NAEP Science Assessment Framework: Possibilities and Priorities

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 assessments 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.

Last 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, along with a summary of specific points raised by major theme, were included in the November Board meeting materials and can be accessed here.

November 2021 Board Meeting Discussion

During the November 2021 Board meeting, there was a plenary discussion on the public comments received and potential implications for policy guidance that the Board may want to provide to framework panels, if the Board proceeds with updating the 2028 NAEP Science Framework.

The ADC proposed that the Board consider providing policy guidance to address the following questions:

- How should NAEP be informed by state science standards?
- Should content from the NAEP TEL Framework be incorporated into the NAEP Science Framework?
- To what extent should maintaining the trend lines be prioritized relative to other factors?
- How should the NAEP Science Framework define and reflect the Board's continuing commitment to equity?
- Are there any special considerations for grade 4?

The Board discussion of the public comments received and potential policy considerations highlighted a particular need to focus on the following questions for the next Board meeting:

1. Should maintaining the existing trend lines be the highest priority for the science framework update? The answer to this question has the potential to constrain all decisions made about other potential policy considerations.
2. What is the current state of science education/standards/instruction/assessment, and where are these likely heading over the next decade?
3. How can the Board create a consistent statement (to apply to all frameworks) about what equity means within an assessment context for NAEP?

March 2022 Board Meeting Science Framework Panel

Board staff assembled the following panel of science experts to write short papers (attached) and participate in a moderated discussion with the Board during the upcoming meeting on March 4th:

Aneesha Badrinarayan, *Learning Policy Institute*
Michael Heinz, *Council of State Science Supervisors*
Eileen Parsons, *National Association for Research in Science Teaching*
James Pellegrino, *NAEP Validity Studies Panel*
Eric Pyle, *National Science Teaching Association*

Additional biographical information about the panelists is attached.

Panelists were asked to address the following questions in their short papers:

1. What degree of change is necessary for the 2028 NAEP Science Framework?
2. What would be lost if comparisons cannot be made to results from prior assessments?
3. What would be lost if potential changes to the framework are limited by the need to maintain trend?
4. From your primary perspective, where is the current state of science education, standards, instruction, or assessment, and where is it heading?

The panelists' written responses to these questions are attached. During the plenary session on March 4, ADC member Christine Cunningham will moderate a panel discussion based on the content in the written responses. Board members will have an opportunity to ask questions of the panelists as well.

March 2022 Board Meeting Small Group Discussions

Following the plenary session described above, Board members will meet in small groups. ADC members Patrick Kelly, Reginald McGregor, and Nardi Routten will facilitate small group discussions for Board members to reflect on the panelists' recommendations and consider policy priorities for the NAEP Science Assessment Framework update, including:

1. Maintaining trend relative to other priorities for the science framework
2. How to think about incorporating equity into the science framework, and into NAEP frameworks generally
3. Initial thoughts on potentially addressing equity in other areas of the Board's work on NAEP

In between the March and May Board meetings, ADC will meet to discuss how to incorporate Board member input on the first two items above into a Board Charge to launch an update of the 2028 NAEP Science Framework, assuming that the Board decides to continue moving forward with the framework update.

Board staff will use input on the third item above to structure future Board sessions to further explore the role of equity in NAEP more generally, beyond the assessment frameworks.

NAEP Science Assessment Framework Panel: Speaker Biographies

Aneesha Badrinarayan, Learning Policy Institute

Aneesha Badrinarayan leads the performance assessment strategies at the Learning Policy Institute. For the last decade, her work has focused on supporting states, districts, and educators to develop and implement student-centered systems of assessment that support all learners. Her portfolio includes leading several complex, multi-state initiatives focused on reconceptualizing and designing new assessment systems; developing criteria and parameters for innovative large-scale and classroom assessments; evaluating assessment system quality; providing professional learning and strategic guidance for state and local leaders; and conducting analyses of state, local, and expert efforts to design and implement performance assessments and systems of assessment, particularly in science. Her passion for coherent and balanced systems of assessment stems from a commitment to high-quality teaching and learning for all and a deep interest in helping practitioners and leaders navigate their systems to achieve that vision. Prior to LPI, she was the Director for Special Initiatives at Achieve, a museum professional, and a neuroscientist.

Michael Heinz, Council of State Science Supervisors

Michael Heinz is the president of the Council of State Science Supervisors (CSSS). His current work focusses on ensuring excellence and equity in science education for K – 12 students in New Jersey. Michael was a member of the Lead State Team that developed the Next Generation Science Standards and was a co-author and contributor to documents published by Achieve to support the implementation of the NGSS.

He is a member of the State Steering Committee for OpenSciEd, a foundation-funded effort to develop freely available evidence-based science instructional materials. He is also deeply involved in the Advancing Coherent and Equitable Systems of Science Education (ACESSE) Project. ACESSE is an NSF project that brings together partners from educational research and practice to promote equity and coherence in science education.

Eileen Parsons, National Association for Research in Science Teaching

Parsons is an established leading scholar in science education, particularly in the areas of cultural inclusivity and racial equity. Several prestigious grants from the American Educational Research Association, Spencer Foundation, and National Science Foundation have funded her research with early work focused on middle school and later research extended to postsecondary and professional contexts. Her scholarship on cultural inclusivity and racial equity from grade 6 to STEM undergraduate pursuits and STEM careers of traditionally underrepresented groups of color is widely published in highly ranked venues in science education and other disciplines. Additionally, she has served as associate editor, editor of a special issue, section editor, and on the editorial boards of the top research journals in science education. She is an American Council on Education Fellow, science policy fellow for the American Association for the Advancement of Science, and Ford Foundation Postdoctoral Scholar—the first science education researcher to receive this honor. She served on the National Academies of Sciences, Engineering, and Medicine (NASEM) Committee that wrote the consensus report on science investigations and engineering design experiences in grades 6-12. Presently, she chairs the newly established NASEM Equity in prek-12 STEM Education committee.

Over the years, she has actively engaged several professional organizations and assisted in developing and implementing strategic initiatives that altered or established new organization-wide directions. She is currently the Immediate Past President of NARST, the premier national and international organization for science education research. Parsons received her PhD and MS degrees from Cornell University, Ithaca NY, and her BS degree from UNC. Early in her career, she taught high school chemistry, physical science, and trigonometry.

James Pellegrino, NAEP Validity Studies Panel

James W. Pellegrino is Liberal Arts and Sciences Distinguished Professor and Founding Co-director of the Learning Sciences Research Institute at the University of Illinois Chicago. He studies children's and adult's thinking and learning and the implications of research and theory for assessment and instructional practice. He has published over 300 books, chapters and articles related to cognition, instruction, and assessment. His research on science education and assessment has been funded by the National Science Foundation, the Institute of Education Sciences, and private foundations. He helped direct the College Board's redesign of curriculum frameworks and assessments for the Advanced

Placement courses in biology, chemistry, and physics. His recent projects have focused on the design of high-quality science assessment and instructional resources for K-8 classrooms. He has chaired several National Academy of Sciences study committees that have issued major reports related to science education, including the Committee for the *Evaluation of the National and State Assessments of Educational Progress*, the Committee on *Learning Research and Educational Practice*, and the Committee on the *Foundations of Assessment* which issued the report *Knowing What Students Know: The Science and Design of Educational Assessment.* Most recently he served on the Committee on *Science Learning: Games, Simulations and Education* and the Committee on a *Conceptual Framework for New Science Education Standards*. He chaired the Committee on *Defining Deeper Learning and 21st Century Skills and* co-chaired the Committee on *Developing Assessments of Science Proficiency in K-12*. He is a lifetime member of the National Academy of Education and the American Academy of Arts and Sciences. He currently serves on the NAEP Validity Studies Panel and on the Technical Advisory Committees for state assessment programs including those of the District of Columbia, Illinois, Maine, New York, Rhode Island, Texas, and Vermont.

Eric Pyle, National Science Teaching Association

Eric J. Pyle is a professor of geology at James Madison University, specializing in geoscience education and teacher preparation. He has published on science teacher preparation and professional development as well as instructional materials development and evaluation. He has served in the leadership of five NSF-funded projects, including grants for GK-12 Teaching Fellows, GeoEd, and the Robert C. Noyce program. He was a member of the Earth & Space Science (ESS) Design Team for *A Framework for K-12 Science Education* and was a primary reviewer for all drafts of the *Next Generation Science Standards.* He teaches coursework in Earth materials, contemporary Earth issues, and planetary geology, as well as joint courses in secondary teaching methods. Currently serving as the President of the National Science Teaching Association (NSTA; 2021-22), he previously served on the Board of Directors for NSTA heading the Preservice Teacher Preparation Division from 2014-2017. He is a past president of the Eastern Section of the National Association of Geoscience Teachers, the West Virginia Science Teachers Association (WVSTA) and the Virginia Association of Science Teachers (VAST). He received a BS cum Laude in Earth science from UNC-Charlotte (1983), an MS in Geology from Emory University (1986), and a PhD in Science Education from the University of Georgia (1995).

