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 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.

IES 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.

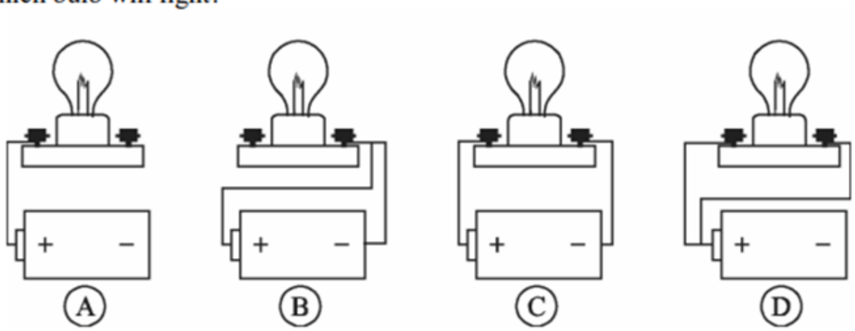
Ambitious content

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

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

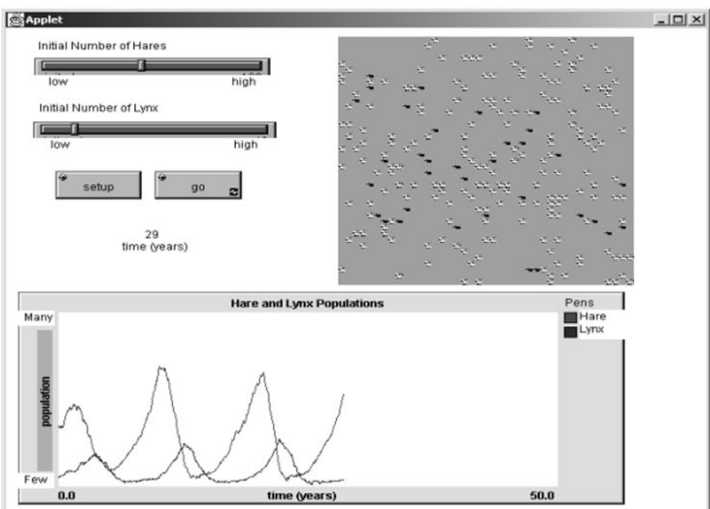
below from pages 69-71 of the Science Framework, was heavily debated among NCES's science standing committee[1] on whether this illustrative item assessed any content knowledge.

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.

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.

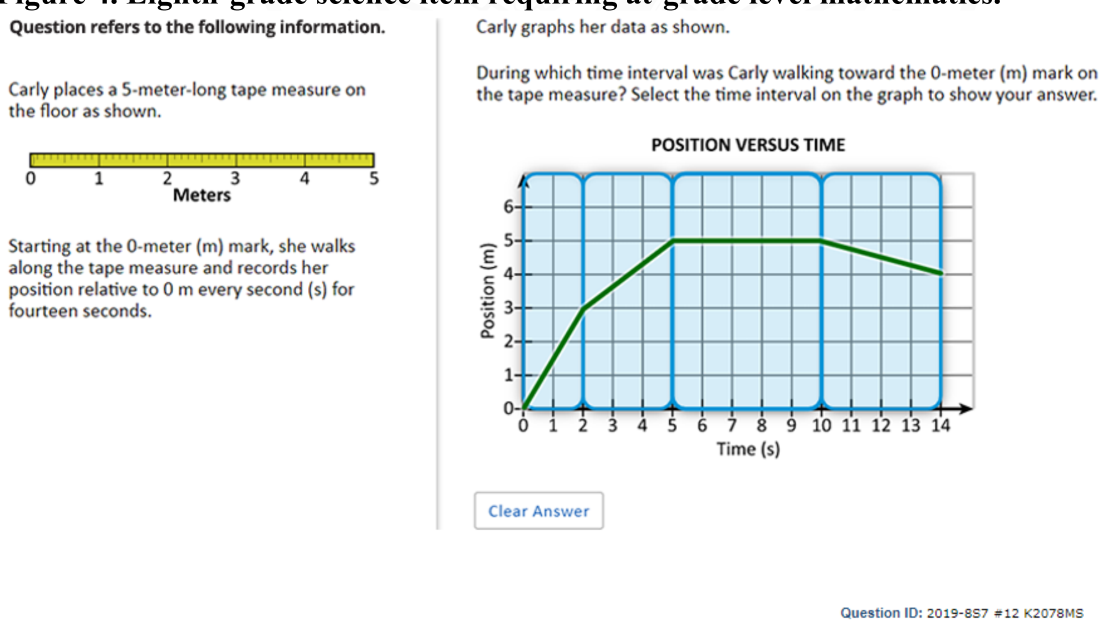
This performance requires students to use knowledge of the content statement and the experimental results to accomplish a practical goal (in this case, a park with grass growing well in areas that receive varying amounts of sunlight).

