20 Development Panelists, this new advisory committee would be a means of soliciting additional input in a manner that is less burdensome to teachers than serving on the panel.

PROCEDURES MANUAL will specify details for the composition of this group and their more specific charge.

l) Public comment shall be sought from a broad array of stakeholders and interested members of the public to reflect multiple perspectives on the draft framework recommendations that have been developed. Outreach efforts should directly engage all stakeholder groups identified in the Introduction section.

Commented [A19]: RATIONALE: Reference to public comment moved from previous Principle 2a and further clarified.

m) If the Development Panel or the Board cannot reach consensus on key issues in the framework, the Board may decide to seek further stakeholder input such as through additional public comment and/or independent reviews by content experts on a framework that has been significantly revised following an earlier public comment period. The Board shall determine whether and how any further revisions to a framework shall be made.

Commented [A20]: RATIONALE: To be transparent about additional steps that could be taken if necessary to achieve a broader consensus.

n) The final framework and specifications documents are subject to full Board approval.

Principle 3: Periodic 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.

Guidelines

a) 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. In the review, the ADC shall solicit input from experts to determine if changes are warranted, making clear the potential risk to trends and assessment of educational progress posed by~~of~~ changing frameworks ~~to trends and assessment of educational progress~~. The Board may decide based on the input that the framework does not require revision, or that the framework may require minor or major updates. To initiate updates, the ADC shall prepare a recommendation for full Board approval. Minor updates include clarifications or corrections that do not affect the construct defined for the assessment. Major updates shall include the convening of a ~~Visioning~~Steering Panel (see *Principle 2*). Framework revisions shall also be subject to full Board approval.

Commented [A21]: PROCEDURES MANUAL will further clarify what constitutes a minor update.

b) 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 deliberate on whether such changes warrant an accelerated schedule of updates to a framework and may recommend that ~~determine whether and how changing conditions warrant an update and~~ the Governing Board ~~via recommendation may~~ convene a ~~Visioning~~Steering Panel to revise or replace the framework. Before framework panels are convened, special research and analysis may also be commissioned to inform the updates to be considered.

c) If the ~~Visioning Panel recommends~~Board charge directs a Steering Panel to recommend ~~major~~ framework updates, then a subset of Steering P~~p~~anel members shall continue as the Development Panel to develop the draft framework and assessment and item specifications, in accordance with *Principle 2*. Regular reports will be provided to the ADC and the recommended framework update shall be subject to full Board approval.

d) When a framework update is conducted, framework ~~Visioning~~Steering and Development Panel recommendations shall describe the extent to which adjustments in the achievement level descript~~ors~~ions (see *1.a*) and contextual variables (see *1~~2~~.b~~e~~*) are needed. (See the Governing Board Policy on Achievement Levels and the Governing Board Policy on Collecting and Reporting Contextual Data for additional details.)

~~Principle 4: Resources for the Process~~

Commented [A22]: RATIONALE: With the reorganization of the policy, this information does not need to be presented in a separate principle; instead this information will be included in the Introduction section, Principle 2, and some of the more specific details will be moved to the Procedures manual.

~~**Framework development and update processes shall 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.~~

~~Guidelines~~

~~a) An initial compilation of resources shall summarize relevant research, advantages and disadvantages of the latest developments, and trends in state standards and assessments for the content area. This compilation shall also summarize how stakeholders have used previous NAEP student achievement trends in the assessment area. The compilation may include public comment. Using this compilation as a springboard, framework panel deliberations shall begin by thoroughly identifying major policy and assessment issues in the content area.~~

~~b) The framework panels shall also consider a wide variety of resources as deliberations proceed, including but not limited to 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, if available.~~

~~c) Technical experts shall be involved to uphold the highest technical standards for development of the NAEP framework and specifications. As a resource to the framework panels, these experts shall respond to technical issues raised during panel deliberations.~~

~~d) 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.~~

Principle 4~~5~~: Elements of Specifications

The specifications document shall be developed for use by the National Center for Education Statistics (NCES) as the blueprint for constructing the NAEP assessment and items.

Guidelines

a) The assessment and item specifications shall produce an assessment that is valid, reliable, and based on relevant widely accepted professional standards. The specifications shall also be consistent with Governing Board policies regarding NAEP design, such as groupings of items, test administration conditions, and accommodations for students with disabilities and English language learners. (S~~S~~ee the Governing Board Policy on NAEP Testing and Reporting on Students with Disabilities and English Language Learners). The specifications shall be reviewed by technical experts involved in the process, prior to submission to the Governing Board.

b) The primary audience for the specifications, or assessment blueprint, shall be NCES and the contractor(s) responsible for developing the assessment and the test questions.

c) The specifications shall evolve from the framework and shall be written in sufficient detail so that item writers can develop high-quality questions based on the framework objectives for grades 4, 8, and 12, where applicable. The specifications shall include, but not be limited to detailed descriptions of:
   - the content and process dimensions, including the weighting of those dimensions in the pool of questions at each grade;
   - types of items;
   - guidelines for stimulus material;
   - types of response formats;
   - scoring procedures;
   - achievement level descriptions;
   - administration conditions;
   - ancillary or additional materials, if any;
   - considerations for special populations;
   - sample items, including a substantial number and range of sample items with scoring guidelines for each grade level; and
   - any unique requirements for the given assessment.

d) Special studies, if any, to be ~~conducted as part of the assessment~~recommended in support of the framework shall be described in the specifications. This description shall provide an overview of the purpose and rationale for the study~~, the nature of the student sample(s), and a discussion of the instrument and administration procedures~~.

Principle 5~~6~~: Role of the Governing Board

The Governing Board, through its ~~Assessment Development Committee~~ADC, shall monitor all framework development and updates. The result of this process shall be recommendations for Governing Board action in the form of two~~three~~ key documents: the framework and~~;~~ assessment and item specifications~~; and contextual variables that relate to the subject being assessed~~.

Guidelines

a) The ~~Assessment Development Committee (~~ADC~~)~~ shall be responsible for monitoring framework development and updates that result in recommendations to the Governing Board on the content and format of each NAEP assessment. The ADC will provide direction to the framework panels, via Governing Board staff. This guidance shall ensure compliance with the NAEP law, Governing Board policies, Department of Education and government-wide regulations, and requirements of the contract(s) used to implement the framework project.

b) In initiating a framework update, the Governing Board shall balance needs for stable reporting of student achievement trends against other Board priorities and requirements.

Commented [A23]: QUESTION: This section has typically been a summary of the Board's role as described in various places throughout the policy statement. Would it be better to move this to the beginning as an overview of the Board's role rather than a summary?

Commented [A24]: RATIONALE: This guideline was moved from below to be consistent with the order of activities.

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.

~~a)~~c) When a framework ~~Visioning~~Steering Panel is to be convened, the ADC shall develop a charge for the panel, and the charge shall be subject to full Board approval (See *2.b.*). ~~The charge will outline any special considerations for an assessment area.~~

d) The ADC shall review panelist nomination materials and develop a recommended slate of panelists, and the panelist recommendations shall be subject to Executive Committee approval.

e) The ADC shall receive regular reports on the progress of framework development ~~and updates~~.

~~b)~~ The full Board shall receive periodic updates about how the Board charge is being implemented and any additional policy considerations that arise during the development process, including from public comment.

~~c)~~f)

~~a) 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.~~

~~d)~~g) At the conclusion of the framework development or update process, the Governing Board shall take final action on the recommended framework and~~,~~ specifications~~, and contextual variables~~. The Governing Board shall make the final decision on the content and format of NAEP assessments. In addition to the panel recommendations, the Board may take into account other pertinent considerations on the domain and scope of what should be assessed, such as the broader policy context of assessment in the subject area under consideration.

~~e)~~ Following adoption by the Governing Board, the final framework and~~,~~ specifications~~, and contextual variables~~ shall be provided to ~~the National Center for Education Statistics (~~NCES~~)~~. These documents, which include the achievement level descriptions for *NAEP Basic*, *NAEP Proficient*, and *NAEP Advanced* ~~performance definitions~~ and recommendations for contextual variables in the subject area, are provided to NCES to guide development of NAEP test questions and questionnaires.

~~f)~~h)

Commented [A25]: RATIONALE: To be more explicit that the Board is not bound by the panel recommendations, analogous to a statement that appears in the Board policy on achievement level setting.

Discussion of Initial Public Comment on Current NAEP Science Framework

The NAEP Assessment Schedule indicates that the Board will consider whether updates to the NAEP Science Framework are needed for the administration of the 2028 assessment and beyond.

Current NAEP Science Framework

The current framework was adopted by the Board in 2005 and implemented beginning with the 2009 NAEP science assessment at grades 4, 8, and 12. The framework includes two dimensions: content and practices.

The science content for NAEP is defined by a series of statements that describe key facts, concepts, principles, laws, and theories in three broad areas:

- Physical Science
- Life Science
- Earth and Space Sciences

Physical Science deals with matter, energy, and motion; Life Science deals with structures and functions of living systems and changes in living systems; and Earth and Space Sciences deal with Earth in space and time, Earth structures, and Earth systems.

The second dimension of the framework is defined by four science practices:

- Identifying Science Principles
- Using Science Principles
- Using Scientific Inquiry
- Using Technological Design

These practices can be combined with any science content statement to generate student performance expectations, and assessment items can then be developed based on these performance expectations.

The framework specifies that 50 percent of the assessment time should be devoted to multiple choice items and the remaining 50 percent should be constructed response items. For each grade level, the constructed response items are intended to include at least one hands-on performance task and at least one interactive computer task.

Trends in State Science Standards

The Board's Framework Development policy calls for using information about trends in state standards as one resource in the decision-making process of whether and how a framework should be updated. In 2016, the American Institutes for Research (under contract to the National Center for Education Statistics) conducted a comparison study of the Next Generation Science Standards (NGSS) and the NAEP Science, Technology and Engineering Literacy (TEL), and Mathematics frameworks. The degree of overlap between the NGSS and NAEP varied across grades and depending on whether the NGSS were compared to the NAEP Science Framework only or whether the TEL and/or Mathematics frameworks were also included. The summary and conclusions are detailed on PDF pages 103-108 of the technical report.

Earlier this year, Board staff commissioned an additional study under a previous contract with the Human Resources Research Organization (HumRRO) to better understand how the NAEP Science Framework overlaps with state standards for the states that did not fully adopt the NGSS – including states that partially adopted the NGSS and states that did not adopt the NGSS. As with the study of NAEP and NGSS, there was some overlap and some important differences between NAEP and state science standards, with variation across grades and content areas. The discussion and conclusions appear on PDF pages 35-36 of the report.

Public Comment

Under the leadership of the Assessment Development Committee (ADC), the Board has been discussing how to strengthen existing processes and procedures for updating NAEP frameworks. One proposed improvement is to conduct a public comment period on the current assessment framework to seek broad input upfront on whether and how the current framework should be updated. Consequently, the Board conducted an initial public comment on the current NAEP Science Framework from August 20 – October 15, 2021. Commenters were asked to address three questions:

- Whether the NAEP Science Assessment Framework needs to be updated
- If the framework needs to be updated, why a revision is needed
- What a revision to the framework should include

The purpose of seeking public comment on the current framework is to surface a broad range of views related to a given subject at the outset of the framework development process. This initial comment then can inform initial Board direction and the selection of panelists to represent diverse perspectives on the issues that are of most importance to the Board.

Thirty submissions were received from a variety of individuals, groups of individuals, and organizations. In addition, Board staff sought input from the National Center for Education Statistics (NCES) on operational issues and challenges associated with the current framework and assessment; a memo was submitted by NCES to summarize their feedback. The raw comments are attached, along with a summary of specific points raised by major theme.

Discussion

The purpose of the November plenary discussion is:

- To identify what information is needed for the Board to make a determination of whether and how the NAEP Science Framework should be updated;
- To identify the key issues/topics for which the Board may want to provide policy guidance to the framework panels; and

- To identify what additional input and expertise (e.g., commissioned white papers, expert panels) is needed to inform the policy guidance to be set forth in a Board charge to the framework panels

ADC Chair Dana Boyd and ADC member Christine Cunningham will facilitate the discussion. Following the November Board meeting, Board staff will commission targeted expert input on the key issues identified to inform future Board decisions during spring 2022 on whether and how to update the NAEP Science Framework.

Summary of Public Comments Received on the Current NAEP Science Framework[1]

November 4, 2021

Contributors

Spurlock, Holly	National Center for Education Statistics
Pellegrino, James	University of Illinois Chicago (NAEP Validity Studies Panel white paper)
Petersen, Anne	Virginia Department of Education
Moulding, Brett	Retired Utah State Office of Education Curriculum Director and Instruction Former NAEP Science Advisory Committee Member
Sneider, Cary	Former NAGB Member
Gordon, David	CAST (originally Center for Applied Special Technology)
Finn Jr., Chester E.	Thomas B. Fordham Institute
Murphy, Stephen	Cognia
Heinz, Michael	Council of State Science Supervisors
Murphy, Danielle	*Affiliation not provided*
Reid, Ann	National Center for Science Education
Foster, Jacob	STEM learning Design, LLC
Huntoon, Jacqueline	Michigan Technological University
Barber-Lester, Kelly	University of North Carolina Pembroke
Wray, Kraig	Pennsylvania State University
Looy, Mark	Answers in Genesis
Lowry, Michael	The McCallie School
Wysession, Michael	NSF's Earth Science Literacy Initiative Earth and Space Science for the NRC's Framework for K-12 Science Education Earth and Space Science for the Next Generation Science Standards Washington University St. Louis
McCarthy, Michelle	Montana Office of Public Instruction
Multiple Authors	Georgia State University
Haverly, Christa Marie Marshall, Stephanie Kayumova, Shakhnoza Cheuk, Tina Basile, Vincent McDonald, Scott Taylor, Jonte' C.	Northwestern University University of Minnesota Twin Cities University of Massachusetts Dartmouth California Polytechnic State University San Luis Opispo Colorado State University Pennsylvania State University Pennsylvania State University
	National Science Teaching Association—Statement endorsed by the Council of State Science Supervisors and the National Science Education Leadership Association

[1] This summary was produced by Dr. Arthur Thacker of the Human Resources Research Organization under subcontract to the Manhattan Strategies Group as part of contract 919995921F0002, Technical and Logistical Services.

	National Science Education Leadership Association
Settlage, John	University of Connecticut
Schwartz, Renee	National Association for Research in Science Teaching (NARST) Georgia State University
Badrinarayan, Aneesha	State Performance Assessment Learning Community (SPA-LC) coordinated by the Learning Policy Institute
Sterling Burnett	Heartland
Codere, Susan	Multiple Literacies in Project Based Learning
Keller, Tom	STEM Education Strategies, LLC
Thomas Tretter	*Affiliation not provided*
Bryan, Carl	Wisconsin Department of Public Instruction

Overall Summary

Twenty eight of the 31 submitted comments recommended some level of revision for the NAEP Science Framework. Most of those comments focused on bringing the framework into alignment with state standards (including but not limited to the Next Generation Science Standards (NGSS)) and improving equity and fairness for all tested students. There were also several comments regarding assessment design and accessibility for all students. Suggested revisions ranged from minor editorial comments to significant overhaul of the framework. *(Note that not all submitters responded directly to the question of "Whether the 2019 NAEP Science Framework needs to be updated." The count is based on the content of the submissions and whether the submitters recommended changes to the current framework.)*

Alignment to NGSS/National Academies Framework (three dimensional standards)

Fifteen of the 31 submitted comments focused, either fully or in part, on updating the NAEP Science Framework to better align with the National Academies Framework and NGSS. Most comments centered around current changes in state standards and teaching and learning and concerns that NAEP assessments would not accurately reflect student performance due to a misalignment between what NAEP tests and what is happening in classrooms. Several of these comments suggest including content from the NAEP Technology and Engineering Literacy (TEL) Framework in the science assessment. A couple of comments suggest merging science and TEL, but there are cautions provided in the full text for that suggestion as well. Conversely, there were three comments cautioning the Governing Board not to make substantive changes in the framework (one specifically indicating that the Board's mission is not to follow NGSS). Summary comments follow in bullet form.

Specific comments received:

- The NAEP Science Framework does not approach science as three dimensions, Science and Engineering Practices (SEP), Disciplinary Core Ideas (DCI), and Crosscutting Concepts (CCC). Revisions should include a clear alignment to the National Academies Framework for K-12 Science Education.

- Merging content from the TEL would improve alignment to NGSS. The TEL might be eliminated, and engineering practices (and technology) incorporated into what is considered science. NGSS includes much of the first two TEL components—designs and systems and technology and society. The third, communications technology, is more closely related to English Language Arts (ELA) than science.

- Attend to shifts in grade levels for content learning progressions. This is especially relevant if NAEP adopts a three-dimensional framework, where the interactions among DCI, SEP, and CCC could potentially cross grade levels for a given phenomenon. It is vital that the assessment items measure constructs that are appropriate for the intended grade level.

- Consider changing the assessed science grade from 4 to 5. The NGSS organized elementary standards for grades K-5, middle school standards for grades 6-8. Many states administer their assessments in grade 5. This might make NAEP science results more comparable and relevant for states.

- Tease out research since the science framework was updated. States have largely changed their standards.

- Frameworks must redefine content, practices, and crosscutting concepts to align to the way they are operationalized in the NGSS. Framework practices overlap NGSS practices, but are too broad to focus on specific expectations of current science instruction.

- Crosscutting concepts in the current NAEP Framework are anchored in the content statements themselves. NGSS and more recent literature refer to crosscutting concepts in a more theme-based way, like the NAEP Science Framework did from 1996-2005. The NAEP framework should adopt the seven crosscutting concepts included in the NGSS, or relabel the current crosscutting content if more substantial revisions are not made.

- Two consensus studies of the National Academy of Science include *Taking Science to School: Learning and Teaching Science in Grades K-8* (2007) and a *Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas* (2013). Forty-four states (representing 71% of U.S. students) have science standards influenced by the Framework for K-12 Science Education.

- Assessment can drive instruction forward or backward. Coherence between federal and state assessment will provide state leaders with another tool to improve science instruction for all students.

- The current NAEP Science Framework has two separate components, science content and science practices. Framework for K-12 Science Education also defines distinct practices, core ideas, and crosscutting concepts—the difference is the expectation that they are integrated in instruction and assessment.

- Integration of science practices and content is vital and may require attention to the measure of each construct independently, plus a measure of the integrated abilities of students.

- The current framework is too differentiated by discipline. Interesting problems in science are less and less likely to be confined to one particular discipline.

- Frameworks for NAEP Science and NAEP TEL were developed before the NRC Framework and NGSS. All drew upon bodies of theory, research, and practice regarding the knowing, learning, and teaching of science and technology available at the time of their development. There are significant similarities, and substantial differences between the two NAEP frameworks and the NGSS.

- Alignment differences between NAEP and NGSS are magnified as grades increase from 4 to 8 to 12. NGSS is more interdisciplinary across grade levels, while NAEP shifts toward physical science in grades 8 and 12, especially grade 12.

- NGSS science practices are more demanding than NAEP practices and focus more on "doing science" rather than knowing science.

- NGSS performance expectations are viewed to demand more than NAEP performance expectations in terms of application of disciplinary content. This leads to misalignment even if the science content covered by both frameworks is similar.

- Combining NAEP Science and TEL might improve alignment to state standards, but the two NAEP frameworks are quite different. If content from the TEL is to be included in science, the high variability of overlapping content by grade must be accounted for. Items/tasks would also need to be redesigned as TEL tasks intentionally omit relevant science content. An assessment aligned to NGSS would look substantially different from assessments aligned to either NAEP Science or NAEP TEL.

- Given state science standards adoptions, the current NAEP Science Framework and assessment may be substantially at variance with a relatively pervasive national perspective on what is desired for students to know and be able to do in science at grades 4, 8, and 12 and how they could be expected to show proficiency via large-scale assessment.

- Evidence shows that adoption of the new science standards has been staggered across time since 2013, as has been the design and implementation of state large-scale assessments aligned to those new standards. The latter invariably lag two or more years behind adoption of new state standards. The most recent national survey of science education (2018) suggests that little changed between 2012 and 2018 in science instructional practice. Results from the NAEP science assessment from 2009 to 2019 also show little in the way of change in student performance across time.

- If substantive review of the frameworks is completed to better align with NRC and NGSS, then the meaning of science proficiency should also be considered. The ability to integrate content and practice knowledge consistent with the separate but related considerations of science and engineering content is key.

- Consider inclusion of technology and engineering content similar to the TEL and whether it would be appropriate to merge the science and TEL frameworks.

- Integrating the NAEP Science and TEL assessments would have benefits in terms of cost savings and alignment, so the Governing Board may wish to consider merging the two frameworks.

- Remove the silos represented by traditional course disciplines in life, physical, and earth science and address the cross-fertilization that is currently happening in STEM (as found in NGSS).

- Emphasize the scientific practices modeling and argumentation. New assessment items should be heavily connected to the modeling process. Argumentation can foster students' abilities to evaluate claims using evidence and consider concepts like confirmation bias and other fallacies.

- Current standards are based on research that originated before 2005. It should be updated to reflect the more current understanding of science education described by the NGSS.

- The NAEP framework is broad but needs to more accurately reflect the depth of learning and application that is now expected of students.

- Given the likely scope of a revision to the NAEP Science Framework and the implications for the 2028 assessment, as well as the possibility of incorporating aspects of TEL in the new framework and assessment, it seems highly likely that preserving the science or TEL trend through 2028 will not be feasible or advisable. Priority should go to insuring the validity of the revised science framework and assessment for 2028 and beyond. Doing so should not be compromised in a possibly misguided effort to preserve trend at all costs.

Equity/Diversity

The second most prevalent comment topic regarding potential framework revisions had to do with ensuring equity among diverse populations of students. Fourteen of the 31 submissions included equity/diversity as a major theme in their comments. The comments ranged from general concerns about the ways that NAEP reports data on student subgroups, to very specific concerns regarding students' opportunities to learn and the representation of the majority group (higher socioeconomic white students) in the content of the test items. Several comments focused on ensuring that the represented science was not taken out of context, but that context be included to make the phenomena and problems more genuine for students.

Specific comments received:

- Lack of physics courses/teachers, especially during year one of high school, and especially for minority and high-poverty student populations, may conflate performance and opportunity to learn first year physics concepts.

- The COVID-19 pandemic shined a spotlight on inequities and unjust public education practices. As an organization that is not constrained by limitations created by statewide policies, the Governing Board should position itself to take up that work and to exemplify how large-scale assessments can provide equitable opportunities for all students to make their thinking visible.

- The following words and phrases are completely absent from the NAEP Science Frameworks—equity, equality, inequality, racism, bias, scientific racism, prejudice, sexism, or ethics. The term race is only used for tracking subgroup performance, and culture is limited to the role of science in influencing cultures. There is no discussion of bias or the mitigation of bias—a well-established and ongoing concern in education.

- The framework presents a vision and version of science as objective, neutral, and divorced from context, despite its unquestionably troubled history (and present) as it pertains to issues of inequity broadly, and specifically racism and sexism.

- Update references and acknowledge advances in understanding of student diversity and cultural relevance.

- Expand the meaning of diversity (beyond students with disabilities and English learners) consistent with more recent NAEP resources (e.g., NAEP TEL Framework).

- Emphasize diversity, equity, accessibility, and inclusion to support learning, increase engagement, and provide visible representation in content with a goal to improve diversity in representation of underrepresented groups in science fields of study and the workplace.

- Make students the focal point of the assessment and include meaningful feedback loops with the community as reflected by the students' contexts and communities.

- Create a practice for understanding diverse learners and connecting them to science activities, including outreach and engagement with family community members. This would inform assessment development, curriculum integration, and solving real problems.

- Adopt a "growth mindset" strategy for revisions that promotes self-efficacy and motivation to learn from mistakes, then expand scientific skills centered on real world/life problem solving and knowledge.

- Connect the performance expectations to students' lived experiences (e.g., relevant phenomena). Equitable and inclusive performance expectations guide the development of assessment items and tasks.

- Develop assessments that reflect the mindsets and habits of professionals in the field and that “this shift from students as consumers of information to practitioners of field knowledge is especially significant for Black, brown and Indigenous students, signaling that they belong to a larger intellectual community” (Safir and Dugan, 2021). The assessments that students encounter should include tasks that elicit authentic student performance to the extent practicable.

- Expand the definition of “assessment of design” to include other considerations beyond scientific principles (e.g., economic, social) to better engage students with more relevant problems based on their lived experiences and social justice.

- Incorporate cross-sectional views of item DIF (e.g., low SES Black females). Real differences may be being washed out by the ways student subgroups are currently defined.

- Include representatives from traditionally underrepresented subgroups in all development processes—from developing the frameworks to developing test blueprints, selecting phenomena for testing, item writing, and development of scoring rubrics/criteria.

- New research outlined in research like How People Learn II: Learners, Contexts, and Cultures (2018) provides further input regarding integration of content and practice for improved and more equitable outcomes. Students do not use their knowledge of content, practice, and cross-cutting concepts in isolation of one another. The knowledge interacts in ways that provide scaffolding for recall, integration and problem solving in the context of a novel or repeat phenomenon(a). As noted by the Achieve Framework for evaluating cognitive complexity, artificially separating these cognitive processes in assessment does not provide us with an accurate or equitable measure of student proficiency in science. It is in our best interest to align our measures with instructional practice.

- The new framework should endeavor to focus on interpretations within communities and populations based on opportunity to learn (OTL) metrics while also maintaining an ‘asset’ orientation in all interpretations, rather than traditional ‘deficit’ views that have been associated with large-scale assessments, such as NAEP, and the reporting of outcomes.

- OTL metrics must consider how students are given experiences to connect their science learning experiences through “forms of knowledge and ways of using language from their everyday experiences in families and communities.” This means broadening the collection of OTL data from districts, communities, and schools.

- Interrogate the assumptions about science knowledge embedded in the standards (i.e., whose histories and narratives are and are not included in this body of knowledge and practices).

- Update the technical aspects of the assessments themselves to be more inclusive of historically marginalized student populations.

- Invite people to participate in this review process, including on the expert panel, who are multilingual, of color, differently abled, and so on; leverage their expertise and lived experiences; and provide them with authority and agency to make substantive changes to the program.

- NAEP should stop fostering deficit explanations about achievement gaps via NAEP science results. NAEP should proactively develop reporting approaches that redirect media, political, and layperson discussions in ways that disrupt widespread beliefs that demographics dictate destinies. Requires more disaggregation and should point toward discussion toward remedies rather than promote ideas about gap inevitability.

- Support secondary research on equity and diversity in science education by allowing access to data and promoting relevant studies on the intersections of student gender, race, and social class.

- When NAEP does include cases where concepts are embedded in context, the contexts (e.g., hares in state park) feature the lived experiences of the dominant groups in U. S. society (e.g., upper middle class).

Accessibility

In addition to comments about equity and diversity generally, there were several comments specifically about accessibility. These comments were mostly about ensuring access to the NAEP assessments for all students. There is concern that NAEP does not assess students with the most severe cognitive disabilities. There were also comments requesting that accessibility be built into all aspects of NAEP test development, from adoption of frameworks through reporting of results.

Specific comments received:

- Incorporate principles of Universal Design throughout the framework. Adopt an inclusive validity framework that considers construct irrelevant factors that learners bring to testing. Include additional accessibility features for all students (including Els, SWDs, and non-identified students).

- Find a way to include students with the most significant cognitive disabilities (reference on the frameworks and include in testing).

- Young students may have insufficient access to and training in computer use for fair inclusion in digital assessments.

- Communities in digital deserts may have insufficient access to broadband services to support digital assessment.

- A major tenet of fairness, as conceptualized in the testing standards, is that assessment administrators must provide access for all examinees in various populations, particularly in allowing for accommodations and modification for learners with different cognitive, linguistic, and physical abilities (AERA/APA/NCME, 2014).

- Sample NAEP science items are laden with dense language and vocabulary, particularly in context-driven items. More consideration for English learners, beyond the current statements, must be put into practice in the development of NAEP science.

- It would be very useful for NAEP to develop equity indicators with respect to achievement and school and community factors, like those used in international assessments. Intentional attention to equity and social justice within the science curriculum and instruction are essential for developing scientific literacy.

- There are interactions between item difficulty and a student access to demonstrate knowledge of science practices. A large proportion of students score in the "Below Basic" performance category, and the large amount of contextual information may limit their ability to demonstrate what they can do. More items in the lower range of difficulty are needed to assess lower ability students.

Cautions Regarding Wholesale Revisions

While most of the received comments requested revisions to the Science Frameworks, there were a few (3) that promoted maintaining the framework as is. These comments posited that the current frameworks were of high quality and that NAEP functions as it is intended currently. There were concerns about maintaining trend and about tracking subgroup performance. Others commented that changes should be made in moderation to maintain the parts of the frameworks that are functioning well (e.g., the inclusion of sample items, focusing on scientific phenomena).

Specific comments received:

- 2012 comparisons between the NAEP Science Framework and state standards conducted by the Fordham Institute determined that the NAEP framework was of very high quality compared to most state standards. Minor updates may be required, but more substantive changes should only be made if absolutely necessary.

- NAEP should continue to include sample test items and complete explanations regarding what those items measure, how they are scored, and how they fit into the larger measurement construct in any revisions.

- The NGSS are already nine years old. Any revisions to NAEP frameworks should include a current literature review to ensure that a new NAEP framework is not outdated before it comes into use.

- Continue to ground assessment items in science phenomena and engineering design problems. A focus on sense making is what we now aspire to for our students.

- The NAEP Science Framework faces a precarious challenge: standardizing the instrument across time to identify longitudinal patterns, while accommodating changes in science education.
- The stated purpose of NAEP science assessment is to evaluate trends in scientific literacy overall and by demographic group. The current content, practices, and test design accomplish this goal. NAEP's purpose is not to mirror NGSS.

Editorial Updates

Editorial updates were included in many of the submitted comments, including a "marked up" version of the current framework. The bullets in this section are examples, but do not constitute the full range of edits, corrections, and clarifications submitted.

Example comments received:

- Eliminate references to NCLB and update to reflect current legislation (e.g., ESSA).
- Eliminate the term "special needs" and replace with "students with disabilities."

Addressing Controversial Subject Matter

Comments about controversial subject matter were inconsistent. They included: a call for NAEP to lead states in teaching socially, but not scientifically, controversial subjects; a request to omit controversial topics from the framework; and a request to ensure that minority views (e.g., creation science) are allowable in science teaching. Specific comments received:

- Special attention should be given to socially but not scientifically controversial topics. These specifically include evolution, climate change, and vaccination, as well as to the nature of science. It is counterproductive to make allowances for states that have chosen to under-educate or miseducate their students.
- A general framework should avoid discussion of scientifically disputed or politically charged issues such as anthropogenic climate change or embryonic stem cell research. If climate change is included, address the controversy regarding the quality of scientific evidence available to support the widely held conclusions.
- Inclusion of controversial ideas in the teaching of science is both legal and beneficial, particularly criticisms of evolution, the earth's age, and the reliability of dating methods. Teachers should not be required to teach creation science of ideas that support a younger age of the earth, but they should have the academic freedom to teacher alternative ideas—even if they happen to be in the minority.

Assessment Design

This section includes comments made regarding the assessment design. The interactions among framework objectives, tested content, and score reports are reinforced by the comments provided here.

Specific comments received:

- NAEP developers must be extremely transparent and explicit about the interpretations—and non-interpretations—of the assessment results based on the methodology in comparison to each particular state's standards and approach.

- Pay close attention to cognitive complexity—as a revision of the frameworks will require more complex items to effectively address the intended measurement construct.

- Increase emphasis on innovative item types, especially constructed response items and "predict, observe, and explain (POE)" items. Items may need to be clustered to address science concepts.

- Include and expand hands-on performance tasks, as these are fundamental to doing science and necessary to demonstrate the application of science.

- Include and expand the use of interactive computer tasks (ICT).

- Illustrative NAEP questions are too narrow in scope and tend toward acquisition of principles and facts. Broader test items should mirror our expectations for science teaching and learning in classrooms, assessing students broader understanding, integration, and use of scientific knowledge.

- NAEP should lead the way in designing science assessments that go beyond traditional large-scale multiple-choice tests. New approaches to science instruction allow many opportunities for informal assessment as student engage in investigations, create representations, and discuss evidence. Meaningful formal assessments will require careful articulation of the desired learning goals and how students can demonstrate that they have achieved them.

- The revision should include:

    1. Modeling as a practice. Students should be asked to create, evaluate, and/or revise models, and use them to predict the result of changes to system components. The development of explanatory models can help students make their thinking visible and can be an equalizer for English Language Learners.
    2. Planning investigations. Students should be able to identify independent and dependent variables and to design scientifically valid investigations.
    3. Analyzing data. Students should be able to analyze complex, real-world data using graphing and graphing analysis tools.
    4. Engaging in argument from evidence. Students should be assessed on their ability to use evidence to construct and justify a scientific claim.

- Measuring of two dimensions (content and practice) are ambiguous. In many cases, the experiences of the student dictate whether they access learned content knowledge or engage in science practice when interpreting an item's content (familiarity with the content/context dictates how the student approaches the problem). Items must have greater specificity regarding the nature of exactly what they are measuring.

- Hands-on Performance Tasks (HOTs) may need to be changed to hybrid models and included as interactive computer tasks due to practical and logistical considerations. Further research is required to determine if they can replace HOTs in terms of psychometrics and content validity.

- Prioritize students' active engagement in phenomena and sense making (figuring out) as the mechanism for science teaching, learning, and assessment.

- Allow for deeper exploration of phenomena by having sets of multiple items digging into a particular phenomenon.

NCES Comments Summary

NCES submitted comments relating to challenges and considerations presented by the current NAEP Science Framework for operationalizing the science assessments. Their issues are categorized into:

1. Ambiguous Content
2. Ambitious Content
3. Standardized Assessment Constraints
4. Implementation Considerations

1. Ambiguous guidance

   Learning progressions (LPs) are referenced heavily in the Science Framework. LPs are not clearly explicated, and their development has not been sufficient to cover the intended science content. Currently, cognitive demands and science practices proved the mental model and structure for measuring student progression in understanding science.

2. Ambitious Content

   Measuring two dimensions (content and practice) is a requirement for science items. There is not enough specificity around expectations for measuring two dimensions. The example items in the current framework show varying approaches, but do not provide guidance on what is acceptable or preferable. In fact, whether a student approaches an item from a content or practice perspective may depend on that student's lived experiences and science background. Several examples are provided.

   There is also concern that the NAEP items are too difficult for many of the test takers. Given how large the proportion of Below Basic students there are, the number of items in that range of the score scale is low. This issue is complicated by the inclusion of language-heavy context provided with items. The context may be needed by lower ability students, but may also contribute to issues with cognitive load and fatigue.

Quantitative reasoning in science. The Science Framework indicates that students' mathematics knowledge should be 1-2 grade levels below their current grade in science. However, the quantitative reasoning may require much higher math skills than even their current grade. As an example, fourth graders must interpret multiple distributions of data on a graph. Further examples from the released items are provided.

3. Standardized Assessment Constraints

   Concept maps require more time than is reasonable given a 30-minute cognitive block. Many students do not reach the end of the task. This is true for partial concept maps as well (on 8^{th} and 12^{th} grade).

   There are design limitations with hands-on performance tasks (HOTs). The 30-minute block, space allotted to the student, and limitations on the materials provided mean that students cannot truly freely design an experiment. Experimental hybrid hands-on performance tasks (HHOTs), administered digitally and completed virtually show promise (especially in terms of speededness). These items will need to be researched to ensure content validity and psychometric soundness.

4. Implementation Considerations

   Hybrid hands-on performance tasks (HHOTs) are resource intensive. Task development is intense, plus these items require kit materials. They also require additional training for administrators.

   Alignment with future NAEP Innovations (like multi-stage testing, online, device agnostic, and reduced contact administration) may require substantial changes. These may include a designated staff administrator to monitor HHOTs. Scenario-based tasks like ICTs and HHOTs may require additional bandwidth. There are currently few easy items in the item pool and item development constraints make them challenging to create, which may limit how lower-difficulty stage adaptive item blocks can be developed.

   Increasing the number of HHOTs and ICTs may require increasing the number of printed booklets and, because they are often paired, may require increasing the required sample size. Increasing the number of these items may create challenges for monitoring trend. An increase in these items types should be implemented over several cycles.

   Further guidance on grade or skill progressions for scientific inquire would be helpful. There is no guidance in the framework for how scientific inquiry skills, like design, conduct, analyze, or draw conclusions from investigations may differ across grades.

Institute of Education Sciences

550 12th Street SW
Washington, DC 20202
Office: 202 123 4567

To: Sharyn Rosenberg, Ph.D.
Assistant Director for Assessment Development
National Assessment Governing Board

From: Holly Spurlock, Ph.D.
Branch Chief for National Assessment Operations
National Center for Education Statistics

Date: October 15, 2021

Subject: Implementation Challenges with the Current Science Framework

This memo summarizes implementation challenges and considerations presented by the current Science Framework for operationalizing a science assessment. The issues can be divided into several categories: ambiguous guidance, ambitious content, standardized assessment constraints, and additional implementation considerations. In addition, attached is a NAEP Validity Studies (NVS) Panel white paper titled "Revision of the NAEP Science Framework and Assessment".

Ambiguous guidance

Learning progressions. Learning progressions (LPs) are referenced heavily in the Science Framework as part of the cognitive and mental models that should be used to measure students successive understanding of complex science principles. While there are no rigid requirements of the framework to assess science content and knowledge using Learning progressions, NCES has not implemented LPs to the extent expected by the framework. This is an area where the field of science assessment development has not caught up with the forward-thinking nature of the science framework. In the field of science, LP development in science assessment development has been uneven and insufficient to fully cover framework content, and existing LPs are still being developed and validated by the science assessment field. Further, there are differing approaches to measuring LPs in a standardized assessment. The science framework views LPs as a mental model for how knowledge matures over time regardless of grade, while other assessment standards focus on grade-level progressions. Instead, NCES relies heavily on the cognitive demands and science practices outlined by the framework to provide the mental model and structure for measuring student progression in understanding science principles.

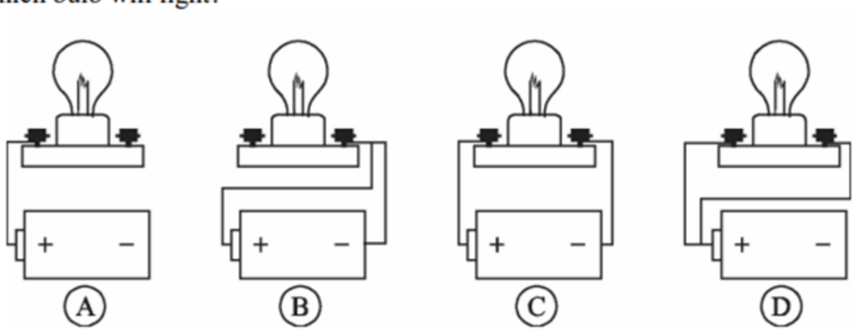
Ambitious content

Measuring two-dimensions (content and practice). The Science Framework requires that each item generate performance expectations for the integration of science content and practice knowledge. That is, each item must measure two-dimensions; "knowing" science and "doing" science. However, the framework does not provide enough specificity around the performance expectations for measuring two-dimensions (i.e., content and practice) for assessment developers and various stakeholders. The example items in the current framework show varying approaches that reflect debates among stakeholders, but it does not provide guidance on which approaches are acceptable or preferable. The example shown below from pages 65-66 of the Science Framework illustrates the challenge with measuring domain knowledge (i.e., content) and application of science skills (i.e., practice), as the latter can depend heavily on the former.

Figure 1. Illustrative item for measuring Using Science Principles (pages 65-66, Science Framework).

Illustrative Item

The pictures show a light bulb connected to a battery.
Which bulb will light?



Key: C

P4.11, Using Science Principles
Source: TIMSS 2003, Grade 4.

Student responses to this item are open to two interpretations. If students have had a great deal of exposure to these types of circuit representations, their responses would fall under Identifying Science Principles. However if, these circuit representations are relatively new for students, then they would need to apply more reasoning and their responses would fall under Using Science Principles.

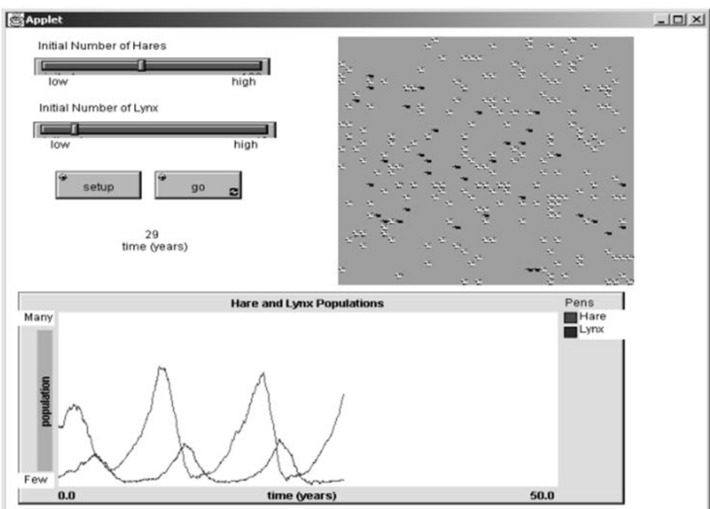
The distinction between how students apply their content knowledge (e.g., "Identifying Science Principles" science practice or "Using Science Principles" science practice in Figure 1) depends heavily on the prior content knowledge students bring to the item. Further, there is not sufficient guidance for how much content knowledge should be measured in scientific-inquiry focused discrete items, hands-on performance tasks and interactive computer tasks – a topic that is heavily debated among the scientific assessment development community. The example shown below from pages 69-71 of the Science Framework, was heavily debated among NCES's science standing committee[1] on whether this illustrative item assessed any content knowledge.

Figure 2. Illustrative Item for measuring Scientific Inquiry (pages 69-71, Science Framework).

Illustrative Item

This interactive computer task is one module in an extended assessment of students' abilities to use a range of technologies to investigate a complex problem, "Should lynx be reintroduced into a national park?" Students accessed, organized, and analyzed data on the number of hares in the park over a 25-year period, researched factors that would impact the population, and created a graph to analyze the trend. (See appendix C for a description of the full task.)

This module allows students to interact with a simulated predator/prey (lynx/hare) population model. Students use the modeling tool to observe population trends that result from different parameter values for the lynx and hare populations. The screen shot below is an example of what students see after they have selected parameters and run the simulation. Note that it is a single screen shot and represents only a small subset of the many screens that students actually see when engaged in this interactive computer task. After students have run the modeling software, they are asked a series of questions (e.g., size of the hare population over time).



L8.6, Using Scientific Inquiry
Source: Quellmalz et al. 2004.

Content experts could (and did during the Science item development process) argue that the illustrative item in Figure 2 measured how well students can manipulate variables to collect data **without** expectations for understanding content knowledge related to the interdependence of species.

---

1 NCES's item development contractor utilizes subject-area standing committees composed of teachers and other content experts, state and local education agency representatives, and content area researchers, to review new item development.

IES Institute of Education Sciences

550 12th Street SW
Washington, DC 20202
Office: 202 123 4567

Greater specificity in future frameworks about approaches and examples demonstrating a consistent approach (or expected and clearly indicated range of approaches) for how to assess content and practice would be helpful. The framework does include a section on the Summary of Practices (page 76) with two examples of clarifications on sample performance expectation for two content statements. For brevity, only the Life Science example is included here.

Figure 3. Clarification: Sample Performance Expectations for a Life Science Content Statement (pages 77-78, Science Framework).

Clarification: Sample Performance Expectations for a Life Science Content Statement

The examples below are all related to the following grade 8 Life Science content statement:

L8.4: Plants are producers—they use the energy from light to make sugar molecules from the atoms of carbon dioxide and water. Plants use these sugars along with minerals from the soil to form fats, proteins, and carbohydrates. These products can be used immediately, incorporated into the plant's cells as the plant grows, or stored for later use.

All examples are also related to a specific situation:

Two different varieties of grass—one better adapted to full sunlight and one better adapted to shade—are each grown in sunlight and in shade.

The results of a controlled experiment along these lines might resemble the following:

Condition	Grass Type A	Grass Type B
Sunlight	Better growth*	Less good growth*
Shade	Less good growth*	Better growth*

* Several variables could be used to indicate growth: mass or dry mass of plants, thickness of stems, number of new sprouts, etc.

<u>Identifying Science Principles</u>

1. State from where a plant's food originates.
2. Classify the grass plants as producers or consumers.

The first performance calls for students to repeat information found in the content statement with little or no modification. The second performance asks students to use the definition of producers given in the content statement to classify or identify the plants.

<u>Using Science Principles</u>

1. Predict whether sugar will move up or down the stems of the grass plants and explain your prediction.
2. Explain where the mass of the growing grass originates.

These performances require students to use principles in the content statement to predict or explain specific observations (growing grass in this case). The content statement itself does not provide the answers to the questions.

IES Institute of Education Sciences

550 12th Street SW
Washington, DC 20202
Office: 202 123 4567

Using Scientific Inquiry

1. Given a data table showing the mass of grass plants of each type grown in the sunlight and shade, draw conclusions about which variety of grass is better adapted to each condition.
2. List other variables that should be controlled in order to feel confident about your conclusions.

The first performance is related to the content statement in that the importance of light for plant growth is useful background information for students. However, the performance requires interpretation of new information (the data table) that has to do with differences among types of plants, while the content statement contains generalizations about all plants. Thus, the performance requires students to use the data to develop new knowledge that they did not have previously. The second performance is in part an assessment of the students' understanding of experimental design. However, good answers would also require knowledge of this and related content statements to identify variables that are relevant to plant growth.

Using Technological Design

1. Given experimental results on the growth of different varieties of grass plants under sunlight and shade conditions, develop a plan for using different types of grass seed in different parts of a partially shaded park.

This performance requires students to use knowledge of the content statement and the experimental results to accomplish a practical goal (in this case, a park with grass growing well in areas that receive varying amounts of sunlight).

While these examples in the Science Framework and Specifications documents are not actual items, they provide considerations for how items can target different science practices. This would make it easier for assessment developers to know what expectations are, for example, for how much content knowledge should be measured in tasks, or whether content as context is sufficient. This would also be helpful in determining how a collection of two-dimension items across item types (DIs, ICTs, and HOTs) can cover the breadth and depth required by the framework.

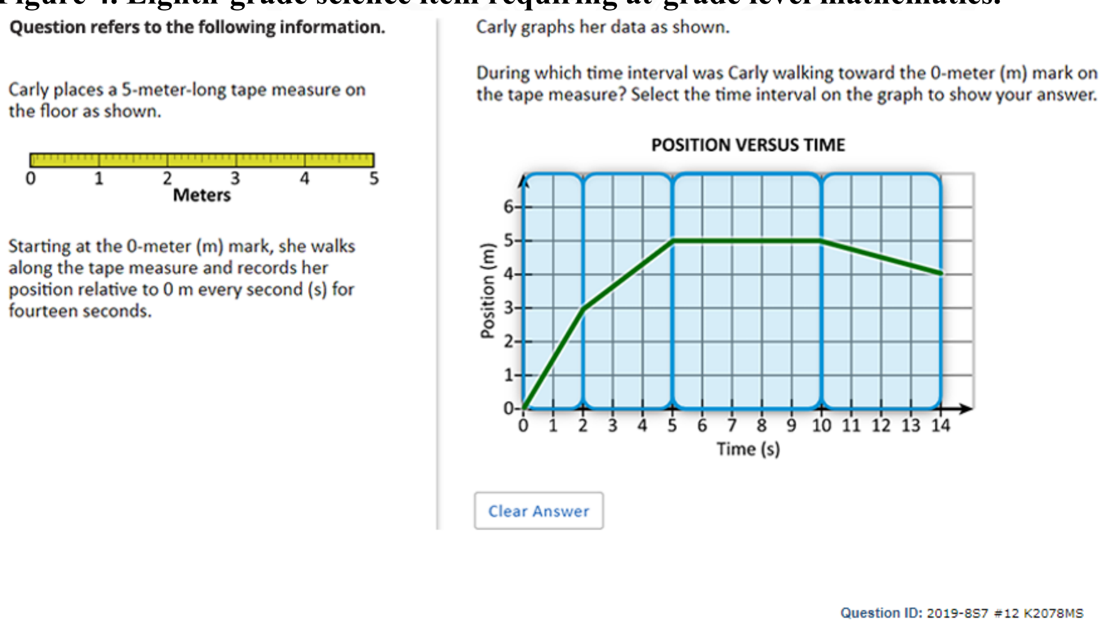
Item difficulty. The Science Framework includes grade-level achievement level descriptors for each science content area and general statements about the science practices for *NAEP Basic*, *NAEP Proficient* and *NAEP Advanced*, and suggests that these descriptions can be used to develop a broad range of items for each achievement level. However, the framework also expects students to be exposed to challenging subject matter, e.g., "[In designing hands-on performance tasks] the NAEP assessment should provide students with a challenging problem… Hands-on performance tasks should be "content rich" in that they require knowledge of science principles to carry them out (Science Framework, pages 106-107)." Given the framework performance expectations for breadth and depth of content knowledge and its integration with practices, it is a challenge to develop items in the easier range while maintaining item rigor and measuring authentic knowing and doing science. If expectations for content knowledge are too high,

550 12th Street SW
Washington, DC 20202
Office: 202 123 4567

students may not be able to demonstrate what they can do (i.e., science practice). However, if they are too low, the measurement of a practice may not be considered valid. Results from the 2019 Science assessments illustrate this point further: 27% of 4th graders, 33% of 8th graders, and 41% of 12th graders fall below *NAEP Basic*, however we have fewer items that measure these students compared to *NAEP Basic*, *NAEP Proficient* and *NAEP Advanced*. Further, the amount of contextual information that students must be given within an item in order to meaningfully engage with the content and practices can lead to higher cognitive load and burden, particularly for lower ability students who may need that context more so than higher ability students. While recent attempts have been made to identify and measure more basic scientific content and skills to develop easier items, the Science item pools continue to be difficult and may reflect a rigorous Science Framework.