National Assessment Governing Board Panel on NAEP Science Framework[1]

Aneesha Badrinarayan, Learning Policy Institute

February 11, 2022

The following responses were developed to support considerations for revising the NAEP Science Framework and Assessment.

Questions posed:

- What are the major relevant trends in terms of science standards and implementation?
- What does the current landscape of science assessments look like? Where is it going?
- What is the role of equity considerations in science assessment conversations?
- How important is maintaining trend?

Major relevant trends regarding science standards adoption and implementation.

Standards. By February 2022, a vast majority of states have adopted new science standards based on *A Framework for K-12 Science Education,* including 20 states that have adopted the Next Generation Science Standards and over 25 states and territories that have adopted other standards that are similar in scope, structure, and expected shifts in teaching, learning, and assessment. These standards represent a careful reconsideration of the specific disciplinary core ideas (DCIs), science and engineering practices (SEPs), and crosscutting concepts (CCCs) that k-12 students should be supported in developing; some shifts that are particularly relevant to the NAEP Science Framework include:

- Specification of the science ideas to be mastered, and the grade-levels/bands at which they should be developed (notably, these are different than how many state standards and nationally-available instructional materials had previously approached content goals)
- Specification of both knowledge and application goals for science ideas, practices, and cross-cutting concepts along learning progressions (e.g., practices are more than an inquiry-based approach to pedagogy or a set of skills, but have knowledge about how and when to use them associated; similarly, evidence of DCIs comes from their meaningful use in context, not from simply "knowing" the science facts)
- Expectation that ideas that have value across scenarios and science domains like patterns and cause and effect (i.e., CCCs) are explicit learning targets across k-12.
- Expectation that DCIs, SEPs, and CCCs are developed and assessed together, in a range of combinations that reflect growing sophistication across all three dimensions.
- Inclusion of engineering core ideas and practices as part of the science expectations.

Implementation. Given how fundamentally different new science standards are, standards implementation—if defined by the degree to which classroom practice has shifted to reflect and prepare students for new expectations—has been expectedly slow as systems supports (e.g., instructional materials, professional development, classroom and large-scale assessments, local accountability, etc) have themselves needed to be developed. In the years since the release of the

[1] This paper was produced under Governing Board contract number 919995921F0002 to Manhattan Strategy Group, with subcontract to the Human Resources Research Organization.

standards, and particularly since 2016, there has been a marked uptick in these supports available, including:

- Quality criteria (EQuIP for Science) and established vetting processes for instructional materials designed for the NGSS (Achieve/NextGenScience EQuIP reviews, EdReports)
- Availability of high-quality instructional materials for middle school (e.g., OpenSciEd, Amplify) and high school science (e.g., inquiryHub) designed to be implemented with curriculum-based professional learning.
- Science professional learning standards
- Quality criteria for large-scale and classroom assessments designed for *Framework*-based standards
- Availability of aligned classroom assessments
- Transition to increasingly aligned large-scale assessments
- Networks of states actively seeking to advance NGSS implementation

Importantly, all of the system supports described above have been developed in intentionally coherent ways. For example, a range of state science leaders have been intentionally involved in the design of all quality criteria described above, which positions those criteria to shape state-wide implementation efforts. At the same time, development processes for criteria and materials described above also involve classroom and district-level educators, to ensure buy-in and awareness, establish local champions, and ensure materials are developed such that they can actually be implemented. Assessment criteria and instruments were developed to intentionally connect with features of high-quality instructional materials. These intentional efforts toward coherence across disparate parts of the system suggest (1) that as implementation continues, we can likely expect to see increasing uptake and shifts in classroom practice and student performance, and (2) that the widespread commitment to new science standards make it more likely that these standards are largely here to stay.

State Science Assessment Landscape.

Common features of new science assessments. Since 2014, states have been developing assessments—both classroom and large-scale—that reflect and are better able to monitor progress toward the expectations established in new science standards. This has been a rapidly evolving space, with many models, trials, and lessons learned. While the science assessment landscape continues to evolve as our understanding of how students learn and demonstrate science changes, there are some key shifts in science assessment features and best practices that the majority of states, developers, and experts have converged upon:

1. **Shift toward tasks vs. items.** As states began developing new science assessments, it became clear that new science standards were too complex to be appropriately represented exclusively through stand-alone selected response items. In many states, this has led to the development of more comprehensive tasks that include a grounding common scenario with a series of questions/prompts designed to surface different aspects of the targeted standard(s).

2. **Require sense-making with multiple dimensions.** Given the shift toward performance expectations that sit at the nexus of science ideas and practice, assessments have shifted from assessing content and practice in isolation, and toward requiring their use, together, within

tasks. In many state assessments, this functionally means that individual items are expected to require the use of multiple dimensions; entire tasks are intended to surface evidence of all dimensions assessed; and scoring/reporting is a reflection of multidimensional performances. Importantly, assessments are expected to surface their integrated use in the service of sense-making (e.g., using science ideas as part of reasoning; using practices to figure out relationships) rather than rote engagement in definitions or skills. Namely, nearly all states designing new standards hold this expectation explicitly for at least the DCIs and SEPs.

3. **Center making sense of compelling phenomena- and problem-based scenarios that highlight authentic uncertainty ("problematization").** A hallmark of NGSS/*Framework*-based teaching and learning is the centrality of phenomena and problems as a mechanism to engage students in the classroom science activities needed to develop the targeted ideas and practice. This extends to science assessments: phenomena/problem provide the observations around which students can make their thinking with the targeted dimensions visible. An important feature of assessment tasks that reliably require and surface meaningful, grade-appropriate sense-making is that they use a relevant, engaging phenomenon that highlights an authentic uncertainty for students to figure out.

4. **Consider more expansive views of equitable science assessment design.** While traditional approaches to equitable assessment design may focus on bias and sensitivity, new science assessments also attend to some features of student engagement, particularly in light of more involved task design. Common equity features expected as part of high-quality large scale assessment design include the use of relevant and engaging phenomena, the use of multiple modalities to convey information, attention to coherence from the student perspective, attention to language use and requirement, decentering vocabulary in favor of conceptual understanding, etc.

It should be noted that while there has been growing consensus on the features described above, there are a number of issues in science assessment design and use that are less clear or that states are approaching in intentionally different ways, and may be relevant to NAEP science discussions. These issues include:

- How to surface evidence of crosscutting concepts within large-scale assessment design.
- Reporting for multidimensional standards.
- Aligning to individual performance expectations or "bundles" of standards.
- Appropriate decisions about balancing depth and breadth in test design.
- Incorporating other features of equitable assessment practices, drawing on evidence from classrooms
- Grade-level expectations across all dimensions vs. emphasizing grade-level DCIs

An emerging movement toward performance-based systems. Given the depth, breadth, and increasingly student-centered implementation focus of new science standards, there is growing recognition that current approaches to large-scale assessment are insufficient to monitor and support student progress in ways that are consistent with how and what students are expected to learn. An increasing number of states are considering how to leverage performance assessments as part of their statewide assessment programs. The shifts many states are making toward the inclusion of sense-making-focused tasks with some open-ended/constructed response prompts represents one step in this direction; in other contexts, states are exploring the use of more extensive simulations and technology-enhanced items; the use of extended tasks as part of the on-demand assessment; and the use of curriculum/classroom-embedded performance tasks as part of the evidence for student progress.

In these cases, states are seeking to develop these performance-based assessments to be more consistent with instruction. States are exploring a number of strategies for developing assessments that are more intentionally coherent with high-quality instruction in science, including how features of high-quality instructional models and curriculum can become part of state assessment system design in ways that incentivize high-quality teaching and learning, attend to opportunity to learn, and provide a more valid measure of student performance relative to standards.

A note about momentum around reconsidering how equity is represented in science assessments. One major driver for continued innovation in science assessment is dissatisfaction with how assessments support equitable engagement and outcomes from a wide range of diverse learners. Ideas about equitable assessments that move beyond the features described above are largely concerned with features of high-quality, equitable classroom instruction and assessment practices, and how to navigate the tension when large-scale assessments are inconsistent with those features. Questions states are grappling with related to their state assessment programs include:

- How can state assessment programs attend more intentionally to scaffolding and other entry points and task-embedded supports for student engagement and sense-making?
- How can state assessment programs contribute to—or not detract from—culturally-sustaining practices while also serving their progress monitoring role?
- How should state assessments negotiate ideas about engaging and relevant phenomena with more traditional thinking about reducing the emotional valence of anything presented on large-scale assessments?
- How can state assessments account for and reflect the understanding that learning is a sociocultural endeavor? What is the role of features like science talk/discourse, collaborative argumentation and modeling, etc?
- How are student interest, identity, and agency included in assessment system design, implementation, and interpretation?
- How can state assessment programs become part of systems focused on supporting learning and oriented toward improvement rather than labeling and shaming students, educators, and schools? [shift from surfacing inequity toward a tool for enacting equity]
- How can state assessments be used to disrupt ableism? Center racial justice?
- How can state assessment systems intentionally support and surface the assets of diverse sense-making routines?