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, 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 (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.

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

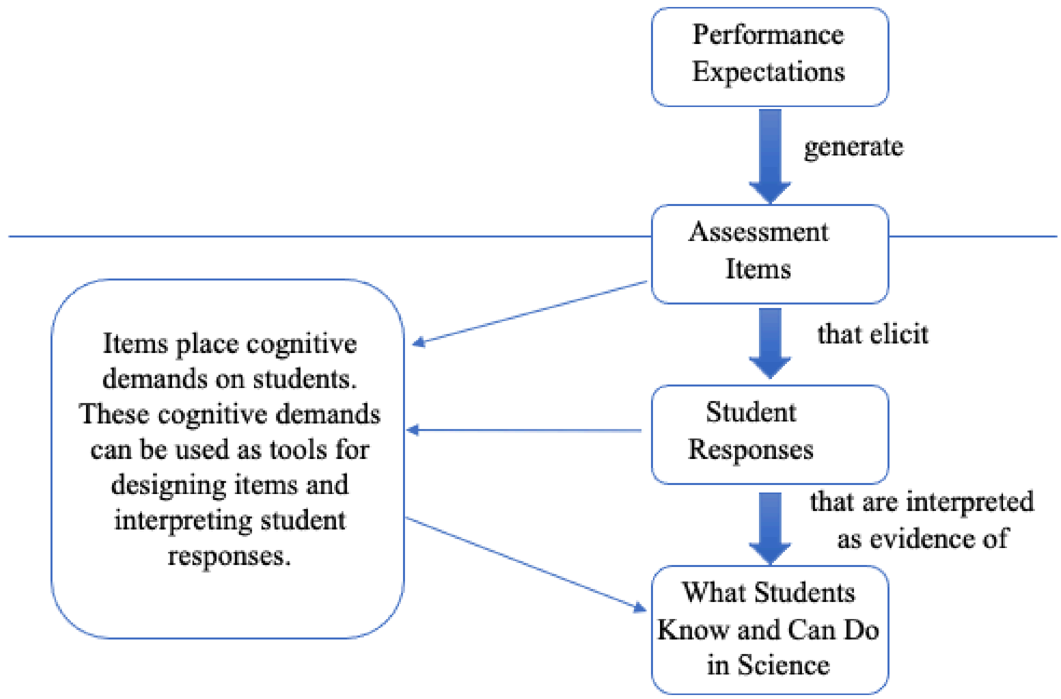
	Practice Label	Practice Applications			
←Communicate accurately and effectively→	Identifying Science Principles	Describe, measure, or classify observations.	State or recognize correct science principles.	Demonstrate relationships among closely related science principles.	Demonstrate relationships among 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 ▪ 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	**Crosscutting Concepts** ▪ 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).

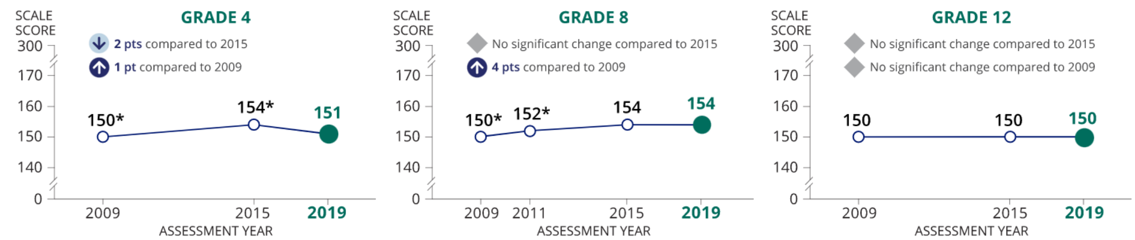
Perhaps not surprising is that substantial changes in science instructional practices were not observed in the 2018 NSSME survey relative to 2012 and that aspects of the vision for science teaching and learning embodied in the NRC framework and NGSS were less well represented in teacher beliefs and instructional 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