Quantitative reasoning in science. The Science Framework Specifications state that the mathematics content required for quantitative reasoning in science content and practice knowledge should be 1-2 years below grade level (Science Framework Assessment and Item Specifications, page 21). However, NCES has had to use at- or above-grade level mathematics content knowledge in some science items to validly measure students' quantitative reasoning in science. For example, the NAEP Mathematics Framework does not expect fourth graders to read or interpret multiple distributions of data. However, displaying multiple distributions of data on a graph may be needed to assess fourth graders scientific inquiry skills of interpreting data and drawing conclusions from an experiment with two or more conditions, e.g., a graph with two or more lines. Figure 4 provides another example from a released eighth-grade science item.

Figure 4. Eighth-grade science item requiring at-grade level mathematics.



550 12th Street SW
Washington, DC 20202
Office: 202 123 4567

The eighth-grade science item in Figure 4 asks students to interpret a line graph that describes Carly's position relative to a 5-meter-long tape measure for 14 seconds. This aligns with the science content objective, P8.14 "An object's position can be measured and graphed as a function of time" (Science Framework, page 34). However, students are not typically introduced to line graphs of this nature until eighth grade according to the common core state standard 8F.B.5 "Describe qualitatively the functional relationship between two quantities by analyzing a graph (e.g., where the function is increasing or decreasing, linear or nonlinear). Sketch a graph that exhibits the qualitative features of a function that has been described verbally." Further, the updated NAEP Mathematics Framework permits this type of graph at eighth grade, but it is not permitted at fourth grade. The examples in this section demonstrate the need to use at-and-above grade level mathematics content knowledge to validly measure students quantitative reasoning in science. This challenge is similarly expressed in the NVS white paper on "Revision of the NAEP Science Framework and Assessment", which states "NGSS performance expectations in science and engineering would likely require students to use some mathematics that is beyond the corresponding grade level".

Standardized assessment constraints

Timing constraints with concept maps. The framework recommends that each assessment include at least one concept-mapping interactive computer task (ICT) at eighth grade and twelfth grade. However, it is not feasible to develop authentic concept-mapping items that allow students to show the process of transferring their mental models into conceptual models as concept maps within a 30-minute cognitive block. NCES developed an ICT that included a partial concept-mapping task for the 2009 science assessment where students were asked to read and synthesize information from animal cards (i.e., habit and diet) to finish a partially constructed food web. However, 51% of students were not able to reach the final item of the task during pilot testing. Edits were made to the task to remove most of the concept-mapping portion so that students were only asked to fill in two missing organisms and their connecting arrows in the food web, but still 22% of students did not reach the end of the task. Given that prior attempts to develop a concept-mapping task within 30-minutes were not successful, NCES has not implemented concept-mapping in the Science assessments.

Design limitations with hands-on performance tasks (HOTs). The framework states that students should be able to freely design the experiment for HOTs, particularly given past criticism that the previous science framework allowed for prescriptive or "recipe"-like HOTs. However, the structure of a HOT and the materials a student can use are limited by assessment timing (i.e., 30-minute cognitive block), space allocated to the student on assessment day, safety, and what is provided in the kit materials. With the migration to hybrid hands-on tasks (HHOTs) for the 2015 pilot, where students were given digital instructions and could record their answers digitally, NCES developed tasks that allowed students flexibility in designing hands-on

550 12th Street SW
Washington, DC 20202
Office: 202 123 4567

experiments and running multiple experimental trials. However, the 2015 pilot showed that hybrid hands-on tasks were speeded, and that speededness varied by grade and task. All three grade 4 and all three grade 12 tasks were speeded, from 23% to 72% of students not reaching the final item. Two out of three grade 8 tasks were speeded, from 75% to 81% of students not reaching the final item. After making considerable edits to constrain the experimental design of the hands-on tasks, the 2019 operational data shows that the HHOTs were much less speeded, ranging from 10% to 28% of students not reaching the final item. Development of hands-on tasks requires careful balance of the amount or depth of directions provided so that all students can engage in the task while designing and carrying out an experiment that can fit within the 30-minute assessment time and materials provided. There is the potential for hands-on tasks to become entirely virtual simulations as part of interactive computer tasks (ICTs). Further research is needed to investigate psychometric and content validity considerations to determine if ICTs can fully replace HOTs to measure scientifically inquiry.

Additional Considerations

Hybrid hands-on performance tasks (HHOTs) are resource intensive. HHOTs incur more expenses, additional resources and level of effort compared to any other item type found in NAEP. Extra resources are required prior to, during and post-data collection to develop and administer HHOTs alongside other science content. Below are some examples of the extra work required:

- In addition to rigorous task development that can cost more than discrete item development, item developers must also perform parallel processes to design and develop the associated kits (e.g., prototyping and testing). Once the kits are finalized, approved, and manufactured, additional quality assurance efforts are required to ensure that the digital tasks and the kits are in sync for a cohesive student experience and smooth administration.
- HHOTs require kit materials, which creates additional resources and costs for the Materials, Distribution, Processing and Scoring contractor to purchase, package and ship the kit materials to field staff. Further, some kit materials can be difficult or expensive to modify after piloting if changes are required.
- The Sampling and Data Collection contractor must hire an additional field staff member to the sample that includes HHOTs so they can monitor the students use of the kits and support the HHOT administration. This requires specialized administrator training and additional staffing to:
    - Receive and inventory kits
    - Distribute kits at appropriate time
    - Monitor kit use
    - Respond to questions in a standardized manner
    - Clean up after the kits

550 12th Street SW
Washington, DC 20202
Office: 202 123 4567

- After the administration, administrators are asked to sort kit materials into goodie bags and waste to offer reusable materials for school use.
- Extra effort is required to develop scoring rubrics and training materials to support scoring of HHOTs. Scoring guides can be intensive given the open-ended nature of student responses to items assessing scientific inquiry.

However, providing students with opportunities to demonstrate their understanding of scientific inquiry and experimentation through designing, implementing, and drawing conclusions is an important part of the Science Framework. NCES continues to investigate ways of replacing hands-on activities with alternative, less-costly designs.

Alignment with future NAEP innovations. In recent years, the NAEP program has expressed an interest in moving towards more innovative and less costly administration models, like multi-stage testing and online, device agnostic and reduced contact administration. There are several aspects of the current Science that should be considered as NAEP moves towards these future innovations. Below are some examples.

- Having a designated field staff administrator to monitor HHOTs must be accounted for as NAEP program goals shift to a reduced contact and contactless administration model. In the reduced contact and contactless models, school staff will serve as administrators and may need further staff and training to accommodate administration of HHOTs.
- Scenario-based tasks, like ICTs and HHOTs, may require additional bandwidth to run resource-heavy science inquiry simulations. This may be challenging for online and device agnostic delivery models that require assessments to run on school internet with limited bandwidth and school devices with reduced processing speeds (e.g., RAM).
- As previously mentioned, the difficulty of the science item pools prohibits implementing adaptive design for the Science assessment as there are insufficient items to support development of easy, or even moderately easy targeted blocks. If there is a desire to implement adaptive design, there are also challenges associated with how to handle HHOTs and ICTs in an adaptive design (e.g., most HHOTs and ICTs target one science subscale).

Design constraints with increasing the number of hybrid hands-on tasks (HHOTs) and interactive computer tasks (ICTs). While it is difficult to predict what impact increasing the number of HHOTs and ICTs will have on measurement validity and reliability in the future, NCES anticipates several operational challenges that should be considered. Analyses from the 2019 science results indicate that a higher proportion of HHOTs and ICTs could have had a larger impact on group scores and consequently an impact on trend reporting. Further, increasing the number of HHOTs and ICTs would add more blocks to the assessment and consequently more booklets since HHOTs and ICTs should be paired, or linked, with each other and with discrete blocks according to balanced incomplete block (BIB) design. Increasing the number of

550 12th Street SW
Washington, DC 20202
Office: 202 123 4567

booklets might increase the sample size requirement for some analyses and potentially increase the level of effort and resources needed to manage a larger item pool. Ultimately, if there is a desire to increase the number of HHOTs or ICTs in the science assessment, then NCES recommends that this increase be implemented gradually over several assessment cycles.

Grade or skill progressions for scientific inquiry. The Science framework does not provide any information as it relates to the application of science inquiry across grade levels and skill progressions. There is no guidance from the framework for how scientific inquiry skills, e.g., design, conduct, analyze or draw conclusions from investigations, may differ for fourth-graders, eighth-graders and twelfth-graders. NCES created evidence centered design (ECD) models to guide grade-level development of items and tasks that assessed scientific inquiry, but further guidance on this area would be helpful.

Enclosure: NAEP Validity Studies White Paper: Revision of the NAEP Science Framework and Assessment

NAEP Validity Studies White Paper: Revision of the NAEP Science Framework and Assessment

James W. Pellegrino
University of Illinois Chicago

October 2021
Commissioned by the NAEP Validity Studies (NVS) Panel

The NAEP Validity Studies Panel was formed by the American Institutes for Research under contract with the National Center for Education Statistics. Points of view or opinions expressed in this paper do not necessarily represent the official positions of the U.S. Department of Education or the American Institutes for Research.

The NAEP Validity Studies (NVS) Panel was formed in 1995 to provide a technical review of NAEP plans and products and to identify technical concerns and promising techniques worthy of further study and research. The members of the panel have been charged with writing focused studies and issue papers on the most salient of the identified issues.

Panel Members:

Keena Arbuthnot
Louisiana State University

Peter Behuniak
Criterion Consulting, LLC

Jack Buckley
American Institutes for Research

James R. Chromy
Research Triangle Institute (retired)

Phil Daro
Strategic Education Research Partnership (SERP) Institute

Richard P. Durán
University of California, Santa Barbara

David Grissmer
University of Virginia

Larry Hedges
Northwestern University

Gerunda Hughes
Howard University

Ina V.S. Mullis
Boston College

Scott Norton
Council of Chief State School Officers

James Pellegrino
University of Illinois at Chicago

Gary Phillips
American Institutes for Research

Lorrie Shepard
University of Colorado Boulder

David Thissen
University of North Carolina, Chapel Hill

Gerald Tindal
University of Oregon

Sheila Valencia
University of Washington

Denny Way
College Board

Project Director:

Sami Kitmitto
American Institutes for Research

Project Officer:

Grady Wilburn
National Center for Education Statistics

For Information:

NAEP Validity Studies (NVS) Panel
American Institutes for Research
1333 Broadway Ave, Suite 300
Oakland, CA 94612
Email: skitmitto@air.org

CONTENTS

TABLES

FIGURES

OVERALL PURPOSE AND ORGANIZATION

The purpose of this white paper is to consider issues related to the scope and focus of a possible new framework for National Assessment of Educational Progress (NAEP) Science (hereafter, NAEP science), including its possible expansion to include aspects of what is represented in NAEP Technology and Engineering Literacy (TEL) (hereafter, NAEP TEL). The goal is to provide the NAEP Validity Studies (NVS) Panel and the NAEP program with input about possible directions for the future and the rationale for choosing among them. Five major sections comprise this paper.

Section I sets the stage for the sections that follow by providing brief background information about the history and projected future uses of the NAEP Science Framework and Assessment as well as the NAEP TEL Framework and Assessment. It also summarizes the National Center for Education Statistics (NCES) and the National Assessment Governing Board (NAGB) timeline for consideration of possible revisions to the NAEP science framework in anticipation of its use to guide the NAEP Science Assessment scheduled for 2028.

Section II contains information on analyses comparing the current NAEP science framework and the NAEP TEL framework to the overall science and technology framework and related set of standards that emerged in the United States in the early part of the last decade. The section begins with a brief synopsis of the content and focus of the NAEP Science and TEL frameworks followed by a brief synopsis of the National Research Council (NRC) *Framework for K–12 Science Education* (NRC, 2012) (hereafter, NRC framework) and the derivative *Next Generation Science Standards* (NGSS) (NRC, 2013). Following that, results are presented from an extensive study comparing the alignment between NAEP Science and NAEP TEL and NGSS (Neidorf et al., 2016). In doing so, the section also considers some of the implications regarding assessments aligned with each reference source.

Section III focuses on the status of science standards and assessments in individual states since the publication of the NRC framework and the NGSS. It reviews the current status regarding state adoptions of science standards that are either identical to NGSS or that are partially aligned with the NGSS (i.e., NRC framework and NGSS "alike"), as well as states with science standards that have no claimed alignment with either the NGSS or NRC framework. For those states with science standards that are NRC framework/NGSS alike, results are summarized from a study examining content alignment between those state standards and the NAEP science framework (Dickinson et al., 2021). The section also includes a summary of the status of the design and implementation of state science assessments relative to their currently adopted standards. This consideration is limited to states that have adopted the NGSS and those whose adopted standards are NRC framework/NGSS alike. The section includes a brief review of the status of the implementation of curricular and instructional practices in states relative to the NRC framework and NGSS. Results are based on the most recent (2018) National Survey of Science and Mathematics Education. The section concludes with a consideration of trends in NAEP science performance for the last 12 years and some possible implications for future NAEP science assessments.

Section IV provides a brief discussion of advances in technology as related to the assessment of science and engineering knowledge and skills. It considers how various developments in digital technologies should be considered in reviewing the existing NAEP Science framework and assessment and envisioning possibilities for their updating. Discussion focuses on the affordances of technology with respect to the constructs that could be included in a revised framework and the associated task design, data capture, and data analytic issues involved in an assessment aligned to an updated framework. The section concludes with a brief discussion of practical and equity concerns related to digitally based assessment of science and technology proficiency.

Section V contains a set of conclusions and recommendations as input to the NCES and NAGB process of reviewing the NAEP science framework and considering possible revision. Conclusions and recommendations are based on the major findings presented in the prior sections.

SECTION I: BACKGROUND, TIMELINE, AND INPUTS

Relevant History: NAEP Science and NAEP TEL

NAEP Science

NAEP science is based on a framework that was adopted in 2005 for the 2009 assessment (NCES, 2009, 2014). That framework was used for the 2015 and 2019 administration of science at grades 4, 8, and 12. It will be used once more for the 2024 (originally 2023) administration of science at eighth grade only. The 2028 (originally 2027) operational administration of the science assessment at grades 4 and 8 at the national, state, and large urban district levels is supposed to be based on an updated science framework.

NAEP TEL

The NAEP TEL assessment is based on a framework developed for grades 4, 8, and 12 in the 2011–2012 period for the 2014 assessment at grade 8. That framework was used for the 2018 TEL administration for grade 8. It will be used twice more for the 2024 (originally 2023) and 2028 (originally 2027) TEL administrations for grade 8. Both planned TEL administrations overlap with NAEP science administrations: 2024 overlaps with the current science framework and assessment, and 2028 overlaps with the new science framework and assessment.

NAEP Science and TEL—Possible Merger

Discussions have been held within NAGB about possibilities for combining NAEP science and TEL, especially because both are now digitally based assessments. Doing so may make logical sense given overlaps in conceptual coverage with contemporary U.S. science and technology frameworks. Another benefit could be cost savings realized by having a single assessment representing key aspects of knowledge and skill for science and technology. Such a merger clearly would be most beneficial for the planned 2028 administration of both science and TEL. NAGB therefore may wish to consider developing a single 2028 assessment based on a new integrated science and technology framework.

Status and Plans for Review, Update, and/or Revision of the NAEP Science Framework

NAGB has started the process needed to consider updating the science framework for application in the design of the 2028 grades 4 and 8 science assessment. Given the current timeline, it appears that a decision about the need for and the scope of a science framework revision will be completed during 2022. Work toward making such a decision includes:

- Detailed information available in an NCES report issued in 2016 titled *A Comparison Between the Next Generation Science Standards (NGSS) and the National Assessment of Educational Progress (NAEP) Frameworks in Science, Technology and Engineering Literacy, and Mathematics* (Neidorf et al., 2016). Information about the results of this study is presented in Section II.

- A recently completed study by HumRRO titled *Comparative Analysis of the NAEP Science Framework and State Science Standards* (Dickinson et al., 2021) in which content overlap was examined between the NAEP science framework and the science standards of individual states. Classification of state standards was based on information from the National Science Teachers Association (NSTA) specifying which states have current standards that are identical to NGSS, partially NGSS, or non-NGSS. The focus for the analysis was on alignment between the NAEP science framework and the standards of the partial NGSS and non-NGSS states. Information about the results of this study is presented in Section III.

- Input from a group of five or more experts, each of whom would consider the information derived from the two studies mentioned above—the 2016 AIR comparison of NAEP to NGSS (Neidorf et al., 2016) and the more recent HumRRO analysis of state standards relative to NAEP (Dickinson et al., 2021)—as well as other factors given the expert's experience in the field of science education, to present their thoughts on whether the framework needs to be changed and why.
- NAGB recently issued a public call for input on the NAEP science framework regarding its revision. NAGB requested responses from interested parties by October 15, 2021.

NAGB is scheduled at its March 2022 meeting to consider whether to move ahead with a revision of the science framework for application in the design of the 2028 science assessment. The board also will consider the input received from the various sources mentioned above. The timing of these activities should NAGB choose to recommend a science framework revision would easily extend into 2023 if not beyond. Given existing statutes, NAGB will convene two panels based on their policy (NAGB, 2018a, p. 5):

- The **Framework Visioning Panel** shall formulate high-level guidance about the state of the field to inform the process, providing these in the form of guidelines. The major part of the Visioning Panel work will be at the beginning to provide initial guidance for developing a recommended framework. The Visioning Panel shall be composed of the stakeholders referenced in the introduction above. At least 20 percent of this panel shall have classroom teaching experience in the subject areas under consideration. This panel may include up to 30 members with additional members as needed.
- The **Framework Development Panel** shall develop drafts of the three project documents and engage in deliberations about how issues outlined in the Visioning Panel discussion should be reflected in a recommended framework. As a subset of the Visioning Panel, the Development Panel shall have a proportionally higher representation of content experts and educators, whose expertise collectively addresses all grade levels designated for the assessment under development. Educators shall be drawn from schools across the nation, who work with students from high-poverty and low-performing schools, as well as public and private schools. This panel may include up to 15 members, with additional members as needed.

The timeline for initiating and completing the work of the panels remains to be specified, and because the work of the development panel follows from the work of the visioning panel, its work would end sometime in 2023 or later, pending public review of a draft framework and commentary with subsequent revision and then final adoption by NAGB. A revised framework would be used to develop the design and tasks for the 2028 NAEP science assessment.

SECTION II. ANALYSIS OF THE NAEP SCIENCE FRAMEWORK RELATIVE TO OTHER CONTEMPORARY SCIENCE AND TECHNOLOGY FRAMEWORKS

This section examines how the NAEP science framework and assessment and NAEP TEL framework compare with the NRC *Framework for K–12 Science Education* (hereafter, NRC framework) and the derivative *Next Generation Science Standards* (NGSS). It begins with a brief description of key elements of each of the four reference sources and is followed by a summary of results from a detailed study of the correspondences between the two NAEP frameworks and the NGSS. Highlighted in the summary are important areas of similarity and dissimilarity and some of the implications relative to assessment.

Overview of the NAEP Science Framework and Assessment

As noted earlier, the current NAEP science assessment is based on a framework originally developed for the 2009 assessment administration at grades 4, 8, and 12. That framework also was used for the 2011 administration at grade 8 and the 2015 and 2019 administrations at grades 4, 8, and 12. The framework is scheduled to be used once more for the 2024 administration for eighth grade only. The scheduled 2028 operational administration of science for grades 4 and 8 is supposed to be based on an updated science framework.

The current NAEP science framework (NAGB, 2008, 2014) was developed approximately 4 years before the 2009 administration and incorporated ideas from contemporary theory and research on science learning and assessment including synthesis volumes from the NRC: *How People Learn: Brain, Mind, Experience and School* (Bransford et al., 2000); *Knowing What Students Know: The Science and Design of Educational Assessment* (Pellegrino et al., 2001); Systems of State Science Assessment (Wilson & Bertenthal, 2005) and *Taking Science to School* (National Research Council, 2007). The framework included important ideas about the learning and knowing of both science content and science practices with a particular emphasis on their integration as discussed below.

Science Content. The science content for NAEP is defined by a series of statements that describe key facts, concepts, principles, laws, and theories in three broad areas: physical sciences, life sciences, and Earth and space sciences. Table 1 shows the major topics and subtopics within each of the three major science domains. The nature of the specific content knowledge changes in both scope and sophistication across the three grade levels.

Table 1. NAEP science content areas and topics

Physical sciences	Life sciences	Earth and space sciences
Matter • Properties of matter • Changes in matter	Structures and functions of living systems • Organization and development • Matter and energy transformations • Interdependence	Earth in space and time • Objects in the universe • History of Earth
Energy • Forms of energy • Energy transfer and conservation		Earth structures • Properties of Earth materials • Tectonics
Motion	Changes in living systems	Earth systems

• Motion at the macroscopic level • Forces affecting motion	• Heredity and reproduction • Evolution and diversity	• Energy in Earth systems • Climate and weather • Biogeochemical cycles

SOURCE: National Assessment Governing Board, 2014, Exhibit 4, p. 19. Reprinted with permission.

Science Practices. The second dimension of the framework is defined by four science practices: Identifying Science Principles, Using Science Principles, Using Scientific Inquiry and Using Technological Design. In the NAEP science framework, the first two practices (Identifying Science Principles and Using Science Principles) generally are considered as "knowing science," and the last two practices (Using Scientific Inquiry and Using Technological Design) are considered as the application of that knowledge to "doing science" and "using science to solve real-world problems."

Table 2 provides a high-level description of the nature of each specific practice in terms of the types of cognitive demands placed on students as they engage in a practice as applied to a topic from a specific science content area.

Table 2. NAEP science practices: General labels and specific applications

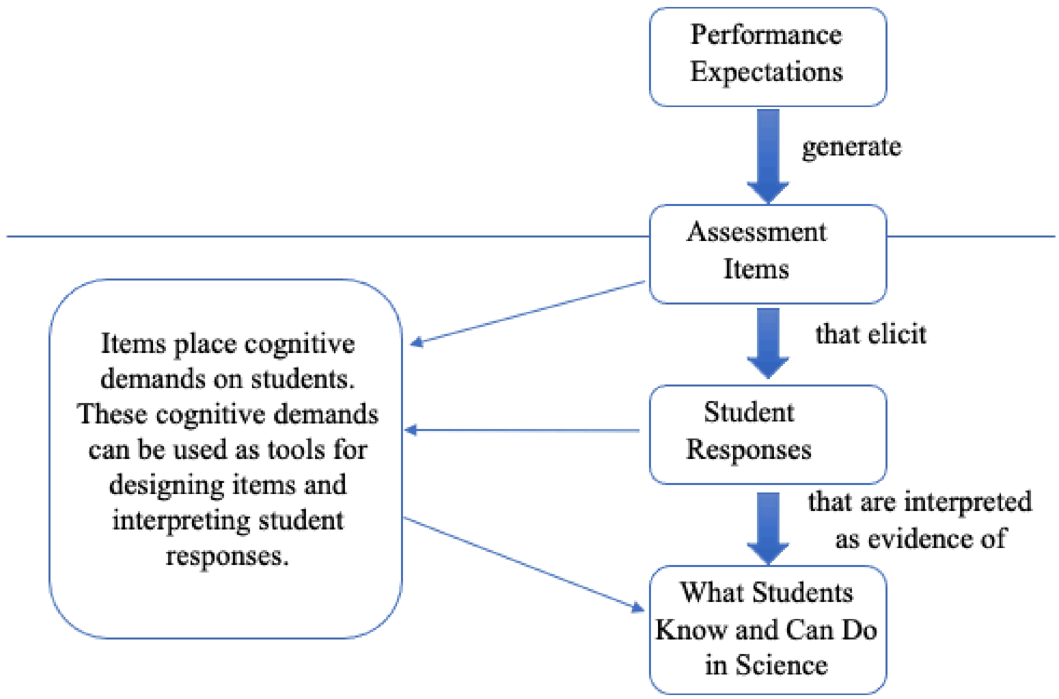
	Practice Label	Practice Applications			
←Communicate accurately and effectively→	Identifying Science Principles	Describe, measure, or classify observations.	State or recognize correct science principles.	Demonstrate relationships among closely related science principles.	Demonstrate relationships among different representations of principles.
	Using Science Principles	Explain observations of phenomena.	Predict observations of phenomena.	Suggest examples of observations that illustrate a science principle.	Propose, analyze, and/or evaluate alternative explanations or predictions.
	Using Scientific Inquiry	Design or critique aspects of scientific investigations.	Conduct scientific investigations using appropriate tools and techniques.	Identify patterns in data and/or relate patterns in data to theoretical models.	Use empirical evidence to validate or criticize conclusions about explanations and predictions.
	Using Technological Design	Propose or critique solutions to problems given criteria and scientific constraints.	Identify scientific tradeoffs in design decisions and choose among alternative solutions.	Apply science principles or data to anticipate effects of technological design decisions.	

SOURCE: National Assessment Governing Board, 2014, Exhibit 13, p. 76.

Performance Expectations—Combining Content and Practices. The design of the NAEP science assessment is guided by the framework's descriptions of both the science content and science practices to be assessed but with the key assumption that the practices are to be combined with a science content statement to generate specific student performance expectations that serve as the target for assessment. Assessment items are then developed based on the description of each specific performance expectation.

Using the logic of specific performance expectations as a guide for item development processes, items are then designed to vary the cognitive demands of tasks, a process that then influences the conclusions to be made about student performance. Such a process of item development can be represented schematically as shown in Figure 1.

Figure 1. NAEP assessment item development model



SOURCE: National Assessment Governing Board, 2014, Exhibit 2, p. 12.

In 2009, 2011, and 2015, NAEP science was administered as primarily a paper-and-pencil test. In 2019 a major shift occurred when NAEP science was administered for the first time as an entirely digitally based assessment. The Nation's Report Card (2019) provides a description of the new digital assessment:

> The NAEP digitally based science assessment consisted of standalone, discrete questions, and scenario-based tasks comprising a connected sequence of questions. Scenario-based tasks were designed to engage students in scientific inquiry through hands-on activities and computer simulations set in real-world contexts. The tasks provided students opportunities to demonstrate their knowledge and skills in each of three science content areas and four science practices. The science assessment included two types of scenario-based tasks:
>
> - Interactive computer tasks (ICTs). ICTs use real-world simulations to engage students in scientific investigations that require the use of science inquiry skills and application of scientific knowledge to solve problems.

- Hybrid hands-on tasks (HHOTs). Students perform hands-on scientific investigations using materials in kits provided by NCES. The "hybrid" in HHOTs denotes that these tasks combine hands-on investigations with digital activities. Students use NCES-supplied tablets to view kit instructions, record results and data, and answer assessment questions.

Overview of the NAEP TEL Framework and Assessment

As noted earlier, a TEL framework was developed for the first TEL assessment in 2014 at grade 8 and was used again for the 2018 TEL at grade 8. It is scheduled to be used twice more for the 2024 and 2028 TEL administrations at grade 8.

The development of this framework and assessment was motivated by several factors. In the science education community, a call for preparing students with technology and engineering literacy has been long awaited. The *Science for All Americans* report (American Association for the Advancement of Science, 1990) explicitly suggested that science education should incorporate technology and engineering as a form of scientific inquiry. Bybee (2010) proposed an advance to STEM education by integrating technology and engineering with science and mathematics education. He argued that "there are very few other things that influence our everyday existence more [than technology] and about which citizens know less" (Bybee, 2010, p. 30). Bybee suggested extending traditional information communication technology education by integrating ICTs with other subjects. He further pointed out that involving students in engineering activities could promote their abilities for both problem solving and innovation. He also acknowledged that engineering as typically presented in schools was inconsistent with its careers and contributions to society, and thus authentic scenarios needed to be developed for both learning and assessment (Bybee et al., 2009).

The NRC report, *Education for Life and Work: Developing Transferable Knowledge and Skills in the 21st Century*, identified information literacy and ICT literacy as two of the most frequently mentioned critical competencies for students to succeed in the 21st century (Pellegrino & Hilton, 2012). That report discussed various foundations for education, and STEM education in particular, including preparing future entrants to the labor market with the ability to adapt to technological changes in society rather than simply acquiring static bits of knowledge. Similarly, another 2012 NRC report, the *Framework for K–12 Science Education* (NRC, 2012), framed one of the overarching goals of science education as the development of students who "are careful consumers of scientific and technological information related to their everyday lives" (p. 1). The framework explicitly includes "Engineering, Technology, and Applications of Science" as one of four disciplinary core ideas and describes "defining problems, design solutions, and using computational thinking" as critical components of science and engineering practices. Further discussion of the NRC framework follows this section on TEL.

These and other trends related to technology and engineering literacy spurred the development of a TEL framework and inclusion of the TEL assessment as part of the NAEP program. The goal of TEL has been to obtain information about students' understanding of technology and its effect on our society and environments, as well as students' ability to design solutions to solve real-world problems. The TEL framework describes TEL as the "capability to use, understand, and evaluate technology as well as to understand technological principles and strategies needed to develop solutions and achieve goals" (NAGB, 2013, p. xi). Specifically, the framework identified three interconnected areas to be assessed (NAGB, 2018b, p. xii) as follows:

- *Technology and Society* deals with the effects that technology has on society and the natural world and with the sorts of ethical questions that arise from those effects. Knowledge and capabilities in this area are crucial for understanding the issues surrounding the development and use of various technologies and for participating in decisions regarding their use.
- *Design and Systems* covers the nature of technology, the engineering design process by which technologies are developed, and basic principles of dealing with everyday technologies, including maintenance and troubleshooting. An understanding of the design process is particularly valuable in assessing technologies, and it can also be applied in areas outside technology, since design is a broadly applicable skill.
- *Information and Communication Technology* includes computers and software learning tools, networking systems and protocols, hand-held digital devices, and other technologies for accessing, creating, and communicating information and for facilitating creative expression. Although it is just one among several types of technologies, it has achieved a special prominence in technology and engineering literacy because familiarity and facility with it is essential in virtually every profession in modern society.

Students taking the TEL assessment are expected to succeed in the following three types of thinking and reasoning practices:

- *Understanding technological principles* focuses on students' knowledge and understanding of technology and their capability to think and reason with that knowledge;
- *Developing solutions and achieving goals* refers to students' systematic application of technological knowledge, tools, and skills to address problems and achieve goals presented in societal, design, curriculum, and realistic contexts; and
- *Communicating and collaborating* centers on students' capabilities to use contemporary technologies to communicate for a variety of purposes and in a variety of ways, working individually or in teams. (NAGB, 2018b, pp. 3-2–3-3)

The TEL assessment has developed scenario-based tasks designed to engage students in multimedia environments to gauge students' understanding of technological and engineering principles and their ability to apply such principles to determine design solutions. Most of TEL's assessment tasks are computer simulation problems involving technology and engineering scenarios.

Overview of the NRC Science Education Framework and Next Generation Science Standards

Based on multiple sources of evidence and discussions about the knowing and learning of science, the nature of science education as it had been practiced in the United States, and evidence of relatively poor student achievement in science across K–16+, agreement emerged during the early part of this century about the need for substantial change in science standards, instruction, and assessment, including what we expect students to know and be able to do in science, how science should be taught, and how it should be assessed.

> Recognition of this science education problem can be found in reports spanning elementary, secondary, and postsecondary education (K–16+). These reports present a consistent description of the nature of competence in science and include NRC reports on K–8 science education in formal and informal learning environments (NRC, 2007, 2009); curriculum and assessment frameworks for Advanced Placement (AP) science courses (e.g., College Board, 2011a, 2011b); and even revisions in the nature of the science knowledge required for entry to medical school and assessed on the Medical College Admissions Test (e.g., American Association of Medical Colleges, 2012). (Pellegrino, 2016, p. 5)

Reconceptualization of the nature of science competence emergent from these many and diverse sources was captured to some extent in the College Board's standards for success in high school science (College Board, 2009). Their most complete expression for all K–12 science education was presented in the 2012 NRC report, *A Framework for K–12 Science Education. Practices, Crosscutting Concepts and Core Ideas.* The NRC framework report contains many important key ideas, including articulation of three interconnected dimensions of science competence as denoted in the report's title. The three dimensions are Disciplinary Core Ideas (DCIs), Crosscutting Concepts (CCCs), and Science and Engineering Practices (SEPs). The NRC framework provides detailed descriptions of each dimension, the concepts that each dimension encompasses, and the rationale for their inclusion. Figure 2 provides a list of the dimensions and their associated high-level concepts.

DCIs are the big ideas associated with a discipline, like life science, and which are essential to explaining phenomena. CCCs are ideas like systems thinking that are important across many science disciplines and provide a unique lens to examine phenomena. SEPs are the multiple ways of knowing and doing science and engineering, like developing models and constructing explanations that scientists and engineers use to study the natural and designed world. The framework focuses on the need for the integration of these three dimensions in science and engineering education. The knowledge associated with each of the three dimensions must be integrated in the teaching, learning, and doing of science and engineering, and in assessing what students know and can do. The framework emphasizes research indicating that learning about science and engineering "involves integration of the knowledge of scientific explanations (i.e., content knowledge) and the practices needed to engage in scientific inquiry and engineering design" (NRC, 2012, p. 11). The disciplinary core ideas, crosscutting concepts, and science and engineering practices serve as thinking tools that work together to enable scientists, engineers, and learners to design solutions to problems, reason with evidence, and make sense of phenomena. When learners engage in science and engineering practices integrated with DCIs and CCCs to make sense of compelling phenomena or design solutions to complex problems, they build new knowledge about all three dimensions and come to understand the nature of how scientific knowledge and engineering solutions develop.

Figure 2. The three dimensions of the NRC framework

Scientific and Engineering Practices	Crosscutting Concepts
▪ Asking questions (for science) and defining problems for engineering ▪ Developing and using models ▪ Planning and carrying out investigations ▪ Analyzing and interpreting data ▪ Using mathematics and computational thinking ▪ Constructing explanations (for science) and designing solutions (for engineering) ▪ Engaging in argument from evidence ▪ Obtaining, evaluating, and communicating information	▪ Patterns ▪ Cause and effect: Mechanism and explanation ▪ Scale, proportion, and quantity ▪ Systems and system models ▪ Energy and matter: Flows, cycles, and conservation ▪ Structure and function ▪ Stability and change
Disciplinary and Core Ideas *Physical Sciences* PS 1: Matter and its interactions PS 2: Motion and stability: Forces and interactions PS 3: Energy PS 4: Waves and their applications in technologies for information transfer *Earth and Space Sciences* ESS 1: Earth's place in the universe ESS 2: Earth's ecosystems ESS 3: Earth and human activity	 *Life Sciences* LS 1: From molecules to organisms: Structures and processes LS 2: Ecosystems: Interactions, energy, and dynamics LS 3: Heredity: Inheritance and variation of traits LS 4: Biological evolution: Unity and diversity *Engineering, Technology, and the Applications of Science* ETS 1: Engineering design ETS 2: Links among engineering, technology, science, and society

SOURCE: NRC 2012, Box S-1, p. 3.

The rationale for the choice of the specific DCIs is important to note here relative to other previous standards and frameworks. One criticism of U.S. K–12 science curricula relative to those of other countries was that they were "a mile wide and an inch deep" (Schmidt et al., 1997, p. 62). The same concerns about breadth versus depth were made in an NRC Report on advanced study of science in U.S. high schools (NRC, 2002). In reaction, the framework focused on core ideas in each of the four content domains with the directive that students should continue to be exposed to these core ideas with increased levels of complexity and explanatory power relative to a range of phenomena and problem contexts throughout their schooling.

While each of the three dimensions matters, a central argument of the framework is that proficiency is demonstrated through *performances* that require the integration of all three dimensions. Such demonstrations are labeled *Performance Expectations (PEs)* because they specify what students at various levels of educational experience should know and be able to do. The Next Generation Science Standards (NRC, 2013) are an expression of the integrated knowledge vision contained in the framework, and provide a set of standards expressed as performances expectations for students from Kindergarten to 12th grade. The NGSS appear as clusters of performance expectations related to particular aspects of a core disciplinary idea (see Figure 3 for an example at grade 4). Each performance expectation requires students to draw upon knowledge of a specific practice and a crosscutting concept in the context of specific elements of disciplinary core knowledge. Across the set of performance expectations at a given grade level or grade band, each practice and crosscutting concept appears in multiple standards. A student demonstrates grade-level proficiency by completing performances that demonstrate that they can make use of their knowledge. To truly know and understand science is to be able to use the three dimensions of scientific knowledge together to explain compelling phenomena and/or provide solutions to complex problems.

Figure 3. NGSS Performance Expectations for Grade 4 Life Science 1: From molecules to organisms: Structures and processes

4-LS1 From Molecules to Organisms: Structures and Processes

PERFORMANCE EXPECTATIONS

Students who demonstrate understanding can:

4-LS1-1. Construct an argument that plants and animals have internal and external structures that function to support survival, growth, behavior, and reproduction. [*Clarification Statement*: Examples of structures could include thorns, stems, roots, colored petals, heart, stomach, lung, brain, and skin.] [*Assessment Boundary*: Assessment is limited to macroscopic structures within plant and animal systems.]

4-LS1-2. Use a model to describe that animals receive different types of information through their senses, process the information in their brain, and respond to the information in different ways. [*Clarification Statement*: Emphasis is on systems of information transfer.] [*Assessment Boundary*: Assessment does not include the mechanisms by which the brain stores and recalls information or the mechanisms of how sensory receptors function.]

Science and Engineering Practices	Disciplinary Core Ideas	Crosscutting Concepts
Engaging in Argument from Evidence Engaging in argument from evidence in 3–5 builds on K–2 experiences and progresses to critiquing the scientific explanations or solutions proposed by peers by citing relevant evidence about the natural and designed world(s). • Construct an argument with evidence, data, and/or a model. (4-LS1-1) • Use a model to test interactions concerning the functioning of a natural system. (4-LS1-2)	**LS1.A: Structure and Function** • Plants and animals have both internal and external structures that serve various functions in growth, survival, behavior, and reproduction. (4-LS1-1) **LS1.D: Information Processing** • Different sense receptors are specialized for particular kinds of information, which may then be processed by an animal's brain. Animals are able to use their perceptions and memories to guide their actions. (4-LS1-2)	**Systems and System Models** • A system can be described in terms of its components and their interactions. (4-LS1-1), (4-LS1-2)

SOURCE: NRC, 2013, p. 38. Reprinted with permission.

An important issue relative to the present paper's discussion of NAEP Science and NAEP TEL is the NRC framework's emphasis on the connections among science, engineering, and technology. While these connections are somewhat separate across NAEP Science and TEL, key practices and ideas from engineering are included in the NRC framework because of important interconnections between science and engineering and because evidence shows that engaging in engineering design can help leverage student motivation and increase learning in science. One goal of including ideas related to engineering, technology, and the applications of science in the framework for science education is to help students understand the similarities and differences between science (the natural world) and engineering (the designed world) by making the connections between the two fields explicit and by providing all students with an introduction to engineering.

The NGSS expanded upon the framework's adoption of the logic of learning progressions to describe students' developing proficiency in the three intertwined domains across grades K–12, noting that "If mastery of a core idea in a science discipline is the ultimate educational destination, then well-designed learning progressions provide a map of the routes that can be taken to reach that destination" (NRC, 2012, p. 26). The stress on learning progressions is supported by research on science knowing and learning described in the 2005 NRC report *Systems of State Science Assessment,* the 2007 NRC report *Taking Science to School* and in other documents describing research on the progression of student learning and understanding in science (e.g., Alonzo & Gotwals, 2012; Corcoran et al., 2009). The framework built in the idea of a developmental progression of student understanding across the grades by specifying grade band end point targets at grades 2, 5, 8, and 12 for each component of each disciplinary core idea. For the practices and crosscutting concepts, the framework also provided sketches of possible progressions for learning each practice or concept but did not indicate the expectations at any particular grade level. The NGSS built on these suggestions and developed tables that define what each practice might encompass at each grade level. The NGSS also defined the expected uses of each crosscutting concept for students at each grade level.

The NRC framework and NGSS stand in sharp contrast to prior generations of U.S. science standards (e.g., American Association for the Advancement of Science, 1992; NRC, 1996, 2000) that treated content and inquiry as separate strands of science learning. Unfortunately, both instruction and assessment followed suit. The form the standards took contributed to this separation: Content standards stated what students should know, largely in the form of declarative knowledge, and inquiry standards stated what they should be able to do, largely in the form of procedural knowledge. Consequently, instruction often separated content learning from inquiry and vice versa. Science education often was often criticized as "lots of hands on but not much minds on." In a similar fashion, assessments separately measured content knowledge in the absence of application or inquiry practice components in the absence of content concerns. Thus, the NGSS idea of an integrated, *multidimensional science performance* represents a different way of thinking about science proficiency. Disciplinary core ideas and crosscutting concepts serve as thinking tools that work together with scientific and engineering practices to enable learners to solve problems, reason with evidence, and make sense of phenomena. Such a view of competence signifies that measuring proficiency solely as the acquisition of core content knowledge or as the ability to engage in general inquiry processes is neither appropriate nor sufficient.

In the context of assessment, the importance of this integrated perspective of what it means to know science is that one should be attempting to assess where a student can be placed along a sequence of progressively more "scientific" understandings of a given core idea and successively more sophisticated applications of practices and crosscutting concepts. This idea is relatively unfamiliar in the realm of science assessments, which more often have been viewed as simply measuring whether students know or do not know particular grade-level content (Pellegrino, 2013). To support an integrated and developmental approach to science learning, the framework explains that assessment tasks "must be designed to gather evidence of students' ability to apply the practices and their understanding of the crosscutting concepts in the contexts of specific applications in multiple disciplinary areas" (NRC, 2012, p. 218). Assessments must strive to be sensitive both to grade-level-appropriate understanding and to those understandings that may be appropriate at somewhat lower or higher grades. This is particularly important for assessment materials and resources to support ongoing classroom instruction. The challenges of designing such multidimensional assessments for classroom and large-scale assessment use are substantial. Potential approaches and solutions were discussed in detail in another NRC report, *Developing Assessments for the Next Generation Science Standards* (Pellegrino et al., 2014).

Comparing the NAEP Science and TEL Frameworks and NGSS

Given the brief descriptions provided above, it should be clear that there are multiple similarities and overlaps as well as differences between the NAEP science framework and the NGSS and between NAEP TEL and NGSS. Even though the NAEP science framework predates the 2012 NRC framework and the derivative 2013 NGSS, overlapping content exists, each has a description of science practices, and both make use of the idea of performance expectations that involve the intersection of content and practice. The NAEP TEL framework was developed about the same time as the NRC framework and overlaps with the latter's highlighting of engineering practices alongside science practices, and its inclusion of Engineering, Technology, and the Application of Science as one of the four disciplinary areas.

Although some of the ideas that are part of the NRC framework and NGSS have found their way over time into the NAEP Science assessment and NAEP TEL assessment, including the design of scenario-based tasks in both NAEP assessments and enacted through technology, neither NAEP framework is reflective of the more dramatic shifts found in the NRC framework and NGSS. NAEP TEL focuses on various aspects of technology and engineering literacy and shares certain things in common with the NRC framework and NGSS. In addition, when it was developed and implemented as a technology-based assessment, TEL included more innovative scenario-based item types than the paper-and-pencil NAEP science assessment. The 2019 digitally based NAEP science assessment has moved in a similar direction. Interestingly, when the NRC framework and NGSS were published, NCES leadership often used TEL items as illustrations of performance tasks in NAEP of the type implied by the NGSS, in part because the paper-and-pencil NAEP science assessment did not include such items at the time.

The most significant difference between NAEP science and NAEP TEL and the NRC framework and NGSS is the singular focus of the latter two on the idea of *knowledge in use*—that competence is demonstrated by being able to use DCI and CCC conceptual knowledge in the context of one or more SEPs to solve problems, explain phenomena, and/or design solutions to challenging problems (Harris et al., 2019). Thus, a major concern regarding the future of the NAEP science and TEL assessments is the nature and degree of the alignment between current NAEP frameworks and the NGSS, especially if most states have adopted NGSS or NRC framework/NGSS alike standards and have implemented state assessments aligned with those standards. A related question is whether states, districts, and schools have accordingly modified curricular choices and instructional practices in ways consistent with their own standards (NRC framework or NGSS) and assessments. If a serious misalignment between NAEP science and the science and technology instruction and assessment practiced in schools exists, the validity and value of the NAEP science assessment results for the 2024 or 2028 administrations could be seriously questioned.

The remainder of this section includes the results from a detailed examination of the alignment between each of NAEP science and TEL frameworks with NGSS.[1] These data are critical in thinking about whether changes are needed in NAEP to better align with contemporary U.S. frameworks and standards as well as the extent to which a single assessment framework more like the NGSS would suffice to create a NAEP science and technology assessment rather than two NAEP science and technology assessments as is currently the case. Section III examines the situation with respect to (a) state science standards relative to the NGSS, (b) state science assessments relative to their current standards, and (c) implementation of new science standards in terms of curricular choices and instructional practices in the field.

Comparative Study of the NAEP Science and TEL Frameworks and NGSS

The main purpose of *A Comparison Between the Next Generation Science Standards (NGSS) and the National Assessment of Educational Progress (NAEP) Frameworks in Science, Technology and Engineering Literacy, and Mathematics* (Neidorf et al., 2016) was "to determine the extent to which the NGSS performance expectations are aligned with the content objectives and definitions of practices in the NAEP science and TEL frameworks. An additional purpose was to determine the extent to which the NGSS performance expectations involving mathematics-related practices are aligned with the content objectives in the NAEP mathematics framework." (Neidorf et al., 2016, p. 2). [2]

A comparison of the NGSS with the NAEP STEM frameworks can yield multiple important outcomes with potential implications for a revision of NAEP science and a possible merger of NAEP science and TEL. Neidorf et al. (2016) listed the following (p. 2):

- For the science comparisons, similarities suggest areas where NAEP may provide useful science assessment examples and national achievement data on the student understandings in the natural sciences described in the NGSS. Differences suggest areas where NAEP and NGSS-based science assessments may each provide unique contributions.
- The TEL comparisons augment these findings by identifying additional areas of overlap with the engineering and technology content and practices in the NGSS. Together, these comparisons explore how completely the full range of content and practices in the NGSS are covered by the NAEP science and TEL frameworks as well as the unique aspects of each.
- The mathematics comparisons, while more limited, explore the degree of alignment between the mathematics-related performance expectations in the NGSS and the NAEP mathematics framework. The NGSS are not intended to guide mathematics

[1] The NAEP Science framework and assessment also can be compared to international large-scale science assessment programs in terms of content focus, assessment practices, and future directions. Doing so is beyond the scope of this paper, but for those interested in the PISA and TIMSS science assessment programs, such information is available in a forthcoming chapter on large-scale science assessment (Zhai & Pellegrino, in press).
[2] The Neidorf et al. (2016) study was conducted prior to the adoption of the 2019 math framework for administration in 2026.

assessments, and the performance expectations in science and engineering do not specify explicit mathematics requirements. However, the mathematics students may need to use in responding to items developed to assess these performance expectations can be inferred and compared to the mathematics included in NAEP across grades. Thus, such comparisons can provide information on how assessments based on the NGSS might compare with NAEP in terms of the level of mathematics and quantitative skills that would be required of students.