While these issues are not new, there is a growing call to figure out how to design, implement, and use state assessment systems that more directly address the relationships and trade-offs between these features and large-scale assessment. While most existing frameworks that tackle these features in science assessment exist for classroom assessment practices, emerging frameworks seeking to unpack “equitable large-scale assessments” tend to do so through the lenses of system design, student experience, and reporting/use. It should be noted that many states' interests in performance-assessments as part of their science assessment system design are motivated by a desire to better address these equity features.

Recommended changes to the NAEP Science Framework.

As the science standards continue to be implemented, the existing gaps (e.g., as identified by Neidorf et al., 2016) between the NAEP Science Framework and the expectations of new state standards will become increasingly apparent in student performance. This will substantially limit the validity and utility of both the framework and student scores on the NAEP science assessment, if not addressed in the upcoming opportunities for framework revision. At minimum, the framework should be revised to address the particularly glaring content misalignments, including:

- Distribution and emphasis of content across both science domains and grade-levels, to ensure that what NAEP is measuring is consistent with what students are expected to develop and learn, at the time in their k-12 experience when they are expected to do so.
- How science practices are represented in terms of both rigor/depth, engagement in the act of "doing science", and distribution
- The inclusion of crosscutting concepts
- Definitions and conceptions of science proficiency/achievement to be more consistent with the use of grade-specified science content and practices to make sense of meaningful real-world phenomena and problems.

In addition to these basic content alignment and interpretation issues, the science framework and associated assessment should be updated to explicitly support key shifts in assessment design.

Considerations should include:

- Incorporating aspects of the NAEP TEL Framework and assessment as part of science.
- 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.
- Ensuring proper alignment to updated framework goals.
- Developing tasks that center making sense of appropriate and compelling phenomena as their foundational basis.
- Attending to advances in equitable assessment that include and expand beyond attention to bias and sensitivity considerations.
- Alternative cognitive complexity models to address multidimensionality of items and item sets.

Maintaining Trend.

Given the misalignment between the current NAEP Science Framework and new science standards adopted by the majority of states and territories, **priority should be given to updating the framework to better align with what students are (and will be) learning and held accountable to within districts and states. This update should not be constrained by an attempt to maintain trend.** While trend data can be an important progress monitoring indicator, a focus on maintaining trend assumes that the goals monitored by the assessment have remained largely consistent, and that the trend line is interpretable. This will increasingly not be the case as new science standards implementation continues, and students' opportunity to learn science prepares them intentionally for a very different set of outcomes than those measured by the current NAEP science assessment. At best, this challenge to the validity of the NAEP Science Framework, assessment, and resulting measures of student proficiency will render the data largely useless to states and districts seeking to use NAEP as part of their progress monitoring system; at worst, NAEP science assessment results may underestimate science education progress and be used to justify misguided claims and programmatic decisions related to science standards implementation.

3-Dimensional Science Standards and the 2028 NAEP Science Assessment Framework[1]

Michael Heinz, President
Council of State Science Supervisors
February 11, 2022

Introduction

The purpose of this paper is to provide guidance on whether maintaining trend with the current National Assessment of Educational Progress (NAEP) Science Assessment Framework should be prioritized when considering potential goals for updating the science framework for 2028. In this paper I address the following questions:

- What is the state of science standards and where are they heading?
- What degree of change is necessary for the 2028 NAEP Science Framework?
- What would be lost if comparisons cannot be made to results from prior assessments?
- What would be lost if potential changes to the framework are limited by the need to maintain trend?

What is the state of science standards and where are they heading?

The NGSS and other Contemporary State Science Standards

The *Next Generation Science Standards: For States, by States* (National Research Council, 2013) (NGSS) are based on the of the National Research Council's consensus study report titled *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas* (National Research Council, 2012) (hereafter, NRC *Framework*). The NGSS embody a vision of what it means to be proficient in science. They rest on a view of science as both a body of knowledge and an evidence-based, model and theory building enterprise that continually extends, refines, and revises knowledge. This evidence-based vision for what it means to be proficient in science has transformed the work of educators, curriculum designers, assessment developers, state and district science administrators, professionals responsible for science teacher education, and science educators working in informal settings (National Research Council, 2012).

The NGSS identify what students should know and be able to do at the end of instruction in assessable performance expectations (PEs). Each PE represents the integration of three "dimensions" of science and engineering: Science and Engineering Practices, Disciplinary Core Ideas, and Crosscutting Concepts.

- **Science and Engineering Practices** reflect the major practices that scientists and engineers use to investigate the world and design and build systems. They are both a set of skills and a set of knowledge to be internalized.
- **Disciplinary Core Ideas** are the most important ideas in the physical sciences; life sciences; Earth and space sciences; and Engineering, Technology and Applications of Science.

[1] This paper was produced under Governing Board contract number 919995921F0002 to Manhattan Strategy Group, with subcontract to the Human Resources Research Organization.

- **Crosscutting Concepts** are a way of linking the different domains of science. They provide an organizational schema for interrelating knowledge from various science fields into a coherent and scientifically-based view of the world (National Research Council, 2013).

The integration of the three dimensions to create performance expectations has transformed science standards. Table 1 provides a side by side comparison of traditional and NRC *Framework* informed standards. The left-hand column contains science standards that were developed prior to the publication of the NRC *Framework*. The right-hand column contains examples from NGSS. Each of the dimensions is color coded. Each PE is comprised of a SEP, in blue text; a DCI, in orange text, and a CCC, in green text.

Table 1: Comparison of Science Standards Pre and Post NRC Framework

By the end of Grade 4, students: 1. Sort materials based on physical characteristics that can be seen by using magnification. 2. Observe that water can be a liquid or a solid and can change from one form to another to form the other and the mass remains the same. 3. Recognize that water, as an example of matter, can exist as a solid, liquid, or gas and can be transformed from one state to another by hearing or cooling. 4. Show that not all materials respond in the same way when exposed to similar conditions (NJDOE, 2004)	Students who demonstrate understanding can: 5-PS1-1: Develop a model to describe that matter is made of particles too small to be seen. 5-PS1-2: Measure and graph quantities to provide evidence that regardless of the type of change that occurs when heating, cooling, or mixing substances, the total weight of matter is conserved. 5-PS1-3: Make observations and measurements to identify materials based on their properties. 5-PS1-4: Conduct an investigation to determine whether the mixing of two or more substances results in new substances (National Research Council, 2013).

Since the beginning of the standards movement, educators have used a number of strategies to figure out what a standard meant. To help alleviate this burden, the Lead States included a number of features in the NGSS that reduce any ambiguity about the meaning of a PE.

Many of the performance expectations are accompanied by Clarification Statements and Assessment Boundaries. The Clarification Statements examples and other information that is useful in refining an understanding of the PE. Assessment Boundaries provide educators with information about what is out of bounds in a statewide assessment. See Appendix A for the complete text 5-PS1-1.

While the performance expectations can stand alone, a more coherent and complete view of what students should be able to do comes when the performance expectations are viewed in tandem with the contents of the Foundation Boxes that lie just below the performance expectations. These three boxes include the SEPs, DCIs, and CCC that were used to construct this set of performance expectations. See Appendix A to see the Foundation Boxes and Connections Boxes for 5-PS1-1.

Below the Foundation Boxes are three Connection Boxes. These are designed to support a coherent vision of the standards by showing how the performance expectations in each standard connect to other PEs in science, as well as to common core state standards for mathematics and English language arts.

Evidence Statements have been written for each PE. The Evidence Statements provide greater detail about what an educator should be able to observe when a student is proficient with that PE. See Appendix B to read the Evidence Statements for 5-PS1-1.

The NGSS offered four innovations:

Three-Dimensional Learning: There are three equally important, distinct dimensions to learning science included in the NGSS: Scientific and Engineering Practices, Crosscutting Concepts, and Disciplinary Core Ideas. The NGSS connect all three dimensions. To prepare students for success in college and 21st century careers, the NGSS also connect scientific principles to real-world situations, allowing for more engaging and relevant instruction to explore complicated topics.