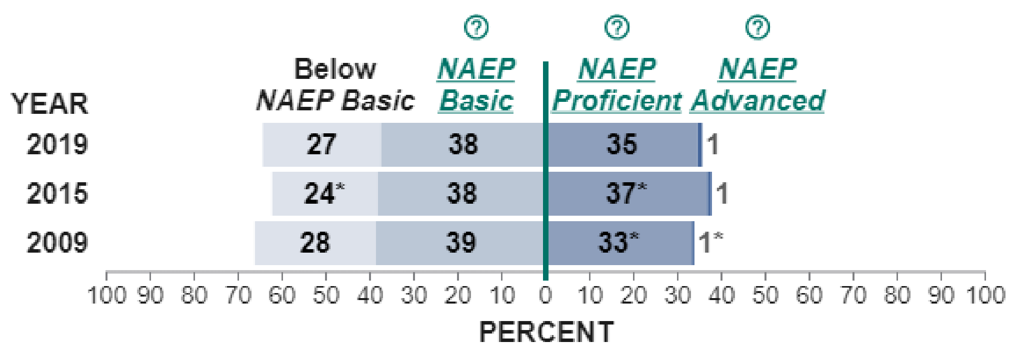


*Significant different ($p < .05$) from 2019.

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

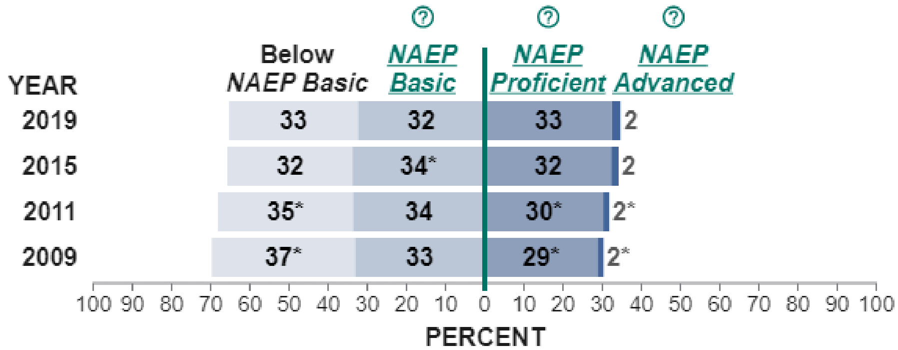


* Significantly different ($p < .05$) from 2019

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

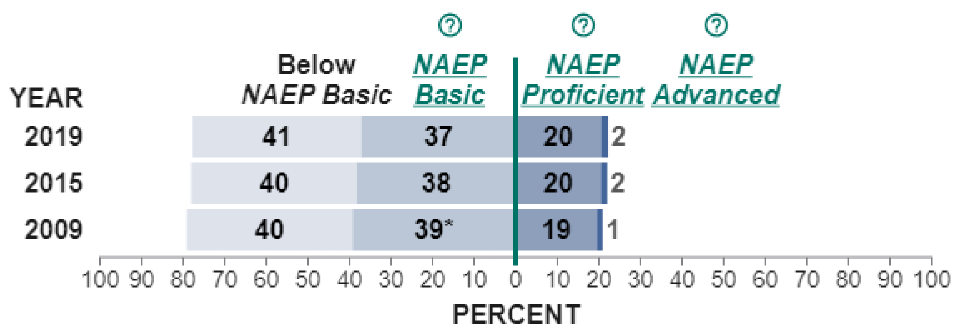


* Significantly different (*p* < .05) from 2019.

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 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

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…

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.

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.

cognia™

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 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.