Three research questions guided this comparison study (Neidorf et al., 2016, p. 3):

1. *Related to the NAEP science framework*: How similar (or different) are the NGSS performance expectations in physical sciences, life sciences, and Earth and space sciences to the content and practices in the NAEP science framework at the corresponding grade levels?
2. *Related to the NAEP TEL framework*: How similar (or different) are the NGSS performance expectations in engineering, technology, and applications of science to the content and practices in the NAEP technology and engineering literacy framework at the corresponding grade levels?
3. *Related to the NAEP mathematics framework*: To what extent are the mathematics-related NGSS performance expectations and practices aligned with the content and skills specified in the NAEP mathematics framework, and at which grade(s)?

Major Findings

The report discusses multiple ways in which the NAEP science and TEL frameworks and the NGSS were compared and contrasted, including different directions and forms of comparison. A plethora of findings are reported and what follows is excerpted from a summary of the major results of those comparisons. It is taken directly from the AIR report.

> There was a moderate to substantial degree of ***content overlap*** between the NGSS and the NAEP science and TEL frameworks. About half of the NGSS performance expectations in the upper elementary grade band (grades 3–5) covered content that overlaps with NAEP science or TEL at grade 4. In contrast, there was much less content in NAEP science that overlapped with the NGSS at grade 4 (and in TEL that overlapped at any grade).
>
> Ninety percent or more of the NGSS performance expectations at the middle school and high school levels covered content that overlaps with NAEP science or TEL at grades 8 and 12, respectively. A somewhat lower, but still substantial, percentage of content in NAEP science at grades 8, and 12 (from 74 to 88 percent) overlapped with the NGSS.
>
> Because of differences in the depth, breadth, detail, or focus of the overlapping content, *content alignment* was lower than *content overlap* when the NGSS was compared to the NAEP science and TEL frameworks together. Moreover, when relevant performance expectations in the natural sciences (physical sciences, life sciences, and Earth and space sciences) and in engineering, technology, and applications of science (ETS) were compared to the NAEP science and TEL

frameworks individually, content alignment differed by grade and by content domain.

Across frameworks, content alignment of the NGSS with the NAEP science and TEL frameworks was moderate. Roughly half of the NGSS performance expectations aligned to NAEP (science or TEL) at each grade level. At grades 3–5, 38 percent of performance expectations were aligned with the science framework and 13 percent with the TEL framework, with 2 percent in the sciences aligned with both NAEP and TEL. At the middle school level, 44 percent of performance expectations were aligned with the science framework and 13 percent with the TEL framework, with 3 percent in the sciences aligned with both. At the high school level, 44 percent of performance expectations were aligned with the science framework and 13 percent with the TEL framework (with no performance expectations aligned with both).

When looking only at the performance expectations in science, the content alignment of the NGSS with the NAEP science framework was low at grade 4 (36 percent) and moderate at the middle school and high school levels (about 50 percent at each grade level). Comparing NAEP science to the NGSS, alignment at grades 4 and 8 was similarly low (23 percent) and moderate (56 percent), respectively; at grade 12, the alignment of NAEP to the NGSS was substantial (71 percent).

Across grades, the greatest degree of alignment between the NGSS and the NAEP science framework was in life sciences and the lowest was in physical sciences, based on the content similarity ratings at both the objective level and at the content area level as a whole. From 48 to 54 percent of NGSS performance expectations in life sciences were aligned with NAEP objectives compared to from 29 to 42 percent of NGSS performance expectations in physical sciences. Looking at the content areas as a whole, life sciences was the only content area rated as similar at two grades (grades 8 and 12) whereas physical sciences was rated as similar only at grade 12, and Earth and space sciences only at grade 8. None of the content areas as a whole were rated as similar at grade 4.

When looking only at the performance expectations in engineering, technology, and applications of science (ETS), content alignment to the NAEP TEL framework was strong for NGSS performance expectations in engineering design (at least 75 percent at each grade level), but weaker for those in the sciences with connections to ETS, especially at the upper grades (as low as 38 percent). The alignment of NAEP TEL with the NGSS, in contrast, was weak at all grade levels, because many more assessment targets are in NAEP TEL as well as assessment areas or subareas that do not have corresponding disciplinary core or component ideas in the NGSS. In addition to engineering design at all three grade levels, both the NGSS and NAEP TEL include the effects of technology on society and the natural world at the middle and high school levels.

The NGSS and NAEP science framework emphasize some content at different grades. That is, some content that was not similar at the corresponding grade level was aligned at a higher or lower grade level in the other framework. In general, the percentage of objectives aligned at a different grade was low—representing no more than one fifth of the objectives. The one exception was for NAEP science at grade 4, where 59 percent of content statements were aligned at a lower of higher grade in the NGSS. The percentage aligned at a different grade decreased over the grade levels for both the NGSS and the NAEP science framework.

Notably, the NGSS and NAEP objectives at middle school/grade 8 that were aligned to other grades were only aligned at the higher grade level in the other framework (high school/grade 12)—i.e., none of the middle school performance expectations were aligned with NAEP grade 4 content statements in science, and none of the NAEP grade 8 content statements in science were aligned with NGSS performance expectations in grades K–5. In addition, some objectives at high school/grade 12 in both the NGSS and NAEP were aligned at the middle school/grade 8 level in the other framework. Thus, the difference between the NGSS and NAEP science framework at grade 8 was more in terms of what content is emphasized in middle school versus high school.

Both the NGSS and the NAEP science and TEL frameworks include objectives at each grade level that cover *unique content*. This reflects nongrouped objectives covering content that is in one framework but not in its counterpart at any grade. (Examples are given in exhibits 10–12 for science and exhibit 13 for TEL). The unique content, together with content that overlapped but was not aligned at any grade in the counterpart framework, represented between 43 and 48 percent of NGSS performance expectations in science and between 18 and 28 percent of NAEP science content statements. Unique content also represented between 14 and 55 percent of NGSS performance expectations in ETS and between 72 and 87 percent of NAEP TEL assessment targets. Unique content reflects areas where each program can contribute different information about student outcomes.

Practices alignment was uniformly strong, but the emphasis of NGSS performance expectations across the NAEP science and TEL practices differed from the emphases specified in the NAEP frameworks.

Ninety-nine percent of NGSS performance expectations in science were aligned with NAEP science practices and 81 percent of performance expectations in ETS were aligned with NAEP TEL practices.

The NGSS performance expectations in science were more strongly concentrated in the NAEP science practice of *using science principles* (60 percent across grades) than was specified in the NAEP science framework (30 to 40 percent across grades). In contrast, very few of the NGSS performance expectations aligned with *identifying science principles* (4 percent across grades) compared to the 20 to 30 percent specified for NAEP across grades. The emphasis on *using scientific inquiry* (22 percent) and *using technological design* (13 percent) was more comparable to NAEP science (30 and 10 percent, respectively, across grades).

> The NGSS performance expectations in ETS were strongly concentrated in the NAEP TEL practice of *developing solutions and achieving goals* (62 percent across grades), which was greater than what is specified in the NAEP TEL frameworks (40 percent across grades). Only small percentages of NGSS performance expectations aligned with NAEP's *understanding technological principles* (12 percent) and *communicating and collaborating* (7 percent) (compared to 30 percent in each practice across grades in NAEP TEL).
>
> **However, despite some strong indications of alignment between the NGSS and NAEP content and practices dimensions separately, when both content and practices were considered together, the NGSS and NAEP science framework were found to be not aligned at the *overall framework level*. That is, at each grade level, the two frameworks were rated as not similar. This was generally because panelists thought that the individual NGSS performance expectations often went beyond what would be expected based on the descriptions of the practices in the NAEP framework when they are applied to specific content statements, even if the science content covered was similar to that in the NGSS. (Neidorf et al., 2016, pp. 94–97, emphasis added)**

Major Conclusions and Implications

The AIR report (Neidorf et al, 2016) also included a set of major conclusions about the relationships among the NAEP science and TEL frameworks and NGSS based on all the various comparisons executed in the study and the judgments made by experts. It focused on implications regarding possible similarities and differences in the demands of assessments aligned to each of the three reference sources. The following is taken directly from the AIR report.

> Together, the results from the various components of the comparison study suggest that NGSS-based assessments and NAEP science and TEL assessments would be aligned to some degree, but each would also have unique content and different emphases in terms of science and TEL practices. This is because some of the grouped NGSS and NAEP objectives with overlapping content—those that were aligned—would likely lead to similar assessment items, but some were different enough that they would likely lead to assessment items with a different content focus. Additionally, those objectives that were not grouped (and either aligned at a lower or higher grade or not aligned at all) would represent unique content at the given grade.
>
> For example, content alignment of an NGSS-based assessment with the NAEP science assessment would likely be low at grade 4—moderate if the entire upper elementary grade band was considered—and moderate at the middle and high school levels. The lower alignment at grade 4 relates to the greater breadth of content in NAEP (evidenced by the greater number of nongrouped objectives) and the fact that some of the content in NAEP at grade 4 may be covered at a different grade in the NGSS's upper elementary grade band.

An NGSS-based assessment also would likely have a much greater emphasis—over half the assessment—on *using science principles* and a much lesser emphasis on *identifying science principles* than a NAEP science assessment—only 4 percent. This is not surprising given that NAEP explicitly includes declarative knowledge in this latter practice, where the NGSS emphasize the application of science knowledge.

Another implication looking across the study is that the content and practices embodied in NGSS performance expectations that involve engineering design are not fully covered by either the NAEP science or NAEP TEL framework, despite strong alignment with the engineering design assessment targets in NAEP TEL. This includes both performance expectations in engineering design and those in the sciences that involve design applications. Thus, assessment tasks involving engineering design could look quite different in the two programs despite these areas of overlap.

The NAEP science framework—which specifies the practice of *using technological design* (with which many of the NGSS performance expectations in science that involve design applications aligned)—is restricted to the consideration of scientific criteria, constraints, and trade-offs in making design decisions. This is in contrast to the NGSS (and NAEP TEL), which more fully reflect the engineering design process and include a broader range of considerations such as social and economic factors (excluded in NAEP science). Additionally, the NAEP TEL framework and assessments do not expect prior science content knowledge, in contrast to the NGSS, which require the application of science concepts. NAEP TEL, rather, provides the background on the science concepts needed to be successful on the items and tasks measuring the engineering design process.

A final implication is that the tasks that could be developed to assess the NGSS performance expectations in science and engineering would likely require students to use some mathematics that is beyond the corresponding grade level in the NAEP mathematics framework; in contrast, the NAEP science and TEL assessments require mathematics at or below the corresponding grade. In other words, some of the mathematics that could be required in an NGSS-based assessment would be at a higher level than what is required in NAEP science and TEL assessments. (Neidorf et al., 2016, pp. 98–99)

SECTION III. ANALYSIS OF THE NAEP SCIENCE FRAMEWORK AND ASSESSMENT RELATIVE TO STATE SCIENCE POLICY AND PRACTICES: STANDARDS, ASSESSMENTS, AND CLASSROOM INSTRUCTION

This section examines how the NAEP science framework aligns with science standards and assessments that have been adopted and implemented in the states. Three main questions are of interest: (1) Since publication of the NRC framework and the NGSS, how many states have adopted the NGSS or standards that are similar in nature? (2) How do the standards of those states that have not completely adopted the NGSS align with NAEP? and (3) For those states that have adopted the NGSS or similar standards, what is the status of the design and implementation of their state assessments relative to their standards? The section then seeks to establish what the states are doing in the way of instruction as related to the NRC framework and NGSS. It closes with an examination of trends in NAEP science assessment performance between 2009 and 2019 and what those results might imply about the current state-of-science education. Overall, the information provided in this section has substantial implications for considering where states are likely to be in science instruction and assessment by the time the current NAEP science assessment is administered in grade 8 in 2024 and when the updated science assessment is administered in grades 4 and 8 in 2028.

NAEP, NGSS, and State Science Standards Comparisons

Since the publication of the NRC framework and NGSS states, 21 states have explicitly adopted the NGSS as their state science standards and 24 other states have adopted standards that NSTA has designated as partial NGSS in that they are multidimensional standards like the NGSS. In such cases they have based their standards development on the NRC framework and have typically adhered to the central idea of integrated performance expectations based on two or more dimensions as in the NGSS.

In February 2021, HumRRO published a report for NAGB entitled *Comparative Analysis of the NAEP Science Framework and State Science Standards* (Dickinson et al., 2021).

> The method used to conduct this comparative study relied heavily on obtaining experts' judgments regarding the overlap of subject matter between the NAEP science framework and states' science standards…. The comparative analysis included only the standards from states that did not fully adopt the NGSS (i.e., 6 states) and those that partially adopted the NGSS (i.e., 24 states, including the Department of Defense schools). The science standards from the partial NGSS adopting states, which are based on the NRC framework, were included in the study. However, NGSS performance expectations were excluded from the analysis, given the previous study comparing NAEP and NGSS. (Dickinson et al., 2021, p. 1.)

Table 3 below shows which state's standards were included in the analysis.

To execute this analysis. the HumRRO team started by pulling out all content statements, objectives, and performance expectations outside NGSS. The focus was on the content overlap and not the practice overlap. They did some preliminary distillation by matching state and NAEP content statements to look at state and NAEP content side by side to rate the overlap. Also, they identified content-related practices in state statements. They then developed a consensus statement to give the overall impression of where states are doing things differently. They tried to include only statements in the science domains and cut out technology and engineering statements if easy to do so. They did not look explicitly at the TEL framework. An important point to note is that in conducting this work, the comparison of NAEP to state standards is based on an aggregation of all the states' standards rather than a state-by-state individual comparison. Thus, the comparison paints a very broad picture of overlap between the NAEP framework and the partial NGSS and non-NGSS states as a whole. Further details about the methodology and specific sets of outcomes can be found in the complete report.

Table 3. Non-NGSS, partial NGSS, and full NGSS adopting states

Non-NGSS Adopting States	Partial NGSS Adopting States	Full NGSS Adopting States
Florida	Alaska	Arkansas
North Carolina	Alabama	California
Ohio	Arizona	Connecticut
Pennsylvania	Colorado	Delaware
Texas	Department of Defense Education Activity	District of Columbia
Virginia	Georgia	Hawaii
West Virginia	Idaho	Illinois
	Indiana	Iowa
	Louisiana	Kansas
	Massachusetts	Kentucky
	Minnesota	Maine
	Missouri	Maryland
	Mississippi	Michigan
	Montana	Nevada
	North Dakota	New Hampshire
	Nebraska	New Jersey
	New York	New Mexico
	Oklahoma	Oregon
	South Carolina	Rhode Island
	South Dakota	Vermont
	Tennessee	Washington
	Utah	
	Wisconsin	
	Wyoming	

SOURCE: Dickinson et al., 2021, p.12.

The following conclusions, based on the analyses completed by both the HumRRO staff and the outside experts, were offered in the report. They are reprinted here verbatim from that document (Dickinson et al., 2021, pp. 6–7).

1. When examining the content covered by the full set of states' science standards (with any NGSS performance expectations removed), there are many state statements that do not overlap in content with any NAEP statement.
   - At grade 4, 31 percent of all state content statements reviewed by HumRRO experts and external science experts were rated as not overlapping a NAEP content statement.

- At grade 8, 32 percent of all state content statements reviewed by HumRRO experts and external science experts were rated as not overlapping a NAEP content statement.
- At grade 12, 55 percent of all state content statements reviewed by HumRRO experts and external science experts were rated as not overlapping a NAEP content statement.

2. Considering only the state content statements that the experts reviewed, all NAEP statements at least partially overlap in content with at least one state statement. In most cases, NAEP statements overlap in content with multiple state statements. Finally, in some cases, NAEP content statements are fully reflected in a combination of multiple state content statements.
   - For each NAEP content statement HumRRO identified multiple state content statements with overlapping content. Review by external experts verified content overlap with at least one of these pairings for each NAEP content statement.
   - Experts noted that there were instances where a combination of state content statements would fully cover the content in a NAEP content statement.
3. Experts rated the least amount of content overlap between NAEP and states' standards at grade 12.
   - Overall, at grade 12, 19 percent of state content statements reviewed by expert panelists were rated as having no content overlap with a NAEP content statement.
4. As with the NAEP-to-NGSS comparison, experts rated the least amount of overlap in content between NAEP and states' standards for the Physical Science domain, especially at grades 8 and 12.
   - At grade 8, 9 percent of state Physical Science content statements reviewed by expert panelists were rated as not overlapping a NAEP content statement.
   - At grade 12, 25 percent of state Physical Science content statements reviewed by expert panelists were rated as not overlapping a NAEP content statement.
5. Science experts identified the grades 4 and 8 state content statements to most frequently reflect NAEP's Identifying Science Practices and the grade 12 state content statements to most frequently reflect NAEP's Using Science Practices. The experts least frequently identified the states' content statements to reflect NAEP's Using Technological Design.
   - At grades 4 and 8, 54 percent of all state content statements reviewed by expert panelists were rated as reflecting NAEP's *Identifying Science Practices*.
   - At grade 12, 51 percent of all state content statements reviewed by expert panelists were rated as reflecting NAEP's *Using Science Practices*.
   - Across the grade levels, between 1 percent and 5 percent of all state content statements reviewed by expert panelists were rated as reflecting NAEP's *Using Technological Design*.
6. Science experts noted that states whose standards are based on the NRC K–12 Science Framework have more in common with NAEP than states whose standards are not based on that framework.
   - Consensus statements developed by both the grade 8 and grade 12 expert panels included assertions that they observed more content overlap between NAEP and the science standards of states who based their standards on the NRC K–12 Science Framework.

State Science Policy and Practices: Standards, Assessments, and Classroom Instruction

Thus far we have established three important findings that bear on a judgment about the validity of results from the NAEP science assessment at the time of its next implementation in 2024 and subsequently in 2028 if substantial revision is not made to both the framework and the derivative assessment before the 2028 administration. First, as described in Section II, major differences exist between the NAEP framework and the NRC *Framework for K–12 Science Education* and the derivative *Next Generation Science Standards* in science content, science and engineering practices, and in their juxtaposition in the form of performance expectations. Second, currently, 45 states (including Department of Defense Education Activity) have either fully adopted the NGSS as their state standards (21) or adopted NGSS-like state science standards (24). Third, when the latter states' standards and those of non-NGSS adopting states (6) are compared with NAEP content, several substantive differences arise. Thus, it seems reasonable to conclude that the current NAEP science framework may be substantially at variance with and lagging a contemporary view of what we want students to know and be able to do in science at grades 4, 8, and 12 and how we would expect them to show proficiency. That view of proficiency has become policy for the preponderance of states and is realized via their state science standards.

How far out of synch the NAEP framework and assessment may be with what instruction and science assessment look like in most states in 2024 and 2028 and with what students know and can do in science depends very much on the following timelines: (a) state adoption of new standards following publication of the NRC framework and NGSS, (b) implementation of new state assessments aligned with those standards, (c) availability of curricular and instructional resources reflecting the new vision of science learning and instruction, and (d) implementation of teacher professional learning programs relative to each of a–c. We provide information relevant to these concerns in the following material.

Time Course for Adoption of New State Standards and Assessments

An article that includes information about adoption of new science standards by Smith (2020) discusses results from the two most recent National Survey of Science and Mathematics Education (NSSME) completed in 2012 and 2018 (see also Banilower et al, 2018). Table 4 shows the pattern of adoption of the NGSS or NGSS-like standards by the states as of 2018. The 16 early adopters did so between 2013 and 2015 while the 24 late adopters did so between 2015 and 2017, and non-adopters had not adopted by spring 2018 when NSSME collected data. Note that there are some differences between Table 4 and the Table 3 shown earlier regarding NGSS adoptions. For example, Florida, North Carolina, Ohio, Pennsylvania, Virginia, and Texas remain nonadopters as of 2021 and they have been joined by West Virginia, which was previously designated as a late adopter. In contrast, Arizona, Alaska, Maine, and Minnesota have moved from the nonadopter group into the late adopter group.

Table 4. Adoption of NGSS or NGSS-like standards – August 2018

Early Adopters	Late Adopters	Non-Adopters
California*	Alabama	Alaska
Delaware*	Arkansas*	Arizona*
District of Columbia	Colorado	Florida
Illinois*	Connecticut	Maine
Kansas*	Georgia*	Minnesota*
Kentucky*	Hawaii	North Carolina*
Maryland*	Idaho	North Dakota
Nevada	Indiana	Ohio*
New Hampshire	Iowa*	Pennsylvania
New Jersey*	Louisiana	Texas
Oklahoma	Massachusetts*	Virginia
Oregon*	Michigan*	
Rhode Island*	Missouri	
South Carolina	Mississippi	
Vermont*	Montana*	
Washington*	Nebraska	
	New Mexico	
	New York*	
	South Dakota*	
	Tennessee*	
	Utah	
	West Virginia*	
	Wisconsin	
	Wyoming	

* Lead state

SOURCE: Data are from Smith, 2020.

One of the many factors driving instructional practice relative to the vision of science teaching, learning, and assessment contained in the NRC framework and state science standards aligned with that vision is the status of each state's large-scale science assessment relative to its adopted standards. Consistent with federal requirements, states that have adopted new science standards are obligated to implement new assessments aligned with those standards having the minimum requirement for at least one assessment in each of the elementary school grade bands (grades 3–5), the middle school grade band (grades 6–8), and the high school grade band (grades 9–12). An analysis for this paper by AIR staff of the 21 states that have fully adopted the NGSS (14 of which are shown as lead adopters in the table above) reveals that all but one of those 21 states, Arkansas, has already developed and in most cases implemented a large-scale science assessment that they claim is aligned with the NGSS. The timeline of assessment implementation varies from 2014 to 2019, with some implementations planned for 2020 but delayed until 2021, given suspension of all large-scale assessments in spring 2020 due to the COVID-19 pandemic. The timelines for implementation of new science assessments for the states classified as partial NGSS are less clear although for the majority of those states their websites indicate that their standards and assessments require integration of the disciplinary core content and practices described in the NRC Framework and many include mention of the third dimension of crosscutting concepts. Some have adopted many if not all the performance expectations from the NGSS. For some states, the timeline for full implementation of new assessments extends to 2025.

Survey Information on Science Instructional Practices: 2018 vs. 2012

NSSME has provided periodic snapshots of K–12 science instruction in the United States for more than 40 years. Study topics include teacher backgrounds and beliefs, professional learning opportunities, course offerings, instructional objectives and activities, resources for instruction, and policies affecting instruction. The two most recent studies were conducted in 2012 and 2018. The 2012 study provides baseline data on multiple indicators prior to publication of the NGSS. From 2013 to 2018, 39 states and the District of Columbia adopted the NGSS or NGSS-like standards. By the time the 2018 survey was conducted, NGSS states accounted for more than two thirds of the nation's K–12 students. The 2018 study provides a snapshot of the state-of-science instruction in 2018 relative to the vision of the NRC framework and the NGSS, including the opportunity to observe any impact on instructional beliefs and practices relative to 2012 in light of the publication of the NRC framework in 2012 and the NGSS in 2013.

Smith's 2020 analysis and discussion of results from the 2018 NSSME (Banilower et al., 2018) shows that states have been slow in the full implementation of their new science standards in terms of making a difference in instructional practice. As discussed by Smith, one reason for the slowness is the lack of good curriculum materials aligned with the new standards. Another reason for the slowness is the need for substantial teacher professional development related to understanding the science and engineering practices as well as the meaning and manifestation of integration of the multiple dimensions expressed by the performance expectations. Related to the latter, valid, high-quality assessments reflecting the kinds of performances expected from students also have been lacking. In general, during the period in question there was a paucity of such examples for classroom use as well as at the large-scale state assessment level given the timeline for implementation of new NGSS-aligned assessments as described above from the analysis of state websites by AIR staff.

Regarding professional development, Smith (2020) reports that roughly four of five secondary science teachers (i.e., middle school and high school) participated in science-focused professional development in the preceding 3 years, in contrast to three of five elementary science teachers. Only about half of schools or districts offered any science-focused professional development in the preceding 3 years, and participation data were largely unchanged since 2012. About a third of secondary teachers participated in more than 35 hours of professional development in the 3 years preceding 2018, and more than 4 in 10 elementary teachers had none. As Smith notes, even 35 hours, spread over 3 years, is not much considering prominent instructional practices and the shifts that the framework and NGSS entail.

Among the other results summarized by Smith were results regarding data on instructional practices and emphases in elementary, middle school, and high school classrooms (see Smith 2020, Table 1). Most importantly, in 2018 the most frequent "heavy emphasis" instructional objective reported by Science teachers was "understanding science concepts," particularly in middle and high schools (47 percent of Science teachers in elementary schools, 77 percent in middle schools, and 76 percent in High schools). In contrast, the second most frequent objective with a heavy emphasis reported by teachers was "learning how to do science" but only in 26 percent of Science classes in elementary schools, 46 percent in middle schools, and 41 percent in High schools. Smith concluded that:

> Despite widespread adoption of the NGSS and NGSS-like standards, data from the NSSME+ point to few differences in science instruction compared to 2012. Further, the data from teachers in adopting states vary little from those in non-adopting states. Among the few differences, we do see encouraging signs. Among them, classes in adopting states were more likely to emphasize learning how to do engineering, and they were less likely to emphasize learning vocabulary and facts. In terms of instructional activities, classes in early-adopting states were less likely to rely on lecture and more likely to have students do hands-on activities. However, the data overall suggest that much work lies ahead to achieve the vision laid out in the framework and the standards themselves (Smith, 2020 p. 608).

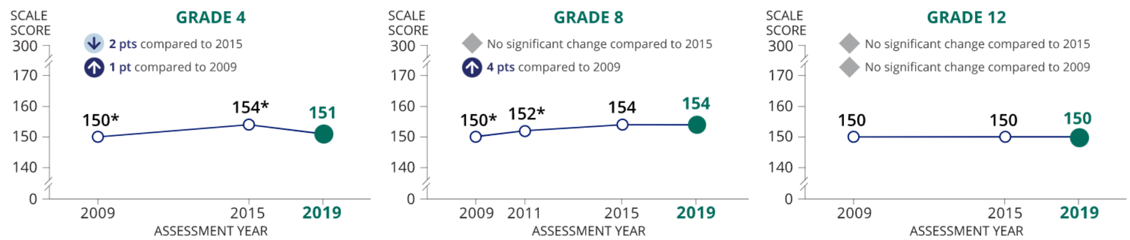
Perhaps not surprising is that substantial changes in science instructional practices were not observed in the 2018 NSSME survey relative to 2012 and that aspects of the vision for science teaching and learning embodied in the NRC framework and NGSS were less well represented in teacher beliefs and instructional practices. As noted by Smith (2020), 5 years may not be enough time. Many of the critical factors needed to spur change are only now becoming more prominent with further changes on the horizon during the next 2 years when NAEP science is set to be administered again for grade 8 only. Among the drivers of change are new state science assessments reflecting the NGSS or similar science standards. In addition, growth in both commercially available and open education resources (OER) aligned with the NGSS has been significant. One of the largest of the OER curricular initiatives is the foundation-funded OpenSciEd project (https://www.openscied.org/about/), which has generated instructional units covering all the middle school NGSS performance expectations and is working on similar materials for other grade levels. At the classroom level, assessment resources have been developed to support formative and summative assessment practices in ways aligned with the multidimensional assessment vision described in the 2014 NRC report, *Developing Assessments for the Next Generation Science Standards* (Pellegrino et al., 2014). See for example the materials available from the *Next Generation Science Assessment Project* (http://nextgenscienceassessment.org) and from the Stanford *NGSS Assessment Project* (https://scienceeducation.stanford.edu/assessment).

NAEP Science Performance Changes Over Time

One final source of information about possible changes in science education in the United States over time might be gleaned from an examination of performance on the NAEP science assessment for the period from 2009 when the new science framework and assessment were first implemented to 2019 when NAEP science was delivered as a digitally based assessment, in contrast to prior years. These data track student performance both before and after the NRC framework and NGSS.

The 2019 NAEP science scale score results are shown in Figure 4 for each of the grade levels in comparison to prior administrations back to 2009. As can be seen in Figure 4, the average science score for the nation at grade 4 was lower by 2 points compared to 2015, whereas average scale scores at grades 8 and 12 did not significantly differ from 2015. At grades 4 and 8, average scale scores were higher when compared to 2009, while the average scale score at grade 12 was not significantly different across years.

Figure 4. Average scores in NAEP science, by grade: 2009–2019

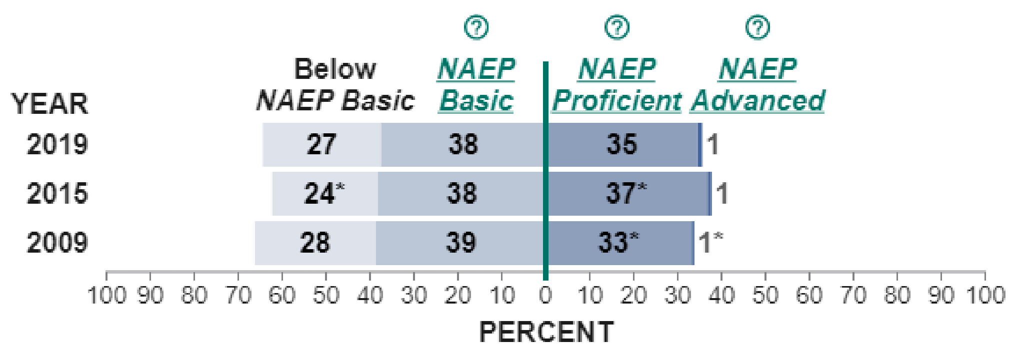


*Significant different ($p < .05$) from 2019.

SOURCE: The Nation's Report Card, 2019. Reprinted with permission.

Although the absolute levels of the scale scores and the trends in those scores are important indicators of student performance, of particular significance is the reporting of results in terms of achievement levels. As shown below in Figures 5, 6, and 7, the rates by which students were classified into the achievement levels varied across the grades with the highest rate of *Proficient* classifications occurring in grade 4, slightly lower levels of proficiency at grade 8 and substantially lower student proficiency classifications at grade 12. Note that at all three grade levels, there is a very low level of classification of student performance at the *Advanced* level. This finding holds across years.

Figure 5. NAEP achievement-level results in NAEP science for fourth-grade students: 2009, 2015, and 2019

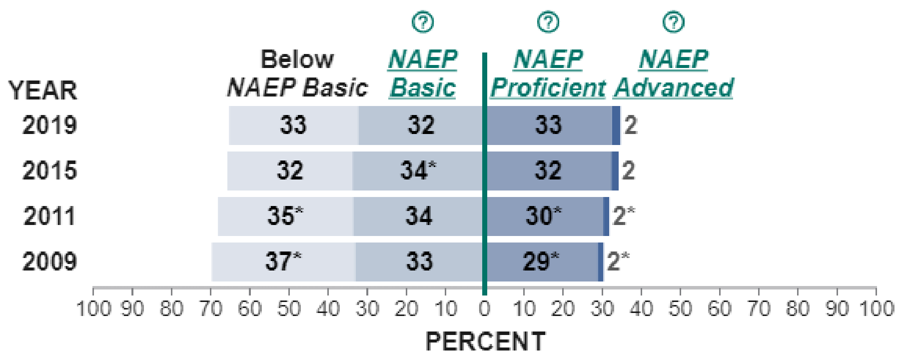


* Significantly different ($p < .05$) from 2019

Note: NAEP achievement levels are to be used on a trial basis and should be interpreted and used with caution.

SOURCE: The Nation's Report Card, 2019. Reprinted with permission.

Figure 6. NAEP achievement-level results in NAEP science for eighth-grade students: Various years, 2009–2019

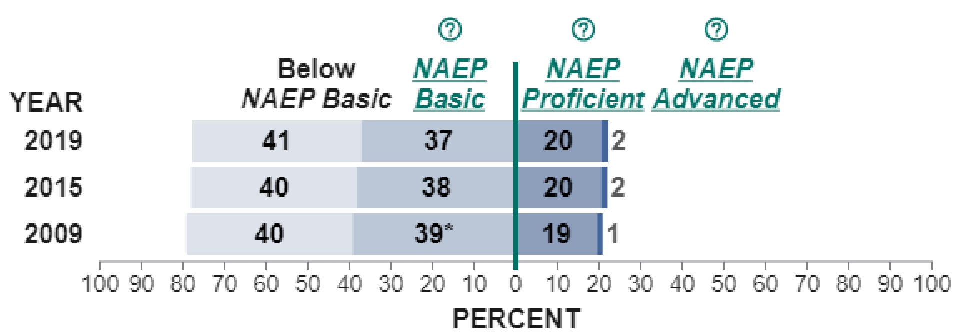


* Significantly different (*p* < .05) from 2019.

Note: NAEP achievement levels are to be used on a trial basis and should be interpreted and used with caution.

SOURCE: The Nation's Report Card, 2019. Reprinted with permission.

Figure 7. NAEP achievement-level results in NAEP science for twelfth-grade students: 2009, 2015, and 2019



* Significantly different (*p* < .05) from 2019.

Note: NAEP achievement levels are to be used on a trial basis and should be interpreted and used with caution.

SOURCE: The Nation's Report Card, 2019. Reprinted with permission.

Perhaps there are two major takeaways from this examination of the NAEP science assessment results. First, not much has changed over time implying that science instruction also has not changed substantially despite the existence and adoption of new standards with higher expectations about what students are supposed to know and be able to do. Despite their differences in content and format of science assessment, the most recent trend results from the PISA science assessment and the TIMSS science assessment largely corroborate the lack of change in U.S. science performance during the last decade. Second, those new standards are much needed because science performance across the grade bands is relatively poor and only declines across grades. The vast majority of students are below *Proficient* as defined by the NAEP achievement levels.

The real concerns then are threefold: (1) whether instruction aligned with the new standards will take hold in ways envisioned by the NRC framework and NGSS and change performance, (2) whether the NAEP science assessment can track the impact of those changes given the differences between the NAEP framework, the NGSS and the majority of state science standards, and (3) whether NAEP science and/or TEL will have sufficient instructional sensitivity to reveal what has and has not happened over time when next administered in 2024 or 2028.

SECTION IV. TECHNOLOGY IMPLICATIONS FOR NAEP SCIENCE

This section briefly considers how various developments in digital technologies need to be considered in reviewing the existing NAEP science framework and assessment and envisioning possibilities for their updating. The discussion that follows focuses on the affordances of technology regarding the constructs that could be included in a revised framework and the associated task design, data capture, and data analytic issues involved in an assessment aligned to an updated framework. The section concludes with a brief discussion of practical and equity concerns related to digitally based assessment of science and technology proficiency.

Technology and NAEP Assessment

During the last two decades, much has been written and speculation made about the power of technology to both improve and transform assessment across a range of assessment contexts and purposes (e.g., Behrens et al., 2019; Bennett, 2008; Drasgow, 2016; Gane et al., 2018, Pellegrino & Quellmalz, 2010; Pellegrino et al., 2001). Although technology's potential for improving and transforming assessment has yet to be fully realized, the vast majority of national-, international-, and state-level assessments of science and technology have moved almost entirely to digital presentations of materials accompanied by technology-based data capture for purposes of scoring, analysis, and reporting. Within the past decade, PISA (2015, 2018), eTIMSS (2019), NAEP Science (2019), and NAEP TEL (2014, 2018) have been delivered via technology using various types of devices including laptops, tablets, and desktops.

Not only has technology changed assessment delivery, response capture, and scoring, it also has had a significant effect on assessment design. This includes the types of tasks and situations that can be presented to students with the goal of tapping into various forms of scientific thinking and reasoning aligned with the practices of science and engineering as found in the NAEP science and TEL frameworks and NGSS. For the NAEP program, some of the newer task types that take advantage of some of technology's affordances were briefly described in Section II, including the scenario-based tasks added to the NAEP science assessment in 2019. The latter were modeled to a great extent after the digitally based tasks were first introduced in NAEP TEL in 2014. The literature on NAEP has considered a number of the affordances of technology for the assessment program, including implementation and analysis of the types of scenario-based tasks in science piloted by NAEP in 2015 and included as part of NAEP 2019, including analyses of student response data (e.g., Bennett, 2008; Bergner & von Davier, 2019; Duran et al., 2020; Lee at al., 2019; Mullis, 2019). The purpose of the discussion that follows is to briefly highlight some of the possibilities for the future of NAEP science as related to both the framework and the assessment.

Opportunities and Possibilities for NAEP Science

As discussed in prior sections of this paper, conceptions of scientific and technological competence have evolved during the last 10–15 years, some of which align with the current NAEP framework and assessment while others go beyond both. Thus, in considering possible changes for the design of the 2028 administration of the science assessment, it will be important to consider how some of the affordances of technology discussed below may influence the nature of the competencies included in the framework, the design of the assessment tasks needed to provide evidence of those competencies, and the associated measurement and interpretive challenges, especially in light of goals for reporting the results. The assessment as evidentiary reasoning argument presented in the NRC report *Knowing What Students Know: The Science and Design of Educational Assessment* (Pellegrino et al., 2001) frames the discussion. In Chapter 7 of that report many of the affordances of technology for advancement of assessment design and practice are discussed in terms of the three interconnected components of the assessment triangle: *Cognition*, *Observation,* and *Interpretation*. As argued in that report:

> The role of any given technology advance or tool can often be differentiated by its primary locus of effect within the assessment triangle. For linking *cognition* and *observation*, technology makes it possible to design tasks with more principled connections to cognitive theories of task demands and solution processes. Technology also makes it possible to design and present tasks that tap complex forms of knowledge and reasoning. These aspects of cognition would be difficult if not impossible to engage and assess through traditional methods. Related to the link between *observation* and *interpretation*, technology makes it possible to score and interpret multiple aspects of student performance on a wide range of tasks carefully chosen for their cognitive features, and to compare the resulting performance data against profiles that have interpretive value. (Pellegrino et al., 2001, p. 252)

The discussion that follows elaborates on these general ideas regarding NAEP science. It focuses is on the constructs that could be represented in an updated framework, the ways in which those constructs could be realized in the assessment environment, and some of the interpretive challenges and solutions associated with doing so for purposes of measurement and reporting.

The *Cognition* vertex of the assessment triangle. What matters in assessment is what we are trying to reason about – the contemporary conception of student *Cognition* in a domain like science that matters to scientists, educators, and society. A contemporary view of multidimensional proficiency in science includes the expectation that learners should be able to use their disciplinary core knowledge to engage in a variety of science practices in the service of explaining phenomena and designing solutions while answering challenging questions (NRC, 2012). As the conception of student cognition changes and expands in terms of what students are supposed to know and be able to do, as has been the case for science, technology affords opportunities for substantially changing and extending the *Observation* and *Interpretation* components of the assessment triangle in order to more adequately represent and provide evidence about the constructs of interest. Doing so enhances the entire evidentiary reasoning process and the validity of the NAEP science assessment given its intended interpretive use as an index of trends in U.S. science achievement.

The *Observation* vertex of the assessment triangle. Technology provides opportunities for presentation of dynamic stimuli (e.g., videos, graphics, 2- and 3-D simulations) that can be interacted with in the service of eliciting relevant sets of responses from students. Simultaneously, technology enables the generation and capture of a variety of response products, including situations in which students generate responses using multiple modalities (e.g., drawing and writing). In general, *technology-enhanced assessments* are defined by their capacity to provide novel stimuli and/or responses that would not be possible with traditional, paper-and-pencil assessment formats. Technology-enhanced assessments such as those included in NAEP science 2019 and NAEP TEL enable engagement with a variety of science and engineering practices (e.g., generating models, planning and carrying out investigations, engaging in computational thinking) by opening the door to interactive stimulus environments and response formats that better match the intended reasoning and response processes that form the basis for desired claims about student proficiency (Gorin & Mislevy, 2013).

Students' interactions with these technology-enhanced assessments can be logged to provide data on how they engage in particular processes. In certain applications such as engineering or experimental design, the process by which one completes the activity can be as important a piece of information about knowledge and skill as the final product. In these cases, understanding the operations that students performed in the process of creating the final product may be critical to evaluating students' proficiency. Log data offer the opportunity to reveal these actions, including where and how students spend their time, and what choices they make in situations like using a simulation. Such applications offer the potential to provide large volumes of "click-stream" and other forms of response process data that might be useful for inferences about student thinking as discussed by Ercikan and Pellegrino (2017). Such data can be complex, however, and must be segmented and analyzed in construct-relevant ways if they are to be reliable and valid for a given interpretive use. An ongoing challenge is identifying how to take massive volumes of log data and distill it into actionable information to make judgements about students' knowledge, skills, and abilities (e.g., Bergner & von Davier, 2019).

The *Interpretation* vertex of the assessment triangle. Technology offers significant opportunities for enhancement of the reasoning-from-evidence process given the types of observations described above. Collecting the types of data just mentioned in the discussion of observations makes little sense unless there are ways to reliably and meaningfully interpret them. This can evolve through mechanisms such as automated scoring of responses and application of complex parsing, statistical and inferential models for response process data. Much has been written recently about the opportunities of student-response-process data for capturing what students are doing when they solve problems and answer questions related to science and technology (see Ercikan & Pellegrino, 2017). Such data include the time taken to perform various actions, the actual activities chosen, and their sequence and organization. The potential exists for examining the global and local strategies students use while solving assessment problems and the implications, including how such strategies relate to the accuracy or appropriateness of final responses. Although capturing such data in a digital environment is "easy," making sense of the data is far more complicated. The same can be said for capturing data to constructed response questions where students may be expressing in written and/or graphical form an argument or explanation about some scientific problem or phenomenon, describing the design of a scientific investigation, or representing a model of some structure or process.

The data capture contexts described above are challenging regarding scoring and interpretation. It is here that AI and machine learning may play a significant role in future science assessments. Machine learning mimics human scoring processes by first "learning" from scoring by human experts to develop algorithmic models and then applying those models to automated scoring of new student responses (Zhai, Yin, et al., 2020). Advances have been made in the automated scoring of short, written, constructed responses for various topics and content in science and other subjects (see Beggrow et al., 2014; Nehm et al., 2012; Williamson et al., 2012). However, automated scoring of other types of constructed response products, such as the features that might be included in drawings and other forms of graphical representation associated with a practice like modeling, has not yet been explored in-depth (see Gerard et al. [2016] for one promising attempt). For both written and graphical responses, well-designed task models that define the features of responses that matter for scoring are needed. This likely will have a considerable impact on the development of automated scoring systems that are both reliable and practical for implementation across a variety of assessment contexts.

Developments in machine learning also may allow researchers to analyze complex response process data of the type described above (Zhai, 2021). Traditional statistical methods are often difficult or inappropriate to apply to such data. Machine learning, however, might assist in analyzing these types of data to reveal patterns that provide important insights into students' cognitive processes in problem solving (Zhai, Haudek, et al., 2020; Zhai, Yin, et al., 2020). Such data may prove to be especially informative about student thinking and reasoning and thus add greatly to the knowledge gained about student competence from large-scale assessments like NAEP that go beyond the performance accuracy data they now provide. An interesting example was provided in a recent study by Pohl et al. (2021). The authors showed that differences in student response processes, of the type described above, when combined with scoring methods, can significantly change the interpretation of a country's performance on a large-scale assessment such as PISA. Their study findings showed that current reporting practices in PISA confound differences in test-taking behavior with differences in competencies and can do so in a different way for different examinees, threatening the validity and fairness of comparisons. Thus, their argument is that test-taking behavior is not a confounding factor introducing construct-irrelevant variance, but that it is something that provides important information on how examinees approach tasks, which can be meaningful outside the testing situation. Disentangling and reporting all these factors as part of a performance portfolio could result in fairer comparisons across groups and enables a better understanding of student competencies and important possible causes of variations in performance. Explorations of the analysis and interpretation of response process data have been initiated for some of the NAEP science tasks (Bergner & von Davier, 2019; Lee at al., 2019) and the results suggest that this is a fertile area for future exploration, albeit taking into consideration some of the cautions mentioned below.

Areas of Concern for NAEP Science

Assessments that can tap into and measure multidimensional knowledge take the form of *knowledge-in-use* tasks (Harris et al., 2019). Technology can make practical the design, administration, and scoring of such tasks. An area of concern is that technology by itself is not enough: Technology cannot fix assessments that are poorly designed or misaligned with the desired learning targets. Instead, technology considerations need to be integrated with assessments through a transparent and principled design process. As the targets of assessment become more conceptually complicated, with demands such as jointly measuring science practices and conceptual knowledge, a principled design process is essential for developing relevant and valid assessment tasks (Gorin & Mislevy, 2013; Pellegrino et al., 2014). A principled design process like *Evidence Centered Design* (Mislevy, 2018; Mislevy & Haertel, 2006; Mislevy & Riconscente, 2006) that identifies task and response features that matter can also move the scoring process from a black box statistical approach to one that is more transparent and defensible. Explicit task and response models with defined response features can lead to improved human scoring as well. A caveat, in a general sense, for NAEP science is that if NAEP wants to capture more complex forms of scientific thinking and reasoning using digital environments, this cannot be done by simply applying technology to the sense-making process "after the fact," which seldom is well done or efficient. Thus, a very deliberate design process needs to be used for task design and data capture that takes into consideration the relevant forms of evidence and the means for interpretation of that evidence throughout the task design, task refinement, and task validation processes.

Although technology can enhance many aspects of large-scale assessment, concerns have arisen about the equity and fairness of digitally based assessment. An area of concern is comparability of results and validity of inferences derived from performance obtained across different modes of assessment, especially for varying groups of students (see Berman et al., 2020). As NAEP science has moved from paper-and-pencil assessment to digitally based assessment, the general focus has been on mode comparability and concerns about student familiarity and differential access to the hardware and software used (see Way & Strain-Seymour, 2021). As the digital assessment world advances, a significant issue for future large-scale science and technology assessments is determining how student background characteristics including language, culture, and educational experience influence performance on different types of tasks and innovative assessment designs that leverage the power of technology. As the tasks become more innovative, equity and fairness concerns may become even more important than general mode comparability effects.

Another area of concern relates to cost, efficiency, and feasibility. Complex, scenario-based tasks such as those found in NAEP science and TEL are challenging to design well and costly to create relative to more conventional tasks. They typically also take significant amounts of time for students to complete. Given the nature of the scenarios, they also tend to be memorable because they depict interesting, engaging, and often realistic problem-solving situations. They exemplify and perhaps magnify many of the challenges that have long been noted about the inclusion of performance tasks in large-scale testing programs such as NAEP. Davey et al. (2015) provided an excellent discussion of the many challenges associated with development and deployment of performance assessments for constructs represented in science standards such as the NGSS. Their report included a discussion of many of the measurement and statistical challenges associated with the interpretation and reporting of performance data. Thus, NAEP science will have to consider tradeoffs associated with inclusion of technology-based assessment tasks relative to adequate representation and sampling of the constructs of interest. The fact that NAEP science uses a matrix-sampled block design for selection and administration of tasks may mitigate some of the many concerns noted by Davey et al. (2015). NAEP can offer leadership to the large-scale science assessment field in providing a vision and examples of how science and technology competence can and should be assessed and reported.

SECTION V. CONCLUSIONS AND RECOMMENDATIONS

The purpose of this white paper is to consider the need for a revised NAEP science framework and its possible scope and focus including expansion to aspects of what is represented in NAEP TEL. The goal is to provide the NAEP NVS Panel and the NAEP program input about possible futures for NAEP science. As such, the paper can also serve as input to NAGB's deliberations in 2022 about the need and possible directions for a revision of the science framework that would in turn serve as the basis for development of the NAEP science assessment scheduled for 2028.