All three dimensions build coherent learning progressions: The NGSS provide students with continued opportunities to engage in and develop a deeper understanding of each of the three dimensions of science. Building on the knowledge and skills gained from each grade - from elementary through high school - students have multiple opportunities to revisit and expand their understanding of all three dimensions by the end of high school.

Students engage with phenomena and design solutions: In instructional systems designed for the NGSS, the goal of instruction is for students to be able to explain real-world phenomena and to design solutions using their understanding of the Disciplinary Core Ideas. Students can achieve this goal by engaging in the Science and Engineering Practices and applying the Crosscutting Concepts.

Engineering and the Nature of Science is integrated into science: Some unique aspects of engineering (e.g., identifying problems) are incorporated throughout the NGSS. In addition, unique aspects of the nature of science (e.g., how theories are developed) are also included throughout the NGSS as practices and crosscutting concepts.

Science is connected to math and literacy: The NGSS not only provide for coherence in science instruction and learning but the standards also connect science with mathematics and English Language Arts. This meaningful and substantive overlapping of skills and knowledge affords all students equitable access to the learning standards (NGSS Fact Sheet, 2022).

To date, 20 states and the District of Columbia have adopted the NGSS as their state’s science standards. The NGSS adoption states are home to approximately 35% of K – 12 students in the United States (NSTA, 2022).

Twenty-four other states have developed their own multi-dimensional science standards. These state science standards are also 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 (Pellegrino, 2021). Approximately 36% of K – 12 students live in a state that have NRC Framework informed science standards (NSTA, 2022).

Where are Science Standards Heading?

It was 15 years between the publication of the *National Science Education Standards* (National Research Council, 1996) and the NGSS. The evolution of the science standards was in response to both the lessons from 10 years of implementing standards-based education, and a growing body of research on learning and teaching in science. The development of the NGSS was informed by existing documents that outlined the major ideas for K-12 science education, including the *National Science Education Standards* (NRC, 1996), the *Atlas of Science Literacy* (AAAS, 2007), the *Benchmarks for Science Literacy* (AAAS, 2009), the *Science Framework for the 2009 National Assessment of Educational Progress* (NAGB, 2008), and the *Science College Board Standards for College Success* (2009) (NRC, 2012, p. 13). There is much more to learn and, eventually, those new understandings will inspire a new group of experts to update the NGSS.

The NGSS were published in 2013. The science education community is currently focusing on their implementation. Transforming science education systems in ways that are consistent with the multi-dimensional science standards is challenging work. In many states this work is focusing on inventing systems that will better meet the needs of students who we have historically not been served well. Projects such as the Advancing Coherent and Equitable Systems of Science Education (ACESSE, or "access") and OpenSciEd appear to be the vanguard for what is to come in science education. ACESSE brings together partners from educational research and practice to promote equity and coherence in science education. OpenSciEd is developing freely available instructional materials that center equity in their Design Specification. Evidence-based strategies that support learners who come from non-dominant communities are baked into the instructional materials rather than being treated as an add on to lessons. Both are national projects, both are staffed by individuals who were either an author of the NRC *Framework*, NGSS, or both.

If changes are made to contemporary science standards, it will likely occur at the state level. Each state education agency periodically reviews, revises as appropriate, and (re)adopts their academic standards. Several NGSS adoption states have completed their review process. A survey of state science supervisors indicated that their review process resulted in either no changes or readoption with slight edits.

What degree of change is necessary for the 2028 NAEP Science Framework?

The NAEP framework needs considerable revisions. Just as previous NAEP Science Assessment Frameworks have been revised in response to literature, so should the 2028 NAEP Science Framework. The 2028 NAEP Science Framework needs to:

- Result in the development of 3-dimensional assessments. Approximately 71% of K – 12 students in the U.S. should be experiencing a 3-D science education. Maintaining the current framework will result in gathering evidence that is not relevant in states that have 3-dimensional science standards.
- Improve the content alignment of the NAEP Science Framework with the NRC *Framework* and the NGSS.

- Improve the coherence of the mathematics required in assessment items with the mathematics standards in place in the vast majority of states (e.g., the Common Core State Standards or similar standards).
- Add Crosscutting Concepts to the framework. These are important tools for students to use in figuring out phenomena, designing solutions, and making their thinking visible.
- Integrate some of the NAEP Technology and Engineering Literacy Framework with the 2028 NAEP Science Framework. Engineering, Technology, and Applications of Science are of equal importance as physical science, life science, and Earth & space sciences.
- Change the 4th grade assessment to a grade 5 assessment. The shift to the end of grade 5 coincides with the grade banding of the SEPs, DCIs, and CCCs in the NRC *Framework* and contemporary science standards
- Include features of equitable assessment. 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.

What would be lost if comparisons cannot be made to results from prior assessments?

The loss of being able to compare results between assessments would have little impact on teaching and learning science. It would be far preferable for the NAEP Science Assessment data provide a snapshot of 3-dimensional student proficiency.

Breaking the data trends would have little consequence for state education agencies. It is the Student Group Score Gaps that provide a snapshot of how well science curricula are meeting the needs of all students.

Researchers would be impacted more significantly than educators or state education agencies. However, there are a variety of data collection designs and data analysis procedures that can be used to achieve the linkages between two assessments (Dorans, 2008).

What would be lost if potential changes to the framework are limited by the need to maintain trend?

States need data that provides insights into how well students can use SEPs, DCI, and CCC, in an integrated way, to explain phenomena or to design solutions to problems. Assessing how well students know key facts, concepts, principles, laws and theories and how well students can identify and use the NAEP science practices would not be useful in states that have adopted the NGSS or embraced 3-dimensional teaching and learning.

Prioritizing trends would prevent the Governing Board from being an innovative leader in science assessment. NAEP example science assessment items have been held up as models for designing local assessment. This is no longer true in states that have adopted the NGSS or multidimensional state science standards.

Conclusions

The NGSS and other state science standards that were also based on the NRC *Framework* drive the science education for 71% of K – 12 students in the United States. This level of coherence in science education is unprecedented. These standards are unlikely to be revised in the near future.

The 2028 NAEP Science Framework needs to be substantially revised to create science assessments that are coherent with the vision of science education described in the NRC *Framework* and embodied in science standards derived from it.

Little would be lost if comparisons cannot be made to results from prior assessments. That loss would be mostly experienced by researchers. Providing relevant data is much more important to policy makers and educators.

The purpose and usefulness of NAEP science assessment data would be lost if maintaining trends is prioritized over coherence with the NRC *Framework* and the science standards that were derived from it.

Works Cited

American Association for the Advancement of Science. (1993, 2009). *Benchmarks for Science Literacy.* Washington, DC: American Association for the Advanement of Science.

American Association for the Advancement of Science. (2007). Atlas of Science Literacy, Volumes 1 and 2. Project 2061. Washington, DC

American Association for the Advancement of Science. (2009). *Benchmarks for Science Literacy.* Project 2061.

College Board. (2009). *Science College Board Standards for College Success*. New York, NY.

Dorans, N. J. (2008, August). The Practice of Comparing Scores on Different Tests. *R&D Connections*.

NAEP. (2022, February 3). *Explore Results for the 2019 NAEP Science Assessment*. Retrieved from NAEP Report Card: Science: https://www.nationsreportcard.gov/science/?grade=4

National Assessment Governing Board (2008). *Science Framework for the 2009 National Assessment of Educational Progress*. Washington, DC.

National Research Council. (1996). *National Science Education Standards.* Washington, DC: The National Academies Press.

National Research Council. (2012). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas.* Washington, DC: National Academies Press.

National Research Council. (2013). *Next Generation Science Standrds: For States, By States.* Washington, DC: National Academies Press.

NGSS Lead States. (2022, February 10). *5-PS1-1 Evidence Statements June 2015 asterisks.pdf.* Retrieved from NextGen Science : https://www.nextgenscience.org/sites/default/files/evidence_statement/black_white/5-PS1-1%20Evidence%20Statements%20June%202015%20asterisks.pdf

NJDOE. (2004). *Core Curriculum Content Standards.* Trenton, NJ: New Jersey Department of Education.

NSTA. (2022, January 31). *About the Next Generation Science Standards*. Retrieved from National Science Teaching Association (NSTA): https://ngss.nsta.org/About.aspx

WestEd and the Council of Chief State School Officers. (2019). *Science Framework for the 2019 National Assessment of Educational Progress.* Washington, DC: National Assessment Governing Board.

5-PS1-1 Matter and Its Interactions

Students who demonstrate understanding can:

5-PS1-1. Develop a model to describe that matter is made of particles too small to be seen. [Clarification Statement: Examples of evidence supporting a model could include adding air to expand a basketball, compressing air in a syringe, dissolving sugar in water, and evaporating salt water.] [Assessment Boundary: Assessment does not include the atomic-scale mechanism of evaporation and condensation or defining the unseen particles.]