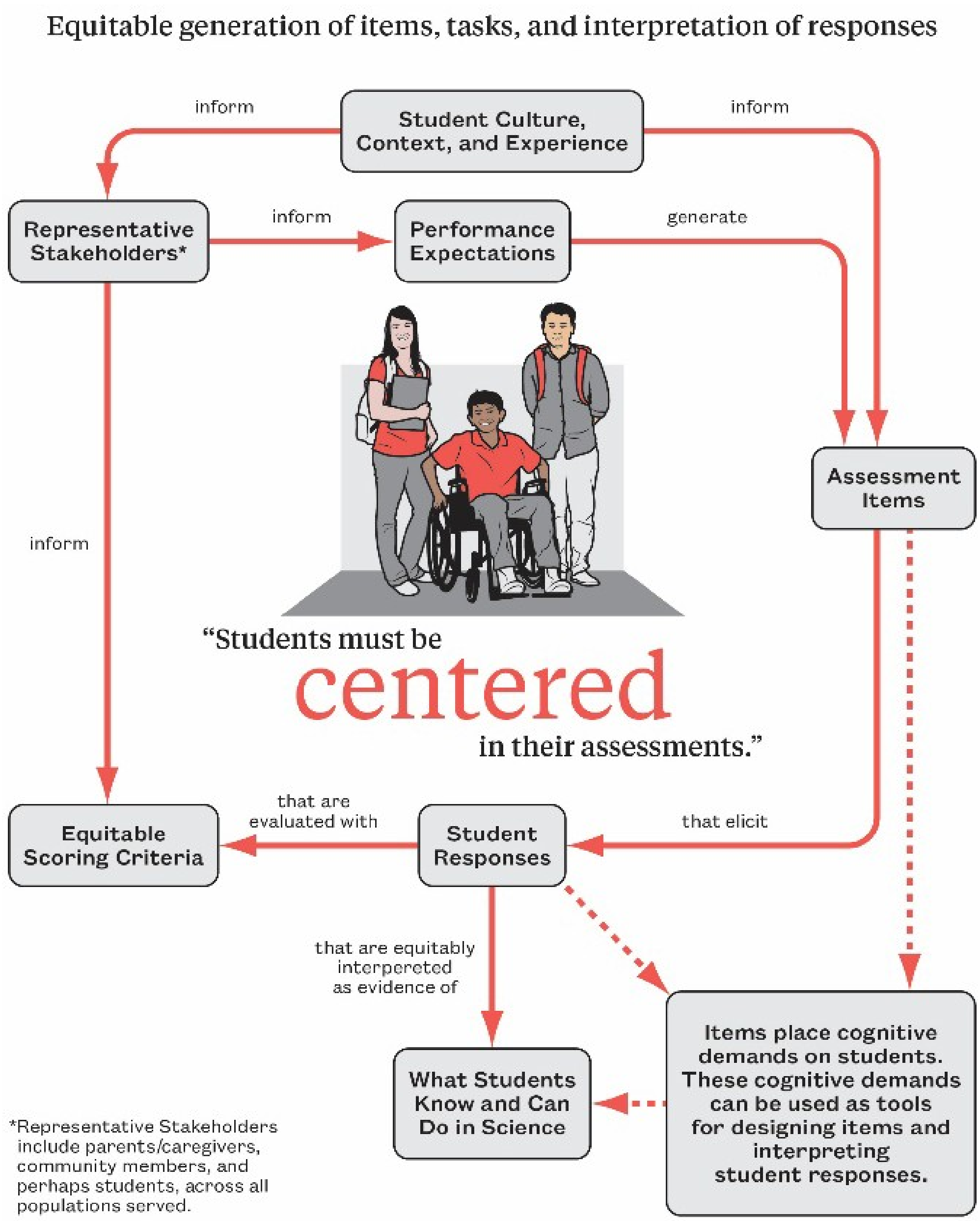


Exhibit 1: Student centered assessments.

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.

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

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.

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.

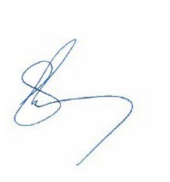
- In adopting recommendations made in this commentary, the distribution of content areas and cognitive complexity in the assessment will have to be revised as well. The NGSS has a different weighting of content in the standards by grade level, and we have already provided reasoning around revising cognitive 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.

  - 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

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

UNC PEMBROKE

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

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. Perils of Supporting Deficit Explanations via NAEP Science Results. 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. Supporting Equity and Diversity Research in Science Education. 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.

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

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