Topics Covered Across Sections I–IV

- A brief history of the current NAEP science and TEL frameworks and assessments and their projected use over the next seven years through 2028
- Brief descriptions of the content and focus of the NAEP science and NAEP TEL frameworks and assessments as well as the National Research Council's *Framework for K–12 Science Education* (NRC, 2012) and the derivative *Next Generation Science Standards* (NRC, 2013)
- Results from an extensive comparison of the content and focus of both NAEP frameworks with the NGSS
- Information on the timeline and status of state adoptions of the NGSS or similar science standards derived from the NRC framework
- Results from a study comparing the content of state science standards with the NAEP science framework for states with science standards similar but not identical to the NGSS together with states with standards unrelated to the NGSS or NRC framework
- Information about the status of development and implementation of standards-aligned, large-scale state science assessments for those states that have either adopted the NGSS or similar standards
- Information about the conditions of science instruction based on the 2012 and 2018 National Survey of Science and Mathematics Education
- Trends in NAEP science assessment performance for 2009–2019 for students at grades 4, 8, and 12
- A discussion of the affordances of technology for consideration in refinements and revisions to the NAEP science framework and assessment

Conclusions and Implications

Alignment of NAEP Science and NAEP TEL With Other Frameworks and Standards

The frameworks for NAEP science and NAEP TEL were developed before the NRC framework and NGSS and all within a window of approximately 6–7 years. All four drew upon bodies of theory, research, and practice regarding the knowing, learning, and teaching of science and technology available at the time of their development. Given time lags among them, it should come as no surprise that there are both significant similarities between the two NAEP frameworks and the NGSS and substantial differences as determined by a 2016 AIR comparison study (Neidorf et al., 2016).

Conclusion 1. Overlap exists between NAEP science and NGSS in terms of the focal science content areas—physical science, life science, and Earth and space science—and subtopic areas within each domain, but substantial differences exist in specific content. The differences are magnified in the movement from grade 4 to grade 8 to grade 12. One reason for the pattern of differences across grade levels is that the NGSS is based on a set of four disciplinary core ideas (DCIs) in each domain of science, and each DCI is elaborated across grades in terms of knowledge expectations. This was a deliberate design decision in the NRC framework that is replicated in the NGSS. In contrast, the NAEP framework changes content emphasis and focus across grades 4, 8, and 12 with an increasing emphasis on physical science content at grades 8 and 12, especially at grade 12.

Conclusion 2. Overlap exists between the NAEP framework and NGSS regarding the concept of science practices that describe ways of thinking about and reasoning with science content. The NAEP science practices and the NGSS science practices are different in at least two ways, however. Two of the four NAEP practices are considered to be more focused on "knowing science" in contrast to the other two that are more focused on "doing science." In contrast, the NGSS includes eight specific science and engineering practices, each of which fall under the category of science inquiry ("doing science") and/or engineering design. In general, the NGSS science and engineering practices are more demanding than at least two of the NAEP practices, and this is especially apparent when the practices are combined with content to form performance expectations as noted below.

Conclusion 3. Although both NAEP and NGSS express the targeted knowledge and skills for students in the form of performance expectations, the NGSS performance expectations are considered to demand much more in the way of application of disciplinary content knowledge to answer a question involving a science practice to demonstrate proficiency. Regarding the latter point, the 2016 AIR comparison study concluded: "… despite some strong indications of alignment between the NGSS and NAEP content and practice dimensions separately, when both content and practices were considered together, the NGSS and NAEP science framework were found to be not aligned at the *overall framework level*. That is, at each grade level, the two frameworks were rated as not similar. This was generally because panelists thought that the individual NGSS performance expectations often went beyond what would be expected based on the descriptions of the practices in the NAEP framework when they are applied to specific content statements, even if the science content covered was similar to that in the NGSS" (Neidorf et al., 2016, p. 97).

Conclusion 4. The NGSS includes a fourth dimension in its content framework—engineering, technology, and the applications of science as well as two engineering practices—defining problems and designing solutions. The AIR comparison study (Neidorf et al., 2016) showed that the NGSS has overlap with both NAEP science and NAEP TEL with respect to certain aspects of engineering, technology, and design. The overlap is highly variable, however, depending on grade level and direction of comparison. A significant difference between NGSS and TEL is that NGSS performance expectations related to technology and design require science content knowledge, which is not true of the TEL assessment that provides relevant science content in the task situation.

Conclusion 5. Given differences between NAEP science, NAEP TEL, and the NGSS in terms of content, practices, and performance expectations, the AIR study (Neidorf et al., 2016) concluded that an assessment aligned to the NGSS could look substantially different from assessments aligned with either NAEP science or NAEP TEL. Much of this difference is associated with the demands of the NGSS performance expectations for science DCIs, as noted above. The same concern applies to performance expectations for the DCI designated as engineering, technology, and applications of science as well as performance expectations involving the engineering practices when combined with science disciplinary content. For the most part, the NGSS performance expectations likely would lead to more challenging assessment tasks than those found in either NAEP science or NAEP TEL.

Status of State Science Standards, Assessments, and Instruction

Given substantial differences between the NAEP science and NAEP TEL frameworks and the NGSS, an obvious question is the degree to which states have adopted the NGSS or similar standards and the status of implementation of policies and practices associated with those standards. Included among the latter is implementation of state large-scale assessments aligned to their current standards. A related concern is penetration of the NRC framework's vision for science learning, teaching, and assessment at the level of classroom practice. Such information has implications for the validity of results from the NAEP science assessment when it is re-administered in grade 8 in 2024 and when an updated science assessment is administered in grades 4 and 8 in 2028.

Conclusion 6. Currently, 45 states (including the Department of Defense Education Activity) have either fully adopted the NGSS as their state standards (21) or adopted NGSS-like state science standards (24; Dickinson et al., 2021). These states represent a substantial proportion of the total U.S. student population across grades K–12. When the standards of states that have adopted NGSS-like standards (24) and those of non-NGSS-adopting states (6) are compared to the NAEP framework based solely on content, several differences arise. Such differences are not surprising given that standards based on the NRC framework are likely to show results that are highly similar to those obtained directly from comparison of content from the NAEP science framework with the NGSS. As mentioned above, the NRC framework and NGSS include a specific set of disciplinary core ideas that remain constant across grade levels while growing in depth and sophistication. State standards based on the NRC framework are likely to show the same pattern of content similarities and dissimilarities with NAEP within and across grades that were revealed in the AIR study (Neidorf et al., 2016) comparing NAEP and NGSS. Results reported in the HumRRO 2021 study of state content standards vis-à-vis NAEP are very similar in that regard (Dickinson et al., 2021). The implication is that at least at the policy level, significant differences exist between NAEP's view of science proficiency and its assessment and the view that has become policy for the preponderance of states and realized via their officially adopted state science standards. Given state science standards adoptions, the current NAEP science framework and assessment may be substantially at variance with a relatively pervasive national perspective on what is desired for students to know and be able to do in science at grades 4, 8, and 12 and how they could be expected to show proficiency via large-scale assessment.

Conclusion 7. The pace at which standards reflecting the NGSS or the NRC framework affects classroom teaching, learning, and assessment has been slow, perhaps not unexpectedly. Evidence shows that adoption of the new standards has been staggered across time since 2013, as has been the design and implementation of state large-scale assessments aligned to those new standards. The latter invariably lag two or more years behind adoption of new state standards. The most recent national survey of science education conducted suggests that little has changed between 2012 and 2018 in science instructional practice (Smith, 2020). Results from the NAEP science assessment from 2009 to 2019 also show little in the way of change in student performance across time (The Nation's Report Card, 2019). One major factor in the slow penetration at the classroom level appears to be limited availability and implementation of professional learning programs for teachers. Although state implementation of large-scale assessments aligned with the NGSS or NRC framework has progressed, and classroom instructional and assessment resources aligned with the NRC framework's vision of teaching, learning, and assessment have become more readily available, the current and future state of classroom practice remains to be determined. Regarding the latter, the National Academies of Science, Engineering, and Medicine (NASEM) is convening a two-day summit in October 2021 at which time the status of implementation of science standards with a focus on areas where additional work may be needed will be discussed. In summary, how far out of alignment the NAEP science framework and assessment may be with science instruction and assessment in most states in 2024 when the current assessment is to be used remains to be seen. It seems reasonable to conclude, however, that significant differences likely will exist in 2028 if the NAEP science framework and assessment are not updated and revised.

Technology and NAEP Science

Conclusion 8. Technology already has had a substantial impact on the NAEP program—and particularly on NAEP science. Both NAEP science and NAEP TEL currently are delivered as digitally based assessments and include new types of tasks that take advantage of some of the affordances of technology for task design, presentation and interaction, data capture, scoring, and analysis. Possibilities exist for capitalizing on the multiple affordances of technology in updating and revising the NAEP science framework and assessment. These include consideration of additional science and technology proficiencies that should be included in the framework, the capacity for their realization in the assessment in the form of tasks and situations that require particular forms of scientific and engineering reasoning, and opportunities for analysis and reporting of those proficiencies in ways that go well beyond overall accuracy. In general, innovative uses of technology offer NAEP science the possibility of leadership in the large-scale science assessment field by providing a vision and examples of how science and technology competence can and should be assessed and reported. Further movement in this direction must take into consideration design and analytic challenges together with equity, cost, and feasibility concerns.

Recommendations

Given the findings described, serious concerns exist about the capacity of the NAEP science assessment to fulfill its mission to provide valid and reliable information about the status of science achievement in the United States in 2028 and beyond unless a detailed review and revision of the NAEP science framework is recommended by NAGB in 2022 and then pursued by an appropriate framework visioning panel followed by a framework development panel.

The major threat to the validity of NAEP science involves adoption by a preponderance of states of science and technology education standards that differ substantially from the NAEP science framework. Assuming continued implementation of assessments, curriculum materials, instructional practices, and professional learning opportunities aligned with those standards, whether the NAEP science assessment can track the impact of those changes on science achievement and whether NAEP science and/or NAEP TEL will have sufficient instructional sensitivity to reveal what has and has not happened over time when administered in 2028, and even quite possibly beforehand in 2024, is questionable.

Two broad recommendations consistent with these concerns and the related findings contained in this paper follow. For each recommendation, additional commentary is provided regarding matters that should be considered in acting upon each recommendation.

Recommendation 1

The NAEP Validity Studies Panel recommends that the NAEP science framework should be reviewed and revised to reflect contemporary changes in science standards, instruction, and assessment.

In reviewing and suggesting revisions to the science framework:

A. The panels should consider the distribution and focus of the content included in the framework regarding two factors. The first factor involves consideration about whether there should be continuity in the content foci within each domain of science across the grades, in ways similar but not necessarily identical to the disciplinary core ideas in life science, physical science, and Earth and space science described in the NRC framework. The second factor is related to the first and involves the specific set of topics included in each domain and across grades. A shift to this organization of content may allow the NAEP science assessment to provide important trend information across grades in the development of core knowledge in prioritized areas of each of the three major science disciplines.

B. The panels should consider NAEP's current science practices relative to a set of science and engineering practices that may be most important for students to understand and use. Such practices should be articulated in the framework as well as their implications for assessment at each grade level and across grades. Such a consideration includes the extent to which they emphasize active engagement with science and engineering practices, as articulated in the NRC framework, that is, the doing of science and engineering, when applied to science content rather than just knowing about those practices but not necessarily being able to use them.

C. The panels should consider the meaning of science proficiency and how that is expressed via performance expectations that integrate content and practice knowledge consistent with the separate but related considerations of science and engineering content and practices discussed above. Particular attention needs to be given to the demands of those performance expectations and how they could be represented in assessments that make use of the affordances of technology.

D. The panels should consider the inclusion of technology and engineering content and practices, similar to their inclusion in the NRC framework and NAEP TEL. Further comments on technology and engineering in the NAEP science framework are included below under Recommendation 2.

E. The panels should gather the most recent information on the status of implementation and impact of current state science standards and projections for the remainder of this decade. The panels should seek information on these matters from the Board on Science Education from NASEM, the National Science Teacher Association, the Council of State Science Supervisors, the Science SCASS of the Council of Chief State School Officers, and the American Association for the Advancement of Science.

Recommendation 2

The NAEP Validity Studies Panel recommends that in reviewing and revising the NAEP science framework, consideration should be given to the possible merger of aspects of the TEL framework with the science framework to create an integrated science and technology framework and assessment for administration in 2028.

The NAEP TEL framework and assessment have served useful purposes since their development and initial implementation in 2014. As noted earlier, NAEP TEL is due to be administered twice more at grade 8—in 2024 and again in 2028. Given the representation and integration of technology and engineering with science content domains in contemporary science frameworks and standards, as well as the partial overlap of the latter with the NAEP science and TEL frameworks and assessments, worth considering is whether the most important aspects of the NAEP TEL framework could be included in a revised NAEP science framework.

While the NAEP TEL Framework covers grades 4, 8, and 12, the TEL assessment has been developed only for grade 8. In addition to the limitation of the assessment to a single grade, the TEL construct representation and focus on technology literacy may have lost some of its currency and value in the intervening decade. A review of the complete grades 4–12 framework and the grade 8 assessment seems warranted especially considering existing state standards that include integrated content and practice knowledge focused on technology, engineering, and applications of science across grades 4–12.

A. In reviewing and suggesting revisions to the science framework, the panels should consider NAEP TEL's current content, practices, and forms of assessment for possible inclusion in an updated NAEP science framework and assessment.

B. In considering inclusion of NAEP TEL content and practices in an integrated science and technology framework and assessment, the panels should simultaneously consider what important aspects of the NAEP TEL framework and assessment would be lost if the assessment was discontinued after 2024 and whether continuation of NAEP TEL through 2028 is advisable even if a combined science and technology framework is developed for the 2028 NAEP science assessment.

Considerations of Trend

One hallmark of the NAEP program is its focus on monitoring progress over time and the analysis and reporting of trends in performance. The NAEP science trend extends back to 2009 and NAEP TEL to 2014. Assuming implementation of both current assessments in 2024, there will be 15 years of trend data for science and 10 years for TEL. Given the likely scope of a revision to the NAEP science framework and the implications for the 2028 assessment, as well as the possibility of incorporating aspects of TEL in the new framework and assessment, it seems highly likely that preserving the science or TEL trend through 2028 will not be feasible or advisable. Whether breaking trend in either case in 2028 is both warranted and necessary demands careful attention in deliberations that ensue in NAGB's decisions about revisions to both NAEP science and TEL and their futures. In such deliberations, priority should go to insuring the validity of the revised science framework and assessment for 2028 and beyond. Doing so should not be compromised in a possibly misguided effort to preserve trend at all costs.

REFERENCES

Alonzo, A. C., & Gotwals, A. W. (Eds.). (2012). *Learning progression in science: Current challenges and future directions*. Sense.

American Association for the Advancement of Science. (1990). Science for all Americans: A Project 2061 report on literacy goals in science, mathematics, and technology. *Bulletin of Science, Technology & Society*, 10(2), 93–101.

American Association for the Advancement of Science. (1993). *Benchmarks for science literacy*. Oxford University Press.

American Association of Medical Colleges. (2012). *MR5 fifth comprehensive review of the Medical Colleges Admission Test (MCAT): Final MCAT recommendations*. https://www.aamc.org/download/273766/data/finalmr5recommendations.pdf

Banilower, E. R., Smith, P. S., Malzahn, K. A., Plumley, C. L., Gordon, E. M., & Hayes, M. L. (2018). *Report of the 2018 NSSME+*. Horizon Research, Inc

Beggrow, E. P., Ha, M., Nehm, R. H., Pearl, D., & Boone, W. J. (2014). Assessing scientific practices using machine-learning methods: How closely do they match clinical interview performance? *Journal of Science Education and Technology, 23*, 160–182.

Behrens, J. T., DiCerbo, K. E., & Foltz, P. W. (2019). Assessment of complex performances in digital environments. *The Annals of the American Academy of Political and Social Science*, *683*(1), 217-232.

Bennett, R. E. (2008). *Technology for large-scale assessment*. ETS Report No. RM-08-10. Educational Testing Service.

Bergner, Y., & von Davier, A. (2019). Process data in NAEP: Past, present and future. *Journal of Educational and Behavioral Statistics, 44(6),* 706–732.

Berman, A. I., Haertel, E. H., & Pellegrino, J. W. (Eds.) (2020). *Comparability of large-scale educational assessments: Issues and recommendations*. National Academy of Education.

Bransford, J., Brown, A., Cocking, R., Donovan, S., & Pellegrino, J. W. (2000). *How people learn: Brain, mind, experience and school (expanded edition)*. National Academy Press.

Bybee, R., McCrae, B., & Laurie, R. (2009). PISA 2006: An assessment of scientific literacy. *Journal of Research in Science Teaching, 46*(8), 865–883.

Bybee, R. W. (2010). Advancing STEM education: A 2020 vision. *Technology and Engineering Teacher, 70*(1), 30.

College Board. (2009). *Science: College Board standards for success*. https://secure-media.collegeboard.org/apc/cbscs-science-standards-2009.pdf

College Board. (2011a). *AP biology curriculum framework 2012–2013*. https://secure-media.collegeboard.org/digitalServices/pdf/ap/10b_2727_AP_Biology_CF_WEB_110128.pdf

College Board. (2011b). *AP chemistry curriculum framework 2013–2014*. https://secure-media.collegeboard.org/digitalServices/pdf/ap/IN120085263_ChemistryCED_Effective_Fall_2013_lkd.pdf

Corcoran, T. B., Mosher, F. A., & Rogat, A. (2009). *Learning progressions in science: An evidence-based approach to reform*. Columbia University, Teachers College, Consortium for Policy Research in Education, Center on Continuous Instructional Improvement.

Davey, T., Ferrara, S., Shavelson, R., Holland, P., Webb, N., & Wise, L. (2015). *Psychometric considerations for the next generation of performance assessment*. http://www.ets.org/Media/Research/pdf/psychometric_considerations_white_paper.pdf

Dickinson, E. R., Gribben, M., Schultz, S. R., Spratto, E., & Woods, A. (2021). *Comparative Analysis of the NAEP Science Framework and State Science Standards: Technical report*. Unpublished manuscript. Retrieved from the National Assessment Governing Board

Drasgow, F. (Ed.) (2016). *Technology and testing: Improving educational and psychological measurement*. Routledge.

Duran, R., Zang, T., Sanosa, D., & Stancavage, F. (2020). *Effects of visual representations and associated interactive features on student performance on National Assessment of Educational Progress (NAEP) pilot science scenario-based tasks*. NAEP Validity Studies Panel.

Ercikan, K., & Pellegrino, J. W. (Eds.). (2017) *Validation of score meaning for the next generation of assessments*. Taylor & Francis.

Gane, B. D., Zaidi, S. Z., & Pellegrino, J. W. (2018). Measuring what matters: Using technology to assess multidimensional learning. *European Journal of Education, 53*(2), 176–187.

Gerard, L. F., Ryoo, K., McElhaney, K. W., Liu, O. L., Rafferty, A. N., & Linn, M. C. (2016). Automated guidance for student inquiry. *Journal of Educational Psychology, 108*(1), 60–81.

Gorin, J. S., & Mislevy, R. J. (2013). *Inherent measurement challenges in the Next Generation Science Standards for both formative and summative assessment* [Paper presentation]. ITS Invitational Research Symposium on Science Assessment, Washington, DC, United States.

Harris, C. J., Krajcik, J. S., Pellegrino, J. W., & DeBarger, A. H. (2019). Designing knowledge-in-use assessments to promote deeper learning. *Educational Measurement: Issues and Practice, 38*(2), 53–67.

Lee, Y.-H., Hao, J., Man, K., & Ou, L. (2019). How do test takers interact with simulation-based tasks? A response-time perspective. *Frontiers in Psychology, 24*.

Mislevy, R. J. (2018). *Sociocognitive foundations of educational measurement.* Routledge.

Mislevy, R. J., & Haertel, G. D. (2006). Implications of evidence-centered design for educational testing. *Educational Measurement: Issues and Practices, 25*(4), 6–20.

Mislevy, R. J., & Riconscente, M. M. (2006). Evidence-centered assessment design: Layers, concepts, and terminology. In S. Downing & T. Haladyna (Eds.), *Handbook of test development* (pp. 61–90). Erlbaum.

Mullis, I. V. (2019). *White paper on 50 Years of NAEP use: Where NAEP has been and where it should go next.* NAEP Validity Studies Panel. https://www.air.org/sites/default/files/2021-06/50-Years-of-NAEP-Use-June-2019.pdf

National Assessment Governing Board. (2008). *Science framework for the 2009 National Assessment of Educational Progress.* https://www.nagb.gov/content/dam/nagb/en/documents/publications/frameworks/science/2009-science-framework.pdf

National Assessment Governing Board. (2013). *Technology and engineering literacy framework for the 2014 National Assessment of Educational Progress.* https://www.nagb.gov/content/dam/nagb/en/documents/publications/frameworks/technology/2014-technology-framework.pdf

National Assessment Governing Board. (2014). *Science framework for the 2015 National Assessment of Educational Progress.* https://www.nagb.gov/content/dam/nagb/en/documents/publications/frameworks/science/2015-science-framework.pdf

National Assessment Governing Board. (2018a). *Revision of framework development policy for NAEP assessments.* https://www.nagb.gov/content/dam/nagb/en/documents/what-we-do/quarterly-board-meeting-materials/2017-11/12-framework-policy-revision.pdf

National Assessment Governing Board. (2018b). *Technology and engineering literacy framework for the 2018 National Assessment of Educational Progress.* https://files.eric.ed.gov/fulltext/ED594359.pdf

National Research Council. (1996). *National science education standards*. The National Academies Press.

National Research Council. (2000). *Inquiry and the National Science Education Standards: A guide for teaching and learning.* The National Academies Press.

National Research Council. (2002). *Learning and understanding: Improving advanced study of Mathematics and Science in U.S. high schools*. The National Academies Press.

National Research Council. (2007). *Taking science to school: Learning and teaching science in grade K-8*. The National Academies Press.

National Research Council. (2009). *Learning science in informal environments: People, places, and pursuits*. The National Academies Press.

National Research Council. (2012). *A Framework for K–12 Science Education: Practices, crosscutting concepts, and core ideas*. The National Academies Press.

National Research Council. (2013). *Next Generation Science Standards: For states, by states*. The National Academies Press.

The Nation's Report Card. (2019). *See how U.S. fourth-, eighth-, and twelfth-grade students performed in science*. https://www.nationsreportcard.gov/highlights/science/2019/

Nehm, R. H., Ha, M., & Mayfield, E. (2012). Transforming biology assessment with machine learning: Automated scoring of written evolutionary explanations. *Journal of Science Education and Technology, 21*, 183–196.

Neidorf, T., Stephens, M., Lasseter, A., Gattis, K., Arora, A., Wang, Y., Guile, S., & Holmes, J. (2016). *A comparison between the Next Generation Science Standards (NGSS) and the National Assessment of Educational Progress (NAEP) Frameworks in science, technology and engineering literacy, and mathematics. Technical report.* National Center for Education Statistics. https://nces.ed.gov/nationsreportcard/subject/science/pdf/ngss_naep_technical_report.pdf

Pellegrino, J. W. (2013). Proficiency in science: Assessment challenges and opportunities. *Science, 340*, 320–323.

Pellegrino, J. W. (2016). *21st Century science assessment: The future is now.* SRI International.

Pellegrino, J. W., Chudowsky, N, & Glaser, R. (Eds). (2001). *Knowing what students know: The science and design of educational assessment.* The National Academies Press.

Pellegrino, J. W., & Hilton, M. L. (Eds.) (2012). *Education for life and work: Developing transferable knowledge and skills in the 21st Century.* The National Academies Press.

Pellegrino J. W., & Quellmalz, E. S. (2010). Perspectives on the integration of technology and assessment. *Journal of Research on Technology in Education, 43*(2), 119–134.

Pellegrino, J. W., Wilson, M. R., Koenig, J. A., & Beatty, A. S. (Eds.) 2014. *Developing assessments for the Next Generation Science Standards.* The National Academies Press.

Pohl, S., Ulitzsch, E., & von Davier, M. (2021). Reframing rankings in educational assessments. *Science, 372*(6540), 338–340.

Schmidt, W. H., McKnight, C. C., & Raizen, S. A. (1997). *A splintered vision: An investigation of U.S. science and mathematics education.* Kluwer Academic.

Smith, P. S. (2020). What does a national survey tell us about progress toward the vision of the NGSS? *Journal of Science Teacher Education, 31*(6), 601–609.

Way, D., & Strain-Seymour, E. (2021). *A framework for considering device and interface features that may affect student performance on the National Assessment of Educational Progress*. NAEP Validity Studies Panel.

Williamson, D. M., Xi, X., & Breyer, F. J. (2012). A framework for evaluation and use of automated scoring. *Educational Measurement: Issues and Practice, 31*(1), 2–13.

Wilson, M. R., & Bertenthal, M. W. (Eds.). (2005). *Systems for state science assessments*. The National Academies Press.

Zhai, X. (2021). Practices and theories: How can machine learning assist in innovative assessment practices in science education. *Journal of Science Education and Technology, 30*(2), 1–11.

Zhai, X., Haudek, K. C., Shi, L., Nehm, R., & Urban-Lurain, M. (2020). From substitution to redefinition: A framework of machine learning-based science assessment. *Journal of Research in Science Teaching, 57*(9), 1430–1459.

Zhai, X., & Pellegrino, J. W. (in press). Large-scale assessment in science education. *Handbook of research on science education* (Vol. III). Routledge.

Zhai, X., Yin, Y., Pellegrino, J. W., Haudek, K. C., & Shi, L. (2020). Applying machine learning in science assessment: A systematic review. *Studies in Science Education, 56*(1), 111–151.

From:	Petersen, Anne
To:	NAGB Queries
Cc:	Waybright, Tyler; Shelley Loving-ryder; Melody Bushley; Gregory Macdougall; Thayer Myra zup95782
Subject:	NAEP Science Framework
Date:	Monday, September 27, 2021, 10:01:27 AM
Attachments:	2019-science-framework_tdw.pdf

Thank you for the opportunity to review the NAEP Science Framework. The assessment and instruction teams at the Virginia Department of Education have independently reviewed the document and a summary of the comments are provided both in the text below and embedded in the attached document.

Recommendations

The NAEP framework was changed so that it aligns to national standards and that alignment remains. The edit recommendations and concerns indicated below and in the attached document do not necessitate a rewrite of the framework by themselves. The framework appears sufficient to achieve the goals of the NAEP program.

Concerns

Virginia twelfth grade students have not participated in the grade 12 NAEP assessment; however, the inclusion of physics content typically covered in a first year high school physics course may cause a public relations issue to those states that do participate in the assessment. Student performance on the physics content of the NAEP may not be an indicator of student mastery of physics concepts; instead, it may reflect an equity issue. At this time, 59% of schools with 80% of the student population consisting of Black, Lantinola, and Indegienous students do not have first year physics coursework as part of their course options (National Academy of Science, 2021). In addition, 90% of schools that are considered high poverty do not offer physics (National Academy of Science, 2021).

A second concern with the inclusion of the physics content on the 12th grade assessment is that there is currently a critical shortage of physics teachers in the United States (EdSource, 2019).

The Virginia Department of Education recognizes that physics coursework should be accessible to all students and that a robust understanding of physics concepts can prepare students for higher education and future careers; however, reporting student performance on high school first year course physics concepts may cause public confusion as to the complex issues involved with K-12 physics education. Lower student performance on the physics content in 12th grade may be an indicator of a lack of opportunity versus poor performance.

Possible Edits to NAEP CF (see attached document for specific suggested edits)

The NAEP framework was reviewed by VDOE assessment staff and made 3 types of edits:

1. Simple grammatical edits like “Earth” or “the Earth.” (most of the edit suggestions made were this edit)

2. Content clarifications and changes in science through time. (there were only a few)

3. Notes for VDOE staff as to the degree of alignment with VA CF.

Please feel free to reach out to VDOE if you have any questions on the feedback provided.

Anne Petersen
Tyler Waybright

--
Anne Petersen, Ph.D.
Science Coordinator
Virginia Department of Education
101 N.14th St., Richmond

Page number	Location on Page	Excerpt from NAEP Science Framework	Recommended Edit / Comment
p. 19	Second paragraph (after bullet points), last sentence	However, using three broad content areas as an organizer helps ensure that key science content is assessed in a balanced way.	not done on NAEP grade 12
p. 20	Fifth bullet point	A deliberate attempt was made to limit the breadth of science content to be assessed so that some important topics could be measured in-depth. Once core content was identified in each science area, additional content statements could be added only if others previously included were eliminated.	not completely sure what accuracy means to VA here. We may teach things at different times.
p. 21	Exhibit 4 title	Exhibit 4. NAEP science content topics and subtopics	hope to see more content subtopics than this but the intro does state that NAEP have been “paired” down. NAEP seems similar to VA in this case. the “benchmark” expectation is quite high.
p. 22	Second paragraph, last sentence	The content statements form the basis for explaining or predicting naturally occurring phenomena. For example, the above content statement about objects in motion can be used to explain and predict the motions of many different specific objects (e.g., an ice skater, an automobile, an electron, or a planet).	i disagree NAEP will not explain (maybe partially) or predict movements of electrons or planets. “Benchmark” level could possibly do this.
p. 23	Exhibit 6 title	Exhibit 6. Commentary on a Physical Science content statement	I feel that VA is a bit more rigorous here than is shown by Exhibit ^
p. 24	Exhibit 6 title continued.	Exhibit 6 (continued). Commentary on a Physical Science content statement	seems to be on par with VA CF except for last bullet
p. 24	First bullet point, Exhibit 6	Some waves are transverse (water seismic) and other waves are longitudinal (sound, seismic).	water is both VA struggles with the same problem
p. 24	Second bullet point, Exhibit 6	In transverse waves, the direction of the motion is perpendicular to the disturbance.	“direction of wave propitiation” In transverse waves, the direction of the motion is perpendicular to the ~~disturbance~~.
p. 24	Third bullet point, Exhibit 6	In longitudinal waves, the direction of motion is parallel to the disturbance.	In longitudinal waves, the direction of motion is parallel to the ~~disturbance~~.
p. 24	Fourth bullet point, Exhibit 6	Waves (e.g., light waves) traveling from one material to another undergo transmission, reflection, and/or changes in speed.	Marked but no comment

Page number	Location on Page	Excerpt from NAEP Science Framework	Recommended Edit / Comment
p. 24	Third bullet point after Exhibit Box	Empty cells in the content statement tables denote that a particular subtopic is not recommended for assessment at that grade level.	Very true
p. 24	Fourth bullet point after Exhibit Box	Retention of foundational knowledge from one grade to the next is assumed; however, if the relevant content statement does not appear in a succeeding grade level, it should not be assessed.	This is no small point. VA folks do not believe in this notion. VA folks say this is not fair. Like the NAEP 12 grade test having LS and most VA kids took it in 10th. I believe the test is designed to test student "residual" knowledge of the three content domains and it can do but VA may not participate in grade 12
p. 25	First paragraph under **Physical Sciences** heading	Familiar changes	
p. 25	First paragraph under **Physical Sciences** heading	Erosion of mountains	Not sure these are familiar
p. 28	Second paragraph in textbox	Understanding the substance of water requires knowledge across the Physical Science categories of Matter, Energy, and Motion.	Understanding ~~the substance of water~~ requires knowledge across the Physical Science categories of Matter, Energy, and Motion. "Properties of" probably ok as is
p. 28	First paragraph after textbox, last sentence	The Periodic Table demonstrates the relationship between the atomic number of the elements and their chemical and physical properties and provides a structure for inquiry into the characteristics of the chemical elements (grade 12).	The Periodic Table ~~demonstrates~~ the relationship between the atomic number of the elements and their chemical and physical properties and provides a structure for inquiry into the characteristics of the chemical elements (grade 12). illustrates
p. 30	First paragraph, last sentence	The Sun as the main energy source for the Earth provides an opportunity at all grade levels to make important connections between the science disciplines (see the following textbox).	The Sun as the main energy source for ~~the Earth~~ Earth provides an opportunity at all grade levels to make important connections between the science disciplines (see the following textbox).
p. 30	Last paragraph, second sentence	As the diver falls, her speed (kinetic energy) increases as her potential energy decreases.	their, they

Page number	Location on Page	Excerpt from NAEP Science Framework	Recommended Edit / Comment
p. 32	Fourth sentence	The Earth and an airplane do not need to be in contact…	~~The~~ Earth and an airplane do not need to be in contact…
p. 33	Exhibit 8 title	Exhibit 8. Physical Science content statements for grades 4, 8, and 12	these learning progressions are very familiar and similar to VA
p. 33	P4.5	P4.5 Magnets can repel or attract other magnets. They can also attract certain nonmagnetic objects at a distance.	not sure we stress this as much as they seem to do
p. 33	Footnote	Although this content statement generally holds true, some compounds decompose before boiling.	not needed for this audience but ok
p. 35	P12.8	P12.8 Atoms and molecules that compose matter are in constant motion (translational, rotational, or vibrational).	Holy cow, NMR this is organic
p. 35	P8.9	P8.9 Three forms of potential energy are gravitational, elastic, and chemical. Gravitational potential energy changes in a system as the relative positions of objects are changed. Objects can have elastic potential energy due to their compression, or chemical potential energy due to the nature and arrangement of the atoms.	much stronger than VA cf
p. 35	P8.10	P8.10 Energy is transferred from place to place. Light energy from the Sun travels through space to Earth (radiation). Thermal energy travels from a flame through the metal of a cooking pan to the water in the pan (conduction). Air warmed by a fireplace moves around a room (convection). Waves (including sounds and seismic waves, waves on water, and light waves) have energy and transfer energy when they interact with matter.	P8.10 Energy is transferred from place to place. Light energy from the Sun travels through space to Earth (radiation). Thermal energy travels from a flame through the metal of a cooking pan to the water in the pan (conduction). Air warmed by a fireplace moves around a room (convection). Waves (including sounds and seismic waves, waves on water, and light waves) have energy and transfer energy ~~when~~ as they interact with matter.
p. 36	P8.13	P8.13 Nuclear reactions take place in the Sun. In plants, light from the sun is transferred to oxygen and carbon compounds, which, in combination, have chemical potential energy (photosynthesis).	P8.13 ~~Nuclear~~ Fusion reactions take place in the Sun. In plants, light from the sun is transferred to oxygen and carbon compounds, which, in combination, have chemical potential energy (photosynthesis).

Page number	Location on Page	Excerpt from NAEP Science Framework	Recommended Edit / Comment
			Comment: This should probably be singular
p. 38	Exhibit 8 Continued title	Exhibit 8 (continued). Physical Science content statements for grades 4, 8, and 12	PS is way above level of VA CF
p. 38	P12.22	P12.22 Gravitation is a universal attractive force that each mass exerts on any other mass. The strength of the gravitational force between two masses is proportional to the masses and inversely proportional to the square of the distances between them.	way above VA cf
p. 38	P12.23	P12.23 Electric force is a universal force that exists between any two charged objects. Opposite charges attract while like charges repel. The strength of the electric force is proportional to the magnitudes of the charges and inversely proportional to the square of the distance between them. Between any two charged particles, the electric force is vastly greater than the gravitational force.	way above
p. 39	Second paragraph, first sentence	Understanding principles in Life Science is inextricably linked with understanding principles in Physical Science and Earth and Space Sciences.	theres that word again
p. 41	Text box, last sentence	Therefore, although synthesis and breakdown are common to both plants and animals, photosynthesis (the conversion of light energy into stored chemical energy) is unique to plants, making them the primary source of energy for all animals.	Anne is “primary” enough to allow inclusion of thermal vent chemotrophs?
p. 42	Second paragraph, third sentence	In these grand-scale cycles, the total amount of matter and energy remains constant, even though their form and location undergo continual change.	In these ~~grand-scale~~ cycles, the total amount of matter and energy remains constant, even though their form and location undergo continual change. Comment: not sure why this is here is it a technical or statistical term?
p. 44	First paragraph under Evolution and Diversity, third sentence	The modern concept of evolution, including natural selection and common descent, provides a unifying principle for understanding the history of life on	The pencil mark is over “principle” but no written comment.

Page number	Location on Page	Excerpt from NAEP Science Framework	Recommended Edit / Comment
		Earth, relationships among all living things, and the dependence of life on the physical environment.	
p. 45	L12.1	L12.1 Living systems are made of complex molecules (including carbohydrates, fats, proteins, and nucleic acids) that consist mostly of a few elements, especially carbon, hydrogen, oxygen, nitrogen, and phosphorous.	teach to this level in bl?
p. 45	L12.3	L12.3 Cellular processes are regulated both internally and externally by environments in which cells exist, including local environments that lead to cell differentiation during the development of multicellular organisms. During the development of complex multicellular organisms, cell differentiation is regulated through the expression of different genes.	this also sounds on level with VA CF
p. 46	Exhibit 10 (continued) title	Grade 12	much of this content is taught in VA
p. 46	Footnote	The statement "they use the energy from light" does not imply that energy is converted into matter or that energy is lost. See textbox "Crosscutting Content: Uses, Transformations, and Conservation of Energy," p. 42.	I really do not think this is needed
p. 47	Exhibit 10 continued title	Exhibit 10 (continued). Life science content statements for grades 4, 8, and 12	Table is very similar to VA in most respects
p. 47	L4.4	L4.4 When the environment changes, some plants and animals survive and reproduce; others die or move to a new location.	change. eg. seasons
p. 48	L8.10	L8.10 The characteristics of organisms are influenced by heredity and environment. For some characteristics, inheritance is more important; for other characteristics, interactions with the environment are more important.	VA goes into Mendel
p. 48	L12.9	L12. 9 The genetic information encoded in DNA molecules provides instructions for assembling	nice!

Page number	Location on Page	Excerpt from NAEP Science Framework	Recommended Edit / Comment
		protein molecules. Genes are segments of DNA molecules. Inserting, deleting, or substituting DNA segments can alter genes. An altered gene may be passed on to every cell that develops from it. The resulting features may help, harm, or have little or no effect on the offspring's success in its environment.	
p. 49	L8.11 (last sentence)	L8.11 Extinction of a species is common; most of the species that have lived on the Earth no longer exist.	L8.11 Extinction of a species is common; most of the species that have lived on ~~the~~ Earth no longer exist.
p. 49	L8.12 (last sentence)	L8.12 Biologists consider details of internal and external structures to be more important than behavior or general appearance.	this may not prove to be true in the see "canis" and "the species problem"
p. 49	L12.13	L.12.13 Evolution is the consequence of the interactions of (1) the potential for a species to increase its numbers, (2) the genetic variability of offspring due to mutation and recombination of genes, (3) a finite supply of the resources required for life, and (4) the ensuing selection from environmental pressure of those organisms better able to survive and leave offspring.	Interesting!
p. 50	First paragraph, third sentence	This concept of Earth as a complex and dynamic entity of interrelated subsystems implies that there is no process or phenomenon within the Earth system that occurs in complete isolation from other elements of the system.	This concept of Earth as a complex and dynamic entity of interrelated subsystems implies that there is no process or phenomenon within ~~the~~ Earth system that occurs in complete isolation from other elements of the system.
p. 50	Last paragraph, third sentence	Other Web-based programs allow students to view and process satellite images of Earth, to direct a camera on board the Space Shuttle, and to access professional telescopes around the world to carry out science projects.	a little dated at this point
p. 50	Footnote	Earth is capitalized, rather than referred to as "the earth," in order to recognize it as one of the planets in the solar system.	see gregg

Page number	Location on Page	Excerpt from NAEP Science Framework	Recommended Edit / Comment
p. 51	Second paragraph under *Earth in Space and Time* heading	Applies to entire paragraph	"the" earth is removed here, as it should be
p. 51	First paragraph under **Objects in the Universe** heading	"the Sun and the Moon"	remove "the" if one is going to capitalize the proper name?
p. 52	First paragraph, second sentence	However, it is now known that the Sun is the central and largest body in the solar system, which includes Earth and other planets and their moons as well as other objects such as asteroids and comets.	Ok no the here. this should be fixed one way or the other
p. 52	First paragraph, second sentence under **History of Earth** heading	Initially, there was no life and no molecular oxygen in the atmosphere.	or water
p. 52	Third paragraph, second sentence under **History of Earth** heading	Some changes are due to slow processes, such as erosion and weathering and others are due to rapid processes such as volcanic eruptions, landslides, and earthquakes (Grade 4).	cosmic impacts
p. 53	First paragraph under **Properties of Earth Materials** heading	Earth materials that occur in nature include rocks, minerals, soils, water, and the gases of the atmosphere. Natural materials have different properties that sustain plan and animal life (grade 4).	nice
p. 53-54	Last sentence on page 53 going into 54	The current explanation is that the outward transfer of Earth's internal heat propels the plates comprising Earth's surface across the face of the globe, pushing the plates apart where magma rises to form mid-ocean ridges, and pulling the edges of plates back down where the Earth materials sink into the crust at deep trenches (grade 12).	The current explanation is that the outward transfer of Earth's internal heat propels the plates comprising Earth's surface across the face of the globe, pushing the plates apart where magma rises to form mid-ocean ridges, and pulling the edges of plates back down where the Earth materials ~~sink~~ subducted into the crust mantel at deep trenches (grade 12).
p. 54	First paragraph, second sentence under **Energy in**	The Sun is the major source of energy for phenomena on Earth's surface.	we use "our" instead of "the" but we do not caps sun

Page number	Location on Page	Excerpt from NAEP Science Framework	Recommended Edit / Comment
	Earth Systems heading		
p. 55	First paragraph, last sentence under **Biogeochemical Cycles**	For example, carbon occurs in carbonate rocks such as limestone, in coal and other fossil fuels, in the atmosphere as carbon dioxide gas, in water as dissolved carbon dioxide, and in all organisms as complex molecules that control the chemistry of life (grade 12).	nice!
p. 56	Textbox heading	Crosscutting Content: Biogeochemical Cycle	This is great stuff
p. 56	Second paragraph, first sentence	Essentially fixed amounts of chemical atoms or elements cycle with the Earth system, and energy drives their translocation of matter(e.g., changes of state, gravity)	Essentially fixed amounts of chemical atoms or elements cycle with ~~the~~ Earth system
p. 56	Third paragraph	Biogeochemical cycles are described more fully in the Earth Systems section of exhibit 12, Earth and Space Science Content Statements for Grades 4, 8, and 12.	Biogeochemical cycles are described more fully in ~~the~~ Earth Systems section of exhibit 12, Earth and Space Science Content Statements for Grades 4, 8, and 12.
p. 58	E8.3	E8.3 Fossils provide important evidence of how life and environmental conditions have changed in a given location.	not sure we go this far
p. 58	E8.4	E8.4 Earth processes seen today, such as erosion and mountain building, make it possible to measure geologic time through methods such as observing rock sequences and using fossils to correlate the sequences at various locations.	pretty heavy into fossils here more so than VA CF
p. 59	Grade 12 header at top of table (note that comment refers to Grade 8)	Grade 12	the grade 8 material here is above VA CF
p. 60	Grade 8 header at top of table	Grade 8	pretty high level compared to VA CF
p. 61	E12.10	E12.10 Climate is determined by energy transfer from the Sun at and near Earth's surface. This energy transfer is influenced by dynamic processes such as	we should have this is VA CF

Page number	Location on Page	Excerpt from NAEP Science Framework	Recommended Edit / Comment
		cloud cover, atmospheric gases, and Earth's rotation, as well as static conditions such as the positions of mountain ranges, oceans, seas, and lakes.	
p. 62	Title of Exhibit	Exhibit 12 (continued). Earth and Space Sciences content statements for grades 4, 8, and 12	NAEP might be interpreted as being more rigorous in 12
p. 62	E4.10	E4.10 The supply of many Earth resources such as fuels, metals, fresh water, and farmland is limited. Humans have devised methods for extending the use of Earth resources through recycling, reuse, and renewal.	Nice!
p. 62	E12.11	E12.11 Earth is a system containing essentially a fixed amount of each stable chemical atom or element. Most elements can exist in several different chemical forms. Earth elements move within and between the lithosphere, atmosphere, hydrosphere, and biosphere as part of biogeochemical cycles.	nice
p. 68	First illustrative item	The Earth's Moon is A. always much closer to the Sun than it is to the Earth. B. always much closer to the Earth than it is to the Sun. C. about the same distance from the Sun as it is from the Earth. D. sometimes closer to the Sun than it is the Earth and sometimes closer to the Earth than it is to the Sun.	~~The~~ Earth's Moon is A. always much closer to the Sun than it is to ~~the~~ Earth. B. always much closer to ~~the~~ Earth than it is to the Sun. C. about the same distance from the Sun as it is from ~~the~~ Earth. D. sometimes closer to the Sun than it is ~~the~~ Earth and sometimes closer to ~~the~~ Earth than it is to the Sun.
p. 73	Footnote	In addition, 12th graders at the Advanced level are expected to be able to identify a scientific question for investigation. See appendix B for achievement level descriptions.	this seems odd shouldn't this be done at all levels
p. 75	Second paragraph, last sentence	After students have run the modeling software, they are asked a series of questions (e.g., the size of the hare population over time).	They have had this since 2009. VA should be ashamed…

Page number	Location on Page	Excerpt from NAEP Science Framework	Recommended Edit / Comment
p. 79	Comment is on the graphic	Forest succession graphic	we have this art
p. 83	Exhibit 14 title	Exhibit 14. Generating examples of grade 8 performance expectations	mailing the table to PEM and ETS
p. 85	E8.2	E8.2 Gravity is the force	Gravity is ~~the~~ a, or one of the forces
p. 86	First bullet point in Using Scientific Inquiry sectoin	Using scientific Inquiry: • Arrange a set of photographs of the Moon taken over a month's time in chronological order and explain the order in terms of a model of the Earth-Sun-Moon system.	Arrange a set of photographs of the Moon taken over a month's time in chronological order and explain the order in terms of a model of ~~the~~ Earth-Sun-Moon system.
p. 87	Second Items to Assess Using Science Principles	Items to Assess Using Science Principles Illustrative Item A space station is to be located between the Earth and the moon at the place where the Earth's gravitational pull is equal to the Moon's gravitational pull.	A space station is to be located between ~~the~~ Earth and the moon at the place where the Earth's gravitational pull is equal to the Moon's gravitational pull.
p. 89	Item Suggestion 1	NASA wants to launch a spacecraft with rockets from Earth so that it will reach and orbit Mars. Which of the following statements about this flight is WRONG: A. In the first phase of the flight, the forces acting on the spacecraft are the thrust of the rocket engine, gravity, and friction from the Earth's atmosphere. B. When the rocket engine shuts off, the only force acting on the spacecraft is the force of gravity. C. Once the spacecraft is above the Earth's atmosphere and the rocket engine is off, it will travel at a constant speed since there is no gravity in space.	Comment: falcon heavy (VDOE) is a better cluster than this Edits: A. In the first phase of the flight, the forces acting on the spacecraft are the thrust of the rocket engine, gravity, and friction from ~~the~~ Earth's atmosphere. B. When the rocket engine shuts off, the only force acting on the spacecraft is the force of gravity. C. Once the spacecraft is above ~~the~~ Earth's atmosphere and the rocket engine is off, it will travel at a constant speed since there is no gravity in space.
p. 104	Illustrative Items	Illustrative Items	What causes days and night? A. ~~The~~ Earth spins on its axis. (66%)

Page number	Location on Page	Excerpt from NAEP Science Framework	Recommended Edit / Comment
		What causes days and night? A. The earth spins on its axis. (66%) B. The earth moves around the Sun. (26%) C. Clouds block out the Sun's light. (0%) D. The earth moves into and out of the Sun's shadow. (3%) E. The Sun goes around the Earth. (4%) The main reason for its being hotter in summer than in winter is: A. The earth's distance from the Sun changes. (45%)	B. ~~The~~ Earth moves around the Sun. (26%) C. Clouds block out the Sun's light. (0%) D. ~~The~~ Earth moves into and out of the Sun's shadow. (3%) E. The Sun goes around the Earth. (4%) The main reason for its being hotter in summer than in winter is: ~~The~~ Earth's distance from the Sun changes. (45%)
p. 133	Last paragraph, first sentence	*In the Earth and space science*, students at the *NAEP Proficient* level should be able to explain how gravity accounts for the visible patterns of motion of the Earth.	*In the Earth and space science*, students at the *NAEP Proficient* level should be able to explain how gravity accounts for the visible patterns of motion of ~~the~~ Earth.
p. 135	Third paragraph	*In the physical sciences,* students at the *NAEP Basic* level should be able to … critique data that claim to show how gravitational potential energy changes with distance from the Earth's surface	*In the physical sciences,* students at the *NAEP Basic* level should be able to … critique data that claim to show how gravitational potential energy changes with distance from ~~the~~ Earth's surface
p. 137	First paragraph	…and evidence for human effects on the Earth's biogeochemical cycles	and evidence for human effects on ~~the~~ Earth's biogeochemical cycles

From:	Moulding, Brett
To:	NAGB Queries
Subject:	Comments on the NAEP Science Assessment Framework
Date:	Friday, August 27, 2021 9:12:56 AM

NAGB Leadership,
Comments on the future revision of NAEP Assessment Framework for Science

Whether the NAEP Science Assessment Framework needs to be updated.
The NAEP Science Assessment Framework needs to be revised.