The performance expectations above were developed using the following elements from the NRC document *A Framework for K-12 Science Education*:

Science and Engineering Practices	Disciplinary Core Ideas	Crosscutting Concepts
Developing and Using Models Modeling in 3–5 builds on K–2 experiences and progresses to building and revising simple models and using models to represent events and design solutions. • Develop a model to describe phenomena. (5-PS1-1)	**PS1.A: Structure and Properties of Matter** • Matter of any type can be subdivided into particles that are too small to see, but even then the matter still exists and can be detected by other means. A model showing that gases are made from matter particles that are too small to see and are moving freely around in space can explain many observations, including the inflation and shape of a balloon and the effects of air on larger particles or objects. (5-PS1-1)	**Scale, Proportion, and Quantity** • Natural objects exist from the very small to the immensely large. (5-PS1-1)

Connections to other DCIs in fifth grade: N/A

Articulation of DCIs across grade-levels: **2.PS1.A** (5-PS1-1); **2.PS1.B** (5-PS1-2); **MS.PS1.A** (5-PS1-1)

Common Core State Standards Connections:

ELA/Literacy –

RI.5.7 Draw on information from multiple print or digital sources, demonstrating the ability to locate an answer to a question quickly or to solve a problem efficiently. *(5-PS1- 1)*

Mathematics –

MP.2 Reason abstractly and quantitatively. *(5-PS1-1)*

MP.4 Model with mathematics. *(5-PS1-1)*

5.NBT.A.1 Explain patterns in the number of zeros of the product when multiplying a number by powers of 10, and explain patterns in the placement of the decimal point when adecimal is multiplied or divided by a power of 10. Use whole-number exponents to denote powers of 10. (*5-PS1-1)*

5.NF.B.7 Apply and extend previous understandings of division to divide unit fractions by whole numbers and whole numbers by unit fractions. (*5-PS1-1)*

5.MD.C.3 Recognize volume as an attribute of solid figures and understand concepts of volume measurement. *(5-PS1-1)*

5.MD.C.4 Measure volumes by counting unit cubes, using cubic cm, cubic in, cubic ft, and improvised units. *(5-PS1-1)*

(NGSS Lead States, 2022)

**Unless otherwise specified, "descriptions" referenced in the evidence statements could include but are not limited to written, oral, pictorial, and kinesthetic descriptions.*

Table 1: Evidence Statements for 5-PS1-1

Observable features of the student performance by the end of the grade:		
1	Components of the model	
	a	Students develop a model to describe* a phenomenon that includes the idea that matter is made of particles too small to be seen. In the model, students identify the relevant components for the phenomenon, including:
		i. Bulk matter (macroscopic observable matter; e.g., as sugar, air, water).
		ii. Particles of matter that are too small to be seen.
2	Relationships	
	a	In the model, students identify and describe* relevant relationships between components, including the relationships between:
		i. Bulk matter and tiny particles that cannot be seen (e.g., tiny particles of matter that cannot be seen make up bulk matter).
		ii. The behavior of a collection of many tiny particles of matter and observable phenomena involving bulk matter (e.g., an expanding balloon, evaporating liquids, substances that dissolve in a solvent, effects of wind).
3	Connections	
	a	Students use the model to describe* how matter composed of tiny particles too small to be seen can account for observable phenomena (e.g., air inflating a basketball, ice melting into water).

(NGSS Lead States, 2022)

2028 NAEP Science Framework Contemplations[1]

Eileen Parsons, Immediate Past President
National Association for Research in Science Teaching
February 11, 2022

Many science teacher preparation courses for prospective practitioners and professional development activities for practicing science teachers include nature of science (NOS) among the curriculum topics. Instruction, when based upon science education research, often features the tentative and iterative nature of scientific knowledge among NOS elements. Although scientific knowledge is durable, it is not certain. New evidence emerging from advanced technologies, novel scientific breakthroughs, altered understandings resulting from revised theoretical perspectives, and other influences linked to dynamic contexts can subject what was once canonical to further scrutiny and ultimate modification.

Consequently, this lack of absoluteness inevitably impacts constructs and domains founded upon or directly connected to scientific knowledge; scientific literacy for citizenship and for STEM workforce preparation, a goal of science education for many decades, is not an exception. Iteration and lack of absoluteness also apply to learning as a phenomenon, evinced in decades of research conducted in education, psychology, and sociology. It is upon the previously stated premise, iteration and lack of absoluteness, I base my commentary to address four posited prompts:

1) What degree of change is necessary for the 2028 NAEP Science Framework?
2) What would be lost if comparisons cannot be made to the results from prior assessments?
3) What would be lost if potential changes to the framework are limited by the need to maintain trend?
4) What is the future of science education and what implications does equity have for the NAEP Science Framework/assessment?

Science literacy is featured as an essential goal in documents discussing the National Assessment of Educational Progress (NAEP) Science Framework; in some documents, it is situated as one justification for it. Additionally, the descriptions of a scientifically literate person correspond with the explicit definition of scientific literacy provided by Organization of Economic Co-operation and Development (OECD) in the 2015 Program for International Student Assessment (PISA) Science Framework. "Scientific literacy is the ability to engage with science-related issues, and with the ideas of science, as a reflective citizen" (OECD, 2017, p. 22). Explicitly declared in some venues and implicitly practiced in others, scientific literacy is a vector of sorts, perhaps a tacit standard, in assessing what students know and can do, the purpose of NAEP science assessments. Significant changes in what is required of scientifically literate individuals justify corresponding revisions in NAEP science assessments and the frameworks that guide their development.

Although the core disciplinary science content around which science literacy coalesces is stable, the societal contexts in which an individual exercises it and how are less so. The societies in which citizens live and local, national, and global scientific and technological advancements influence the degree and

[1] This paper was produced under Governing Board contract number 919995921F0002 to Manhattan Strategy Group, with subcontract to the Human Resources Research Organization.

nature of reflectivity citizens need to productively function within and to contribute to communities, nations, and humanity writ large. The contexts in which reflective citizens engage science-related issues and scientific ideas are not static; they are ever-evolving, gradually in slight mutations over time and abruptly in stark, catastrophic, and pronounced ways.

Numerous contextual shifts—by way of scientific breakthroughs, research on learning, and societal events—have transpired since the employment of the current NAEP Science Framework in 2009. Within the short span of two years, a global pandemic; development and distribution of COVID vaccines at record-breaking speed; the increased visibility and elevation of anti-science/ anti-evidence sentiments in the United States (U.S.); and the concerning impact of algorithms in social media platforms upon the behaviors of youth are a few among several society-altering examples. The previously described, and other events not listed here, warrant a re-examination of what is needed to be a reflective citizen capable of engaging science-related issues and scientific ideas in the present time. These society-altering events and corresponding implications for scientific literacy cannot be insulated from the changing demographics in the U.S. In 2017, multilingual learners comprised 10% of the total enrollment in public elementary and high school (U. S. Department of Education, 2019), and in 2020, youth of color comprised 52% of children under age 18 (Jacobsen, 2021). Using past and present conditions to anticipate the U.S. society generations of learners will inherit, inhabit, and reflectively navigate approximately ten years into the future is one important consideration in developing the 2028 NAEP Science Framework; the National Assessment Governing Board's Framework Development Policy explicitly indicates others.

The 2018 Framework Development Policy delineates six principles; principle four suggests sources the National Assessment Governing Board should consult when updating the NAEP Science Framework (Orr, 2021). Widely adopted and implemented standards, curricula and assessments utilized at the state and local levels, and rigorous research are among the recommendations. These sources offer insights to guide and inform the necessary 2028 NAEP Science Framework updates and the extent of them.

Conditions Necessitate Change

According to the National Science Teachers Association (n.d.), the *Framework for K-12 Science Education* (National Research Council (NRC), 2012), a foundation for the *Next Generation of Science Standards (NGSS)* (National Academies Press, 2013), informs the science education standards for 44 states. Twenty states adopted *NGSS*, and 24 used the *Framework* as a blueprint to develop state-specific standards. Even though a systematic and extensive account, utilizing most recent data, of the degree to which these *Framework*-informed standards have permeated curriculum, instruction, assessments, and professional development activities has not yet emerged, research indicates the *Framework* and *NGSS* have influenced K-12 science education. Studies span the gamut—designs from experimental to qualitative investigate a diversity of domains like instruction, student learning, student participation, and identity development to name a few. Consequently, it is reasonable for the *Framework* and *NGSS* to be prominent sources in determining the nature of the 2028 NAEP Science Framework.

Neidorf et al. (2016) conducted independent analyses to compare the current NAEP Science Framework and *NGSS*. The analyses revealed areas in which the current NAEP Science Framework and *NGSS* were concordant. These overlaps may justify slight modifications to the current NAEP Science

Framework to produce the 2028 NAEP Science Framework, but I contend these tweaks would be insufficient. Though situated in similar cognitive and sociocultural traditions, the conceptual and guiding premises of the current NAEP Science Framework and present efforts are not adequately congruent. The different conceptual framings undergird where the current NAEP Science Framework and NGSS diverge. Various stakeholders illuminated several of these divergences in the public comments which are synthesized below.

- NAEP situates knowing science and doing science as separate entities whereas NGSS promotes the integration of knowing and doing science (Codere; Foster; Heinz; Keller; Learning Policy Institute; National Science Education Leadership Association; Spurlock; Tretter).
- NAEP assesses an endpoint of learning whereas NGSS presents a progression of it (Cognia; Foster).
- Crosscutting concepts are embedded in NAEP but are explicit and serve a structuring role equated conceptually to scientific thinking in *NGSS* (Cognia, Council of State Supervisors).
- NAEP emphasizes identification and recall of scientific knowledge which is reflected in multiple choice items whereas *NGSS* requires active engagement and synthesis, an integration of scientific practices and science concepts when making sense of phenomena, which is more adequately assessed by hands-on performance tasks, simulations-based tasks, etc. (Cognia; Council of State Supervisors; Georgia State University, Huntoon; National Science Education Leadership Association; Sneider).
- NAEP is based upon and assesses two dimensions whereas *NGSS* reflects three dimensions and their interactions (iterated in some manner in majority of the public comments).

The conceptual differences in the current NAEP Science Framework and *NGSS* reflect a gap in cognitive complexity; a reframing for the 2028 NAEP Science Framework appears necessary to transverse this chasm. Because the cognitive complexity featured and promoted in *NGSS* is likely to persist and become more compounded over time with the increased centrality of science, technology, engineering, and mathematics (STEM) in the quality of life and for optimal engagement in a democratic society, the potential tradeoffs are worthwhile.

Ramifications of a Revamp

Ramifications will likely result from substantial revisions, such as a conceptual reframing, to the NAEP Science Framework for 2028. I am not a psychometrician so I am unequipped to anticipate what those ramifications might be, but 2009 can be informative. In 2009 a new framework was constructed in response to numerous factors including the emergence of and attention given to the *National Science Education Standards* (NRC, 1996) and *Benchmarks for Scientific Literacy* (American Association for the Advancement of Science, 1993). Lessons learned from the 2009 substantial shift in the NAEP Science Framework may be useful in identifying the ramifications for re-envisioning the 2028 framework. Additionally, an evaluation of what was done in 2009 to mitigate or remedy adverse effects may facilitate and inform proactive efforts. One outcome of the decision to revise the NAEP Science Framework in 2009 was unequivocal: the choice disrupted the trend line associated with comparing assessment results across the 1996, 2000, and 2005 time periods. It is plausible the same disruption will occur from a consequential update of the 2028 NAEP Science Framework. That is, the validity of any comparisons across NAEP science assessments implemented in accordance with a substantially revised framework to assessments administered from 2009 to date would be highly questionable.

2028 NAEP Science Framework Contemplations

Many researchers have used the science assessment results across time as the focal point of studies or as context, establishing the state of science education, for investigations. On one hand, the use of achievement trends has aided progress; on the other, the treatment of these trends has hampered movement forward, especially as it relates to equity. Progress towards equitable science education has been hindered by the acontextual and ahistorical treatment and positioning of the achievement trends. Such a positioning often locates the sole responsibility for achievement outcomes upon individuals—students and their significant others—in lieu of situating achievement outcomes within social, economic, and political ecosystems. Perhaps, the disruption in the trend line will create an opening for the Governing Board to measure contextual factors, beyond the control of students and their significant others, that impact opportunities to learn (additional details in "Opportunity to Elevate Equity"); make the contextual data accessible to researchers, practitioners, and policymakers; and encourage stakeholders to examine the achievement outcomes in tandem with the ecosystems data.

Opportunity to Elevate Equity

Researchers, practitioners, and policymakers conceptualize, and the general populace uses equity in various ways. The World Health Organization's (n.d.) view of equity aligns with my utilization of it. "Equity is the absence of unfair, avoidable, or remediable differences among groups of people…". With respect to equity, I give prominence to underserved groups who have been denied full participation, historically and contemporarily barred from equal opportunity, equal access, and a level playing field for converting opportunity and access to actual advantage. Many of the public comments accentuate equity and advocate centering equity in assessment design, data collection, and the interpretation of results. Even though few provided explicit definitions, the public comments appeared to converge on the idea of unfair differences among groups with groups demarcated by race, gender, disability, and language.

- The NAEP Science Framework does not adequately account for group inequities resulting from access (e.g., no course offerings in the assessed content) and accessibility (e.g., barriers created by the language and discourse practices used in assessments); engagement and expression (e.g., assessment format and mode of elicitation); and representation (e.g., assessment items featuring dominant groups' ways of knowing couched within the dominant culture) (CAST; Cognia; Georgia State University; the Haverly group; Heinz; Learning Policy Institute; NARST; Petersen; Settlage; Wray).
- Equity is marginalized or absent in the purposes of the NAEP Science Framework, in its development, and in the implementation of the assessment it informs. The NAEP science assessment neglects sociocultural and cognitive research highlighting the cultural embeddedness of learning as demonstrated in the scarcity of items that are culturally relevant and culturally responsive to the increasingly diverse population under age 18 (Barber-Lester; Council of State Supervisors; Foster; the Haverly group; Heinz; Murphy; NARST; Settlage; Wray).

Centering equity, a move promoted in many of the public comments, requires the Governing Board to utilize the research on cultural embeddedness of learning and the ways culture influences demonstrations of learning as well as employ culturally responsive assessment practices. Furthermore, to embrace equity it is necessary for the Governing Board to supplement the contextual data points in the current science assessments with metrics to assess opportunities to learn. Even though it is not tailored to science education, *Monitoring Educational Equity* is a useful starting point for indicators specific to the U.S. (National Academies of Sciences, Engineering, and Medicine, 2019). The proposed indicators encompass factors pertaining to the student (e.g., self-regulation), family (e.g., resources), neighborhood (e.g., chronic stress), school (e.g., poverty concentration), community (e.g., environmental quality) and societal structures (e.g., degree of economic and racial segregation) in U.S. society.

Concluding Remarks

Retaining the current NAEP in its present form or a slightly modified version is a viable alternative. No or few inconsequential revisions would maintain the science assessment trend line starting with 2009. Continuity would be preserved, but to what end? If the science assessments do not measure what students need to know and do in the current times and in the near future, then of what value are the results?

References

American Association for the Advancement of Science (AAAS) (1993). *Project 2061: Benchmarks for science literacy*. New York: Oxford University Press.

Jacobsen, L. (2021, December 15). *A demographic overview of children under age 18 in the United States.* The National Academies of Sciences, Engineering, and Medicine Committee on Equity in preK-12 STEM Education open virtual session.

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

National Science Teachers Association (n.d.). *About the Next Generation Science Standards.* Retrieved February 10, 2022 from https://ngss.nsta.org/About.aspx

National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core Ideas*. Committee on a Conceptual Framework for New K-12 Science Education Standards. Board on Science Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

National Research Council (1996). National science education standards. Washington, DC: National Academy Press.

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*. U.S. Department of Education. Washington, DC: National Center for Educational Statistics. Retrieved February 9, 2022 from https://nces.ed.gov/nationsreportcard/subject/science/pdf/ngss_naep_technical_report.pdf

NGSS Lead States (2013). *Next Generation Science Standards*: For states, By states. Washington, DC: The National Academies Press.

OECD (2017), *PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematic, Financial Literacy and Collaborative Problem Solving*, PISA, OECD Publishing, Paris, https://doi.org/10.1787/9789264281820-en.

Orr, C. (2021). *History, policy, and decision points: Developing NAEP assessment frameworks.* https://www.nagb.gov/content/dam/nagb/en/documents/publications/frameworks/History-of-NAEP-Frameworks-Report-Final.pdf

U.S. Department of Education, National Center for Education Statistics (2019). Common Core of Data (CCD), "Local Education Agency Universe Survey," 2000-01 through 2017-18. Retrieved February 11, 2022 from https://nces.ed.gov/programs/digest/d19/tables/dt19_204.20.asp

World Health Organization (n.d.). Health equity. Retrieved February 11, 2022 from https://www.who.int/health-topics/health-equity#tab=tab_1.