If the framework needs to be updated, why a revision is needed.
The current Framework does not identify the science being taught in the majority of our schools. The science NAEP cannot be a report card on science education in the nation if it does not measure the current science being taught in our schools. The current NAEP framework is not consistent with the current research in how students learn.

What should a revision to the NAEP framework include?
The revision should include a clear alignment to the National Academies Framework for K-12 Science Education. The revision should include descriptions of the three-dimensional science performances that need to be assessed. The New NAEP Framework needs to include measurement of students using Practices, Crosscutting Concepts, and Core Ideas consistent with the NGSS approach to science performance expectations.

Thank you,
Brett

Brett Moulding
Retired
Utah State Office of Education Curriculum Director and Instruction
Former NAEP Science Advisory Committee Member

From:	Cary Sneider
To:	NAGB Queries
Subject:	NAEP Science Framework
Date:	Friday, August 20, 2021 2:36:35 PM
Attachments:	A-Cary"s final Comments to NAGB 2019 re TEL&Science.docx

Hello Friends,

When I ended my tenure on NAGB I made the following plea for updating the NAEP Science Framework to be consistent with the Framework for K-12 Science Education: Practices, Crosscutting Concepts and Core Ides (NRC 2012) and the subsequent Next Generation Science Standards (NGSS Lead States 2013). These have now been adopted or adapted by 44 states. Such an updated would essentially be a merger of most of the TEL and an improved NAEP Science Framework. I have attached those comments to this email.

As I've also noted in some of my prior comments during my time on the Board, NAEP has been referred to as a "Gold Standard" and a "North Star." These qualities are not the same. The "Gold Standard" refers to NAEP as a "truth-teller," because of meticulous attention to scientific rigor and detail. The "North Star" means that NAEP also points to a future destination. In this case it means that the updated NAEP Science Framework should not just reflect the two existing documents now being used by most states to guide their own science standards, but blaze the trail for future improvements in what students should know and be able to do in the STEM fields.

Warm regards, Cary

Cary Sneider, PhD

Cary Sneider's parting comments to the full NAGB Board

Friday, August 2, 2019

I'm completing 8 years on the Board, but in a sense, it's been 16 years, since my friend and colleague, Alan Friedman rolled off the Board just before I joined. Alan was a friend and mentor for most of my career. Many of us were very sad when he passed away after a brief illness at age 72.

Part of Alan's legacy to the Board and to me has been the NAEP TEL. I want to spend a few minutes reflecting on that. As a fresh context I'd like to ask how many of you read the story of the New Navy that was referenced in a recent *Staying On Board* newsletter.

There were three parts of that story relevant to the TEL. They correspond to the three phases of the engineering design process, which is the cornerstone of engineering, which is deeply embedded in the Framework for K-12 Science Education (NRC 2012) and the subsequent NGSS (2013). In contrast to prior science standards, the Framework and NGSS emphasize not just what students should know about science, but what skills they need to develop to use what they know to solve meaningful problems.

1) Defining the Problem. In contrast to the old Navy, when the purpose of training was for sailors to learn to do their job right, today's sailors are trained in many different jobs. They have to ask themselves "Am I doing the right job?" Similarly, an essential aspect of engineering, which is now a part of the science standards in 44 states, is "Am I solving the right problem?

2) Generating Creative Solutions. There's an example of creative thinking in which sailors figure out how to secure the ship to the dock using only the materials that were in front of them. That's solving a problem under constraint—one critical aspect about problem solving that students have to learn during 12 years of schooling.

3) Optimize. Once you have met the criteria and constraints of a problem you are not done. You need to refine the solution. We learned from the article that things were going so well with the new Navy that the brass decided to end the experiment early and build more light ships and hire more of the right kinds of people. Then problems cropped up. Problems always crop up with new technologies. Continuing the experiment to refine the solution is an important part of the process. In engineering it's called "optimization."

PEOPLE. The upshot of the New Navy article is that the recruiters need to find "the right people." But as educators, we don't have the luxury of turning away 9 out of 10 kids that show up for our classes. We need to prepare all of them for a rapidly changing world.

They Learn Engineering in School. The data from the context variables on the TEL inform us that more than half of our students take courses in engineering—in addition to the science courses that will—as more schools adopt the new standards—help them learn to define problems, creatively solve them under constraint, and be persistent as they continue to refine and optimize solutions to persistent problems.

In future meetings you'll be considering revision of the Science Framework. When that work is done, if it measures what students are expected to learn, it will incorporate 50% to 80% of the TEL, depending on grade level. **Essentially, that means merging the Science and TEL frameworks.** When that happens, it is my hope that funds previously spent on separate administration of the TEL can be repurposed to support state and TUDA level assessment for science (now more appropriately referred to as STEM) so that educators across the country have a golden meter stick to see how well they're doing. That's the baton I'm passing along from Alan and from me.

Input regarding the NAEP Science Assessment

Cary Sneider, Former NAGB Member

September 4, 2021

In the following paragraphs I will argue that the NAEP Science Framework needs to be updated to include much of what is in the NAEP TEL Framework. Once that is done the TEL can be eliminated and funds saved can be used to conduct science assessments at grades 4, 8, and 12 at the state and TUDA levels.

Does the NAEP Science Framework need to be updated?

Yes.

If the framework needs to be updated, why is a revision needed?

1. The NAEP Science Framework is significantly out-of-date. The NRC's consensus *study A K-12 Science Education Framework: Practices, Crosscutting Concepts, and Core Ideas* (2012) and the subsequent *Next Generation Science Standards* (NGSS Lead States, 2013) has gained traction in 44 states that have adapted or adopted new standards based on these documents. Even states that claim not to base their standards on either of these documents are influenced by them.

An essential innovation of these new standards documents is the inclusion of engineering as a part of science. It is deeply woven into the fabric of the standards, as both a set of practices complementary to science, as well as crosscutting concepts, and even core ideas, which are listed at the same level as the traditional sciences. The reason for including engineering as an essential element of science is stated in the Framework as follows:

> We anticipate that the insights gained and interests provoked from studying and engaging in the practices of science and engineering during their K-12 schooling should help students see how science and engineering are instrumental in addressing major challenges that confront society today, such as generating sufficient energy, preventing and treating diseases, maintaining supplies of clean water and food, and solving the problems of global environmental change. (NRC 2012, p. 9).

Providing a foundation in engineering design allows students to better engage in and aspire to solve the major societal and environmental challenges they will face in the decades ahead. The same document also makes clear distinctions among the important terms science, technology, and engineering.

> In the K–12 context, "science" is generally taken to mean the traditional natural sciences: physics, chemistry, biology, and (more recently) earth, space, and environmental sciences . . . . We use the term "engineering" in a very broad sense to mean any engagement in a systematic practice of design to achieve solutions to particular human problems. Likewise, we broadly use the term "technology" to include all types of human-made systems and processes—not in the limited sense often used in schools that equates technology with modern computational and

> communications devices. Technologies result when engineers apply their understanding of the natural world and of human behavior to design ways to satisfy human needs and wants. (NRC 2012, p. 11-12)

2. NGSS performance expectations have substantial overlap with NAEP Science and NAEP TEL at the **8th and 12th grade levels**.

According to a study by AIR commissioned by NAGB:

> "Ninety percent or more of NGSS performance expectations at the middle school and high school levels covered content that overlaps with NAEP science or TEL at grades 8 and 12, respectively (Neidorf et al. 2016)."

This means that the great majority of students in middle and high school will increasingly have an opportunity to learn what is in the TEL Framework through science instruction. It will be important to monitor implementation of those standards over the next decade—and only a combined Science-TEL framework, administered across states, can do that. While administering NAEP Science and NAEP TEL in a coordinated fashion would provide useful information, a revised NAEP Science Assessment could improve the monitoring function. Also, the science assessment would be fairer to students and teachers, and of greater interest to educational leaders in cities and states if it were consistent with the new standards.

What should a revision to the framework include?

1. What states are currently advocating. The purpose of the NRC's Framework and NGSS, led by the National Governor's Association and Council of Chief State School Officers, was to help all states pull in the same direction. If NAGB is to be the North Star, its essential that a new Framework not attempt to lead in an entirely different direction. In addition to being guided by these two documents, however, it will be important to commission a study of state science standards to ensure that the six states that claim more independence in their science standards are included.

2. **Additional topics from the TEL.** The TEL consists of three parts: Design and Systems, Technology and Society, and Information and Communications Technology. The first two are very strongly represented in the NGSS and Framework, and therefore in the great majority of state standards. The third area is not taught explicitly in most schools. A consolidated framework would therefore consist, in broad strokes, of the first two areas of the TEL and an updated version of the Science Framework. What will be lost is some of the third part of the TEL, which may be more closely related to ELA than to science.

If these recommendations are followed, NAGB would be able to report on accomplishments of our nation's youth in their ability to solve problems, to analyze systems, and understand key issues at the intersection of technology and society as a part of the Science Report Card. NAGB has broken new ground by developing the TEL, the first fully DBA assessment in its portfolio. That was an important accomplishment, but now it's time to consolidate it with Science, so that we can have an efficient assessment that is maximally useful to the states, while at the same time increasing NAGB's efficiency.

3. New topics highlighted by recent world events. If NAGB is to serve as the North Star, the NAEP Framework should also lead, not just follow the states. So, it will be important to consult with a wide variety of experts. Among the considerations should be the experience of a highly stressful pandemic, and the possible inclusion of topics directly related to epidemiology, vaccinations, institutions such as the CDC and WHO, and the nature of science.

CAST

178 Albion St., Suite 210
Wakefield, MA 01880
781.245.2212
781.245.5212
cast.org
@CAST_UDL

National Assessment Governing Board
800 North Capitol Street, NW, Suite 825
Washington, DC 20002

RE: NAEP Science Framework

Submitted *via email* to nagb@ed.gov

Dear Governing Board,

Since 1984, CAST (originally the Center for Applied Special Technology) has worked relentlessly to ensure that our nation is one where learning has no limits for all individuals. CAST pioneered Universal Design for Learning (UDL), a set of principles and guidelines for inclusive design for learning—including curricula, learning goals, materials, instructional methods, and assessments. UDL is now incorporated in key federal education, career training, and workforce laws.[i] UDL provides the basis for innovation and success in expanding and strengthening education across all subject areas (e.g., reading, mathematics, science). When applied to assessments, UDL can ensure that accessible normative and summative assessments are available to all students regardless of any potential learning barrier they may experience whether it be due to socio-economic status, language, or disability status.

CAST is pleased to submit comments and recommendations to the National Assessment Governing Board (NAGB) query regarding the National Assessment of Education Progress (NAEP) Science Framework ("the Framework"). Because universal design is included as a minor reference in the current framework, CAST strongly urges the NAGB to update the Framework to make it consistent with current federal law and documented best practices in the application of inclusive design in student engagement, student learning, assessment design, and assessment application.

CAST leads work funded through grants provided by the National Science Foundation (NSF), U.S. Departments of Education (ED) and Labor (DOL), state education agencies, local education agencies, as well as the private sector. CAST seeks to ensure that the full power of UDL is applied to technology, instructional, and assessment design and practice in order to remove barriers to learning and assessment in digital as well as physical settings. Our UDL initiatives encourage and support the design of flexible learning environments that anticipate learner variability and provide alternative routes or paths to success, as well as provide flexible opportunities for learners to demonstrate their construct-relevant knowledge, skills, and abilities during summative, formative, and diagnostic assessment. UDL acknowledges that the variability of how people learn is the *norm* rather than the exception. UDL provides viable alternatives for *all* learners to access in-person, blended, and online education and assessment, providing a responsive framework to support students and educators in any academic subject, including in science.

In support of our recommendation that NAGB update the Framework, CAST has examined and compared NAEP participation data for students with disabilities and English Learners (ELs) in the science assessment for the years 2009, 2015, and 2019 respectively. While NAEP data show that participation rates do increase between 2009 and 2019 for both groups of students (NAEP Science Assessment data)[ii], the participation rates remain well below NAEP's own 95 percent requirement (NAEP Policy, 2014).[iii]

Additionally, the participation of students with disabilities falls between grades 8 and 12 (NAEP Participation Rate).[iv] Therefore, CAST strongly encourages NAGB to consider our recommendations, which intend to ensure that the [new] NAEP science assessment incorporates from the outset the most modern and inclusive design so that a variable and diverse student population can successfully access and complete the assessment in grades 4, 8, and 12 at a participation rate of at least 95 percent. To help NAGB accomplish these goals, we offer the following:

General Recommendations

- Incorporate the principles of UDL throughout the Framework to support and assure student access to the NAEP science assessment, regardless of literacy level, language, and/or disability status.
- Adopt a validity framework that promotes consideration of the broad range of construct-irrelevant factors learners bring to testing. This framework should be applied from the beginning of test and item design in an effort to reduce reliance on retrofitted accommodations that provide inadequate support and/or compromise construct integrity. Examples of such frameworks, based on principles of UDL, include Dolan et al. (2013)[v] and Almond et al. (2010)[vi], the former of which has been applied in development of next-generation science assessments (e.g., Quellmalz et al., 2016).[vii]
- Eliminate all references to No Child Left Behind and include in a new Framework references and citations consistent with current law, the Elementary and Secondary Education Act currently known as the Every Student Succeeds Act (ESSA).[viii]
- Eliminate use of the term ‘special needs’, replacing such term with ‘students with disabilities’ to ensure consistency with the ESSA and the Individuals with Disabilities Education Act (IDEA).
- Discuss how to include students with the most significant cognitive disabilities in NAEP assessments who take state-designed alternate assessments on alternate achievement standards. Currently these students are not included *in any* NAEP assessment. Recent research has demonstrated the promise of combining learning map model- and UDL-based approaches in evaluating the science knowledge, skills, and abilities of students with significant cognitive disabilities.

Recommendations for the Framework (based on current pages 2-5):

- Add new rationale to ensure the Framework and new NAEP Science assessment:
    - **Inclusive Design**: Incorporate the principles of UDL as an essential component to developing a robust assessment tool from inception and design to roll-out of the assessment.
    - **Student Diversity**: Respond to the growing and increasingly diverse student population in the nation, the inclusion of all types and ages of students in the general curriculum, and the growing emphasis and commitment to serve and be accountable for all students. Such diversity does include students with disabilities and English Learners (ELs); however, the Framework *must assure* the meaning of diversity is expanded [beyond students with disabilities and ELs] consistent with NAEP resources developed in recent years (NAEP Engineering Framework).
    - **Cultural Relevance**: Acknowledge that advances have been made in understanding cultural relevance and its impact on student engagement, learning and assessment.
    - **Access Features**: Include specificity in the need for the assessment to be designed with access features consistent with WCAG 2.1 and UDL recommendations and provide built-in navigation and access supports (e.g., motoric supports, language/glossary, audio, fonts, text size, etc.) without altering the science construct. Such features are increasingly no longer considered ‘accommodations’ and instead are regularly available to all users. The Framework must require and acknowledge their incorporation and encourage/allow for their use for all students.

- **Accessibility and Accommodations**: Ensure full accessibility in the design of test items, including in the availability of standard accommodations for students with disabilities and ELs as required by federal laws (IDEA and Section 508).[ix] The Framework must assure accessibility specifically includes the use and interoperability with any external assistive technology [device/system] required by the student. Consistent with ESSA[x] such accessibility is specifically intended to increase inclusion of formerly excluded groups in assessments, including the NAEP (e.g., students with disabilities and English learners).
- **Computer Skills**: Clarify that recent events show that young students (e.g., grade 4 NAEP test takers) may have insufficient access to and training in computer use for fair inclusion in digital assessments.
- **Access to Broadband**: Make clear that many communities and schools that exist in digital deserts may have insufficient access to broadband services to support access to the assessment across grades 4, 8, and 12.

Recommendation for the Steering Committee (current page 5):

- Provide guidelines to the Steering Committee which clarifies the framework applies UDL in determining assessment content, access features and—when necessary—accommodations consistent with the objectives being assessed. (Rose et al., 2018)[xi]

Recommendations for the Model of Assessment Development and Methods:

- Ensure the methodology outlines how the assessment incorporates inclusive design and is built upon the principles of UDL, and also includes access features including in the use and interoperability with assistive technology
- Describe considerations for English learners and students with disabilities. In particular, that assessment design applies a UDL-based validity framework to help ensure full accessibility, including in the use and interoperability with assistive technology, consistent with ESSA.[xii]

Recommendation: Chapter 4: Students With Disabilities and English Language Learners (Current Pages 114-115)

- Make updates consistent with current research and practice, incorporating the principles of UDL throughout the Framework to support and assure student access to the NAEP science assessment, regardless of literacy level, language and/or disability status. (Rose et al., 2018)[xiii]

Recommendations: Chapter 4: Key Attributes of Effective Assessment (current page 124)

- Takes into account student diversity as reflected in gender, geographic location, language proficiency, race/ethnicity, socioeconomic status, and disability status consistent with NAEP policies (e.g., NAEP Engineering Framework, 2018).[xiv]
- Clarifies the design and implementation is guided by the best available research on assessment item design and delivery:
  - so that it is accessible to all students and whose design minimizes the need for any/standard accommodations for students with disabilities and English Learners.
  - so that students with disabilities and other diverse learners are considered during initial assessment design so they can fully participate and are provided adequate means to demonstrate their construct-relevant knowledge, skills, and abilities, including—but not limited to—the use and interoperability with any needed external assistive technology. (Almond et al., 2010; ESSA; Dolan et al., 2013)[xv]
  - Eliminate the use of the term ‘special needs’.

CAST thanks the NAGB for the opportunity to provide these comments, to advocate for a revision to the NAEP Science Framework, and to provide thoughts on how the Framework can be updated to align with current federal policy and documented best practices in the application of inclusive design in assessment design and application. This will allow the nation to provide all learners the opportunity to demonstrate fairly and accurately their science knowledge, skills, and abilities regardless of any potential learning barrier they may experience, whether it be due to socio-economic status, language, or disability status.

Please contact CAST's Director of Federal Relations Sherri Wilcauskas at swilcauskas@cast.org with any questions or for additional information.

Sincerely,

David Gordon
Interim CEO

---

[i] P.L. 110-315, P.L. 113-28, P.L. 114-95, P.L. 115-224, National Education Technology Plan (2021), U.S. Department of Education.

[ii] National Center for Education Statistics Appendix Tables (2009) at: https://nces.ed.gov/nationsreportcard/pdf/main2009/2011451.pdf; Appendix Tables (2015) at: https://www.nationsreportcard.gov/science_2015/files/2015_Science_Technical_Appendix.pdf; Appendix Tables (2019) at: https://www.nationsreportcard.gov/science/supporting_files/2019_appendix_sci.pdf

[iii] National Assessment Governing Board *Testing and Reporting on Students with Disabilities and English Language Learners Policy Statement*, (2014) at: https://www.nagb.gov/content/dam/nagb/en/documents/policies/naep_testandreport_studentswithdisabilities.pdf

[iv] National Center for Education Statistics Appendix Tables (2009) at: https://nces.ed.gov/nationsreportcard/pdf/main2009/2011451.pdf; Appendix Tables (2015) at: https://www.nationsreportcard.gov/science_2015/files/2015_Science_Technical_Appendix.pdf; Appendix Tables (2019) at: https://www.nationsreportcard.gov/science/supporting_files/2019_appendix_sci.pdf

[v] Dolan, R.P., Burling, K., Harms, M., Strain-Seymour, E., Way, W. (Denny), & Rose, D.H. (2013) *A Universal design for Learning-based Framework for Designing Accessible Technology-Enhanced Assessments* at: http://images.pearsonclinical.com/images/tmrs/dolanudl-teaframework_final3.pdf

[vi] Almond, P., Winter, P., Cameto, R., Russell, M., Sato, E., Clarke-Midura, J., Torres, C., Haertel, G., Dolan, R., Beddow, P., & Lazarus, S. (2010). Technology-Enabled and Universally Designed Assessment: Considering Access in Measuring the Achievement of Students with Disabilities: A Foundation for Research. *The Journal of Technology, Learning and Assessment*, *10*(5) at: https://ejournals.bc.edu/index.php/jtla/article/view/1605

[vii] Quellmalz, E. S., Silberglitt, M. D., Buckley, B. C., Loveland, M. T., & Brenner, D. G. (2016). Simulations for Supporting an d Assessing Science Literacy. In Y. Rosen, Y., Ferrara, S., & Mosharraf, M. (Eds.). (2016). *Handbook of Research on Technology Tools for Real-World Skill Development*. IGI Global at: http://doi:10.4018/978-1-4666-9441-5

[viii] See: P.L. 114-95

[ix] See: P.L. 108-446, Sections 300.105 and 300.324; and 29 U.S.C. 794d

[x] See: P.L. 114-95, Section 1111, (b)(2)(B)(vii)(II)

[xi] Rose & Gravel, (2013); Daley & Rappolt-Schlichtmann, 2009; Rose & Meyer, (2006); Blascovich, Mendes, Tomaka, Salomon, & Seery,(2003); Csiksentmihalyi, (1991)

[xii] See: P.L. 114-95, Section 1111, (b)(2)(B)(vii)(II)

[xiii] Rose & Gravel, (2013); Daley & Rappolt-Schlichtmann, 2009; Rose & Meyer, (2006); Blascovich, Mendes, Tomaka, Salomon, & Seery,(2003); Csiksentmihalyi, (1991)

[xiv] The 2018 NAEP Technology and Engineering Literacy Framework at: https://www.nagb.gov/content/dam/nagb/en/documents/publications/frameworks/technology/2018-technology-framework.pdf

[xv] Almond, P., Winter, P., Cameto, R., Russell, M., Sato, E., Clarke-Midura, J., Torres, C., Haertel, G., Dolan, R., Beddow, P., & Lazarus, S. (2010). Technology-Enabled and Universally Designed Assessment: Considering Access in Measuring the Achievement of Students with Disabilities: A Foundation for Research. *The Journal of Technology, Learning and Assessment*, *10*(5) at: https://ejournals.bc.edu/index.php/jtla/article/view/1605; P.L. 114-95, Section 1111, (b)(2)(B)(vii)(II); Dolan, R.P., Burling, K., harms, M., Strain-Seymour, E., Way, W. (Denny), & Rose, D.H. (2013) *A Universal design for Learning-based Framework for Designing Accessible Technology-Enhanced Assessments* at: http://images.pearsonclinical.com/images/tmrs/dolanudl-teaframework_final3.pdf

From:	Chester E. Finn, Jr
To:	NAGB Queries
Subject:	NAEP Science Framework
Date:	Wednesday, September 8, 2021 3:54:22 PM
Attachments:	2012-State-Science-Standards-NAEP-6.pdf

In response to your request for comments on the current NAEP science framework, I'm pleased to weigh in, both on my own behalf and that of the Thomas B. Fordham Institute. We formally reviewed that framework in 2012 in connection with a wide-ranging Fordham examination of state K-12 science standards. This led to an A-minus grade for the NAEP framework from our reviewers (led by the distinguished biologist Paul Gross). This included a maximum score of 7 out of 7 for the framework's "content and rigor." You can see that review at
http://edexcellencemedia.net/publications/2012/2012-State-of-State-Science-Standards/2012-State-Science-Standards-NAEP.pdf
and I attach a copy with this note.

Here's how we explained our decision to review the NAEP framework side-by-side with the standards of 50 states and DC: "The National Assessment of Education Progress (NAEP) is the most-often used barometer of student learning in science. Results from NAEP are used to compare student achievement across states and to judge states' student proficiency levels. Because NAEP is so central to the conversation on state and national science achievement, we felt it was important to analyze the quality of its implicit standards—embodied in its assessment framework—to see how they compare with the quality of each state's standards."

I should note that most state standards fared dismally in that review--only a handful got top marks.

Which leads me both to underscore the singular importance of NAEP and its frameworks as pacesetters and academic gold standards, and to say that the document you're starting with is very, very strong in its present form. As the old saying goes, if it ain't broke....It may well need some updating but the National Assessment Governing Board should think long and hard before undertaking a wholesale overhaul or replacement.

Thanks for your consideration.

Chester E. Finn, Jr.
Distinguished Senior Fellow & President Emeritus, Thomas B. Fordham Institute, and Senior Fellow, Hoover Institution, Stanford University

1016 16th Street NW, 8th floor, Washington DC 20036
Website: www.fordhaminstitute.org

SCIENCE

NAEP

GRADE: **A-**

SCORES:
Content and Rigor **7/7**
Clarity and Specificity **2/3**

TOTAL SCORE: **9/10**

Overview

The NAEP *Science Framework* for science is an extended statement of science learning expectations at grades four, eight, and twelve. The NAEP assessment is based on the science content, skills, and testing procedures outlined in the *Framework*. Sample questions show how learning expectations discussed in the *Framework* are actualized in the assessment.

Although the *Framework's* design and organization are complex and in a few places difficult to understand, in general the document works well, providing a useful epitome of K-12 science knowledge and related skills.

There are two main issues to be addressed in evaluating this *Framework*. One is length—the number of content expectations that it includes is substantial, even though limited to three grade levels. The second is purpose: How may we evaluate this *Framework*, which is conceived as a design for testing, as a set of standards that can guide curriculum making? Early in its 155 pages, the *Framework* makes this important distinction between content and curriculum:

> **Key *principles as well as facts, concepts, laws, and theories* that describe regularities in the natural world are presented…as a series of *content statements* to be assessed at grades 4, 8, and 12…[T]hese statements comprise the NAEP science content. *They define only what is to be assessed by NAEP and are not intended to serve as a science curriculum framework.* (emphasis added)**

The writers are to be congratulated for having taken the trouble thus to define "content" as used by them. Yet although the *Framework* is not intended as a comprehensive set of standards for K-12 science, it clearly does *imply* such a set. In fact, it is unlikely that state education officials, district administrators, and teachers will ignore its plentiful science content and proposed achievement levels, particularly in light of the strong influence that NAEP and its assessment results carry in American primary and secondary education. Thus, we treat the NAEP *Science Framework* here as a set of expectations for K-12 science knowledge—a.k.a. science content standards.

Organization of the Framework

NAEP sidesteps enduring debates over how to define scientific relationships among themes, principles, content, practices, scientific reasoning, inquiry, and so forth by

Document(s) Reviewed

- *Science Framework for the 2009 National Assessment of Educational Progress.* 2009. Accessed from: http://www.nagb.org/publications/frameworks/science-09.pdf
- *NAEP Science Sample Questions: Grade 4.* 2009. Accessed from: http://nces.ed.gov/nationsreportcard/pdf/demo_booklet/09SQ-O-G04-MRS.pdf
- *NAEP Science Sample Questions: Grade 8.* 2009. Accessed from: http://nces.ed.gov/nationsreportcard/pdf/demo_booklet/09SQ-G08-MRS.pdf
- *NAEP Science Sample Questions: Grade 12.* 2009. Accessed from: http://nces.ed.gov/nationsreportcard/pdf/demo_booklet/09SQ-G12-MRS.pdf

Figure 1. Crossing content and practices to generate performance expectations

		Science Content		
		Physical Science Content Statements	**Life Science Content Statements**	**Earth and Space Sciences Content Statements**
Science Practices	**Identifying Science Principles**	Performance Expectations	Performance Expectations	Performance Expectations
	Using Science Principles	Performance Expectations	Performance Expectations	Performance Expectations
	Using Scientific Inquiry	Performance Expectations	Performance Expectations	Performance Expectations
	Using Technological Design	Performance Expectations	Performance Expectations	Performance Expectations

dividing science knowledge into just two broad categories: principles and practices. The various principles comprise what is usually called science content: facts, concepts, theories, and laws. They are organized into the now-familiar content areas: physical, life, and earth and space sciences.

Next, NAEP identifies four science practices: identifying science principles, using science principles, using scientific inquiry, and using technological design.

Finally, the *Framework* designers assemble all three areas of general content (principles and their expansions) and all four general areas of practice into a matrix. Each resulting cell of this matrix is a potentially large set of performance expectations (see Figure 1). Thus for every general content area, there are four possible (and testable) practices corresponding to the -*ing* actions listed: 1) recognizing, naming, or describing the content; 2) employing the content correctly in one of its contexts; 3) showing skills needed to use that content in answering a scientific question, and 4) applying the content in a design or engineering problem.

Organization of Content Topics

Within the three main content domains (physical, life, and earth and space), how many standards do K-12 students really need to meet? In science education, at present, this is a vexed question. Some say "very few." Others say "enough to display, at least, the *range* of modern science." Still others would answer "a whole lot." NAEP settles somewhere in the middle by expanding its three content areas into eighteen foundational statements: six on physical science, five on life science, and seven on earth and space science. These are then further specified by various detailed explanations encompassing most of the basics at each assessed grade level (four, eight, and twelve), but increasing in number, sophistication, and detail from fourth grade through twelfth grade.

The physical science content area illustrates this complex structure. It is divided into six basic principles: properties of matter, changes in matter, forms of energy, energy transfer and conservation, motion at the macroscopic level, and forces affecting motion. These six principles are represented by fifteen actual content statements in fourth grade, by sixteen statements in eighth grade, and by twenty-three statements in twelfth grade. Therefore, all assessable physical science is represented in this *Framework* by fifty-four short statements of science content.

Moreover, these content statements are amplified at each grade. For example: One of the six principles of physical science is "changes in matter." In fourth grade, this principle is represented by one explicit content standard—that cooling and heating can convert matter from one recognizable state (solid, liquid, or gas) to another. In eighth grade, "changes in matter" expands to two representations, one on the molecular organization of matter and the other on chemical reactions and the conservation of mass in the course of reaction. And by twelfth grade, this principle expands to three (carefully crafted) statements, one on the energetics of state change, a second on atomic structure and electrons in atoms, and a third on chemical bonds and reactions.

In addition to the fifty-four content statements for physical science, there are thirty-two for life science and thirty-nine for earth and space science—a total of 125 explicit content statements. Since all the assessable content of K-12 science is supposed to be covered, that is not an unreasonable number.[1]

Content and Rigor

Physical Science

Content statements for fourth-grade physical science are comprehensive and emphasize properties, states, and transformations of matter. They address adequately the basics of energy and motion in grade-appropriate terms. Content statements for eighth-grade physical science—concerned with physical and chemical change—are more specific and comprehensive than are our own criteria (see Appendix A). For twelfth grade, content is strong except for light treatment of some important advanced topics of twelfth-grade chemistry (reaction mechanisms, acid-base chemistry, chemical bonds in important classes of macromolecules). Overall, the physical science content presented covers the necessary ground with neither critical omissions nor trivialities.

Earth and Space Science

The earth and space science content is well chosen. Content and sequencing concerning Earth's internal structure and plate tectonics—including the key geological evidence from seafloor spreading—are analytical and sufficiently comprehensive. For the principle "earth in space and time," the single fourth-grade expectation appropriately concerns the distinction between slow and catastrophic change. Fossils appear in eighth grade, as do mountain building and erosion. Twelfth-grade expectations expand to include, among other topics, the scale and magnitudes of geologic time. Perfect science standards would give more attention to the earth's age and to stellar evolution (as exemplified in the Hertzsprung-Russell diagram). The *Framework* gives weather and climate unusual prominence, but at the expense of astronomy and cosmology. That said, the development of scientific ideas is generally appropriate throughout the grades, and the few omissions are compensated for by careful presentation of the included content.

Life Science

Life science coverage is broad and reasonably inclusive. Basic themes—such as the mechanisms of heredity—are represented (as they should be) at all three grade levels. But "evolution and diversity," central to modern biology, does not appear until eighth grade—and some even of its simplest elements not until twelfth grade. Even then, there is no mention of the now-indispensable molecular and population genetics relevant to evolution. Somewhat disproportionate attention is paid to ecology and ecosystems (here under the thematic head of "interdependence"), and that comes at the expense—inter alia—of physiology, control systems, and developmental biology. Basic cell biology, on the other hand, is very well covered and is sequenced thoughtfully by grade.

The *Framework's* principles and detailed content statements cover virtually all the expectations spelled out in our review criteria and introduce no significant peripheral matter. A full-credit score of seven out of seven for content and rigor is justified. (See Appendix A: Methods, Criteria, and Grading Metric.)

Clarity and Specificity

This *Framework* document concedes—as it must—that distinctions among its four basic practices are anything but sharp. They are nevertheless convenient for communicating skill expectations and for representing the underlying standards that must guide writers of test questions. The authors are evidently comfortable with the residual ambiguities, perhaps judging that they do not damage the implied standards. They make possible, presumably, the construction of fair and comprehensive tests, which is of course what the *Framework* is about. Nevertheless, while the total number of principles is appropriate, the potentially dense intersections of them and the practices (that is, the total number of principles as expanded grade by grade, multiplied by the four broad and not sharply distinguishable practices) make it difficult for a reader to comprehend a bounded set of expectations. Thus clarity is to some extent compromised by complexity; as such, the *Framework* is awarded a score of two out of three for clarity and specificity. (See Appendix A: Methods, Criteria, and Grading Metric.)

[1] The *Framework* reports that content selection was guided primarily by two national sources: the *Benchmarks for Science Literacy* of the American Association for the Advancement of Science (1993) and the *National Science Education Standards* of the National Research Council (1996), plus follow-up documents. The authors note, however, that those documents do not limit or prioritize content in the form of assessable units. (In fact they are often concerned with history, philosophy, and sociology of science.) The NAEP *Science Framework* concerns itself with "science" as commonly understood. And its tabulated content is justified and supported by clarifications and discussions of "crosscutting"—content relevant to more than one of the three science domains.

9115 Westside Parkway
Alpharetta, GA 30009

888.413.3669
cognia.org

Lesley Muldoon
Executive Director
National Assessment Governing Board
U.S. Department of Education

Dear Ms. Muldoon,

These comments are submitted by Cognia, a global non-profit education company, in response to the request for preliminary public comments for the Science Assessment Framework for the 2028 National Assessment of Educational Progress (NAEP). The comments submitted by Cognia focus on science frameworks and equity in the development of assessments.

Cognia has served as a trusted partner for over 125 years, aiding education providers in providing and advancing the pathways of success for all learners, supporting continuous improvement and accreditation. In addition, for nearly forty years, Cognia has delivered high-quality assessment services in support of student learning and growth, and accountability for both general education students and students with significant cognitive disabilities. Cognia is a leading provider of custom-designed assessments, specializing in a full range of text test development activities.

Cognia's team is diverse and expansive with expertise and experience in assessment, accreditation, certification, systems thinking, continuous improvement, school turnaround, and professional learning to provide comprehensive, aligned, and innovative services. We serve education organizations at every level from state agencies and large school systems to individual schools, leaders, and teachers. Cognia is committed to ensuring every child has equal access to learning opportunities and resources. This process begins with helping our institutions address the complex issues related to diversity, equity, accessibility, and inclusivity through quality of education.

Cognia is leading efforts to address the history and legacy of racism in educational assessment through development of *A Call to Action: Confronting Inequity in Assessment (Lyons, Johnson, and Hinds, 2021)*. Working closely with Lyons Assessment Consulting, several authors from Cognia contributed to this paper, which provides a strong foundation for the work Cognia is doing with respect to diversity, equity, accessibility, and inclusion. *A Call to Action: Confronting Inequity in Assessment* offers deep dives into five opportunities for centering the principles of diversity, equity, accessibility, and inclusion in the design and use of educational assessments. Problems related to equity are not limited to those of racial injustice, but the authors focus this document primarily on race-related issues in the hope that dismantling such structures will provide pathways for addressing other marginalized communities in our society generally and in educational assessment specifically. The Call to Action is designed to foster meaningful conversation and innovative ideas for advancing practice in educational measurement and improving our assessments to help move us toward a more equitable future. As an organization, we are dedicated to supporting our institutions in their improvement of what they do to help students learn.

The comments below have been compiled from our experts in content development, measurement services, and equity and transformation learning services.

Cognia Recommendations for Revisions to the NAEP Science Framework

As a "key measure in informing the nation on how well the goal of scientific literacy for all students is being met," the NAEP Science Assessment should be based upon the standards, instruction, and research in science education most immediately influencing the nation's science classrooms. It should also embody culturally relevant assessment practices, to ensure representation and fair evaluation of all student groups. While we have several clear recommendations for necessary revisions of the content elements of the current NAEP Science Framework (National Assessment Governing Board, 2019), we feel it imperative to begin our recommendations on the point of equity, diversity, accessibility, and inclusion. The necessity of attention first and foremost being placed on creating an equitable science assessment framework cannot be overstated in order to support all students in learning science.

Rationale for an Equitable Science Assessment Framework

A new equitable science framework would emphasize diversity, equity, accessibility, and inclusion to support learning, increase engagement, and provide visible representation in content with a goal to improve diversity in representation of underrepresented groups in science fields of study and the workplace. This framework would consider students as the focal point and include meaningful interactions and feedback loops with the community as reflected by the students' contexts and communities.

An equitable science framework is a commitment to serving *all* students throughout the assessment design, development, and implementation process. This framework would ensure that underrepresented students are visible in curriculum and assessment content and would provide opportunities to create culturally relevant approaches for students from marginalized groups, particularly students of color, students living in poverty, and non-male identified students. Increased student (and community) engagement, especially from underrepresented groups, will expand opportunities for equitable representation in advanced studies in science fields and the workplace.

Culturally relevant assessment practices are supported by the sociocultural perspective on how students learn. Making sense of new learning concepts is developed and maintained by mental schema, and we integrate new knowledge by searching for meaning and relevance, building on our prior understandings organized in mental structures informed by our lived experiences and social interactions (National Academies of Sciences, Engineering, and Medicine, 2018). Culturally sustaining assessment validates the cultural embeddedness of learning and explicitly attends to the sociopolitical reality of students in marginalized populations. It affirms their cultures and identities, creates counter-narratives, and ultimately builds student agency for understanding, critiquing, and confronting systems of social injustice *(Lyons, Johnson, and Hinds, 2021)*. When students are at the center of assessment, students are reflected in the curriculum and assessment content.

Creating a practice for understanding diverse learners and connecting them to science activities includes outreach and engagement with families and community members. This begins with the assessment development process, curriculum integration, and solving real problems. A community issue and/or problem can be framed within the context of an informal or formal learning community that includes multiple stakeholders such as learners, educators, local community members, businesses, and other nonprofit organizations. Embedding this within an equitable framework will increase community connection to scientific practice and data, and support the inclusion of participation from communities that have not had an adequate voice in the scientific educational process.

Growth Mindset Approach

A growth mindset is the belief that learning skills and qualities are on a continuum and can be developed through effort and support from others. A growth mindset can be cultivated in the classroom environment with students and educators, as well as with parents and guardians.

In a recent growth mindset study by PISA (2021), students who present a growth mindset score higher than their peers with a fixed mindset. People who consider their ability to be malleable (a growth mindset) will strive to develop it by setting challenging learning goals. They consider effort an inherent part of the learning process and setbacks to be fruitful experiences to assimilate…This leads them to stretch and expend efforts to reach their full potential whereas people with a fixed mindset are more likely to develop a hunger for approval that restricts them to their comfort zone (Dweck and Yeager, 2019).

Growth mindset can be leveraged as a strategy to support students of color and underrepresented students by reflecting growth mindset approaches in the language used in the framework in order to increase learner self-efficacy and motivation to learn from mistakes, and expand scientific skills centered on real world/life problem solving and knowledge. This also supports centering an approach for encouraging students to engage with science within the context of the framework.

Revising Development Processes to be Centered on Equity

In operationalizing an equity science assessment framework, the development process must be updated to include the long-overdue centering of students in assessment and meaningful engagement of stakeholders who are representative of student populations served by NAEP. Exhibit 1 illustrates an updated process of equitably generating assessment items and tasks and interpreting student responses that includes these commitments. Stakeholders include parents/caretakers, community members, and perhaps high school students and younger students.

9115 Westside Parkway
Alpharetta, GA 30009

888.413.3669
cognia.org

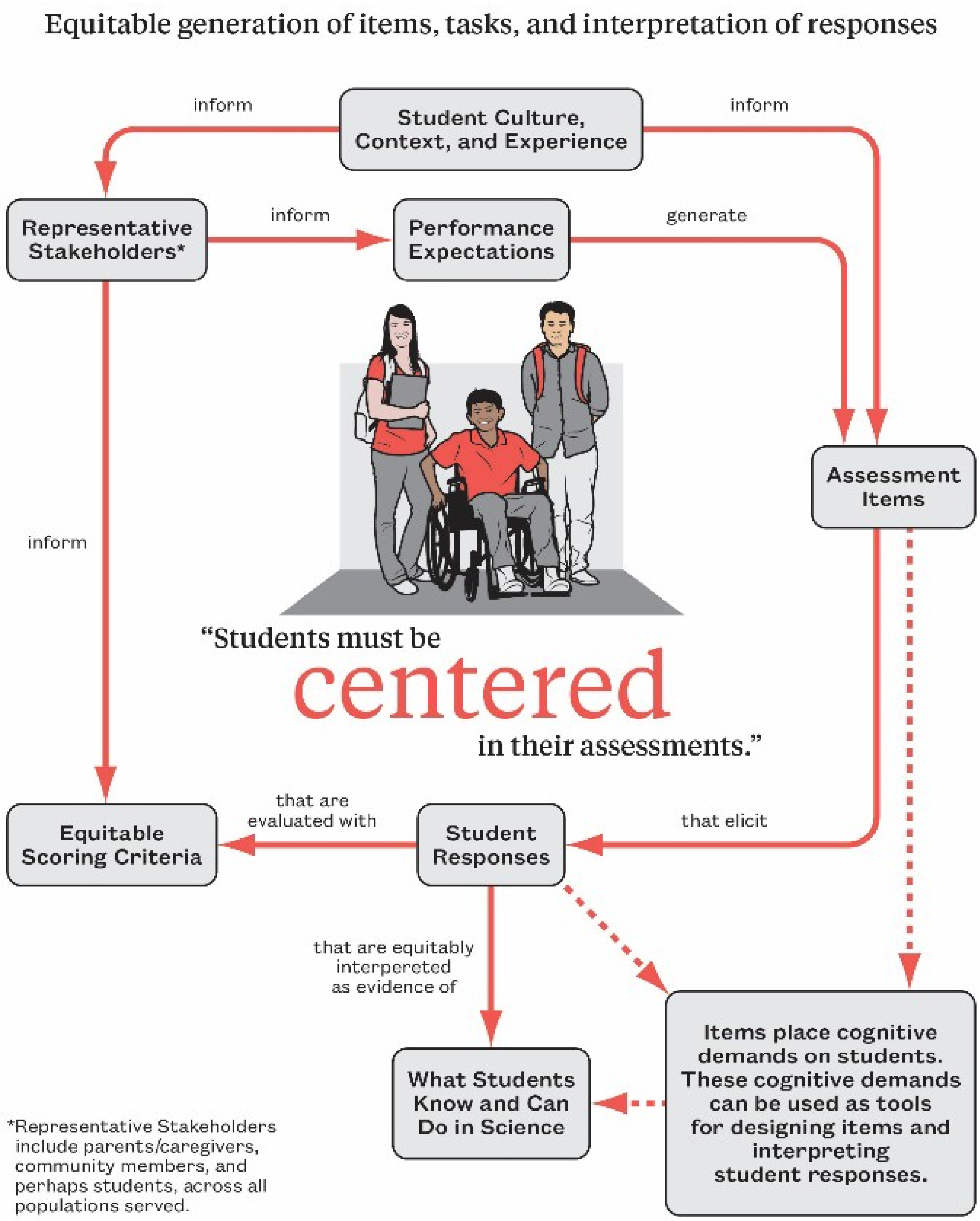


Exhibit 1: Student centered assessments.

An item or task is an individual question or exercise on the NAEP Science Assessment and is used to gather information about students' knowledge and abilities. Items and tasks are anchored in well-informed performance expectations, which describe in observable terms what students are expected to know and do on the assessment.

As shown in Exhibit 1, students must be at the center of any assessment of their progress. Their cultures, contexts, and experiences must inform the development of assessment items and tasks and the understanding and actions of representative stakeholders who are involved in the development process. In turn, representative stakeholders are involved in the creation of performance expectations by providing input regarding the cultural relevance and responsiveness of the expectations, including how to connect the performance expectations to students' lived experiences (e.g., relevant phenomena). These equitable and inclusive performance expectations guide the development of assessment items and tasks.

The cognitive demands and cultural relevance of assessment items and tasks can then be used to interpret students' responses as evidence of what students know and can do in science and how science concepts and skills relate to students' lives. Educators Shane Safir and Jamila Dugan cite the importance of developing assessments that reflect the mindsets and habits of professionals in the field and that "this shift from students as consumers of information to practitioners of field knowledge is especially significant for Black, brown and Indigenous students, signaling that they belong to a larger intellectual community (Safir and Dugan, 2021). The assessments that students encounter should include tasks that elicit authentic student performance to the extent practicable.

The development of scoring criteria for all student-constructed responses to items and tasks also actively involves representative stakeholder engagement, in order to ensure that all student populations are considered and represented in the scoring criteria. Exhibit 1 suggests that assessment development is both a multifaceted and iterative process, with significant consideration given to examining the equitable performance of assessment items across all tested populations as a compulsory part of the piloting process.

In evaluating item performance, in the Call to Action we suggest that examining differential item functioning (DIF) separately by gender, socioeconomic status, and race is now not only insufficient, but counter-productive in that cross-sectional views of item DIF are washing out the within-group intersectional effects (e.g., low SES Black females) (Russell, 2020). Class, race, ethnicity, language, and gender diversity are all possible influences on the manner in which knowledge is acquired and demonstrated on an assessment (Gordon, 1995). The field should be able to quickly move to detecting intersectional effects in estimates of cumulative test bias, or differential test functioning, particularly with the large sampling that NAEP is able to perform (Lyons, Johnson, and Hinds, 2021).

In summary, it is no longer enough to point to diversity, equity, accessibility, and inclusivity solely based on traditional approaches such as universal design, accommodation features, and classic DIF categories. While these approaches have their place, a true shift that starts with and maintains students at the center of the assessment is required for the NAEP Science Assessment to measure and reflect the science achievement of our nation's current students.

Constructs to be Assessed

The conditions that necessitated the revisions resulting in the *Science Framework for the 2019 National Assessment of Educational Progress* – namely publication of new science standards, advances in research, growth in innovative assessment approaches, and the need for increased inclusivity – are the same conditions that point to the need to revise the framework at present. While we assert that prioritization of diversity, equity, accessibility, and inclusion must be the driver of a new framework as the most critical lens for revisions, we have also identified several aspects of the assessed content that need to be reviewed and revised as well.

Since the publication of *A Framework for K-12 Science Education* and the *Next Generation Science Standards* (NGSS), almost all states have adopted the NGSS as their science standards or have developed science standards that are *Framework*- or NGSS-adapted. As was the case with the *Science Framework for the 2019 National Assessment of Educational Progress*, a change in the standards driving science curriculum and instruction clearly necessitates revisions to the framework again. The NAEP Science Framework needs to be updated to reflect the constructs presented in the NGSS, structured around the philosophy of three-dimensional performance expectations. Content, practices, and crosscutting concepts need to be redefined and aligned to match the way they are operationalized in the NGSS. We will elaborate on the considerations for each dimension more specifically in the following paragraphs.

Content (Disciplinary Core Ideas)

In this case, "content" refers to traditional disciplinary-based knowledge. The content in the NAEP Science Framework needs to be crosswalked with the Disciplinary Core Ideas (DCI) presented in the NGSS to redefine the appropriate set of content for the NAEP Science Assessment going forward.

While there is significant overlap for some concepts between the NAEP Science Framework and the NGSS, there are also many differences. Some content in the current NAEP Science Framework is not emphasized to the same degree in the NGSS, and likewise there are some concepts in the NGSS that are missing or sparse in the NAEP framework. As an example, in Physical Science, wave concepts and the connections between speed and energy are two content topics more prominent in the NGSS DCIs than the NAEP Science Framework; as another example, there is a heavy emphasis on motion graphs in the NAEP framework, whereas in the NGSS, motion graphs are not specifically codified into separate DCIs but are a part of the tools for evidence used by students to make claims about an object's motion or forces on an object. Similar examples appear in Life Science and Earth and Space Sciences as well.

Those revising the framework will also need to attend to any shifts in grade levels for content. Learning progressions should continue to underpin the content statements across grades in each domain, just as both the NGSS DCIs and the current NAEP Science Framework have done. To better reflect this in the new framework, we recommend considering a coding scheme that does account for these progressions rather than the sequential numbering currently used in the NAEP Science Framework. Additionally, developers must be mindful in applying those learning progressions in item development to ensure there is understanding of the effect of cognitive complexity, practice, and crosscutting concept influences at each node of content along the progression, such that assessment items measure constructs as appropriate and intended for the grade level.