National Assessment Governing Board Panel on NAEP Science Framework and Assessment

James W. Pellegrino
University of Illinois Chicago
February 11, 2022

The following responses to questions regarding the NAEP Science Framework and the 2028 NAEP Science Assessment draw from materials in: *NAEP Validity Studies White Paper: Revision of the NAEP Science Framework and Assessment*[1] (J. W. Pellegrino, October 2021).

Question 1: How would you evaluate the Validity of the NAEP Science Framework and Assessment if continued in 2028 and beyond?

Given multiple findings described in Pellegrino (2021) and summarized below, I would argue that 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 major revisions of the NAEP Science Framework and Assessment are pursued. The major threat to the validity of NAEP science involves adoption by a preponderance of states of contemporary science and technology education standards that differ substantially from the NAEP Science Framework. Those contemporary standards are either the *Next Generation Science Standards* (NGSS, 2013), or standards that are NGSS-alike given their derivation from the NRC *Framework for K-12 Science Education* (NRC, 2012) that undergirds the NGSS.

A 2016 study commissioned by NCES compared the NAEP Science Framework and the NGSS (Neidorf et al., 2016). That study found that 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. The study also showed overlap between the NAEP Framework and NGSS regarding the concept of science practices that describe ways of thinking about and reasoning with science content. However, the NAEP science practices and the NGSS science practices are different in at least two ways. 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.

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 comparison study concluded:

[1] The NAEP Validity Studies Panel was formed by the American Institutes for Research under contract with the National Center for Education Statistics.

"… 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).

Given substantial differences between the NAEP Science Framework and the NGSS, and the preponderance of states that have adopted the NGSS or similar standards, an obvious question is 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.

The pace at which standards reflecting the NGSS or the NRC Framework affect 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 often lag two or more years behind standards adoption. The most recent national survey of science education suggests that little changed between 2012 and 2018 in science instructional practice (Smith, 2020). One major factor in the slow penetration at the classroom level appears to be limited availability and implementation of professional learning programs for teachers. Since 2018, however, 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.

Assuming continued implementation of assessments, curriculum materials, instructional practices, and professional learning opportunities aligned with the contemporary standards that 45 states have adopted, it is questionable whether the NAEP Science Assessment can validly track the impact of those changes on science achievement in 2028, and even quite possibly beforehand in 2024. It remains to be seen 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. 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.

Question 2: What degree of change is necessary for the 2028 NAEP Science Framework?

Given differences between the current NAEP Science Framework and the standards adopted by most states it is likely that considerable change will be needed to bring NAEP Science in correspondence with K-12 science and instructional practice in the majority of the United States. Assuming the formation of Framework Steering and Development Panels to consider revisions and make recommendations for what should be included in a revised Science Framework for the 2028 Science Assessment, the following revisions to the Science Framework and Assessment should be considered.

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 and used in NGSS. 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 future NAEP science assessments to provide important trend information across grades in the development of core knowledge in prioritized areas of each of the three major science disciplines.

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.

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 earlier. 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.

The panels should consider the inclusion of technology and engineering content and practices, similar to their inclusion in the NRC Framework and NAEP Technology and Engineering Literacy (TEL). 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.

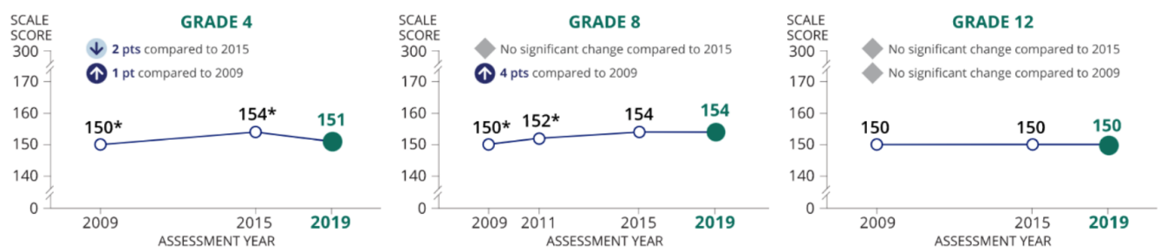
In considering the topics described above, the panels should gather the most recent information on the status of implementation and impact of current state science standards and projections for the remainder of this decade. The panels should seek information on these matters from the Board on Science Education from NASEM, the National Science Teaching 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.

Questions 3 and 4. What would be lost if comparisons cannot be made to results from prior assessments? What would be lost if potential changes to the framework are limited by the need to maintain trend?

One of the hallmarks 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. Assuming implementation of the current Science Assessment in 2024, there will be 15 years of trend data for science. Given the likely scope of a revision to the NAEP Science Framework and the implications for the 2028 assessment, including the possibility of incorporating aspects of TEL in the new framework and assessment, it seems highly likely that preserving the science trend through 2028 will not be feasible or advisable. In the Governing Board's decisions about revisions to NAEP Science 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.

To consider questions about comparing 2028 NAEP Science results to prior administrations it may prove useful to look at the Science trend results. The 2019 NAEP science scale score results are shown in the Figure below for each of the grade levels in comparison to prior results back to 2009. As can be seen in the Figure, 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 did not change 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.

While the absolute levels of scale scores and trends are important indicators of student performance, of particular significance is the reporting of results in terms of achievement levels. The rates by which students were classified into the achievement levels varied across the grades with the highest rate of *NAEP Proficient* classifications occurring in grade 4 (35%), slightly lower levels of proficiency at grade 8 (33%) and substantially lower student proficiency classifications at grade 12 (20%). Consistent with the scale score trends, small changes have been observed in achievement level performance across time at grades 4 and 8. At all three grade levels there has been a consistent and very low level of classification of student performance at the *NAEP Advanced* level across years and the level of students performing below *NAEP Basic* has remained substantial ranging from 27% at grade 4 to 40% at grade 12.

Some important considerations follow from this brief examination of the NAEP Science Assessment results and trends over time. First, not much has changed over time in student performance 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. This conclusion is not inconsistent with the results mentioned above about the slow pace of adoption and implementation of the new standards in terms of state assessments and classroom instructional practice over that same time period. It should also be noted that despite 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.

Imagine what we might see in 2028, and perhaps even 2024, if instruction and assessment aligned with the new science standards has further taken hold since 2018 in ways envisioned in the NRC Framework and NGSS. Trend results based on the current Science Framework and Assessment might be expected to show no growth, and quite possibly some decline, in scale score performance and proficiency estimates given significant differences between what and how NAEP assesses science proficiency relative to assessment and instruction of science in the majority of states. NAEP could seriously underestimate what students know and can do in science at each grade level given content differences that have been shown to increase across grade levels and that are especially large at grade 12. Regardless of which direction the trend results go, will we be able to make sense of them given what has been happening in U.S. science teaching, learning and assessment over the 16-year period since 2012? Without revision to the NAEP Science Framework and Assessment it is doubtful that the NAEP Science Assessment will be able to serve as a valid indicator of what students know and are able to do in science at grades 4, 8 and 12 given the differences between the NAEP Framework, the NGSS and the majority of state science standards. It seems highly likely that NAEP Science will not have sufficient instructional sensitivity to reveal what has and has not happened over time in science education and student competency when administered in 2028, and quite possibly in 2024.

The comparisons that can be made across years and their interpretability will ultimately depend on the nature and scope of the revisions made to the current NAEP Science Framework and Assessment between now and the 2024 and 2028 administrations of NAEP Science, especially the 2028 administration. That said, I would only repeat what was noted at the beginning of this document. Serious concerns exist about the capacity of the NAEP Science Assessment to fulfill its mission to provide valid and reliable information about the status and progress of science achievement in the United States unless major revisions of the NAEP Science Framework and Assessment are pursued.

In Consideration of Changes to the NAEP Science Framework for 2028[1]

Eric J. Pyle, President
National Science Teaching Association
February 11, 2022

It is clear that substantive changes to a document that has such a profound influence on education policy as the NAEP Science Framework is no simple task, nor is it to be considered lightly in terms of cost and potential disruption. That said, there are ample reasons, including the diffusion and adoption of *NGSS* or standards influenced by the majority of states, to consider seriously that change may not only be necessary, but vital in order for the NAEP science assessment to retain the validity necessary for meaningful impact on policy development. It is also important to note that the policy-scape of science education does not exist independently, and that recent social and political disruptions makes the establishment of validity paramount if NAEP assessment results are to fulfill their purpose. It is equally important to note that the science education policy-scape is influenced by a number of factors that extend beyond statements of curricula and assessment congruence, such as the growing understanding of how diverse students learn, how teachers of science are prepared and find professional learning, and the role of persistent challenges in basic school resource and student preparation. Considering the role of each of these elements, and others, defies a simple linear or reductionist approach to addressing inequity and effectively monitoring student growth in science literacy. In this essay, I will attempt to address each of the four posed questions where applicable in the context of contemporary science teaching, with an emphasis placed on the activities of the National Science Teaching Association (NSTA) and its affiliates.