A very significant additional consideration related to grade levels is whether the NAEP elementary assessment grade should be changed from grade 4 to grade 5. While the *National Science Education Standards* organized the elementary grade band K-4, the NGSS created elementary standards by grade for grades K-5 and designated the middle school grade band standards for grades 6-8. A large number of states have redesigned their elementary science assessment to assess students at grade 5 instead of grade 4 in adopting NGSS or NGSS-like standards, and NAEP assessment designers should give serious consideration to doing the same as they examine the content to include in the framework.

In addition to the three traditional content areas of Physical, Life, and Earth and Space Sciences, the NGSS includes Engineering Design as a content domain. While the NAEP Science Framework addresses elements of engineering and technological design, it has been more so through the practices, and the framework revision will need to look at recategorizing and elevating Engineering Design as *A Framework for K-12 Science Education* and the NGSS do.

While the nationwide shift to NGSS-based instruction is argument in and of itself for revising the NAEP Science Framework, the NGSS are also internationally benchmarked standards. In preparing to develop the K-12 Science Framework and the NGSS, Achieve completed an international benchmarking study of ten countries' science standards, including those countries who are consistent high performers on PISA and TIMMS. The current NAEP Framework acknowledges the importance of comparing expectations against international science education achievement expectations.

Practices

In defining the Science and Engineering Practices, the writers of *A Framework for K-12 Science Education* intentionally defined several targeted practices "to better specify what is meant by inquiry in science and the range of cognitive, social, and physical practices that it requires" (National Research Council, 2012). While the current NAEP Science Framework includes "practices," they are simply too broad to focus towards the specific expectations of current science instruction, and new practices need to be defined, aligned to the eight practices of the NGSS.

Some of the expectations within the four NAEP practices overlap with various NGSS practices, e.g., explaining observations and proposing and evaluating alternative explanations within Using Science Principles align with concepts for Constructing Explanations and Engaging in Argument from Evidence;

proposing and critiquing solutions, considering criteria and constraints, and identifying tradeoffs within Using Technological Design align with concepts for Defining Problems and Designing Solutions. However, there is much more interpretation and generality associated with the NAEP practices, which renders them insufficiently aligned to the expectations of current science instruction. Further, the first practice, Identifying Science Principles, would not be considered a practice according to the NGSS, and in fact should not be assessed. The NGSS set expectations for knowledge in use, and simply being able to recognize or recall facts is no longer sufficient for demonstrating proficient science achievement. Also, in regard to engineering practices, the NAEP Science Framework restricts assessment of design to only the science principles associated with the problem and does not include other considerations (e.g., economic, social) for the problem. This, however, contradicts the current need to build more relevant, equitable assessments that do engage students based on their lived experience and social justice. Some other assessments, such as PISA, seem to be more fully engaging with social and global problems, and NAEP assessment designers should do the same for equity, putting students at the center of the assessment.

Crosscutting Content (Crosscutting Concepts)

In the current NAEP Science Framework, "crosscutting content is not represented by abstractions such as 'models,' 'constancy and change,' or 'form and function,' but is anchored in the content statements themselves" (National Assessment Governing Board, 2019). This approach is quite opposite that of *A Framework for K-12 Science Education* as well as the *National Science Education Standards* and *Benchmarks for Science Literacy*, which defined crosscutting concepts (or unifying concepts and processes, common themes in *NSES* and *Benchmarks*, respectively) as more schematic approaches to science thinking, i.e., concepts having explanatory value via "an organizational framework for connecting knowledge from the various disciplines into a coherent and scientifically based view" (National Research Council, 2012). The NAEP Science Framework needs to pivot back to defining theme-based crosscutting concepts, which in fact was how they were represented in the 1996-2005 Framework. This shift is required to provide coherence and consistency between NAEP and current NGSS-based instruction, bringing the third dimension of the performance expectations into alignment. *A Framework for K-12 Science Education* defines seven crosscutting concepts, which should be the basis for redefining crosscutting concepts in the new Science Framework. If for some reason NAEP framework developers choose not to align to this definition of crosscutting concepts, they should name this concept something else in the new framework in order to avoid confusion for the field.

Additional Recommendations for Revising the Science Constructs to be Assessed

As the next set of framework constructs are created, the wording of each statement needs to be carefully reviewed to detect and eliminate bias and to ensure inclusivity. Some current content statements are biased and not inclusive – for example, "manmade," "heavenly body," etc. The new framework needs to clearly avoid such phrasing.

In tandem with updating the constructs to be assessed in the next framework, we encourage NAEP assessment developers to be thorough in updating the accompanying specifications documentation. We recommend including a significant amount of explicit information around clarifications and assessment boundaries, as this level of detail is in our experience extremely useful in ensuring assessment items measure the constructs as intended. Further, we recommend including examples of grade-appropriate phenomena for the assessed content in the specifications, although it should be made clear that the examples are not an exhaustive list and analogous phenomena should also be used in assessment development. Many of those examples, or similar examples, as well as assessment items should continue to be included in the framework itself, to provide direct illustration of how the framework constructs and assessment design will be operationalized.

The framework and specifications should also document clear methodology around the creation of performance expectations for NAEP assessments, given that the crosses of DCIs, SEPs, and CCCs (assuming they are adopted) will yield a far greater number of possible combinations than the crosses of content and practice in the current NAEP framework. At present, states vary on the approach of assessing any possible combination of the foundational dimensions of the standards versus assessing only the specifically crossed performance expectations defined in the NGSS. Given that NAEP has a different purpose than a state accountability assessment does, we propose that continuing to be more generalized may better reflect the variety in format and instruction of the standards across the nation, as well as the holistic way instruction should occur, and would provide the opportunity to measure a range of applied performances that students can do. Whatever methodology is chosen, clear definition of the blueprint that any given NAEP assessment's performance expectations must meet will be paramount in the design and interpretation of the assessment and results. NAEP developers must be extremely transparent and explicit about the interpretations – and non-interpretations – of the assessment results based on the defined methodology in comparison to each particular state's standards and approach.

It will be important for NAGB to select an organization well-versed in the NGSS and the advances in science education research to do the work around construct revisions, and this organization should be continually executing on a strong mission in support of diversity, equity, accessibility, and inclusion. NAGB should also connect with members of the National Research Council of the National Academies for advisement on the status of NGSS implementation and any revision considerations for the NGSS. The time lag between framework revisions and the first NAEP assessment to be aligned to a new framework is significant and given that the NGSS are almost nine years old already, any effort to ensure the NAEP Science Framework is not outdated before it even comes into use, both in terms of science content and student representation, will be extremely important.

Item Types and Assessment Design

Based on the changes we have recommended to the constructs to be assessed, we offer additional recommendations relative to the NAEP assessment design to best support these proposed changes, beginning with overall assessment design principles and progressing to specific blueprint and item type feedback.

The very first steps in a principled approach to assessment design and development are to clearly define the assessment targets (for which we have made recommendations in the previous section) and to define intended score interpretations and uses (SIUs). We recommend, based on the proposed construct revisions for the new NAEP Science Framework and the known variations in the structure and implementation of NGSS-based standards and curriculum across the nation, that NAEP assessment designers take the time to very intentionally and explicitly define the SIUs for the forthcoming NAEP Science Assessments based on the new framework. There must be a clear, common understanding of what the new NAEP assessment is really telling the nation about its students and their achievement in science – accompanied then by transparent, emphasized, public messaging of the SIUs – in order for assessment results to be meaningful and actionable.

An associated piece in these first design steps, which follows defining the assessment targets and coordinates with a model of cognition or learning to guide the assessment design, is considering the framework to be used for cognitive complexity. Achieve has published ideas for reconceiving cognitive complexity for the NGSS (Achieve, 2019), which depart from Webb's Depth of Knowledge model (used by many states, though not by NAEP in science) and press for more depth than the four-level scheme used by NAEP for science. As previously noted, the lowest complexity level that focuses on identification and recall really no longer meets the bar for adequate science literacy and achievement. Items that only assess declarative knowledge should not be included in the assessment, or only included to the most limited extent. Given these considerations, we encourage framework developers to explore new schemes for cognitive complexity. We would also encourage conducting cognitive labs to probe the validity of the chosen new scheme as applied to science assessment items.

After these foundational design steps are completed, we offer the following additional recommendations for more detailed designing of the new framework and assessment:

- Continue to ground all assessment items in science phenomena and engineering design problems. The focus on sense-making around phenomena and designing solutions to problems is the heart of the vision for science education in *A Framework for K-12 Science Education* and is what we now aspire to for our students. Associated with this, there is abundant opportunity to continue to integrate, and even more fully integrate, the Nature of Science into assessment items. Intentional care must be taken to represent this lens and all phenomena in items authentically, however, rather than simply provide "window dressing" to declarative items. The illustrative item on page 97 of the current NAEP Science Framework is a prime example; the response demands of the item are completely separate from the framing of the history and nature of science. The new framework and the assessment items that it directs should require application and sense-making of the stimulus material for the response.

- In adopting recommendations made in this commentary, the distribution of content areas and cognitive complexity in the assessment will have to be revised as well. The NGSS has a different weighting of content in the standards by grade level, and we have already provided reasoning around revising cognitive complexity schemes and weighting in the assessment.
- The item types being used, and the distribution of those item types, must also be reevaluated. Given the increased complexity of the NGSS, a significant reliance on multiple-choice items may no longer be sufficient to fully assess the science constructs as intended. We anticipate the need to place greater emphasis on constructed-response items and leverage more item clusters, POE items, and performance tasks, as well as introduce technology-enhanced items (e.g., drag-and-drop items, graphing interactions). Some additional elaborations on recommendations for various item types are as follows:
    - POE items have significant relevance to NGSS with their strong emphasis on evidence and reasoning. We recommend utilizing POE items to a greater degree.
    - Item clusters, or even two-part items, can be used to assess constructs in greater depth, supporting valid measurement of students' sense-making. Branching items may also be useful to further pursue for this purpose, with potential to gauge depth of understanding and ability to sense-make around a phenomenon. Leveraging the ability online to lock responses and then update those students who cannot move far into a branching set with correct information and allow them to continue on to additional questions may also be an area of measurement innovation to study.
    - We question the utility of concept mapping to some degree, relative to other item types, when considering the demands of the NGSS. Perhaps concept maps can be applied to specific phenomena presented, but we have concerns around the degree of inference that can be made without requiring students to provide evidence and reasoning for the links between concept terms in the map. More research on this item type may be necessary to support continued use.
    - Performance tasks are generally agreed upon as a necessity for authentic assessment of the NGSS. We see value in both hands-on performance tasks and interactive computer tasks. There may be ways to leverage technology to enhance what can be measured with hands-on performance tasks, by controlling what information students provide and when they get additional information to respond to (e.g., students design and carry out an investigation, record information online about their procedure and results, and then responses are locked before students are presented with a correct procedure and result to interpret). Hands-on tasks will be well-suited to assess both scientific investigation and engineering design. Interactive computer tasks will continue to allow assessment of constructs that can't be investigated in a hands-on manner and/or with reasonable economy. We would recommend changing the assessment design parameters to include a task for *all* students in the new science assessment, however, given the highly authentic match to the new constructs that need to be assessed. We also recommend carrying out the previously proposed study to compare the hands-on performance tasks and interactive computer tasks.

cognia™

9115 Westside Parkway
Alpharetta, GA 30009

888.413.3669
cognia.org

  - In considering equity, assessment developers may want to explore what affordances there are for more response modes relative to the item types. Is it possible to leverage technology and administration to support more students in providing responses in a mode that best allows them to show what they know and can do, for example, allowing recording of a spoken response rather than a typed response for a constructed response item?
- Ensure assessment development practices are aligned to the latest industry standards, as updated in the 2014 edition of the *Standards for Educational and Psychological Testing*.

As cited in the current framework, “The NAEP Science Assessment signals the kinds of responses to tasks, problems, and exercises, along with the kinds of knowledge and reasoning, that should be expected of students as a result of what is taught in the science curriculum.” We agree that the NAEP assessment has this impact, and we believe that the next revision of the science framework must therefore reflect the current efforts to center science instruction around *all* students through the NGSS. Throughout the current framework, there are elements that already resonate with and reflect principles that ground the content of the K-12 Science Framework and the NGSS, and the requirement now is to update the framework to be in clear alignment and thus measure science achievement relative to the new vision for science education being implemented across the nation.

Sincerely,

Stephen Murphy

Chief Learning Officer

References

Achieve. (2019). A Framework to Evaluate Cognitive Complexity in Science Assessments. https://www.achieve.org/files/Science%20Cognitive%20Complexity%20Framework_Final_093019.pdf

Dweck and Yeager. (2019). "Mindsets: A View From Two Eras", *Perspectives on Psychological Science*, Vol. 14/3, pp. 481-496 http://dx.doi.org/10.1177/1745691618804166.

Gordon. (1995). Toward an equitable system of educational assessment. Journal of Negro Education, 360–372.

Lyons, Johnson & Hinds. (2021). A Call to Action: Confronting Inequity in Assessment. https://www.lyonsassessmentconsulting.com/assets/files/Lyons-JohnsonHinds_CalltoAction.pdf

National Academies of Sciences, Engineering, and Medicine. (2018). How people learn II: Learners, contexts, and cultures. National Academies Press.

National Assessment Governing Board. (2019). Science Framework for the 2019 National Assessment of Educational Progress. https://www.nagb.gov/content/dam/nagb/en/documents/publications/frameworks/science/2019-science-framework.pdf

National Research Council. (2012). A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas. Washington, DC: The National Academies Press. https://doi.org/10.17226/13165

Russell, M. (2020). Personal communication.

Safir, S. and Dugan, J. (2021). Street Data: A Next Generation Model for Equity, Pedagogy, and School Transformation. Thousand Oaks, CA: Corwin.

From:	Heinz, Michael
To:	NAGB Queries
Cc:	Heinz, Michaekil
Subject:	NAEP Science Framework
Date:	Friday, October 15, 2021 9:33:25 AM

CAUTION: This email originated from outside of the organization. Do not click links or open attachments unless you recognize the sender and know the content is safe.

Thank you for the opportunity to provide comments and recommendations relative to the Science *Framework for the 2019 National Assessment of Educational Progress* (hereafter referred to as the *NAEP Science Framework*). I am submitting this document on behalf of the Board of Directors and the members of the Council of State Science Supervisors.
The Council of State Science Supervisors (CSSS) provides leadership in advancing excellence in P-12 science education at the local, state, and national levels. Our members include state science supervisors who are responsible for academic standards in science and/or statewide science assessments in 48 states. In addition to our state members, our organization includes researchers from institutions of higher education, experts from federal STEM mission-based agencies, and leaders from informal education organizations. Our members work both independently and collaboratively to ensure widespread, consistent, coherent opportunities for high-quality science learning is available to all students across K-12 and that people of all backgrounds are welcomed in science learning environments.

As science education leaders working at the intersection of local, state, and federal policies, we are most aware of the systemic value of coherence between state and federal assessment and the ability of CS3 to facilitate such coherence. Assessment tends to drive instruction and it can drive us forward or backward. Coherence between state and federal assessment will provide state leaders with another tool to improve science instruction for all students.

Recognizing the important role that NAEP science assessment data plays in decision making in states, territories, and at the Department of Defense Education Activity, **CSSS advocates for updating the *NAEP Science Framework*.** In this document we provide evidence to support our recommendation and describe some of the key components that should be a part of the revised framework.

In the announcement soliciting comments and recommendations, we were asked to focus on three questions. In the following section, we provide our responses.

Whether the NAEP Science Assessment Needs to be updated.

CSSS is a proponent for updating the *Science Framework for the 2019 National Assessment of Educational Progress*. Just as previous *NAEP Framework*s have been based on the latest research, so should be the *2028 NAEP Science Framework*. Two consensus studies of the National Academy of Sciences are most relevant to this include *Taking Science to School: Learning and Teaching Science in Grades K*-8 (2007), and *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas* (2013). A consensus study results from extensive research and deliberations by diverse groups including scientists, engineers, mathematicians, learning scientists, educational practitioners, and educational policymakers. The National Academies of Sciences, Engineering, and Medicine are acknowledged as the "Advisors to the Nation."

As of this writing, forty-four states (representing 71% of U.S. students) have science standards influenced by the *Framework for K-12 Science Education.* Quite simply, since the National Academies of Sciences, Engineering, and Medicine are acknowledged as "Advisors to the Nation', these reports are the best information available for how best to instruct our youth. And with a statistic of over 70% of U.S. students being taught using standards influenced by the *Framework for K-12 Science Education,* it makes sense as a focal point of measurement for coherency with American trends in science education.

If the Framework needs to be updated, why is a revision needed?

The current *NAEP Science Framework* has two separate components, science content and science practices. *Framework for K-12 Science Education* also defines distinct practices, core ideas, and crosscutting concepts—the difference is the expectation that they are integrated in instruction and assessment.

The current NAEP Framework is focused on research from the 1990's, upon which we have built considerable information. New research outlined in research like How People Learn II: Learners, Contexts, and Cultures (2018) provides further input regarding integration of content and practice for improved and more equitable outcomes. Students do not use their knowledge of content, practice and cross-cutting concepts in isolation of one another. The knowledge interacts in ways that provide scaffolding for recall, integration and problem solving in the context of a novel or repeat phenomenon(a). As noted by the Achieve Framework for evaluating cognitive complexity, artificially separating these cognitive processes in assessment does not provide us with an accurate or equitable measure of student proficiency in science. It is in our best interest to align our measures with instructional practice.

A second reason that a revision is needed is that *A Comparison Between the Next Generation Science Standards (NGSS) and the National Assessment of Educational Progress (NAEP) Frameworks in Science, Technology and Engineering Literacy, and Mathematics* found that differences in the depth, breadth, detail, or focus of that content resulted in low to moderate levels of content alignment, with differences by grade and content domain (2015).

Alignment with practices was strong, but the emphasis of NGSS performance expectations across NAEP science and TEL practices differed from the emphases specified in the NAEP frameworks.

What should a revision to the framework include?

Recommendation 1: Increased attention to equity. A new framework should include a renewed look at how science assessments reflect and includes features of equitable assessment. The COVID-19 pandemic shined a spotlight on inequities and unjust public education practices. As a result, many states have reconceptualized how they are working to make teaching, learning, and assessments more equitable for all students, including reconceptualizing how assessments are constructed, how diverse student experiences are represented in assessment tasks, and how students are able to make their thinking visible. The NAEP assessments have a long-standing history of representing the best of what is known about disciplinary assessment practices and revising the science framework to better represent equitable science assessment provides NAGB with the opportunity to continue to play this leadership role. As an organization that is not constrained by limitations created by statewide policies, NAGB should position itself to take up that work and to exemplify how large-scale assessments can provide equitable opportunities for all students to make their thinking visible.

Recommendation 2: Align to current shifts in state science standards. A new framework should also be responsive to, and a reflection of what states are doing with academic standards and statewide assessments. For example, there is a low level of alignment between the *NAEP Science Framework* and the disciplinary core ideas for grades K-5 defined in the NRC's *Framework*.

In Closing, a revised *NAEP Science Framework* should provide the nation with data that can be used to evaluate the effectiveness of states' efforts to make science education more equitable and meaningful for each of our approximately 48 million students.

CSSS stands ready to offer our considerable expertise and experience to assist with soliciting stakeholder feedback and to participate on an expert panel to support revisions to the *NAEP Science Framework*., as we did for the 1996-2005 and 2009-2015 NAEP Frameworks. As President of CSSS, I would be pleased to provide names and contact information for individuals to serve the NAGB.

Respectfully,

Michael Heinz
President
Council of State Science Supervisors

From: DANIELLE MURPHY
To: NAGB Queries
Subject: NAEP Science Framework
Date: Monday, September 27, 2021 2:59:18 PM

CAUTION: This email originated from outside of the organization. Do not click links or open attachments unless you recognize the sender and know the content is safe.

Hello,

I definitely think that the updated framework needs to include authentic reference to equity and justice. There is enough research showing that typical science knowledge and standards unfairly favor certain races and genders. To ignore research, and the public cost of doing so, is doing a disservice to students and cannot be considered a fair assessment.

I hope you consider ALL students when designing this assessment.

Sincerely,
Danielle

NCSE National Center for Science Education

September 9, 2021

National Assessment Governing Board
by e-mail

Dear colleagues,

I am writing on behalf of the National Center for Science Education, a non-profit organization affiliated with the American Association for the Advancement of Science and the National Science Teaching Association, with comments on the current NAEP Science Assessment Framework.

In NCSE's view, the NAEP Science Assessment Framework, while valuable in its time, needs to be updated now.

The primary reason to update the NAEP Science Assessment Framework is that its content was largely based on the National Science Education Standards and the AAAS Benchmarks for Scientific Literacy, which were then the most authoritative guides to science education. They have since been supplanted by the NRC's *A Framework for K–12 Science Education* (2012) and the Next Generation Science Standards (2013), both of which are considerably more up-to-date with regard both to science content and pedagogical methods. By now, twenty states (plus the District of Columbia) have adopted the NGSS, which are based on the NRC Framework, and a further twenty-four states have adopted state science standards that are based on the NRC Framework: it is fair to say that a majority of the nation's public school students are learning science more or less in the way envisioned by these documents.

A revision to the Framework should thus align it to the content and structure of the NRC Framework and the NGSS.

In addition, NCSE recommends that special attention be given to socially but not scientifically controversial topics—evolution, climate change, and vaccination in particular—and to the nature of science. For a variety of reasons, these topics are often neglected or inadequately treated in American science education, even in authoritative documents such as the NRC Framework and the NGSS. It would therefore be helpful to consult state science standards that improve on the NGSS's treatment of these topics, such as Massachusetts's with regard to evolution and Wyoming's with regard to climate change, and position statements from relevant professional scientific societies such as the Society for the Study of Evolution and the American Meteorological Society. While it is not realistic to expect students across the country to receive instruction conforming to best practices, it is counterproductive to make allowances for states that have chosen to undereducate or miseducate their students.

Sincerely,

Ann Reid

Ann Reid
Executive Director, NCSE

230 Grand Avenue, Suite 101, Oakland CA 94610 | 510-601-7203 | info@ncse.ngo | www.ncse.ngo

September 20, 2021

Dear National Assessment Governing Board,

Please find below comments relevant to the potential update of the NAEP Science Assessment Framework. I am comfortable with my name and affiliation being included with my comments.

I submit these comments based on my experience as a former state STEM leader at the Massachusetts Department of Elementary and Secondary Education. During my almost 12 years at the state agency, I was a member of a design team for the *Committee on a Conceptual Framework for New K–12 Science Education Standards*, was a Writing Team member for the *Next Generation Science Standards* (NGSS), was a state representative to the Lead State NGSS review process (facilitated by *Achieve*), and I led state STEM standards development and contributed to state assessment development. I also participated in several rounds of alignment reviews between NAEP and emerging or current science standards, including as a member of the NAEP/NGSS Comparison Panel in 2014, facilitated by the National Center for Education Statistics, and more recently between NAEP Science and selected state science standards, facilitated by HumRRO in 2020.

At a broad level, I would encourage a future iteration of NAEP science to maintain and/or enhance the following elements:

- Hands-on performance tasks. Such performance tasks are fundamental to doing science and necessary to provide opportunities to demonstrate the application of science concepts and practices. While a logistical challenge, these are critical and should be continued and even expanded as possible.
- Interactive computer tasks. The tasks have provided for a wider variety of innovative scenarios and contexts for students to apply their knowledge and skills. They are also helping to advance state-level assessment through proven examples of interactive assessment items. These too should be continued and expanded as possible.
- Integration of science content and practices. Science requires integration and application of both science concepts and practices together, not individually. The assessment of these two dimensions within individual items and across assessments is critical. Even as content or practices may be adjusted, and the practical implementation of assessing both dimensions may change, the measure and integration of both these dimensions should be continued.

Based on my experiences with science standards and assessment development in the recent past, I would encourage an update of the NAEP Science Assessment Framework for the following reasons:

- Since the last NAEP science revision, the National Research Council published the Framework for K-12 Science Education, and many states have adopted or adapted NGSS. Both efforts provide an updated framework of what is important to learn in science education, including the set of science concepts and a significantly different set of science practices.
- The NRC and NGSS documents attend to recent research on progression of learning in science education. An updated NAEP assessment framework can both attend to those and potentially contribute to the further study and articulation of science progressions of learning through the generation of data useful to researchers.
- There is a significant need for additional attention to equity, both from a racial perspective and to account other diversity within student populations. We must ensure that future NAEP assessments do not unintentionally disadvantage anyone from demonstrating their ability to perform science.
- An updated Framework provides an opportunity to advance multi-dimensional assessments that account for both concept and practice proficiencies in innovative items, assessment structures, and statistical analyses. More explicit guidance or specifications on item and assessment development should be produced to guide future NAEP administration. In my opinion the integration of the two dimensions of science concepts and practices is a substantial accomplishment; the integration of three dimensions at once (the third being cross-cutting practices, as defined by NRC and NGSS) is confounding to designers and users alike.

The work undertaking with NAEP Science is hugely influential to states across the country, and ultimately to curriculum and classroom practice. As such, I highly encourage an update to the NAEP Science Assessment Framework, and I am very interested in supporting and participating in work to achieve such an update.

Jacob Foster
Founder, STEM Learning Design, LLC (www.stemlearningdesign.com)

From:	Jacqueline Huntoon
To:	NAGB Queries
Subject:	NAEP Science Framework
Date:	Tuesday, September 28, 2021 2:11:30 PM

Please find below responses to the questions that were posed to the science education community:

- Q: Whether the NAEP Science Assessment Framework needs to be updated.
- A: Yes

- Q: If the framework needs to be updated, why a revision is needed.
- A: It focuses too heavily on content and tends to exclude the science and engineering practices and the crosscutting concepts. It should place greater emphasis on students' ability to use tools (which may include data presented to them) to investigate phenomena and design solutions to problems. A different way of saying this might be that it needs to focus on determining whether or not students can USE science as a tool to develop their own understanding.

- Q: What should a revision to the framework include?
- A: It should place more emphasis on applying the practices and crosscutting concepts in a variety of situations. I would also like to see less disciplinary differentiation because the interesting and challenging problems in science are less and less likely to be confined to one particular discipline. Even the example given for 8th grade earth science (gravity and planetary motions) has as much to do with physics as with earth science. I am an admittedly strong proponent of problem-based instruction in which science is taught as an integrated whole rather than as a series of separate disciplines. I am certain the leadership is aware of the National Academies reports on designing assessments in support of the Framework for K-12 Education and the NGSS. Documents such as these could provide good guidance.

Dr. Jacqueline E. Huntoon, PhD, PG
Provost and Senior Vice President for Academic Affairs
Michigan Technological University
www.mtu.edu

From:	Kelly Barber-Lester
To:	NAGB Queries
Subject:	NAEP Science Framework
Date:	Wednesday, September 29, 2021 5:26:33 PM
Attachments:	image004.png

Good afternoon,

In response to the request for feedback that was elicited via email, I am writing to share some input into the updating of the NAEP Science Framework.

Upon reviewing the document found here (https://www.nagb.gov/naep-frameworks/science/science-framework-feedback.html), I was struck and deeply concerned by the fact that the following words and phrases were completely absent:

- Equity
- Equality
- Inequality
- Racism
- Bias
- Scientific racism
- Prejudice
- Sexism
- Ethics

The term "race" is only present insofar as it is used to refer to student demographics for tracking sub-group assessment performance. "Culture" is only found once in the document, in reference to "the role science has played in various cultures"(p. 96). The term "harm" is used almost exclusively to refer to harm that could be caused to environments or ecosystems, and never in reference to the harm that has been caused by scientific pursuits (for example, the ways in which science has been "advanced" by experimenting entirely unethically on specific minoritized populations).

Furthermore, there is no discussion of bias or the mitigation of bias (cultural or otherwise) in terms of assessment, which is a well-established and ongoing concern in the field of education.

As it stands, the framework presents a vision and version of science as objective, neutral, and divorced from context and its unquestionably troubled history (and present) as it pertains to issues of inequity broadly, and specifically racism and sexism.

I hope that you will take these observation into account when updating the framework. Issues of equity must be explicitly included and addressed within this framework. Continuing to teach science devoid of its messy and often uncomfortable intertwining with issues of inequity and oppression may be attractive in its simplicity, especially to those that already see themselves and those like them represented positively in textbooks and in the discipline; that approach, however, ensures that we will continue to struggle with these same issues in science as we move forward.

Best wishes,
Dr. Kelly J. Barber-Lester

Kelly J. Barber-Lester, Ph.D.
Assistant Professor
Pronouns: she/her/hers [Learn more about pronouns.]
School of Education- Office 345
1 University Drive I P.O. Box 1510 I Pembroke, NC 28372O

"The world is before you, and you need not take it or leave it as it was when you came in."
- James Baldwin

From: Wray, Kraig
To: NAGB Queries
Subject: Science framework feedback
Date: Monday, September 20, 2021 11:51:19 AM

In reading the executive summary the things that stand out as important to the NAEP committee are: applying science to students lives, science literacy for all students, participation in society and work, and addressing local, national, and global challenges.   IF this is truly the purpose and primary driving factors for science education and therefore science assessment, I can not see how making sure phenomena, explanation, and understanding of science can exclude cultural and community ways of knowing and applying science. No where does the executive summary mention equity and making the practices relevant to local communities and students. Yet when you think deeply about the items listed above, they necessitate cultural relevance. Having members of the board and other team members that are knowledgeable about multiple ways of knowing, the history of marginalization, and by having these goals be explicit in the mission are essential to the success of the program. If we want students to be successful in science learning and for that learning to be reflected in the NAEP assessment, the development of an assessment with an equity focus is imperative.

Kraig Wray

Kraig A. Wray, Ph.D. (he/him/his)
Postdoctoral Scholar
Pennsylvania State University
Department of Curriculum & Instruction

From:	Mark Looy
To:	NAGB Queries
Subject:	FW: NAEP Science Framework
Date:	Saturday, September 25, 2021 3:30:07 PM

Greetings. I represent a non-profit organization with several staff holding earned doctorate degrees in science from prestigious institutions (e.g., Harvard and Brown). We appreciate the opportunity to suggest revisions to the science framework, especially in building the critical-thinking skills of students when they examine both sides of a scientific debate.

We submit that state and local educators should ensure that their teachers recognize that discussion about controversial subjects can lead to a more robust learning experience. For one, this approach helps hone the critical thinking ability of students. Unfortunately, there is false belief that it is unconstitutional to teach criticisms of topics such as evolution, the earth's age, the reliability of dating methods, etc. In reality, the constitutional approach would not prohibit the censoring of scientific ideas that run contrary to accepted belief, especially when credentialed scientists have opposing views. The teaching of controversial ideas held by dissenting scientists is both legal and beneficial—and with historical success as time and time again the status quo in science has been challenged.

Now, do we believe teachers should be *required* to teach creation science or ideas thatsupport a younger age of the earth? No. Such a policy would be counter-productive, for those positions would likely be taught poorly by most evolutionary instructors. But teachers shouldat least have the academic freedom to teach alternative ideas that are being presented by scientists, even if they happen to be in the minority.

--Mark Looy, CCO, Answers in Genesis

Mark Looy
CCO/Co-founder, Executive Department

From: Michael Lowry
To: NAGB Queries
Subject: Re: NAEP Science Framework
Date: Thursday, September 23, 2021 2:31:35 PM

CAUTION: This email originated from outside of the organization. Do not click links or open attachments unless you recognize the sender and know the content is safe.

Dear Sir/Madam:

I agree the framework should be updated to better reflect where we are as science educators, specifically as it relates to incorporating engineering practice (as found in NGSS) and the cross fertilization that is happening in STEM. The problems we face as scientists and engineers require more than the usual silos of "life science, physical science and earth science." The urgency of climate change should also play a more prominent role in the framework.

Regards,

Michael Lowry

--
Michael J. Lowry, NBCT, PAEMST
Science Department Chair
The McCallie School
500 Dodds AV
Chattanooga, TN 37404

Ancora imparo

Corrections to the NAEP Framework

1) In **E.12.8**, the statement "Plates are pushed apart where magma rises to form midocean ridges, and the edges of plates are pulled back down where Earth materials sink into the crust at deep trenches" is incorrect. The rise of magma at mid-ocean ridges is a passive effect, and not an active one. This statement incorrectly implies that the magma is rising up from the mantle and is actively pushing the two sides of the oceanic plates apart. The opposite is true. Other forces are pulling the plates apart, creating a low-pressure zone along the axis of the spreading center, and this pulls up mantle rock from below to fill the void. Because of the phenomenon of *pressure release*, as the hot rock is pulled up from below, certain minerals exceed their solidus temperature and exsolve from the solid mantle peridotite rock, rising up to the surface as more fluid magma with a gabbro/basalt composition, and either erupts on the seafloor as basalt or crystallizes within the crust as gabbro. The evidence for this passive, rather than active, upwelling of mantle rock beneath midocean ridges is multiple. First, there are no deep roots to the thermal anomalies beneath ridges; these are shallow features. Second, the state of stress within oceanic lithosphere is indicative of a significant "ridge-push" force, but this name is somewhat misleading because the magnitude of the ridge-push force is actually zero at the ridge itself and in fact increases away from the ridge, a result of the thermal topographic swell of the warm mid-ocean ridge rock (essentially, the ocean lithosphere "surfs" down the thermal swell from the ridge). Third, repeated geodynamic computer convection modeling has shown that the circulation of mantle convection, of which plate tectonics is the surface expression, is nearly entirely driven by the sinking into the deep mantle of subducted ocean lithosphere, also known as the "slab-pull" force. Basically, because heat is generated internally within the earth through diffuse radiogenic production from a small number of long-lived radioactive isotopes (K-40, U-235, U-238, Th-232), the actual patterns of mantle convection, and therefore plate tectonics, is a result of the cooling and sinking of Earth's surface and not the heating of Earth's interior.

So, to fix this, please change this sentence to:
"Old oceanic plates sink into the mantle at the deep trenches of subduction zones, creating a patterns of tectonic plate movements. Oceanic plates are pulled apart at mid-ocean ridges, allowing magma to rise to form new oceanic crust."

2) In **E12.3**: Change "Stars, like the Sun," to "E12.3: Stars, such as the Sun,"
The word "like" means "similar to," but similes are generally exclusive. A flashlight might appear "like" a star at night, but it is not a star. Here, we want to use "such as" to reiterate that our sun is a star.

3) In **E.8.10**: Change "Earth's magnetic field is similar to the field of a natural or man-made magnet with north and south poles and lines of force" to
"Earth's magnetic field is approximately similar to the field of a natural or man-made magnet with north and south poles and lines of force."

In fact, a quick glance at maps of the actual inclination and declination of Earth's magnetic field will show you that, in fact, Earth's magnetic field is actually not at all like the dipolar magnetic field from a simple north-south magnet. This is because Earth's magnetic field actually has significant contributions from higher-order magnetic terms (quadrupole, octupole, etc.). In fact, these terms dominate near the core-mantle boundary, but because they decay more rapidly with distance than the dipolar field, the dipole is more than 90% of the field at Earth's surface. Nonetheless, Earth's magnetic field is MUCH more complex than a bar magnet or solenoid, so we need to qualify this statement with something like "approximately."

4) In **E.12.9** Change "Earth systems have internal and external sources of energy, both of which create heat" to
"Earth systems have internal and external sources of energy, both of which provide heat."

It is misleading to say "create" heat for two reasons. First, heat is the ***transfer*** of energy, distinct from the thermal energy that is a material property of Earth substances. Second, we repeatedly say that energy/mass is conserved, neither created nor destroyed, so it could generated misconceptions to say "create heat."

5) In **E.12.10** Change "This energy transfer is influenced by dynamic processes such as cloud cover, atmospheric gases, and Earth's rotation, as well as static conditions such as the positions of mountain ranges, oceans, seas, and lakes" to
"This energy transfer is influenced by short-term processes such as cloud cover, Earth's rotation, ocean circulation changes, and the distributions of atmospheric gases, as well as long-term processes such as changes in the positions of continents, mountain ranges, ocean basins, and lakes."

This statement is very misleading. There is nothing "static" about the positions of mountain ranges, oceans, seas, and lakes! A good portion of geology addresses how these are all constantly changing over time. Likewise, a large part of research and understanding of climate examines how climate responds and changes to the occurrence and locations of mountain ranges (which increase erosion, removing carbon dioxide from the atmosphere and pushing global climates to be cooler) and ocean basins (which control how heat is circulated around Earth's surface, and is therefore dominant in controlling regional climates). Also, the ocean science community is quite adamant that there is just one ocean (as there is just one atmosphere), so you should avoid saying "oceans" when you really mean "ocean basins." Also, given the prominence of ocean circulation in controlling both regional and global climate changes, you should call out ocean circulation as distinct from the locations of ocean basins.

6) In **E.8.14** Change "Water, which covers the majority of Earth's surface, circulates through the crust, oceans, and atmosphere in what is known as the water cycle" to

“Water, which covers the majority of Earth’s surface, circulates through the geosphere, ocean, and atmosphere in what is known as the water cycle.”

Again, there is only one ocean. More significantly, most of Earth’s water (estimated to be about 5 ocean’s worth) is in the rock of Earth’s mantle. This water is constantly being pumped into the mantle along with the subducting ocean lithosphere. This water in the mantle is critical to Earth’s geology; it significantly lowers the viscosity of mantle rock, actually allowing the mantle to convect. Venus does not have plate tectonics, and this is likely because it is dry and does not have water. This water is constantly reentering the atmosphere and ocean at subduction zone volcanoes after it dehydrates from the sinking lithosphere at depths that begin about 100 km down.

7) **pp. 87-88**: Good gracious! Your whole example of finding a location between the earth and moon that has the same value of gravity is TOTALLY WRONG! The gravity at the surface of the moon is about 1/6 of that at Earth’s surface, but this has LITTLE to do with the equipotential location between them! This is significantly influenced by the different densities within the two bodies (which determines the location of the radius of the surface, which therefore determines the values of gravity at that particular location!) All that matters for the equipotential is their masses!

If we let the distance from the center of Earth to point C be “R,” then we can define the distance from the center of the moon to point C to be some fraction of that, called k*R. The total distance from the earth to the moon is therefore R+kR, or (1+k)R.

To find point C, we need to equate the values of g:

gE = gM
so
GMe / r^2 = GMm / (kr)^2

The G’s and r’s cancel, so we have:
k^2 = Mm/Me = 7.35e22 kg / 5.97e24 kg = 0.0123
so k =0.11 and therefore the distance from the center of the moon to point C is:
= k/(1+k) = 0.11 / 1.11 = 0.10
So point C is very close to being 1/10 of the way from the moon to the earth and NOT 1/6!!!!

So, on page 88, change the “Interpretation” to:
“Interpretation: The correct answer is C. Because the Moon has a mass that is about 1.2% of the mass of Earth, a body that experiences an equal gravitational force from Earth and the moon should be much closer to the moon. Point C is the only point that is closer to the moon. Note: Point C is about one-twelfth of the way between the moon and Earth; it should be one-tenth of the distance.”

[Also note: “the moon” should not be capitalized, just as “the earth” is not capitalized (although “Earth” correctly ***is*** capitalized).]

8) Why is this framework intentionally obsolete? There are lots of references to old and outdated NRC reports, but nothing from the 21st century? Why is the NRC’s *Framework for K-12 Science Education* omitted? Why are the *Next Generation Science Standards* omitted? A total of 45 U.S. states and D.C. are now using K-12 science standards that are adopted or adapted from the NGSS, but the rest (Florida, Texas, etc.) are using the eight Science and Engineering Practices (SEPs) of the NGSS. Why are the NGSS’s eight SEPs omitted and not even mentioned? It is almost as if you are intentionally trying to have this framework be irrelevant upon arrival?

Michael Wysession
Chair, NSF’s Earth Science Literacy Initiative
Chair, Earth and Space Science for the NRC’s Framework for K-12 Science Education
Chair, Earth and Space Science for the Next Generation Science Standards

Professor of Geophysics, Department of Earth and Planetary Sciences
Executive Director, Center for Teaching and Learning
Washington University in St. Louis
St. Louis, MO 63130

Dear Committee,

I have concerns that are listed below.

EXECUTIVE SUMMARY

In the rapidly changing world of the 21st century, science literacy is an essential goal for all of our nation's youth. Through science education, children come to understand the world in which they live and learn to apply scientific principles in many facets of their lives. In addition, our country has an obligation to provide young people who choose to pursue careers in science and technology with a strong foundation for their postsecondary study and work experience. The nation's future depends on scientifically literate citizens who can participate as informed members of society and as a highly skilled scientific workforce, well prepared to address challenging issues at the local, national, and global levels. Recent studies, including national and international assessments, indicate that our schools still do not adequately educate all students in science.

Science seeks to increase our understanding of the natural world through empirical evidence. Such evidence gathered through observation and measurement allows for an explanation and prediction of natural phenomena. Hence, a scientifically literate person is familiar with the natural world and understands key facts, concepts, principles, laws, and theories of science, such as the motion of objects, the function of cells in living organisms, and the properties of Earth materials. Further, a scientifically literate person can connect ideas across disciplines; for example, the conservation of energy in physical, life, Earth, and space systems. Scientific literacy also encompasses understanding the use of scientific principles and ways of thinking to advance our knowledge of the natural world as well as the use of science to solve problems in real-world contexts, which this document refers to as "Using Technological Design."

The National Assessment of Educational Progress (NAEP) and its reports are a key measure in informing the nation on how well the goal of scientific literacy for all students is being met. The *Science Framework for the 2019 National Assessment of Educational Progress* sets forth the design of the NAEP Science Assessment. The 2019 NAEP Science Assessment will use the same framework used in 2009. The 2009 NAEP Science Assessment started a new NAEP science trend (i.e., measure of student progress in science), and the 2019 NAEP Science Report Card will include student performance trends from 2009 to 2019. Trends in student science achievement were reported from 1996 to 2005 as well. However, the trend from 1996 to 2005 was not continued due to major differences between the 2005 and 2009 frameworks. The 2009 – 2019 framework represents a unique opportunity to build on key developments in science standards, assessments, and research. This document is

The most important component of Scientific Literacy is to understand,

reflect upon issues critically and explicitly, empowers the future citizens to engage in critical deliberation on science-based social issues

Scientific literacy for democratic decision-making, Hagop A. Yacoubian, Pages 308-327 | Received 18 Jun 2016, Accepted 19 Dec 2017, Published online: 29 Dec 2017

“in a year-long TCA program, researchers administered attitudinal surveys to understand the program's impact on two important aspects of scientific literacy: students' perceptions of science as important to society and personal decision-making, and student ability to carry out scientific practices.” https://eric.ed.gov/?id=EJ1228452

Engels, Mary; Miller, Brant; Squires, Audrey; Jennewein, Jyoti S.; Eitel, Karla
Electronic Journal of Science Education, v23 n3 p33-58 2019

Comments must be submitted via email to nagb@ed.gov with the email subject header *NAEP Science Framework* no later than 5:00 p.m. Eastern Time on Thursday, September 30.

When providing comment, please indicate if you are not comfortable with your name and affiliation being included with your comments, which may be shared and discussed publicly in upcoming Governing Board meetings and materials.

If the Governing Board decides that an update is needed, the charge to launch the revision process for the NAEP Science Framework is anticipated at the March 2022 quarterly Board meeting. 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.

Michelle

Michelle M McCarthy, M.Ed.
Science Instructional Coordinator
Department of Teaching and Learning
Montana Office of Public Instruction

- Phone: 406.444.3537
- Mobile: 406.860.6619
- Email: MMcCarthy5@mt.gov
- OPI Science Website

Department of Middle and Secondary Education
College of Education & Human Development

Mailing Address:
P.O. Box 3978
Atlanta, GA 30302-3978

In Person:
College of Education & Human Development
6th Floor
30 Pryor Street
Atlanta, GA 30303

Phone: 404-413-8060
Fax: 404-413-8063

October 15, 2021

To the NAEP Governing Board,

We wish to offer our collective feedback for the upcoming revision of the NAEP Science Framework. As a group of colleagues, we represent a diverse range of disciplinary expertise and research interests while also sharing a commitment to the continued improvement of K12 education and teacher preparation. Further, we also share a strong commitment to the ever-increasing importance of both considerations of and actions towards developing equitable classrooms for learners from all communities, prioritizing minoritized communities that for too long have been underserved or relatively abandoned by many elements of the national K12 infrastructure. In light of this shared vision, we offer several broad considerations and relevant literature for the board to review and incorporate into the new NAEP framework.

To begin, we will directly address the three guiding questions offered by the board in their call for public comment during this process. Yes, the NAEP Science Assessment Framework needs updating since the visions, standards, and curricular approaches for science education across the nation have undergone significant restructuring and reorienting in their emphases since the last version was developed. Why should the framework be revised? Although the NAEP Science Assessment must remain "curriculum-neutral", the shift in focus across much of the nation towards visions and standards that emphasize "three-dimensional learning" (National Research Council, 2012; NGSS Lead States, 2013). These significant shifts involve prioritizing students in active learning experiences where they engage in various scientific practices while using important and broadly applicable science concepts to make sense of various real-world phenomena. As such, this fundamental view of learning that is grounded in science practice necessitates assessments that reflect that emphasis as well. The current NAEP framework and the assessment structures that have resulted from it involving mostly conceptual recall multiple-choice questions do not align well with these more active visions of science education nor do the various conceptualizations of 'inquiry' in the previous framework.

Further, the forced nature of assessments that rely heavily on multiple-choice questions does not reflect the wealth of knowledge that has developed over the past few decades regarding Universal Design for Learning (Meyer, Rose, & Gordon, 2014). We agree with broad considerations offered by the CAST organization (2015) that all assessments should “support learner variability through flexible assessments using UDL guidelines” which would also include more variety and flexibility in NAEP assessment item structure and ways of accessing the NAEP Science Assessment for different learners. Following UDL guidelines, assessments should “eliminate unnecessary barriers in assessments” including, for example, thick reading passages that may present greater challenges for multilingual learners and not connect the lived experiences of many groups of learners (AERA/APA/NCME, 2014; CAST, 2015). Finally, assessments should also “assess engagement as well as content knowledge”, which remains necessary for the previously described visions for science education and for developing more equitable assessments (Wiliam, 2010; CAST, 2015).

The final question posed in the call is “What should a revision to the framework include?” The remainder of this letter will offer a broad overview of two critical areas for consideration that any meaningful Science framework revision will include in significant and explicit ways. We also point to several national-level reports and texts along with more specific empirical research and perspective articles that could support the revision teams’ work and the growth of the NAEP Science Framework in beneficial ways.

Science Learning through Science Practices

Reviewing the previous NAEP Science Framework, an obvious but critical change that is necessary involves extensive revision of the language, foci, and structure of the framework and assessment items in ways that more accurately reflect the current visions for science education that guide most districts and states in the country, including the *Framework for K12 Science Education* (National Research Council, 2012) and the *Next Generation Science Standards* (NGSS Lead States, 2013), as well as corollary texts that focus on assessment at all levels (National Research Council, 2014; Schweingruber, Beatty, & NASEM, 2017). Science teachers, researchers, and administrators tirelessly but thoughtfully work to shift the nature of instruction and learning experiences offered to students in science classrooms throughout the country. These shifts emphasize the foundational role of engaging students in a collection of specific practices that reflect the work of scientists as they endeavor to develop and refine scientific understandings of the world and universe.

As emphasis on these practices continues to grow, the distribution of item types and guidance language in the NAEP Science Framework needs to reflect those shifts as well. Such change requires the inclusion of more performance tasks and simulation-based tasks and less knowledge or conceptual recall items (NRC, 2014). Further, efforts in science classrooms and standards aim for students to not simply engage in these practices, but to also learn about how they function in the development of scientific knowledge (Ford, 2015). Therefore, the practices should also be viewed as science “content” so that items could be developed that assess students’ understanding of the function of the practices. For helpful reviews of the nature of these practices and how science education continues to emphasize their role in learning, we recommend Crawford (2014) and Osborne (2014) as supportive reading for the revision team.

Research and curricular innovation of the last decade heavily emphasized two explanatory practices in science, modeling and argumentation. Modeling as one of the central sensemaking processes in science has been well established over the past decade (Miller & Kastens, 2018; Wade-James, Demir, Qureshi, 2018). The development and use of scientific models set the foundation for students to construct scientific understandings of systems as well as predictions about new but related systems, while also affording explicit opportunities to expand students' learning about the nature of science as they engage in modeling (Schwartz, 2019). Many different curricular interventions that have gained popularity in classrooms across the country are grounded in this major scientific practice (Windschitl, Thompson, & Braaten, 2020; Windschitl & Thompson, 2013). Thus, modeling is a primary practice that constitutes an important component of the "content valued by the nation". As such, the development of new assessment items should be heavily connected to the modeling practices. These items can have students engage in interpreting representational and mathematical models while also using developed models to make predictions about systems.

For argumentation, much research and development work has established several considerations for how students and teachers learn through and about the practice (Henderson, McNeill, Gonzalez-Howard, Close, & Evans, 2018; Osborne, 2014). The goal for learning through argumentation involves supporting learners' ability to develop evidence through the analyses of various types of data collected from a range of investigations and phenomena and use core science concepts to reason with that evidence and develop claims in response to compelling investigative questions (Grooms, Sampson, & Enderle, 2018; Sampson, Enderle, & Grooms, 2013). The robust scholarship around scientific argumentation led to the development of several prominent curricular resources and instructional models (Hand, Wallace, & Yang, 2004; McNeill & Krajcik, 2011; Grooms, Enderle, & Sampson, 2015) that have been taken up by districts and schools across the nation, establishing this fundamental practice of science as further "content valued by the nation". Items aimed at assessing students' grasp of argumentation and their ability to engage in it could address evaluating the quality of evidence provided for a claim, evaluating the coordination between evidence and claims, describing appropriate science concepts that have been used to reason through evidence in support of a claim, and considerations of confirmation bias and other fallacies when engaging in arguing from evidence.

Other practices have not received as much research attention but are at the forefront of many science learning experiences, such as computational thinking (Enderle, King, & Margulieux, 2021). Although much debate exists around holistic conceptualizations of this practice, some common elements exist across all of them, including abstraction, algorithmic thinking, and decomposition (Grover & Pea, 2013). These shared conceptual elements could serve as the focus for items that target students' understanding of computational thinking. Although the NAEP Science Framework aims to be "curriculum neutral", the framework does need to be designed in ways that make it flexible and applicable for the next ten years of growth in science education. To achieve this flexibility, attention must be given to the total assemblage of scientific practices being implemented in classrooms across the country, from major ones to those less emphasized.

Equity and the Assessment of Diverse Learners

Reviewing the previous NAEP Science Framework, there is a striking silence when it comes to considerations of diverse learners and equitable assessment. Several of the "Special Studies" identified in the previous framework do take steps towards considering the needs of diverse populations of learners. However, most of these studies focus on technical comparisons of formats and capabilities. The scholarship surrounding the significant influence of students' cultures and communities on their learning has grown tremendously since the publication of the previous framework. A recently published text from the National Academies of Science, Engineering, and Medicine (NASEM, 2018) provides an excellent introduction to this work by highlighting the important role that culture and learning contexts play in every student's learning trajectory, including the influence of culture on learners' biological, motivational, and reasoning development.

The *Standards for Educational and Psychological Testing* developed jointly by American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (2014) should be prominent in the revision of the NAEP Science Framework. These guidelines synthesize a vast body of literature regarding assessment and provide critical insights into many aspects of assessment development, including those of the size and scope of the NAEP. Concerning equity, the *Standards* offer great detail and consideration of the concept of "fairness" in assessments. This particular section of the *Standards* underwent significant expansion in response to the rapidly developing knowledge base surrounding equity and supporting diverse populations of learners, including recognizing this work as foundational to assessments as considerations of validity and reliability. A major tenet of fairness, as conceptualized in this text, is that assessment administrators must provide access for all examinees in various populations, particularly in allowing for accommodations and modification for learners with different cognitive, linguistic, and physical abilities (AERA/APA/NCME, 2014).

Behizadeh (2014) offers examples of how to align large-scale writing assessments with fundamental knowledge generated through sociocultural theories of learning, lenses that elucidate the construct of 'fairness' while highlighting the many challenges assessments, including NAEP, present for students, particularly those from marginalized communities. This work also draws attention to the consequential validity of such assessments. Consequential validity considers the intended and unintended impacts of large-scale assessments on all learners, and such considerations must acknowledge the detrimental impacts that assessment scores have had in the ways they have been used to characterize minoritized communities as deficient. To understand more nuanced concerns about how assessment scores, including NAEP across several disciplines, have been used in oppressive ways towards these communities of learners, we also recommend Love (2019), Muhammad, Ortiz, & Neville (2021), and Stinson (2015). For considerations of fairness and equity in science education across a range of student populations and learning environments, we also strongly recommend the committee consider the seven chapters in Section III: Diversity and Equity in Science Education of the *Handbook of Research on Science Education II* (Lederman & Abell, 2014). Finally, we recommend some of our work to provide insight into ways that students from minoritized populations, including Black girls and students who are deaf or hard of hearing, can be denied access to science through various aspects of the education system (Enderle, Cohen, & Scott, 2020; King & Pringle, 2018; Wade-James, King, & Schwartz, 2021).

Another element of the AERA/APA/NCME construct of fairness emphasizes the need to minimize barriers in accessing assessments, including aligning the design and development of assessment items using the tenets of UDL. As mentioned previously, UDL highlights the need to provide students taking assessments with multiple means of engagement, expression, and representation. Applying these principles to the design of assessment items entails the development of multiple question formats and response options, providing students with choices to enhance access for diverse learners. Further, in the design of all item types, issues that might restrict an examinee's ability to demonstrate what they know (AERA/APA/NCME, 2014) must be removed. Examples in the *Standards* (AERA/APA/NCME, 2014) provide ways to address these issues for various populations of learners. The work of Fine and Furtak (2020) offers insight into ways science assessments can be developed to support, rather than restrict, multilingual learners. Even the most straightforward consideration of minimizing barriers should include commitments to offering the assessment in multiple languages, rather than just English, and supporting students who are deaf or blind with additional video interpretations and audio recordings of assessment items so they all have the opportunity to represent their full understanding of the content.

The final and perhaps most critical element of 'fairness' explored in the *Standards* entails promoting fair test score interpretations. A requirement for fair test score interpretation involves the inclusion of data points and metrics that characterize students' "opportunity to learn" (OTL). Indeed, the *Standards* emphasize the importance of incorporating OTL metrics as causal factors in score interpretations. Such usage necessitates that the new NAEP Science Framework explicitly commits to avoiding traditional and staid comparisons of outcomes across learners from communities varying greatly in OTL metrics. Rather, the new framework should endeavor to focus on interpretations within communities and populations based on OTL metrics while also maintaining an 'asset' orientation in all interpretations (NASEM, 2018), rather than traditional 'deficit' views that have been associated with large-scale assessments, such as NAEP, and the reporting of outcomes.

Haertel, Moss, Pullin, and Gee (2008) assert that thoughtful consideration of OTL metrics extends beyond basic considerations of content resources and instructional practices. OTL metrics must consider how students are given opportunities to personally connect to their science learning experiences through "forms of knowledge and ways of using language [from their] everyday experiences in families and communities" (Haertel et al., 2008, p. 8) and funds of knowledge (Gonzalez, Moll & Amanti, 2005). Practically, to achieve this aspect of fairness, the NAEP Science Framework revision team must work to broaden the collection of OTL data from participating districts, administrators, communities, and schools. We encourage the revision team to consider an example of such nuanced quantitative analyses around a community-based science learning effort offered by King and colleagues (2021). Further, as an example of a thoughtful and broad data collection effort around science education, including community OTL factors, the revision team should also review the work of Banilower and colleagues (2018), who produced the *Report of the 2018 National Survey of Science and Mathematics Education*, as well as the OTL instruments developed for the ATLAST (Assessing Teacher Learning About Science Teaching) project from the same organization, Horizon, Inc.

The United States continues to live through an acute inflection point as a society that highlights the sincere need for continued and sustained discussion and efforts that work to support *ALL* of its citizens, particularly young people. Such support cannot advance the communities where these learners come from without transparent and thoughtful reckoning with how large-scale assessments have shaped those learners' experiences within the national education system and been used to their detriment. Further, such reflection *must* be coupled with deliberate actions that work in direct opposition to the continued use of such harmful practices while also working to expand opportunity, fairness, and equity in our science classrooms. At a minimum, the NAEP Governing Board and those working on the revision of the NAEP Science Framework must explicitly and emphatically assert the importance of equity and fairness throughout the various elements of the framework and the design of the next NAEP Science Assessment.

We provide a full list of references cited throughout our letter in the hopes that the various revision teams will take time to read and reflect on their connections to the new framework. We hope the NAEP Governing Board and those working on the revision teams of the NAEP Science Framework sincerely reflect on the two major issues we have elaborated on above, science learning through science practices, and equity and assessment of diverse learners. Both warrant considerable attention and explicit inclusion in any new assessment framework for science education, particularly if the goal is to "maintain NAEP as the gold standard", including "ensuring that the NAEP frameworks are updated for modern expectations for students" *and* the country's entire K12 education system.

Sincerely,

Patrick J. Enderle, Ph.D.
Associate Professor of Science Education

Renee Schwartz, Ph.D.
Professor of Science Education

Kadir Demir, Ph.D.
Associate Professor of Science Education

Natalie King, Ph.D.
Assistant Professor of Science Education

David Stinson, Ph.D.
Professor of Mathematics Education

G. Sue Kasun, Ph.D.
Associate Professor of Language Education

Christine D. Thomas, Ph.D.
Professor of Mathematics Education

Paula Garrett-Rucks, Ph.D.
Associate Professor of World Languages Education

Nadia Behizadeh, Ph.D.
Associate Professor of Adolescent Literacy

Caroline C. Sullivan, Ph.D.
Associate Clinical Professor, Teacher Preparation and Social Studies

Gertrude Tinker Sachs, Ph.D.
Chair, Department of Middle & Secondary Education

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). Standards for educational and psychological testing. Washington, DC: American Educational Research Association.

Banilower, E., Smith, P., Malzahn, K., Plumley, C., Gordon, E., & Hayes, M. (2018). Report of the 2018 NSSME+. Chapel Hill, NC: Horizon Research, Inc.

Behizadeh, N. (2014). Mitigating the dangers of a single story: Creating large-scale writing assessments aligned with sociocultural theory. *Educational Researcher*, *43*(3), 125-136.

CAST. (2015). *Top 10 UDL tips for assessment.* CAST Professional Publishing. Retrieved from https://slds.osu.edu/posts/documents/top-10-udl-tips.pdf on October 11, 2021.

Crawford, B. (2014). From inquiry to scientific practices in the science classroom. In *Handbook of research on science education, volume II* (pp. 529-556). Routledge.

Enderle, P., King, N., & Margulieux, L. (2021). What's in a Wave?. *The Science Teacher*, *88*(4), 24-28.

Enderle, P., Cohen, S., & Scott, J. (2020). Communicating about science and engineering practices and the nature of science: An exploration of American Sign Language resources. *Journal of Research in Science Teaching*, *57*(6), 968-995.

Ford, M. J. (2015). Educational implications of choosing "practice" to describe science in the next generation science standards. *Science Education*.

Fine, C. G. M., & Furtak, E. M. (2020). A framework for science classroom assessment task design for emergent bilingual learners. *Science Education*, *104*(3), 393-420.

Gonzalez, N.E., Moll, L., & Amanti, C. (2005). *Funds of knowledge. Theorising practices in households, communities and classrooms.* Routledge.

Grooms, J., Sampson, V., & Enderle, P. (2018). How concept familiarity and experience with scientific argumentation are related to the way groups participate in an episode of argumentation. *Journal of Research in Science Teaching*, *55*(9), 1264-1286.

Grooms, J., Enderle, P., & Sampson, V. (2015). Coordinating scientific argumentation and the Next Generation Science Standards through argument-drivenTips inquiry. *Science Educator*, *24*(1), 45-50.

Grover, S., & Pea, R. (2013). Computational thinking in K–12: A review of the state of the field. *Educational researcher*, *42*(1), 38-43.

Hand, B., Wallace, C. W., & Yang, E. M. (2004). Using a Science Writing Heuristic to enhance learning outcomes from laboratory activities in seventh-grade science: quantitative and qualitative aspects. *International Journal of Science Education*, *26*(2), 131-149.

Henderson, J. B., McNeill, K. L., González-Howard, M., Close, K., & Evans, M. (2018). Key challenges and future directions for educational research on scientific argumentation. *Journal of Research in Science Teaching*, *55*(1), 5-18.

King, N. S., Collier, Z., Johnson, B. G., Acosta, M., & Southwell, C. N. (2021). Determinants of Black families' access to a community-based STEM program: A latent class analysis. *Science Education*.

King, N. S., & Pringle, R. M. (2019). Black girls speak STEM: Counterstories of informal and formal learning experiences. *Journal of Research in Science Teaching*, *56*(5), 539-569.

Lederman, N. G., & Abell, S. K. (Eds.). (2014). *Handbook of research on science education, volume II* (Vol. 2). Routledge.

Love, B. L. (2019). *We want to do more than survive: Abolitionist teaching and the pursuit of educational freedom*. Beacon Press.

McNeill, K. L., & Krajcik, J. S. (2011). Supporting Grade 5-8 Students in Constructing Explanations in Science: The Claim, Evidence, and Reasoning Framework for Talk and Writing. *Pearson*.

Meyer, A., Rose, D. H., & Gordon, D. T. (2014). *Universal design for learning: Theory and practice*. CAST Professional Publishing.

Miller, A. R., & Kastens, K. A. (2018). Investigating the impacts of targeted professional development around models and modeling on teachers' instructional practice and student learning. *Journal of Research in Science Teaching*, *55*(5), 641-663.

Muhammad, G. E., Ortiz, N. A., & Neville, M. L. (2021). A Historically Responsive Literacy Model for Reading and Mathematics. *The Reading Teacher*, *75*(1), 73-81.

National Academies of Sciences, Engineering, and Medicine. (2018). How People Learn II: Learners, Contexts, and Cultures. Washington, DC: The National Academies Press. doi: https://doi.org/10.17226/24783.

National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. National Academies Press.

National Research Council. (2014). Developing assessments for the next generation science standards.

NGSS Lead States. (2013). The next generation science standards. Achieve, Inc.

Osborne, J. (2014). Scientific practices and inquiry in the science classroom. In *Handbook of research on science education, Volume II* (pp. 593-613). Routledge.

Sampson, V., Enderle, P., & Grooms, J. (2013). Argumentation in science education. *The Science Teacher*, *80*(5), 30.

Schwartz, R. S. (2019). Modeling Competence in the Light of Nature of Science. In *Towards a Competence-Based View on Models and Modeling in Science Education* (pp. 59-77). Springer, Cham.

Schweingruber, H., Beatty, A., & National Academies of Sciences, Engineering, and Medicine. (2017). *Seeing students learn science: Integrating assessment and instruction in the classroom*. National Academies Press.

Stinson, D. W. (2015). The journal handbook of research on urban mathematics teaching and learning: A resource guide for the Every Student Succeeds Act of 2015.

Wade-Jaimes, K., King, N. S., & Schwartz, R. (2021). “You could like science and not be a science person”: Black girls' negotiation of space and identity in science. *Science Education*.

Wade-Jaimes, K., Demir, K., & Qureshi, A. (2018). Modeling strategies enhanced by metacognitive tools in high school physics to support student conceptual trajectories and understanding of electricity. *Science Education*, *102*(4), 711-743.

Wiliam, D. (2010). What counts as evidence of educational achievement? The role of constructs in the pursuit of equity in assessment. *Review of Research in Education*, *34*(1), 254-284.

Windschitl, M., Thompson, J., & Braaten, M. (2020). Ambitious science teaching. Harvard Education Press.

Windschitl, M., & Thompson, J. (2013). The modeling toolkit: Making student thinking visible with public representations. The Science Teacher, 80(6), 63-69.

From: Christa Marie Haverly
To: NAGB Queries
Cc: Stefanie Marshall; Shakhnoza Kayumova; tcheuk@calpoly.edu; Vincent.Basile@colostate.edu; smcdonald@psu.edu; Dr. Jonte" C. Taylor (JT)
Subject: NAEP Science Framework
Date: Friday, October 15, 2021 10:51:07 AM

To Whom It May Concern:

We write to you as a collective to urge you to update the NAEP Science Assessment Framework, taking into consideration key points as described below. These recommendations account for the dynamic relationship between theories of learning and practice and how approaches to assessment become consequential to what is made (in)visible as knowledge in the classroom. Therefore, we urge NAEP to pay attention to the evidence that has emerged in equity-based scholarship that interrogates dominant ways of knowing in science education, towards recognizing and making visible the epistemological pluralisms that racially and linguistically diverse youth enact in classrooms.

Equitable science education is critical given the increasing racial, cultural, and linguistic diversity in our country; the potential for the fields of science to benefit from the varied perspectives and lived experiences of our current and future PK-12 populations; and the obligation of our country's education system to rigorously prepare all of our students to be scientifically literate. This obligation has become more stark as we watch citizens across our country reject wearing masks or receiving vaccines against COVID-19, actively denying wide scientific consensus of the importance of these measures to protect personal and public health. This obligation has also become more stark as we have watched Black, Indigenous, and other citizens of Color in this country fighting to be treated humanely, with dignity, and equitably, emphasized in the months following George Floyd's murder, but representative of centuries of struggle. Further, this obligation has become more stark as we have watched communities and species be decimated by increasingly harsh natural disasters and habitat loss caused by over a century of preventable and mitigatable changes to our climate.

Now more than ever we need a science education program that serves to broaden participation in the fields of science and consequently broaden the epistemological dimensions of the sciences themselves. We need a science education program that prepares our youth to make critical, life- and planet-saving decisions that are rooted in evidence, not conspiracies. It is therefore essential that the NAEP standards and assessments that measure outcomes of our work with students reflect the research and recommendations that we share with you here recognizing that teaching and learning practices are often shaped by assessments and accountability measures.

We offer four sets of recommendations:

1. Interrogate the assumptions about science knowledge embedded in the standards (i.e., whose histories and narratives are and are not included in this body of

knowledge and practices).

   a. For example, see Morales-Doyle, D., Childress Price, T., & Chappell, M. J. (2019). Chemicals are contaminants too: Teaching appreciation and critique of science in the era of Next Generation Science Standards (NGSS). *Science Education, 103*(6), 1347-1366, and
   b. Rodriguez, A. J. (2015). What about a dimension of engagement, equity, and diversity practices? A critique of the Next Generation Science Standards. *Journal of Research in Science Teaching*, *52*(7), 1031-1051.

2. Update the technical aspects of the assessments themselves to be more inclusive of historically marginalized student populations.
   a. Consider implications and limitations of administering the test solely in English (see work from Guillermo Solano-Flores, Alison Bailey, and Jamal Abedi)
   b. Fund the special studies on “innovative assessment tasks, testing special needs students, and computer adaptive testing” (p. 121 of current NAEP framework).
   c. Develop assessment tools that can guide teachers and researchers to critically examine whether or not the assessments they are using or developing are sensitive to the instruction and the diverse ways students' thinking and knowledge can be embodied and represented.

3. Invite people to participate in this review process, including on the expert panel, who are multilingual, of Color, differently abled, and so on; leverage their expertise and lived experiences; and provide them with authority and agency to make substantive changes to the program.

4. Seek recommendations from the National Academies' Committee on Equity in PreK-12 STEM Education, which will be announced in the coming months.

Sincerely,

Christa Haverly, Ph.D., Northwestern University
Stefanie Marshall, Ph.D., University of Minnesota- Twin Cities
Shakhnoza Kayumova, Ph.D., University of Massachusetts, Dartmouth
Tina Cheuk, Ph.D., California Polytechnic State University, San Luis Obispo
Vincent Basile, Ph.D., Colorado State University
Scott McDonald, Ph.D., Pennsylvania State University
Jonte’ C. Taylor, Ph.D., Pennsylvania State University

Comments to the NAEP Science Assessment Framework
Submitted by the National Science Teaching Association
October 14, 2021

Thank you for the opportunity to address the National Assessment of Educational Progress (NAEP) Science Assessment Framework. The National Science Teaching Association (NSTA) is the world's largest organization promoting excellence and innovation in science teaching and learning for all. We are committed to best practices in teaching science and its impact on student learning. NSTA offers high-quality science education resources and continuous opportunities for learning that help science educators grow professionally and excel in their career.

As requested, we have focused our response on these three questions:

- Does the NAEP Science Assessment Framework need to be updated?
- If the framework needs to be updated, why is a revision needed?
- What should a revision to the framework include (or exclude)?

Working with a group of practitioners from several NSTA standing committees, we have answered these questions through the lens of what science and engineering could look like in 10 years and how technology can and should support more complex and meaningful assessments that reflect how people have been documented to learn science.

Does the NAEP Science Assessment Framework need to be updated?

NSTA strongly believes the NAEP Science Assessment Framework must be updated.

The current framework is extremely outdated. It is antiquated regarding standards for science education and science education research and is predicated on standards that originated before 2005.

Currently, states, districts, and schools are focusing their science curricula and instructional programs on *The Framework for K–12 Science Education: Practices, Crosscutting Concepts, and Core Ideas,* released by the National Academies of Sciences, Engineering, and Medicine in 2012. Twenty states have adopted the *Next Generation Science Standards* (2013). As outlined in the recent report *Call to Action for Science Education: Building Opportunity for the Future,*

> "The *Framework* catalyzed an ongoing transformation of elementary and secondary science education across the United States. The *Framework* provides guidance for improving science education that builds on previous national standards for science education and reflects research-based advances in learning and teaching science. As of April 2020, 44 states and the District of Columbia had developed and adopted science standards that are informed by or

> directly based on the *Framework.* This represents approximately 70% of K–12 public school students. The vision for science education outlined in the *Framework* differs in important ways from how science has traditionally been taught. It emphasizes engaging students in using the tools and practices of science and engineering and providing them with opportunities to explore phenomena and problems that are relevant to them and to their communities."

In conclusion, we emphatically state that the current NAEP Science Framework is woefully outdated, designed for a specific purpose that has largely ceased to exist, and incompatible with contemporary science curricular frameworks.

If the NAEP Science Assessment Framework needs to be updated, why is a revision needed?

Science education in the United States is currently in a state of transition as we move to align classroom teaching practices with *A Framework for K–12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. Teachers are shifting toward helping students employ science and engineering practices rather than solely familiarizing them with scientific principles.

Currently the NAEP science framework has the following item distributions: Science Content, Science Practices, and Items by Type (interactive computer tasks, hands-on performance tasks, and specific question types).

The next NAEP science framework should reflect how we currently teach and project the development of science teaching over the next decade.

The current NAEP science assessment framework does not adequately reflect the computational thinking required for grasping complex scientific issues, as well as the use of large databases. The current framework does not support the explicit nature of science pedagogy, nor does it reflect the shift to three-dimensional thinking needed for sensemaking. Each of these are found in the National Research Council's (NRC) *Framework for K–12 Science Education*. Furthermore, science and engineering design thinking and application should be added to the NAEP framework.

Illustrative NAEP questions are too narrow in scope and tend toward the mere acquisition of principles and facts. For example, representative NAEP questions in 12th-grade physics focus on familiarity with gravitational force equations and relationships between variables, which tend to reward memorization. This type of question should be replaced with a broader assessment of a student's understanding of how gravitational fields can store and transfer energy.

The NAEP range of topics also seems very broad in nature and less in-depth, which results in rewarding memorization and familiarity with specific concepts, but not their application or extension. The NAEP framework should more accurately reflect the depth of learning and application that is now expected of students.

What should a revision to the framework include (or exclude)?

According to the *Call to Action for Science Education: Building Opportunity for the Future,*

> "Science assessments and accountability systems need to be aligned with the vision for high-quality science instruction. Assessing science learning in ways that are aligned to our vision will require approaches that go beyond single tests of factual knowledge. Traditional, large-scale, multiple-choice tests cannot capture the ability of students to engage in the practices of science and reason about evidence. An advantage of the new approach to science instruction is that it provides many opportunities for assessing learning informally (formative assessment) as students engage in investigations, create representations, and discuss evidence. However, designing useful and meaningful formal assessments such as tests will require careful articulation of the desired learning goals and how students can demonstrate that they have achieved them."

To genuinely be forward-looking, future science assessment based on the NRC *Framework* should capture a student's ability to behave like a scientist and to engage in scientific practices to deconstruct and make sense of a situation or phenomenon.

The revision should include the following:

- *Modeling as a practice.* Students should be asked to create, evaluate, and/or revise models, and use them to predict the result of changes to system components. The development of explanatory models can help students make their thinking visible and can be an equalizer for English Language Learners.
- *Planning investigations.* Students should be able to identify independent and dependent variables and to design scientifically valid investigations.
- *Analyzing data.* Students should be able to analyze complex, real-world data using graphing and graphing analysis tools.
- *Engaging in argument from evidence.* Students should be assessed on their ability to use evidence to construct and justify a scientific claim.

Each of these elements should be approached with a recognition that the science experiences of many students are not equitable, inclusive, or reflective of our expanding diversity as a nation.

It is important to note that the recent pandemic has facilitated the shift in science teaching that is unprecedented in its scope and duration. The use of simulations, along with hands-on experiential learning, is much more common than when the current NAEP science framework was adopted. Subsequently, the scope of science teaching has changed to better reflect three-dimensional sensemaking. As a result, the NAEP framework should be modified to include novel approaches that incorporate shifts in science practices that are observed. To this end, a revised NAEP Science Assessment Framework would increase validity by reflecting the shifts that form the foundation of students' sensemaking through the practices, inquiry, nature of science, science content, and crosscutting concepts.

In addition to these ideas, we offer some specific suggestions for changes to the current NAEP science framework:

While the Science and Engineering Practices and the Disciplinary Core Ideas expressed in *NGSS* are evident in the framework, the Crosscutting Concepts need to be more explicitly represented. Hence, summary charts should be included to reflect the current three-dimensional sensemaking supported by the nature of science. Less emphasis should be placed on identifying science principles, and more emphasis should be placed on higher order of reasoning skills. However, the current sample questions focus more on rote knowledge and do not give students opportunities to demonstrate the application of that knowledge to novel situations.

Scientific and Engineering Practices, rather than principles, should be reflected. Science Practices should be expanded to include analyzing and interpreting data; using mathematics and computational thinking; constructing explanations (for science) and designing solutions (for engineering); engaging in argument from evidence; and obtaining, evaluating, and communicating information. When these practices are added, students should be able to demonstrate their science literacy based on performance expectations.

In conclusion, it can be said that the value of any assessment is rooted in the purpose for which it is intended. If one purpose of NAEP is to provide a longitudinal trajectory of how American students are learning science across their compulsory education, then its science assessment framework must reflect the dramatic shifts in the mode of instruction, as well as the curricula upon which that instruction is based.

This statement has been endorsed by the Council of State Science Supervisors and the National Science Education Leadership Association.

October 15, 2021

Dear National Assessment Governing Board,

I am writing to communicate my professional perspectives in response to requests for commentary about the NAEP Science Frameworks. As background, I am a professor of STEM teacher education at the University of Connecticut and Co-Editor of the journal Science Education. As I examine the 2019 Science Assessment Framework document, several aspects caused great concern. Especially given the unique times in which we find ourselves, I want to earnestly communicate the need for major shifts to the NAEP Science Frameworks. In their current condition, I found few positive advances over previous iterations. Given the sea changes in society, and in light of considerable research gains in the learning science, school leadership, and instructional delivery, without dramatic improvements to the NAEP Science Framework, we will miss an opportunity to respond to contemporary challenges. Any efforts to maintain the status quo with the NAEP Science assessment will effectively neglect this unique chance to make positive changes to K-12 science education throughout our nation. Below are several concerns which need your attention:

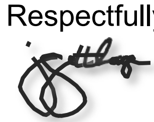
A. <u>Perils of Supporting Deficit Explanations via NAEP Science Results</u>. Even with the Coleman Report clearly demonstrating racial differences in student performance were much stronger within rather than between schools, NAEP continues its pattern of feeding information to the contrary. Decision-making purported to inform policy and practice to support school is overshadowed by data "gaps" that compares states and school urbanicity. For those who accept inequities as challenge worth resolving, the unit of change is known to be at the school level. Responses to questions about WHY science performance gaps exist are greatly influenced by HOW such data are collected and reported. I would submit that NCLB data powerfully influenced achievement gap discourses simply by disaggregating school level data. Seeing disparities in outcomes within specific schools and communities made it much harder ignore the reality the inequities lurk within the places where we send our children and for which we pay taxes. Rather than support deliberations about the presence of science achievement gaps as artifacts of institutional and organizational factors – with an eye toward remedying those disparities – NAEP data will instead perpetuate beliefs about gap inevitability and progress toward closing those gaps is only likely as scores by White students come. Absent from the design is information that might indicate how non-White student performance could be improved. More than recognizing complicity with fostering such narratives, I would submit that NAEP should proactively develop data reporting approaches that could redirect media, political, and layperson discussions in ways that disrupt widespread beliefs that demographics dictate destinies.

B. <u>Supporting Equity and Diversity Research in Science Education</u>. Although the framework expresses the ambition of collecting data suitable for informing policymaking and support secondary forms of research, to date there has been very little research about the results from NAEP Science. We can attribute this to shortcomings of the data collection – weaknesses which have frustrated those of us who would like to do this research. For example, the intersection of student gender, race, and social class are very relevant to building better understandings of science achievement. NAEP Science data has the potential to advance understandings of a variety of equity concerns (and to in turn shape instructional practices) only if more thought is given to making such data available. NAEP's own report cards reduce "Score Gaps" to singular designations without revealing whether Black females and Black males perform similarly. OR similarly multidimension features for NSLP eligibility, English learner status, etc. While some might suggest such analyses are possible (via special access to data), that approach has not proven to be fruitful. There are few to no examples within the demonstration material for NAEP Data Explorer. But the absence of such secondary research for the NAEP Science cannot be blamed on the research community. Instead, the NAEP system itself is not supportive of those types of studies – despite expressed claims that secondary research studies are a goal.

What I hoped to communicate in this letter is the immense potential for NAEP to shape, inform, and improve science education with a potentially national scope. My frustrations are rooted in the fear that such possibilities will be missed. As a consequence, not only would potential advancements be lost but also the likelihood that outdated perceptions of school science would be perpetuated by dubious information. In addition to the concerns about marginalizing equity as expressed above, I am deeply troubled by how outdated the resources are the are being used to shape the NAEP Science Frameworks. Included in this list is the absence of research published with the past ten years, the failure to acknowledge the substance of NGSS, and even the presence of retired and deceased members on your various committees. In some respects, I would advocate that the NAEP Science Frameworks begin with fresh people and perspectives rather than continue moving forward with such a dilapidated foundation. There are admittedly many dimensions of the NAEP Framework process that I cannot fully appreciate. On the other hand, as a research journal editor and participant in national communities of science education research, I can only hope that the NAGP will recognize the real possibility of missing a vital opportunity to improve science education by continuing with the current strategies.

In closing, the NAEP Science Assessment Framework is in profound need of updating. The materials used as the basis for this framework are outdated and fail to make effective use of contemporary understandings of science teaching and learning. Further, the framework's updating must attend to the shifting demographics of America's schools. More than acknowledge the existence of students who are traditionally marginalizing from science learning opportunities as consequences of their race, social class, English fluency, disabilities, gender, and immigration status, such awareness must accompany a strong centering of equity as a singular goal – in the design of the assessments, the structure of the data collection, and the release and reporting of results. Otherwise, it seems inevitable that the status quo procedures will further reify discriminatory assumptions and actions as by-products of the subsequent Science Report Cards.

Respectfully submitted,



John Settlage, Professor
University of Connecticut

From:	Renee Schwartz
To:	NAGB Queries
Subject:	NAEP SCIENCE FRAMEWORK
Date:	Friday, October 15, 2021 6:42:04 PM

Comments re the NAEP Science Framework revision:

The Board of Directors and Executive Committee of NARST [A global organization for improving science education through research] submit the following suggestions regarding the upcoming revision of the NAEP Science Framework:

The NAEP science framework faces a precarious challenge: standardizing the instrument across time to identify longitudinal patterns while accommodating changes in science education. The document thoughtfully addressed the tensions created by these competing goods. Even though some aspects of the framework reflect more current reform in science education (e.g., crosscutting concepts), it is difficult to ascertain the extent to which the NAEP science framework aligns with the more recent emphases put forth by the Framework for K-12 Science Education (NRC,2012) and the Next Generation Science Standards [NGSS]. There are notable differences between how the current NAEP framework and the NGSS define, focus, and recommend science concepts and science and engineering practices. A misalignment may prove problematic when using NAEP science achievement data to better inform decisions in policy and practice. It would be more advantageous for the advancement of K-12 science learning if more items corresponding with current science education reform are developed and included in the forthcoming assessment.

On one hand, the importance of context and its role in learning were primarily absent in the framework. Examples of prospective assessment items were abstract. On the other hand, in the cases in which concepts were embedded in context, the contexts (e.g., hares in state park) featured the lived experiences of dominant groups in U.S. society (e.g., upper middle class). It seemed the science framework did not incorporate decades of sociocultural research on cultural responsiveness and inclusivity in learning and assessment. Additionally, while noting the framework spoke to the need to consider the language demands of test items for English language learners, there were no explicit actions related to considerations of item development responsive to language. Indeed, the sample items shared were laden with dense language and vocabulary, particularly in context-driven items.

Because of the prevalent inequities in the quality of science education in K-12 education, it would be very useful for NAEP to develop equity indicators with respect to achievement and school and community factors, like those used in international assessments. Intentional attention to equity and social justice within science curriculum and instruction are essential for developing scientific literacy.

Sincerely,
Renee' Schwartz, President of NARST
Eileen Parsons, Immediate Past President of NARST
Gillian Roehrig, President-Elect of NARST
Jerome Shaw, Secretary/Treasurer of NARST
Lisa Martin-Hansen, Executive Director of NARST
NARST Board of Directors:
Scott McDonald, Leon Walls, Noemi Waight, Christina Schwarz, Malcolm Butler, Theila Smith, Bhaskar Upadhyay, Knut Neumann, Brooke Whitworth, Sonya Martin
Troy Sadler and Felicia Moore Mensah: Editors of the Journal of Research in Science Teaching
Michael Bowan: NARST Liaison to NSTA
Cynthia Crockett: NSTA Liaison to NARST

Renee' Schwartz, PhD

Professor, Science Education

Georgia State University

President *NARST: A global organization for improving science education through research* [narst.org]

Program Coordinator: *PhD Teaching and Learning, Science Education,* Georgia State University

Department of Middle and Secondary Education

College of Education and Human Development

Office: CEHD, 30 Pryor St. #629

NSELA Response to: Seeking Initial Public Comment Prior to Updating the Science Assessment Framework for the 2028 National Assessment of Educational Progress (NAEP)

The National Science Education Leadership Association (NSELA) is an organization of approximately 600 members in science leadership roles either at the school, district, university, informal science, or state level. Our mission is to catalyze leadership to maximize effective science teaching and learning in a complex and changing environment. We connect and support emerging and experienced leaders by providing high-quality professional development, a collegial network, access to research and resources, and a voice for leaders in science education. As requested by the National Assessment Governing Board, our members have provided feedback to address three questions about the current NAEP Science Assessment Framework:

- Does the NAEP Science Assessment Framework need to be revised?
- If the NAEP Science Assessment Framework needs to be revised, why is a revision needed?
- What should a revision to the NAEP Science Assessment Framework include?

NSELA recommends that yes, the NAEP framework does need to be revised. There have been many new findings from research in science education since the writing of the last NAEP Science Assessment Framework in 2005. The publication *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas* (2012) includes more current research in science education than does *The National Science Standards* (1996) with which the 2005 NAEP Framework is aligned.

The current NAEP Science Assessment Framework is heavily focused on science content knowledge rather than the integration of science content with crosscutting concepts and science and engineering practices. With 44 states having revised their science standards to align with *A Framework for K-12 Science Education,* including 20 states that have adopted the Next Generation Science Standards, the 2005 NAEP Science Assessment Framework does not accurately assess what today's science students know, understand, can do and apply. This creates a misalignment in what is being assessed on the NAEP science assessment and the current research and best practices for students. Although the 2019 NAEP report is very comprehensive and recognizes how science can change, it is still based on antiquated science education research with the intent to create a snapshot of what is being taught in American schools. The following proposed changes will better align the NAEP Assessment Framework with current science education research and practices.

Rather than aligning science content with *The National Science Standards* (1996) and *Benchmarks for Science Literacy* (1993), align the content with the recommendations of *A Framework for K-12 Science Education* (2012) and the *Next Generation Science Standards* (2013). In developing performance expectations and performance assessment items, consider merging not only science content with science practices, but also integrating crosscutting concepts, as recommended in *A Framework for K-12 Science Education*. This change would create a need for a section of the NAEP Science Assessment Framework that focuses on the Crosscutting Concepts to be assessed.

For the Science Content section of the NAEP Science Assessment Framework, consider focusing less on nuggets of knowledge and more on application of that knowledge to make sense of phenomena. To better align with the recommendations of *A Framework for K-12 Science Education* and the *Next Generation Science Standards*, consider aligning the content section of the NAEP Framework with the disciplinary core ideas within these documents.

For the Science Practices section of the NAEP Science Assessment Framework, rather than using the former broad science practices "identifying science principles, using science principles, using science inquiry, and using technological design", instead use some of the science and engineering practices listed within *A Framework for K-12 Science Education* and the *Next Generation Science Standards.* Possible science practices to be assessed might include: Developing and Using Models, Planning and Carrying Out Investigations, Analyzing and Interpreting Data, Using Mathematics and Computational Thinking, and Engaging in Argument from Evidence. The focus should be on using the science and engineering practices to determine whether students can "do" science.

The Assessment Design section of the NAEP Science Assessment Framework needs to be updated to include performance expectations where science content, crosscutting concepts, and science and engineering practices intersect. Assessing all three dimensions (content, concepts, and practices) will require a greater number of performance-type assessment items, either hands-on or computer simulation-based, where students might use multiple data sources to construct reasonable explanations, analyze data, develop scientific arguments, or develop conclusions. Give students a scenario to make sense of that they may actually see in their lives. Look for a development of student thinking to make sense of the scenario - consider multiple questions around this scenario to scaffold and get at student ability to work and think like a scientist.

For the Science Achievement Level Descriptors section of the NAEP Science Assessment Framework, the descriptors need to align with the changes in content strands recommended in *A Framework for K-12 Science Education.* Use the *Next Generation Science Standards* to review appropriate descriptors. Use the grade band endpoints given for 6-8 and 9-12 as no matter what content sequence may be utilized within a state, by the end of grade 8 and 12 all students should have learned the content being assessed.

For the section of the NAEP Science Assessment Framework focused on English Language Learners and Students with Disabilities, first consider changing the term ELLs to Multilingual Students as is more widely utilized today. Ensure grade appropriate language is utilized to assess student proficiency of grade level standards. Provide the opportunity for the test to be read aloud as an option for any child who takes the NAEP to ensure we are offering a level playing field and reading does not hinder the ability to respond. Align NAEP assessment modifications or accommodations with those that are utilized by states across the country.

The purpose of the NSELA recommendations is to better align the NAEP assessment with the current expectations for student learning within science classrooms across the country. Having relevant, meaningful assessment data is important to science education leaders. Aligning the NAEP Science Assessment Framework with current science education research and practice will result in a NAEP assessment that more accurately measures student understanding and application of science.

LEARNING POLICY INSTITUTE
Research. Action. Impact.

To Members of the National Assessment Governing Board:

Thank you for the opportunity to provide feedback regarding the development of the next framework for the National Assessment of Educational Progress in Science. This document shares feedback collected by the State Performance Assessment Learning Community (SPA-LC) from science education communities across the nation in response to the three questions posed by NAGB:

1. Whether the NAEP Science Assessment Framework needs to be updated.
2. If the framework needs to be updated, why a revision is needed.
3. What should a revision to the framework include?

SPA-LC, coordinated by the Learning Policy Institute, represents over 25 states and 10 national partners committed to the development and implementation of meaningful and balanced assessment systems, beginning with science. SPA-LC's members include state commissioners, curriculum and instruction directors, assessment directors, and science leadership within state education agencies as well as local communities. Together, SPA-LC supports within- and cross-state efforts to develop meaningful assessment systems in science through support for better instruments, effective capacity building, and meaningful policies. As such, we find ourselves distinctively positioned to offer relevant input regarding the country's distinguished assessment of scientific learning.

A careful review of the current NAEP science framework and progress in science education--including state standards, foundational research, contextual and environmental shifts, and recent advances in science teaching, learning, and assessment practice was completed by convening three focus groups and collecting information via survey. As a result of this review, **SPA-LC recommends that the NAEP Science Framework be updated in targeted ways to better reflect both the current state of science education across the country, as well as the direction in which we expect science education efforts to shift in the next decade. Specifically, we recommend:**

1. **Update the content and practice included in the framework to align with the most recent research on what students should know and be able to do.**
2. **Prioritize integration of content, practice, and conceptual elements in service of sense-making.**
3. **Ensure the NAEP science framework supports and addresses needed advances in assessment design and use.**

LEARNING POLICY INSTITUTE
Research. Action. Impact.

Below, we outline key shifts that should be addressed in the next science framework. The SPA-LC community stands ready to support any efforts to make these and other needed shifts to ensure that NAEP remains a relevant cornerstone of science assessment systems nationwide.

The need for an update.

According to the National Assessment Governing Board, NAEP frameworks are updated for modern expectations for students and to "address recent standards, curricula, and instruction, research on cognitive development, and the latest perspectives on what students should know and be able to do" (NAGB, 2021). Since the last substantial review of the NAEP science framework, there have been sufficient shifts in science education research and practice to recommend a review and revision of that framework.

Advances in research on how students learn and demonstrate science understanding and practice. Since the NAGB last made substantial changes to the NAEP science framework, the following developments in science education and assessment have initiated a great deal of adaptation in the field:

- Release of the publications How Students Learn Science in the Classroom and Taking Science To school, which together began to push the community to think, "beyond the artificial dichotomy between content and process in science" (TSTS, p viii)
- Development, publication and release of "A Framework for K-12 Science Education: Practice, Crosscutting Concepts, and Core Ideas."
- Supporting cognitive research such as How People Learn II: Learners, Contexts, and Cultures (2018) provide further input regarding integration of content and practice for improved and more equitable outcomes.
- Assessments begin to use sensemaking and cognitive complexity models that incorporate multi-dimensional analysis of student interaction with phenomena such as those illustrated in "A Framework to Evaluate Cognitive Complexity in Science Assessments."
- Substantial efforts to support research-based instructional models that prioritize students' active engagement in phenomena and sense-making ("figuring out") as the mechanism for science teaching, learning, and assessment. This includes materials themselves (e.g., OpenSciEd, inquiryHub, Multiple Literacies Project Based Learning, etc) as well as within criteria for high quality materials (EQuIP, EdReports) and assessment (e.g., Science task screeners, Task Annotation Project in Science, New Meridian Science Assessment Framework, Harris et. al. work to focus assessments on knowledge-in-use)

Substantial shifts in the science standards landscape. The most recent versions of the NAEP science framework have largely attended to and reflected the 1996 National Science Education Standards (NSES). While these standards provided a strong foundation for science education and assessment, the release of *A Framework for K-12 Science Education* led to the development and widespread adoption of new standards such as the Next Generation Science Standards (NGSS) and other, similar standards. These standards, currently adopted in over 45 states and the District of Columbia, reflect key conceptual shifts in standards, teaching, learning, and assessment. Given the widespread use of new standards, a review and revision of blueprint content/practice alignment may be warranted to ensure that what is tested by NAEP is reflective of what students are given the opportunity to learn in their classrooms .

Advances in equitable science assessment design and implementation. As states, districts, and teachers have worked to implement new science standards, there has been a call to redesign science assessments such that they 1) better reflect what we expect students to understand and be able to do in science, and 2) attend to equity in assessment in ways that move beyond traditional conceptions of bias and sensitivity. This includes:

- Centering sense-making and knowledge-in-use as essential elements of aligned science assessment items and tasks
- Leveraging advances in simulations, item sets/clusters, scoring algorithms, and test design to better approximate performance-based tasks and approaches that more authentically represent science learning and mastery
- Attending to features of equity within assessment design and use, including racial equity; culturally responsive assessment practices; and attending to student interest, identity, and agency within assessment design.

Many of these advancements reflect both a desire to develop more valid assessment instruments and reports as well as an effort to ensure that assessments are coherent with instructional and professional learning components of the science educational system. It will be important that the NAEP science framework attend to these shifts in assessment understanding, design, and practices to produce assessment results that both represent the state of science learning in the country as well as serve to lead the way for assessment work of the future.

LEARNING POLICY INSTITUTE
Research. Action. Impact.

What revisions should include.

While there is endless nuance and details that could be addressed, SPA-LC makes three central recommendations for revisions to the NAEP science framework:

1. **Update the content and practice included in the framework to align with the most recent research on what students should know and be able to do.**
2. **Prioritize integration of content, practice, and conceptual elements in service of sense-making.**
3. **Ensure the NAEP science framework supports and addresses needed advances in assessment design and use.**

Recommendation 1: Update the content and practice included in the framework to align with the most recent research on what students should know and be able to do.

Rationale. As described above, science teaching, learning, and assessment have been deeply influenced by *A Framework for K-12 Science Education* and the shifts represented by new standards based on it (e.g., NGSS). Recent analyses of content alignment between current state standards and the NAEP science framework have found substantial differences, including differences in targeted science ideas and how scientific practice is represented. For example, A Comparison Between the Next Generation Science Standards (NGSS) and the National Assessment of Educational Progress (NAEP) Frameworks in Science, Technology and Engineering Literacy and Mathematics (Neidorf et. al., 2016) found:

- At grades 3-5 only 38% of performance expectations were aligned to the [NAEP] Science framework, with 44% alignment at both middle and high school.
- Considering only grade 4 NGSS performance expectations for the grade 4 NAEP 36% of performance expectations were aligned.
- Across all grades the highest degree of alignment was in life sciences (from 48-54%) with the lowest degree of alignment in physical science (29-42%)

Additionally, the existing overlap between the NGSS practices and the practices outlined in the current NAEP framework provides a strong foundation for a meaningful framework and related assessment. A revision to the framework provides an opportunity to consider how the practices are represented in ways that are coherent with other science education efforts. Questions to consider include:

- In what ways can the practices be better integrated as an essential part of sense-making--either through making sense of phenomena or designing solutions to problems?
- In what ways should the existing practices be clustered to both reflect and complement how the practices are used together in instruction and assessment nationally?

- Are the measures used to assess scientific practice in alignment with the goals of science educational practice across the country?

With 20 states (and the District of Columbia and Department of Defense Educational Agency) aligned directly with NGSS and 24 states aligned with the *Framework for K-12 Science Education*, it may be appropriate to revise the NAEP science framework to better align with current state activities. This will ensure NAEP is able to appropriately monitor science learning across states and over time, remaining a vital element of our understanding of how science education is progressing.

Recommendation 2: Prioritize integration of content, practice, and conceptual elements in service of sense-making.

Rationale. According to the *Framework for K-12 Science Education* (p. 218; emphasis added), "Standards and performance expectations that are aligned to the framework must take into account that **students cannot fully understand scientific and engineering ideas without engaging in the practices of inquiry and the discourses by which such ideas are developed and refined [1-3]. At the same time, they cannot learn or show competence in practices except in the context of specific content.**" Research suggests that surfacing student understanding and ability in science requires that they are able to show both the depth of their conceptual understanding of science ideas as well as their ability to engage in scientific practice together. Recent work focused on how to assess student mastery of widely adopted science standards requires a shift toward assessments that ask students to actually engage in using science ideas and practice together in service of sense-making; conversely, assessing students for understanding outside of the context of the integration of content and practice would provide incomplete-- and potentially even inaccurate--information about true student facility with science expectations.

While the current NAEP Science framework and associated assessment specify and assess important aspects of science content and science practice, these are often done separately. Moving forward, it may be appropriate to consider more intentional integration of science core ideas, practices, and crosscutting concepts in both framework and assessment design.

Recommendation 3: Ensure the NAEP science framework supports and addresses advances in assessment design and use.

A primary way the NAEP science framework influences the national science education community is through the NAEP science assessment, which has had a long history of setting the standard for high-quality assessment design in science. For the NAEP science assessment to continue to be both immediately compelling and forward-leading, it will be important for NAGB to consider how revisions to the science framework are accompanied by revisions to the assessment, including:

- Items and forms that can appropriately engage sense-making at the nexus of multiple dimensions, including effective use of performance tasks and technology enhanced items and scoring paradigms.
- Ensure proper alignment to updated framework goals.
- Develop tasks that center making sense of appropriate and compelling phenomena as their foundational basis.
- Attend to advances in equitable assessment that include and expand beyond attention to bias and sensitivity considerations.
- Consider alternative cognitive complexity models to address multidimensionality of items and item sets.

As a measure of educational trends, the NAEP assessment would need to address continuity across tests, requiring innovation in terms of equating and development of linking items from form to form. While this may be a complex undertaking, it is not impossible, and given the large-scale, non-accountability model of the NAEP assessment, the creative use of matrix blocks to achieve the desired outcomes may offer a useful solution.

Conclusion.

The NAEP science framework, and associated assessment, are strong components of current science assessment systems. With key revisions, they stand to continue shining a light on how we can continue supporting effective and meaningful science learning for all students. We stand ready to assist NAEP and the National Assessment Governing Board in support of this effort.

Warm regards,

Aneesha Badrinarayan, Senior Advisor, Learning Policy Institute

on behalf of the State Performance Assessment Learning Community.

Seeking Initial Public Comment Prior to Updating the Science Assessment Framework for the 2028 National Assessment of Educational Progress (NAEP)

Deadline Extended to Oct. 15

Comments should specifically address three things:

Comments must be submitted via email to nagb@ed.gov with the email subject header NAEP Science Framework no later than 5:00 p.m. Eastern Time on Friday, October 15.

1. **Does the NAEP Science Assessment Framework needs to be updated?**
2. **If the framework needs to be updated, why a revision is needed?**

In general, no. The principles and frame work are sound, stressing empirical knowledge and testing. As is appropriate with a general framework, discussion of scientifically disputed or politically charged issues such as anthropogenic climate change or embryonic stem cell research are avoided.

However, given the current political and educational climate, this may change. If it does and climate change becomes a specific focus of discussion in the framework, below we offer a few suggestions to provide a balanced discussion of theories of climate change, and an accurate assessment of climate data versus model projections.

3. **What should a revision to the framework include?**

Any discussion of Climate Change within the framework should be focused on helping students learn how to think through the issue and weigh different types of information. For example, any climate-specific material should teach students the difference between verified objective observations and data versus predictive models.

Regarding specific components of the climate change issue, any climate-specific framework should include:

1. The theory that anthropogenic greenhouse gas emissions are causing catastrophic changes to the climate is not settled science, and this should be acknowledged.
2. Science does not proceed by consensus (which is a political term tantamount to vote counting) but rather be experimentation and discovery, grounded in verifiable data, and independent testing.
3. Myriad factors, many only poorly understood, drive climate changes over the short, medium, and long-terms.
4. Climate model projections of temperature fail to accurately mimic actual temperatures and temperature trends as measured by ground-based weather stations, global satellites, and weather balloons.
5. Projections of climate change impacts are driven by computer model simulations of temperature responses to greenhouse gases and speculative assumptions about climate feedback mechanisms. Simple models that don't include feedback mechanisms better track actual temperature measurements and project less warming with each additional unit of carbon dioxide.
6. Statements regarding worsening weather conditions should note that there have been few if any observed worsening global trends for extreme weather despite decades of speculation that such worsening is imminent. Objective data and measurements show each of these weather phenomenon are well within the range of natural historic variation and most types of extreme weather events show no recent change or a trend of less frequency and severity.

7. Additional carbon dioxide in the atmosphere has contributed to a substantial greening of the earth and record crop production, which has resulted in declining rates of starvation and hunger.
8. Cold conditions result in more premature deaths each year than warm conditions. As the Earth has warmed modesty, the number of deaths attributable to extreme temperatures has substantially declined.

Specific issues in the current Text:

On Pg. 42 (62 incl. preface) box under life sciences should state, "Plants also require light and carbon dioxide to grow."

Pg. 54/55 (74/75) mentions climate, but doesn't discuss the difference between weather and climate. Climate changes aren't measured or determined over the short term of just a few years, but rather over 30-year periods. Modest changes between periods don't signal climate change for a region, only substantial changes do.

Pg. 61/62 (81/82) Boxes discussing changes in earth system and biogeochemical cycle are accurate.

If climate change is discussed in the updated NAEP assessment, it should note the long-term decline in carbon dioxide in the atmosphere prior to the Industrial Revolution. Most plants evolved before the long-term decline began, when carbon dioxide levels were considerably higher than today. It would also note that if carbon dioxide levels dip below 150 ppm, plants can't photosynthesize and begin to die. The Earth came perilously close to that prior to the Industrial Revolution.

Avoid controversial and overly politicized topics related to energy systems, but if it is discussed ensure that students are provided a balanced view of the virtues and drawback of each source of energy generation. All forms of energy have environmental impacts.

Possible Design experiments:

Set up three plants (sets of plants) in greenhouse-like conditions, one with ambient carbon dioxide levels, a second with elevated carbon dioxide, a third with even more elevated carbon dioxide. Study growth rates, mass, fruiting, etc…

Use GIS system to map the greening of the earth.

Pg. 117 (137) Hands-on-Performance vs. Interactive Computer Investigations

Make clear that computer model simulations are only as good as the assumptions built into them. The more complex the phenomenon to be simulated and the farther out in time projections are made, history, research, and data show the less accurate the model simulations are. For climate, many of the factors or forcing mechanisms that impact climate are only poorly understood, and thus attempts by modelers to mathematically capture them are very speculative and error-prone. In the end, when models are run, their outputs should be compared to hard data for phenomenon for which data is available, and if the data and the model outputs conflict, the model outputs are not to be trusted and either the model must be adjusted, or the hypothesis reexamined.

Suggested reading material or supplementary classroom material:

Short pieces or Monographs:

Craig Idso, et al., “Why Scientists Disagree About Global Warming,” Non-Governmental International Panel on Climate Change, 2015; https://www.heartland.org/publications-resources/publications/why-scientists-disagree-about-global-warming

Anthony Watts and James Taylor, "Climate at a Glance: Facts for Climate Realists," The Heartland Institute, 2021, **(insert link here)**

A Global Warming Primer, H. Sterling Burnett (ed), The National Center for Policy Analysis, 2013, http://www.ncpathinktank.org/pdfs/Global-Warming-Primer-updated-reduced-size.pdf

Book Length Discussions:

Gregory Wrightstone, **Inconvenient Facts: The science that Al Gore doesn't want you to know** (Silver Crown Productions, 2017); for purchase on Amazon: https://www.amazon.com/Inconvenient-Facts-science-that-doesnt/dp/1545614105

Bjorn Lomborg, **False Alarm: How Climate Change Panic Costs Us Trillions, Hurts the Poor, and Fails to Fix the Planet** (Basic Books, 2020); for purchase on Amazon: https://www.amazon.com/False-Alarm-Climate-Change-Trillions/dp/1541647467/ref=pd_lpo_3?pd_rd_i=1541647467&psc=1

Steven E. Koonin, **Unsettled: What Climate Science Tells Us, What It Doesn't, and Why It Matters** (BenBella Books, 2021); for purchase on Amazon: https://www.amazon.com/Unsettled-Climate-Science-Doesnt-Matters/dp/1950665798/ref=pd_bxgy_img_2/140-1238615-9822725?pd_rd_w=E89Hq&pf_rd_p=c64372fa-c41c-422e-990d-9e034f73989b&pf_rd_r=G36RP2E13RENSEN00W4W&pd_rd_r=81f9f61d-5348-4d8d-a548-46774737b653&pd_rd_wg=K9EBl&pd_rd_i=1950665798&psc=1

From: Susan Codere
To: NAGB Queries
Subject: NAEP Science Framework
Date: Wednesday, August 18, 2021 12:33:13 PM

CAUTION: This email originated from outside of the organization. Do not click links or open attachments unless you recognize the sender and know the content is safe.

Dear NAGB Science Framework Committee,
Please accept my comments regarding

Solicitation of Public Comments for Updating the Science Assessment Framework for the 2028 National Assessment of Educational Progress (NAEP)

As requested, my comments specifically address:
(a) Whether the 2019 NAEP Science Framework needs to be updated and (b) if the framework needs to be updated, why a revision is needed.
Comment - Yes, the NAEP Science Framework needs to be revised. The current NAEP Science Framework was developed before T*he Framework for K-12 Science Education* and the *Next Generation Science Standards* were completed, and thus does not reflect the focus of the most recent standards considered as the current 'national level' standards guidance documents in the US K-12 system.
and
(c) what should a revision to the framework include?
Comment - The revision should include a restructuring to place value on all 3 dimensions of science learning -- Science and Engineering Practices, Disciplinary Core Ideas, and Crosscutting Concepts in an integrated way and NOT as individual constructs and should not focus on technology applications.
The National Academies Board on Science Education has conducted numerous study sessions and produced publications to guide science assessment. This guidance should be reflected in the new NAEP Science Framework.

Thank you for this opportunity to provide public comment.

Susan Codere
ML-PBL Project Director
ML-PBL website
https://mlpbl.open3d.science/

From:	Tom Keller
To:	NAGB Queries
Subject:	NAEP Science Framework
Date:	Monday, September 13, 2021 10:33:34 AM

Thank you for this initial opportunity to provide comments and recommendations regarding the updating the Science Assessment Framework for the 2028 National Assessment of Educational Progress.

I have been active in science education at the state and national level for thirty years, as a classroom teacher, school leader, state science supervisor in Maine and senior program officer at the National Academy of Sciences. While at the National Academy, I co-directed development of the *Framework for K-12 Science Education,* with a committee of 18 scientists, engineers, educational researchers, cognitive scientists and educational practitioners, including 2 Nobel laureates.

This document is the most recent record of current research on science education, and makes some important advances that are being implemented across the country.

For this reason alone, the NAEP Science Education Framework must be reviewed and updated. The last NAEP Framework was completed prior to the findings listed in the National Academy of Sciences' Framework. The major step forward described in the National Academy's Framework is the melding of science and engineering practices, crosscutting standards and disciplinary core ideas as the fundamental unit of instruction. Separating these three dimensions reverts to past thinking on process versus content.

It is vital that the review of the NAEP Framework include significant participation by members of the Council of State Science Supervisors. As science education leaders working at the intersection of local, state, and federal policies in each state and jurisdiction, they are most aware of the systemic value of coherence between state and federal assessment and have the ability to facilitate such coherence. Assessment tends to drive instruction and it can drive us forward or backward. Coherence between state and federal assessment will provide state leaders with another tool to improve science instruction for all students.

The Council of State Science Supervisors played an outsized role in gathering and collating feedback for the 2005 NAEP Framework. I am sure that they would be happy to once again work with the Framework committee to collect meaningful feedback that represents the nation.

Relative to the three questions posed by the NAGB communication:

- *Whether the NAEP Science Assessment Framework needs to be updated.*

Clearly the NAEP Framework requires updating. The last updating was done in 2005 and this was prior to both the National Academy of Sciences' Framework and other seminal science education consensus studies reported by the Academy.

The National Academy's *Framework for K-12 Science Education* cites the need and power of instructing students in science and engineering practices, crosscutting concepts and disciplinary core ideas as a whole rather than separating science into content and practices as does the current NAEP Framework. This is a major difference for which the current NAEP Framework looks back and the National Academy's Framework looks forward.

- *If the framework needs to be updated, why a revision is needed.*

The current NAEP Framework has two separate components, science content and science practices. This leads to teaching them separately. And we know assessment tends to drive instruction. Many older textbooks have a first chapter on 'the scientific method' and never return to that topic. Science and engineering practices, a much better conceptualization of 'the scientific method', should be experienced repeatedly and the skills to do so should be constantly improved.

Also consistency between the NAEP Framework and what and how science and engineering are taught in schools, most of whom are using standards influenced by the Academy's Framework also makes the case for a revision.

- *What should a revision to the framework include?*

An important consideration is to know how the results will be used. If this truly is the Nation's Report Card and is not intended for any use by states, that brings up a different set of considerations. But if it is to be taken seriously by states, there has to be some value in it for them. So aligning as much as possible to the current science educational frameworks in use – and for most, that is the National Academy of Sciences' Framework, makes the results useful.

It is important that input of state science education leaders who work in this area daily be included in a revision.

Certainly a revision must include the three dimensions described in the National Academy's Framework. NAEP has the capacity to create assessment scenarios and bundles that assess these dimensions in an authentic and reliable way.

In summary, a revision to the NAEP Framework is necessary and I am willing to assist in any process to make that a reality.

Thank you.

Tom Keller

--
Thomas E. Keller, Ed. D.

Founder & Director

STEM Education Strategies LLC
208 N Newcastle Rd
Newcastle, ME 04553

www.stemeducationstrategies.org

Dr. Thomas R. Tretter
Professor of Science Education
Director Center for Research in Mathematics and Science Teacher Development
Director Gheens Science Hall & Rauch Planetarium
University of Louisville

Page xii (executive summary) and throughout document uses the label "***Science Practices***" in a way not completely aligned with NGSS "Science and Engineering Practices" – recommend updating these to the NGSS practices (8 of them, instead of 4) which also part of NGSS vision for the practices to cross science content and "…generate student performance expectations, and assessment items can then be developed based on these performance expectations". NOTE: this implies that all of chapter 3 will need to be revised.

Page xii (executive summary) and throughout. Need to incorporate the ***third dimension of NGSS*** as well – crosscutting concepts. These 3 dimensions (content, practices, crosscutting) then are used to generate performance expectations (detailed in NGSS) which can guide development of assessment items that measure all 3 dimensions. NOTE: May need to add an additional chapter focused on crosscutting concepts (parallel to science practices) OR add this as a primary new section in the updated "science practices and crosscutting concepts"

Page xii (executive summary) "***distribution of items***" needs to be reconsidered in light of NGSS. Both in terms of content emphases (or not) at each grade band, and if any NGSS practices should be emphasized or not.

Page xiIi (executive summary). Consider expanding the formats/types of ***interactive computer tasks***; see examples of what various states are doing in their science assessments. For example, building/modifying scientific models (different from existing 'empirical investigation' or 'simulation'). Also consider making interactive computer tasks a standard part of the assessment for all testtakers rather than a subset, given the widespread availability of computers and/or internet access (especially post-COVID pandemic when school systems across the world had to figure out how to instruct online – and make those resources accessible to all students).

Page 5 (and elsewhere) – update to indicate "***framework informed by NGSS***" (which have replaced the prior Benchmarks and NSES). Aligned with many of the comments above about updates to align with NGSS.

Will need to update "***Descriptions of NAEP Basic, NAEP Proficient, and NAEP Advanced*** must be as clear as possible" so that the NAEP levels are aligned with all 3 dimensions of NGSS thinking that would be assessed… so that for example 'basic' still includes descriptions about the level of skill/understanding that students bring to using practices, or using crosscutting concepts as a sense-making lens.

WISCONSIN DEPARTMENT OF
Public Instruction

Jill K. Underly, PhD, State Superintendent

September 30, 2021

Lesley Muldoon, Executive Director
National Assessment Governing Board
800 North Capitol Street, NW, Suite 825
Washington, DC 20002

Document Number: 2021-17676

Dear Ms. Muldoon:

The Wisconsin Department of Public Instruction (WDPI) appreciates the opportunity to provide public comment on preliminary guidance by the National Assessment Governing Board (NAGB) in updating the Assessment Framework for the 2028 National Assessment of Educational Progress (NAEP) in Science. Please find the WDPI's feedback in response to the NAGB's updates to the Science Assessment Framework for the 2028 NAEP below.

The 2019 NAEP Science Framework does not need to be updated.

The stated purpose of the NAEP in Science is to evaluate trends in scientific literacy overall and by demographic group. The current content, practices, and test design adequately accomplish this goal. The focus on phenomena and content linked to practice mirror the National Research Council's (NRC) Framework for K-12 Science and the Next Generation Science Standards (NGSS). While that mirroring is not a strong alignment, that is not the purpose of the NAEP.

Further, a review would likely result in relatively small changes that will not significantly change the impact this framework and test have on the field. Changes are unlikely to affect student learning. Instead, they are more likely to perpetuate the unhelpful focus on a practice referred to as gap gazing[1], which highlights achievement gaps instead of focusing on real systems change.

If a committee is formed, this could be an opportunity to expand innovative approaches to the NAGB's work. The WDPI suggests that the NAGB dedicate some time and capacity to developing materials and guidance that support systems of assessment and effective implementation of those systems.

If a revision is going to happen, a few ideas should be considered.

The WDPI believes that if the NAGB updates the 2019 NAEP Science Framework, the following suggestions must be taken into consideration:

1. Replace the Depth of Knowledge - Level One items that rely on memorization skills with items that test the student's skills in application, evaluation, and analysis of concepts.

[1]https://www.researchgate.net/publication/227252559_Beyond_Gap_Gazing_How_Can_Thinking_About_Education_Comprehensively_Help_Us_Reenvision_Mathematics_Education

2. Allow for deeper exploration of phenomena by having sets of multiple items digging into a particular phenomenon.
3. Create phenomena or contexts that would interest students and engage them in a real-life scenario that requires critical societal thinking and would better reflect scientific literacy instead of looking at phenomena that are disconnected from any meaningful context (e.g., random food webs).
4. Involve learners by engaging them in the practices of modeling, asking questions, and critiquing evidence or scientific practice, which could support more effective sensemaking and prompt scientific literacy development.
5. Align the NAEP Science Framework completely to the 2012 NRC Framework for K-12 Science and the NGSS, which would provide a more coherent signal and system for the field.

Thank you for the opportunity to comment. If you have any questions, please contact Viji Somasundaram, Director, Office of Educational Accountability, at visalakshi.somasundaram@dpi.wi.gov.

Sincerely,

Carl J Bryan

Carl Bryan
Administrative Rules Coordinator

CB:vs

Strategic Vision 2025
Annual Progress Report
November 2021

Introduction

With the Strategic Vision unanimously approved in September 2020, Governing Board staff drafted implementation plans. With the guidance of standing committees, staff created plans cross-functional in nature, ensuring that cross-committee input drives implementation. The work plans also establish a process by which staff collaborate. The Executive Committee monitors the Board's progress and provides oversight in implementation.

The Board received quarterly reports via committees at the March, May, and August meetings. In November each year, the Board will receive an annual progress report. The purpose of this annual report is two-fold: to review the Board's accomplishments over the past year and to preview the year ahead.

Strategic Pillars and Priorities

Strategic Vision 2025 is organized by three pillars: *Inform, Innovate, and Engage*. Housed under the three pillars are eight strategic priorities. Since adoption of the Strategic Vision, staff have initiated work plans that reflect priorities led by each of the standing committees.

Just as the Board set forth on the path toward Strategic Vision implementation, the COVID-19 pandemic struck, closing schools across the country. The Board and the National Center for Education Statistics (NCES) turned its attention to the urgent set of decisions facing NAEP and whether it would be administered in 2021 as planned. Ultimately, the Board requested and received a Congressional waiver to postpone 2021 administration to 2022. On the heels of this decision, the Board was engaged heavily in revision of the NAEP Reading Framework, the first time the framework had been revised in nearly two decades. These two issues, alongside complications of the ongoing pandemic, took significant time for the Board and staff alike. However, noteworthy progress was made on other activities as described below.

Inform and Engage

Under the *Inform* and *Engage* pillars, the Board's vision prioritizes addressing the needs of stakeholders and increasing and improving their use of NAEP data. By enhancing the use and relevance of NAEP data, stakeholders can inform education policy and practice in ways that lead to improving student achievement.

The foundation for this work is an engagement strategy through communications and partnerships, that is, effectively communicating compelling stories with NAEP data to motivate stakeholders to take action.

The first step to engaging with NAEP stakeholders is building awareness of key findings from NAEP, such as the growing gulf between high- and low-performers over the last decade or the declines in NAEP reading and math scores nationally and in many states. Once awareness is developed, NAEP stakeholders can start gathering and evaluating data to understand factors potentially related to the findings on NAEP. From there, stakeholders and practitioners can share best practices that address those factors and eventually, hopefully, help solve the issues.

Year 1 Implementation

In Year 1, significant accomplishments in the engagement strategy included adopting new ways of conducting NAEP release events, collaboration with partners to further use of NAEP data and results, and highlighting areas of interest and opportunity for NAEP. Release events have traditionally been held in person but transitioned to the virtual space due to the pandemic. While driven by circumstances, the Board seized the moment to make events more inclusive and more attractive to participants. The virtual events also created an opportunity to showcase a wide range of partnerships.

Release Events. Near the end of October 2020, the Governing Board led a virtual release of the NAEP Grade 12 results in reading and mathematics. Chair Haley Barbour kicked off the release event, after which then-Commissioner of NCES Dr. Lynn Woodworth set the context for the grade 12 data. Next, Dr. Peggy Carr, then-Associate Commissioner of NCES, shared highlights from the data. During the second half of the event, Board members Alberto Carvalho (district superintendent representative), Paul Gasparini (secondary school principal representative), and Reginald McGregor (business representative) addressed questions from NAEP stakeholders in the education and business communities, including Siemens, Lumina, the Fordham Institute, and Education Trust-West. More than 500 attendees tuned into the live-streamed event. Following this release, the Board produced a series of graphics spotlighting the grade 12 data that elicited attention on the Board's social media channels, with focus on two- and four-year college application and acceptance rates, student confidence in reading and math skills, and performance of students with disabilities.

In November 2020, the Governing Board partnered with NCES, Education Trust and the Fordham Foundation for a well-attended Twitter chat. Participants discussed score and contextual data from the Grade 12 report, including reading and subgroup performance, as well as implications for policymakers and educators in such areas as course taking and college preparation.

In May 2021, the Governing Board hosted the release of the 2019 NAEP Science results. Nearly 600 attendees joined to hear then-Acting Commissioner Carr's presentation of the findings and a policy-focused conversation with Board member Christine Cunningham and Board alumnus Cary Sneider. The release event featured stakeholders in the science education community, fostering a network of NAEP-savvy experts within policy and advocacy groups, such as the National Science Teachers Association (NSTA), National Association of Elementary School Principals (NAESP), and the National Association for Research in Science Teaching (NARST).

Next, in June, the Governing Board partnered with the questioners from the National Science Teachers Association featured at the May release event and Stephen Pruitt of the Southern Regional Education Board (SREB) for a popular Twitter chat. Activities like these solidify strong working partnerships with stakeholders in NAEP Science, which are particularly useful ahead of the Board's review of the NAEP Science Framework. Finally, to accompany the release event and the Twitter chat, the Board produced videos about the subscales measured by the NAEP science assessment and graphics highlighting results. These circulated on the Board's social media channels.

Just last month, the Board released the 2020 NAEP Long-Term Trend results for 9-year-olds and 13-year-olds. As these data measure very different skills from main NAEP with different populations (age-based vs. grade-based), the release itself assumed a different tone. Thus, the release took the form of a video, produced in close collaboration with NCES, which highlighted and explained results of most relevance. Media and policy advocates grasped the main findings readily, an important goal of the Board's Reporting and Dissemination Committee (R&D) and disseminated the findings widely. Plans for follow-up activities are underway. In addition, 9-year-olds will participate in the 2022 NAEP Long-Term Trend, so next year's release will lend invaluable insights into the impact of COVID on learning.

Collaboration. Beyond the stakeholder involvement that surrounded 2021 release events, the Board also partnered with the American Association of Geographers (AAG) to highlight NAEP Civics, NAEP U.S. History, and NAEP Geography data. Led by AAG's Michael Solem, this NAEP symposium spotlighted solid research with these data as exemplars for new researchers to follow. Board staff and NCES staff presented at the symposium and assisted in its coordination.

Spotlights. Over the past year, the Board convened discussions and produced materials to reflect areas of need and interest. For instance, at its March 2021 quarterly meeting, Board members heard from three experts about considering equity within the context of assessment generally and NAEP specifically. Highlights from the Board's discussion focused on urging educators to think beyond <u>*what*</u> factors in education work to <u>*how*</u> 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. This discussion was the first of what is anticipated to be a growing area of emphasis for the Board, particularly as discussions of equity in assessment increase among the testing and measurement community and beyond.

As an initial step in this direction, the Board published a full-length feature *Leveraging NAEP Data to Study and Improve Educational Equity*, the first in a suite of multimedia products that will focus on equity. This feature tackles equity from a national perspective on how the education, research, advocacy and policy communities have used NAEP data to understand inequity in education and inform steps to improve equitable outcomes for students. The feature also includes perspectives from experts on how NAEP data from the 2022 administration will be critical in understanding the impact the pandemic has had on student achievement. Future installments will focus on equity efforts in several TUDA districts.

In nascent areas of interest, the Board developed a short video to explain the increasingly divergent trend lines in NAEP Reading and Mathematics results over a decade, an ongoing area of exploration for the Board. To date, the video has been viewed over 1900 times. As it is by far the most watched video the Board has produced, it suggests that interest is high as the Board takes up this body of work.

Finally, in response to immediate needs over the past year, the Board produced collateral that provided timely information to the public about the NAEP 2021 waiver and the NAEP Reading Framework.

Year 2 Implementation

In Year 2, the engagement strategy involves the use of communications and outreach as the "connective tissue" among the Board's priorities. Drawing from committee and full board discussions over the past year, as well as issues raised by stakeholders, preliminary planning at the staff level is focused on several key areas of interest. Those areas may include:

- Communicating 2022 NAEP reading and math results in relevant, useful, and actionable ways, particularly given nationwide concerns about students' learning loss during the pandemic;
- Examining the growing divide between NAEP's high- and low-performers and finding ways to effectively communicate this information and catalyze action among stakeholders;
- Exploring how to improve measurement and reporting of students who score below *NAEP Basic*;
- Participating actively in the measurement community's deliberations on how to create more equitable assessments;
- Identifying potential solutions toward a more accurate SES indicator in NAEP; and
- Making policy decisions that will inform and support NAEP's transition to its next generation digital platform.

These and other issues will be points of discussion in upcoming Board meetings. The Board will determine and prioritize the areas of emphasis, providing guidance to staff in responding to NAEP's high priority needs in 2022. It is anticipated that the Board's engagement strategy will require pruning, aiming for more depth in a few key areas and less breadth across a wide range of topics. The engagement strategy will be less about stand-alone events and activities and more about sustained efforts on key issues. Success in 2022 means making connections between the public's interests and NAEP, strengthening relationships with key stakeholder groups best positioned to use insights from NAEP to improve student achievement, and collaborating with NCES to provide robust, actionable information to the public.

Inform and Innovate

Under the *Inform* and *Innovate* pillars, the Board aims to optimize NAEP through the Schedule of Assessments, achievement levels, and frameworks. The NAEP Assessment Schedule instantiates the Board's policy priorities of frequency, utility, and efficiency. Ensuring that policymakers share those priorities and support them via appropriate funding for NAEP is essential to implementing the Board's vision of assessments that cover a breadth of content areas, provide information frequently enough to allow for insights about student progress, and provide state-level results when possible.

Since the Board's inception, it has been committed to setting high expectations for the achievement of U.S. students and spearheaded the use of policy-driven achievement levels in educational assessment. The NAEP Achievement Levels are an essential feature of NAEP's reporting that communicate to the public expectations for student achievement—and have inspired improvements in states' own standards. Equally important, the Governing Board is responsible for determining the content and format of all NAEP assessments and has carried out this statutory responsibility by engaging a broad spectrum of stakeholders in developing recommendations for the knowledge and skills NAEP should assess in various grades and subject areas. By updating frameworks, NAEP remains a relevant, useful resource to the public.

Year 1 Implementation

In Year 1, accomplishments include ongoing communication and meetings with policy stakeholders about the NAEP assessment schedule, launch of the achievement levels descriptors project in mathematics and reading, and framework updates.

Communication with Policymakers. One of the Board's legislatively mandated responsibilities is to set the NAEP Assessment Schedule. In 2021, the Board communicated with the U.S. Secretary of Education and members of Congress about the Board's commitment to administering the full assessment schedule, which is crucial to understanding what America's students know and can do in various subjects. This action on behalf of the Board came about based on the projected funding flows through 2024, as presented by NCES at 2020 and 2021 quarterly Board meetings. On behalf of the Board, staff conducted meetings with Department of Education staff in the Secretary's office, the Office of Planning, Evaluation, and Policy Development, and the Office of Legislation and Congressional Affairs to express support for an increase to the NAEP budget to maintain the assessment schedule. Ultimately, President Biden's fiscal year (FY) 2022 Budget Request for the U.S. Department of Education included a $15 million increase to the NAEP program that would cover most costs associated with the projected budget deficit. In addition, the House Labor-HHS-Education 2022 appropriations bill included a $40 million increase, which is $25 million more than President Biden's budget request. The additional $25 million would be reserved to conduct a state-level Civics assessment in 2024. The FY 2022 appropriations process is still underway and, to date, no final bill has been passed by Congress and signed into law by President Biden.

Achievement Level Descriptors. In September 2020, the Board awarded a contract to Pearson to review and revise the NAEP Reading and Mathematics achievement level descriptions. This project is intended to provide validity evidence to address the most important recommendation from the 2016 evaluation of NAEP achievement levels. The project will also produce reporting achievement level descriptions to more clearly communicate what students within each achievement level can actually demonstrate on the assessment in accordance with the Board policy on NAEP achievement level setting. Recent activities include a virtual pilot study meeting conducted on October 25-29, 2021, which will inform the operational study planned for February 21-25, 2022 (also to be held virtually).

Frameworks. Since 2018, the Board has been actively engaged in the process to update NAEP Frameworks. Three key activities took place in this area during the last year. First, the Board unanimously approved the 2026 NAEP Reading Assessment Framework in August 2021. Then, in November 2021, the Board will take action on the 2026 Reading Framework Item and Specifications, an accompanying document designed to provide guidance to NCES and its contractors in operationalizing the assessment.

Next, with the recent updates to the frameworks in mathematics and reading—after a decade-long period of not updating existing frameworks—earlier this year the Board initiated review and revision of the framework development policy and procedures. The Assessment Development Committee (ADC) held a joint meeting with the Committee on Standards, Design, and Methodology (COSDAM) to discuss the current framework process and, as part of a continuous improvement effort, consider potential revisions to the process. ADC will bring to the Board in November 2021 a set of recommendations for Board input and discussion.

Related, in fall 2021, ADC introduced a new activity to kick off review of the NAEP Science Framework, conducting an initial public comment period to collect input from the field on the state and relevance of the existing NAEP Science Framework. In November, the Board will review feedback received from this public comment. The initial public comment is intended to help the Board identify the key issues for which the Board may want to provide policy guidance to Science Framework panels (if the Board decides to update the framework) and to identify what additional input and expertise is needed to inform that policy guidance so that white papers can be commissioned and/or expert panels can be convened in early 2022.

Year 2 Implementation

Looking ahead to Year 2, the Board will continue to monitor and manage the NAEP assessment schedule in robust ways, link NAEP to external data sources, enhance framework update processes and procedures, and effectively communicate NAEP achievement levels with external audiences. Further, the Board will turn its attention to exploring NAEP's next generation of digitally-based assessments and improving NAEP's measurement and reporting of lower-performing students.

In 2022, the Board will continue to coordinate efforts with NCES to ensure the NAEP program has the necessary funding to administer the assessment schedule and conduct research and development to modernize the program. Board staff will work with congressional committee staff and other policymakers to communicate funding needs and any changes being considered to the assessment schedule. Additionally, staff will support the Board in establishing the policy priorities and understanding policy implications of NAEP's transition to the next generation of digitally-based assessments.

The Board will make strides in linking NAEP to external data sources by convening a working group of Board members from COSDAM and R&D. In conjunction with NCES, the group will identify policy-relevant findings from existing linking studies and recommend how this work can be highlighted in ways that are actionable to policymakers. Simultaneously, through a contract the staff will begin cataloguing data already linked to NAEP and disseminating this information to stakeholders and partners, developing a plan for increasing linkages, determining which additional linkages to take on, and getting underway conducting the necessary studies to do so.

Guided by the Board's Achievement Levels Work Plan, studies to review ALDs in other subject areas are planned following the completion of the studies focusing on reading and mathematics. Advisory groups will convene to provide recommendations on communicating NAEP achievement levels and, in collaboration with R&D, an interpretive guide for the NAEP achievement levels will be produced. In addition, the Board will explore ways to improve measurement and reporting of students who fall below the NAEP Basic achievement level.

Finally, the Board will continue making progress in the framework update arena, through policy review and revision, a schedule for updates that aligns with the NAEP Assessment Schedule through 2030, and proceeding with revisions to the NAEP Science Framework, as determined by ADC and the Board.

Conclusion

Strategic Vision 2025 was crafted to frame the Board's work over a four-year period. The aims are lofty, but necessary. On behalf of America's students and in service of NAEP the Board must prioritize what information to share and how to share it (inform), build bridges with stakeholders in communicating NAEP's relevance and value (engage), and serve as the touchstone for the policies that drive the future of large-scale assessment (innovate). In doing so, the Governing Board will continue to ensure from its policy position that NAEP remains the "gold standard" assessment of educational achievement in the United States.

Mirrors or Windows: Briefing and Discussion

The working lunch on Friday, November 19th will feature Dr. Ray Hart, Executive Director of the Council of the Great City Schools, who will share findings from the Council's recent report, *Mirrors or Windows: How Well Do Large City Public Schools Overcome the Effects of Poverty and Other Barriers*.

Education is considered an effective path out of poverty, but most educational outcomes are strongly correlated to poverty. How can these apparently contradictory principles be reconciled? Schools can be windows of opportunity and help overcome or mitigate poverty or schools can act as mirrors and reflect society's inequities. This report analyzed the last ten years of NAEP data in reading and mathematics, in grades 4 and 8, to address the question: Are urban public schools, which have the largest numbers and concentrations of poor students in the nation, windows or mirrors?

The report's findings show students in Large City public schools scored higher than predicted on NAEP and than other schools in the aggregate across the nation. The report also showcases the immense value of the Trial Urban District Assessment (TUDA) program of NAEP. Dr. Hart will present highlights from the report, after which members will engage in discussion.

Dr. Raymond C. Hart is the Executive Director of the Council of the Great City Schools. Hart, who has more than 20 years of experience in research and evaluation, was previously the Director of Research for the Council, where his work has spanned policy areas such as post-secondary success and college readiness, professional learning communities and school improvement, teacher effectiveness and value-added analysis, early childhood education, and adult and workforce literacy. He has worked with clients from a number of federal agencies, including the U.S. Department of Education, the U.S. Department of Housing and Urban Development, the U.S. Department of State, the National Science Foundation, and many state and local departments of education.

Hart recently lead the Analytic Technical Support Task for the Regional Educational Laboratory – Mid Atlantic. He served as the Executive Director of Research, Planning and Accountability for Atlanta Public Schools, President and CEO of RS Hart and Partners, which is an evaluation and assessment consulting firm, and an Assistant Professor of Research, Measurement, and Statistics at Georgia State University. Prior to his work as a consultant, Hart served as the Director of the Bureau of Research Training and Services at Kent State University. His career began in 1989 as a program director for African American, Hispanic, and Native American students in Engineering and Science.

Hart holds a Ph.D. in Evaluation and Measurement from Kent State University, a M.Ed. with a focus on Curriculum and Instruction – Educational Research from Cleveland State University, and a bachelor’s degree in Industrial Engineering from the Georgia Institute of Technology.

Planned and Potential Innovations for NAEP

November 2021 Quarterly Board Meeting

Background

NAEP has long been considered the "gold standard" of assessments. The Governing Board focuses on policy-setting and oversight, while the National Center of Education Statistics (NCES) develops, administers, and analyzes NAEP. Recognizing their shared stewardship of the Nation's Report Card, both entities approach changes to NAEP with considerable caution, knowing that **three essential NAEP principles** must be prioritized during any transition:

1. Technical quality, credibility, and trust in the validity of the results must be maintained;
2. Trend lines that allow for comparisons of results from year-to-year and from jurisdiction to jurisdiction must be maintained; and
3. Potential changes to the program should ensure that public dollars are wisely spent while (a) increasing efficiency of the data collection and reporting systems and (b) upgrading the assessment content to reflect 21st century expectations and better understand what American students know across grades and subject areas.

Over the past 20 years, the Governing Board and NCES have successfully implemented several major changes to NAEP:

- In 2001, NAEP began using new testing methods and question types that reflect the growing use of technology in education.
- In 2002, NCES began evaluating the use of a computer-based writing assessment.
- In 2009, NCES began studying Interactive Computer Tasks in science.
- In 2011, the NAEP Writing assessment was the first NAEP test to be administered on digital devices.
- In 2014, the NAEP Technology and Engineering Literacy assessment was first introduced and conducted as a digitally-based assessment.
- In 2015, NAEP mathematics and reading assessments were piloted on digital devices.
- In 2017, NAEP mathematics and reading fully transitioned from paper-pencil to digitally-based assessments (DBA), with other assessments following.

Additional major advancements are now underway: a transition to the **Next Generation eNAEP test platform** and the **modernization of the NAEP Digitally-Based Assessment (DBA) administration model.** NCES has previewed preliminary plans for these transitions in closed sessions with the Governing Board.[1] NCES will continue to regularly update the Governing Board.

NCES's current operational plans to modernize the NAEP DBA administration model include the following planned milestones:

KEY MILESTONES		
By 2024:	**By 2026:**	**By 2028/2030:**
• Operationalizing the Next Generation eNAEP platform • Transitioning from the current "offline" model to **online capability**	• Transitioning to a **device-agnostic approach** to NAEP test administration (i.e., schools use their own devices instead of NAEP bringing devices into schools) • Moving to **"reduced contact" administration** of NAEP, in which fewer NAEP-trained personnel would support the test administration	• Exploring options for **computer-adaptive** testing (i.e., multi-stage adaptive but possibly other models) • Exploring options for a **two-subject design of NAEP** to reduce sample sizes by about one third, reduce costs, and maintain technical quality

Given the three essential principles, NCES will carefully phase in these transitions, enabling pilot testing and research/analysis, with the goal of maintaining NAEP's longstanding trend lines. Based on current plans, each step would be implemented sequentially with a proof-of-concept study and a field test phase before implementing the change operationally.

To further inform NCES' thinking about **the modernization of the NAEP DBA administration model**, the Institute of Education Sciences (IES), the federal agency within which NCES is housed, has commissioned the National Academies of Science, Engineering, and Medicine (NASEM) to consider how the use of digital technology and other major innovations could transform NAEP over the next ten years and beyond. This NASEM study aims to identify opportunities to substantially reduce NAEP costs while largely preserving its technical quality and informative value. The panel's report is expected to be issued in winter or spring 2022. The NASEM study may include recommendations that significantly affect or expand upon current NCES operational planning efforts. Once the report is released, the Governing Board should examine its recommendations and consider whether/how to incorporate them into Board policy development. For more information see: *Opportunities for the National Assessment of Educational Progress in an Age of AI and Pervasive Computation: A Pragmatic Vision*.

[1] The sessions are closed due to budget implications that affect procurements.

The Governing Board's Role

As the body responsible for setting policy and maintaining oversight of NAEP, the Governing Board will work collaboratively with NCES to:
- Set policy goals for this transition;
- Develop or update Governing Board policies, as necessary, to support successful implementation of this modernization effort; and
- Monitor changes to ensure the trust and credibility of NAEP results are maintained throughout the transition.

Broadly speaking, the Governing Board's policy goals for this modernization effort *may* include:
- Ensuring the modernization efforts balance the need for careful, methodical implementation with the rapid pace of change in technology so that NAEP's DBA administration model can continue to adapt over time.
- Ensuring technological innovations not only improve efficiency but also meet policy and assessment design goals for NAEP.
- Ensuring that the Board can exercise its legislative mandates (i.e., to select the subjects to be tested, to establish assessment frameworks and test specifications, to design the methodology of the assessment, and to develop standards and procedures for regional and national comparisons) per P.L. 107-279, Title III.
- Ensuring this transition—and the cost savings it is intended to achieve—allow for the Board's policy priorities for the NAEP Assessment Schedule (i.e., maintain subjects beyond reading and math, increased reporting at the state- and TUDA-level, and increased assessments at multiple grades).
- Maintaining NAEP's trend lines during the modernization of NAEP's DBA administration models.
- Maintaining NAEP's sterling reputation, in terms of technical quality and in terms of credibility and trust in the validity of the results.

In upcoming quarterly meetings, the Governing Board will consider what policy actions should be taken to support the successful transition of NAEP towards these goals. This may include the development of new Board policy and/or updates to existing Board policies (e.g., the Item Development and Review Policy last updated in 2002). It may also include working with NCES to review recommendations from the NASEM study.

November 2021 Quarterly Board Meeting

The Governing Board will hold a plenary session focused on the modernization of NAEP's DBA administration model. This session will include two components:

First, NCES will provide a briefing for the Governing Board on its progress in transitioning to an online, device-agnostic, and "reduced contact" administration model of NAEP. NCES will emphasize key issues that are surfacing in the early stages of implementation and will forecast decisions that will be crucial to the success of this transition.

Second, the Board will host a panel of experts that will share their expertise in transitioning other large-scale assessment programs to online and device-agnostic administration models. The panel will surface key policy issues that the Board should contemplate. These experts include:

- ***Tony Alpert, Executive Director of Smarter Balanced***, who will share relevant lessons learned from the multi-state assessment consortium's implementation of its online, device-agnostic statewide assessments, including how Smarter Balanced has approached comparability of results while maintaining flexibility for states and the interactions between device capabilities and the sophistication of items.
- ***Marianne Perie, President of Measurement in Practice***, who will share reflections from working with multiple states and assessment consortia in the implementation of new online assessments, including the potential threats to trend, the potential impacts of different technology devices used by students, and the implications for accessibility and accommodations for students with disabilities and English learners.

In subsequent quarterly Board and committee meetings, members will have the opportunity to engage with other experts with various perspectives relevant to this transition for NAEP—for example, industry leaders and technology experts who can help the Board and NCES think about the impact that the rapid pace of change in technology will have on the current plans; state and local technology directors who can help prepare for the on-the-ground implementation of these modernization efforts; etc.

In preparation for this session, some potential policy questions the Board may want to consider include:

- To what extent might aspects of this modernization effort impact NAEP's reputation as the "gold standard" in assessment? What approaches can be taken to mitigate potentially negative impacts?
- To what extent might aspects of the modernization effort impact NAEP's ability to report trend from year-to-year?
- To what extent might aspects of this modernization effort affect NAEP's ability to report valid and reliable comparisons among states and TUDAs *within* a year?
- How can NCES and the Board prepare the field for the significant transitions inherent in the modernization effort, when a key "selling point" of NAEP to schools is the ease of implementation?
- What assurances are needed regarding the protection of student and data privacy?
- To what extent will the proposed innovations achieve the cost efficiencies to which we aspire? What implications does that have for the long-term priorities of the program?
- What are the implications of this transition on accessibility and accommodations, particularly for students with disabilities and English learners?

Planned and Potential Innovations for NAEP
Panelist Biographies

Tony Alpert

Tony Alpert serves as Executive Director for the Smarter Balanced Assessment Consortium, which provides a dynamic system of tools that support teaching success. As Executive Director, he collaborates with Consortium members to ensure that the Consortium's tools for statewide improvement and tools to improve teaching and learning meet the needs of their students, parents and policy makers. Prior to assuming the role of Executive Director, Tony served as Smarter Balanced Chief Operating Officer, where he provided oversight for the financial, assessment, and technical operations of the Consortium. Previously, Tony worked at the Oregon Department of Education where he served in several different roles including managing the state's assessment accountability reporting, managing the allocation of the state's school fund, and finally as the director of assessment. Tony also served on the U.S. Department of Education's National Technical Advisory Committee where he provided technical counsel. Tony earned his master's degree at the University of Oregon.

Marianne Perie

Dr. Marianne Perie is the President of Measurement in Practice, LLC, a small education consulting firm focusing on K–12 assessment and accountability. She currently serves on eight state technical advisory committees, two of which are currently designing an innovative approach to a new assessment. As an extension of the advisory work, she has provided testimony to state legislatures and boards of education, evaluated standard-setting workshops, facilitated task-force meetings, and provided professional development on formative evaluation practices and data literacy. She has designed and directed workshops to draft achievement level descriptors and test blueprints. She has also served on advisory panels for NAGB, the College Board, and AICPA. Her areas of research focus on standard setting, validity theory, comparability of large-scale assessment, interim assessment, and alternate assessment for students with the most significant cognitive disabilities. Previously, she was the Director of two educational research centers at the University of Kansas, overseeing two state operational assessment programs, one career pathway assessment, and several grants. Prior to joining KU, she was a Senior Associate with the National Center for the Improvement of Educational Assessment, providing technical assistance to 16 states on accountability and assessment issues related to Federal policy. In her early career, she worked on multiple state and district assessments, the National Assessment of Educational Progress (NAEP), and international assessments as an employee of the Educational Testing Service and the American Institutes for Research.

GROUND TRANSPORTATION OPTIONS

Hilton Washington DC Capitol Hill
525 New Jersey Avenue, NW
Washington, DC 20001
(202) 628-2100

App Based Ride Services

BWI Thurgood Marshall Airport (BWI)
App-based ride services pick up and drop off passengers at the terminal curbs on the Departures/Ticketing Level between doors 9 and 11.

Dulles International Airport (IAD)
Passenger pick-up is located on the ground level outside of Baggage Claim, accessible via Doors 2, 4 and 6. Your driver will communicate the specific arrival door via in-app messaging.

Ronald Reagan National Airport (DCA)
Private vehicle pick-up is located on the third (outer) curb outside Terminal A and on the second (outer) curb outside Terminal B/C Baggage Claim (arrivals level). Passengers coordinate directly with the driver.

Taxi Service

Arrivals and Departures via BWI Thurgood Marshall and Ronald Reagan National Airports
Several taxi companies provide service from BWI Thurgood Marshall Airport (BWI) and Ronald Reagan National Airport (DCA). The one-way trip from BWI takes approximately one hour and the fare is approximately $80 - $120. The one-way fare from Reagan is approximately $25 and travel time is approximately 15 minutes. Taxi stands are located outside the airport and hotel.

Arrivals and Departures via Dulles International Airport
Washington Flyer Taxi Service (703) 661-6655 provides taxi service from Dulles International Airport. The one-way fare is approximately $65 per person and travel time is approximately 35 minutes. Upon arrival at Dulles, proceed to the baggage claim/arrivals area on the lower level of the main terminal and proceed to the Washington Flyer taxi stand. A curbside representative will assist you with coordinating service.

Public Transportation-Metrorail

The Hilton Washington DC Capitol Hill Hotel is accessible by Metrorail via the Union Station metro station on the Red Line. Exit Union Station metro station at Massachusetts Avenue, NE & 1st Street, NW. Walk a short distance SW to 1st Street, NE. Walk south on 1st Street, NE. Turn right on Massachusetts Avenue, NE. Walk approximately 1 block and bear left on North Capitol Street, NW. Walk a short distance south on North Capitol Street, NW and turn right on F Street, NW and left on New Jersey Avenue.

NATIONAL ASSESSMENT GOVERNING BOARD			
Travel Expense Report			
Name:		**FY 2022 Per diem rates (DC)**	
Trip Purpose:	**November 2021 Quarterly Board Meeting**	Daily Per Diem	$79.00
Trip dates – **Start on**		Meal & Incidental Expense Breakdown	$18 Breakfast, $20 Lunch, $36 Dinner, $5 Incidentals
Trip dates – **End on**		**Expenses at a glance**	
Notes		Other Expenses, (e.g. tolls, Internet access)	$0.00
		TOTAL EXPENSES	$0.00

DATE	From (trip origin)	To (destination)	Airline/Train (if purchased own)	Lodging (if not pre-paid)	Per Diem (based on meals provided)	Taxis	# of miles	Total mileage (0.56/mile)	Parking	Other expenses (enter description in line)	TOTAL
					Per Diem : $79 less 25% on travel days plus meals if applicable.						
Arrival		**DC**	$0.00	Pre-paid							
Return											
Detail Other Expenses											
						$0.00	0.00	$0.00	$0.00	$0.00	$0.00
						$0.00	0.00	$0.00	$0.00	$0.00	$0.00
						$0.00	0.00	$0.00	$0.00	$0.00	$0.00
						$0.00	0.00	$0.00	$0.00	$0.00	$0.00
						$0.00	0.00	$0.00	$0.00	$0.00	$0.00
						$0.00	0.00	$0.00	$0.00	$0.00	$0.00
						$0.00	0.00	$0.00	$0.00	$0.00	$0.00

			$1.00	$0.00	$0.00	$0.00	0.00	$0.00	$0.00	$0.00	$0.00	