Question 1: What degree of change is necessary for the 2028 NAEP Science Framework?

As is noted repeatedly in the public comments and the insight of professional organizations in science education, substantial changes are needed for the 2028 NAEP to reflect the performance expectations of current curricula as well as near future developments. The current NAEP Science Framework is derived from the *National Science Education Standards (NSES)* from 1996, but almost as soon as the 2009 NAEP assessment was administered, the *NSES* were supplanted by the development of *A Framework for K-12 Science Education* (abbreviated henceforth as *AFK12SE*) commenced, convening of design teams centered on Life Science, Physical Science, Earth & Space Science, and Engineering. Each team was charged with isolating the essential "big ideas" in each domain. As a result, substantially fewer disciplinary content ideas (DCIs) were expressed than *NSES*, but these were integrated into robust performance expectations in the subsequent *Next Generation Science Standards (NGSS)* with science & engineering practices (SEPs) and cross-cutting concepts (CCCs) to providing relevance and the integration of those ideas and practices, which the 2009 NAEP framework does not directly support as interactive elements. Each dimension (crosscutting concepts, core ideas, and practices) is supported by well-defined learning progressions, defining the expected growth of student learning over time. Science proficiency is not just about science facts that are known, or scientific methods that can be emulated, but rather the sensemaking that can be found through integrating each element as a part of sustained learning experiences.

[1] This paper was produced under Governing Board contract number 919995921F0002 to Manhattan Strategy Group, with subcontract to the Human Resources Research Organization.

Currently, there are a total of 44 states and the District of Columbia that have either adopted or adapted NGSS, or created similar standards based on *AFK12SE*. Non-adopting/adapting states have had the development of their newer standards influenced by NGSS, making them "crypto-adapters." It is important to note that in the context of state legislative code, standards become very stable and resistant to substantive change. They become the drivers for arguably more variable state assessment blueprints and malleable classroom curriculum frameworks. This was anticipated by the developers of the *AFK12SE* and the *NGSS* in their use of the performance expectations -- that there would be different approaches and perhaps more locally meaningful representations of the same practices, ideas, and cross-cutting concepts.

A prime example of this local variation in adoption is the case of Earth & Space Science (ESS) standards. Traditionally, ESS concepts have been primarily represented in the middle grades, and where offered in high school, Earth science courses have largely been limited to 9th grade, with almost no states requiring Earth science as a graduation requirement. The adoption of *AFK12SE* as a driver for state curricular standards presents a special problem, though, for the framers intended the ESS performance expectations to influence the design of capstone science courses in high school. California devised a third option, integrating ESS performance expectations into existing biology, chemistry, and physics coursework in the so-called "three course" model. This in turn creates disruptions in the typical sequence in these other courses and attendant professional development needs for those teachers. Data sources supporting the efficacy of this approach are difficult to find and instructional materials in support are not yet widely available.

Combined, therefore, the historical approach to K-12 science curricula, emphasizing different sciences at different grade bands, is reflected in the current NAEP Science Framework, and has supported the assessment of science content and science processes as separate unintegrated elements. The approach provided by *AFK12SE* and *NGSS*, however, supports instruction all sciences across all grade bands, and in a manner that integrates science content and science & engineering practices with concepts that cross-cut content and practices and provide relevance to larger ideas in science. In addition, *AFK12SE* includes engineering on an equal standing with the sciences, which is absent from the current NAEP Science Framework. Thus, assessments based on the 2009 framework would unable to detect the changes in science classrooms currently seen or anticipated. If the next generation of NAEP instruments is to provide significant value, the 2028 NAEP Science Framework should be updated in anticipation of the future of science classroom, and the complex factors influencing policies, implementation, assessment, and equity.

Question 4: Current and Future Science Teaching

Pedagogy

With respect to the status of classroom science teaching, there is a bifurcation between aspirational pedagogies supported by research and pedagogies constrained by the impact of high-stakes testing at the state level. Over the last three decades, there has been a constant push from leaders in science education (based on research on teaching and learning) to move away from teacher-centered direct modes of science instruction in favor of more student-centered active approaches. "Inquiry" based pedagogies have been promoted since the publication of *NSES*, and continue to be the subject of continued pedagogical research. One constraint to implementing these pedagogies is based on teacher confidence with the subject matter or the nature of science. Innovations such as the BSCS Learning

Cycle, and its multiple variations, presented in preparation and professional learning settings can help build the needed confidence. Continued research into "sensemaking" in science instructional design builds upon and extends inquiry models, with an emphasis on student dialogue and equity. This is an emergent area, and is reflected in the most recent offerings from NSTA (https://www.nsta.org/sensemaking). But while innovative and built upon decades of research into how people actually learn science, there is a strong potential for implementation to be uneven, given the differences in time and financial resources available to aid teachers in adjusting or adapting their pedagogical approaches.

Current science pedagogy is also influenced by state-mandated tests, which in some ways pushes against the trend toward student-centered learning. First, there is a desire to maintain or increase science test scores, and any activity that would hinder that goal is devalued by school administration and may be potentially threatening to teachers, who in many cases have their students' test performance used as a factor in their annual evaluation. Large chunks of instructional time have been focused on preparation for these tests, particularly at the secondary level. This leads to more direct instruction specific to test blueprints. One could argue that "teaching to the test" is merely teaching the curriculum, but it also puts ceiling on possible student performance, as well. Second, if science is not tested, such as in elementary grades, there is a higher likelihood of diminished time dedicated to science teaching. A trend toward less time for science in elementary has implications for NAEP results in 4th grade and may result in poorer performance in middle and high school over the longer term.

Assessment

This gap between aspirational and constrained pedagogies is a confounding factor to NAEP science assessments, in that revising the science framework to more closely align with AFK12SE could exacerbate the performance differences between schools where aspirational pedagogies are favored and those schools that are more constrained by high-stakes state assessments. Currently, these state assessments show differences in student performance between well- and under-resourced schools. While these assessments are not designed to detect differences in teacher professional learning, it is a reasonable inference to make that students of highly-qualified teachers perform better on such tests, and that investment in teacher learning professional is one means by which teacher qualifications can be enhanced.

One cannot underestimate the outsized influence of state assessments on the pedagogical choices that teachers make or are expected by their administrators. Government policy has slowly evolved to accommodate flexibility in high-stakes test design, moving from the commonplace multiple-choice to constructed response items, and from primarily paper-pencil tests to online platforms. Regardless, as curricula move more towards an *NGSS*-based design, the development of assessments that follow a similar design lag behind. This alignment is slowly occurring, but there is a lag between standards adoption, implementation, and assessment across all elements of the system. As a result, classroom practice may not yet be fully transformed, but more progress is likely to occur before 2028, when the new NAEP framework will be used. It would be vital, therefore, to anticipate such changes in the 2028 NAEP Science Framework, rather than lag even further behind the evolution of science teaching.

Preparation Programs

One of the central drivers for the design of science teacher preparation programs has been the accreditation frameworks under which they must operate. "Accreditation," or the assurance of quality, takes on many forms ranging from approval by state departments of education to external accrediting bodies, such as the Council for the Accreditation of Educator Preparation (CAEP). In these circumstances, the role of science preparation is most likely focused on the broad expectations of state curricula. Some states (and colleges) require a deeper review, leading to national recognition of programs by specialist professional associations (SPAs), using content & grade level-specific standards. NSTA has served as the SPA for secondary science programs, although they have also provided national standards for middle grades science preparation programs as well as elementary science specialists. The exact impact of these standards on new teachers is extraordinarily difficult to quantify, but working with available NSTA SPA reports, Title II teacher licensure data, and US Bureau of Labor Statistics, a best estimate of 10% of new secondary science teachers experienced a preparation program that was based on these standards. This estimate does not take into account the many programs that were not required to seek national recognition or state-level approval. In addition, the NSTA Elementary teacher preparation standards have only been available for the last two years, so as preparation programs adjust their curricula, these standards will likely see greater use. NSTA has anticipated that *NGSS* and *AFK12SE* will continue to have uptake and penetration of instructional approaches, and preparation curricula aligned to the NGSS is likely to continue. This underscores the need for the 2028 NAEP Science Framework to change as well, to better represent the learning by the students of multiple cohorts of new teachers prepared by programs informed by *AFK12SE*.

NSTA's SPA review standards were revised several times since their inception, with 2012 the last date before the release of *AFK12SE* and *NGSS*. Because of the incompatibility of these prior standards with the new documents, NSTA and ASTE (Association for Science Teacher Education) initiated a joint effort to update the preparation standards. They were closely matched to *AFK12SE* grade bands at all levels. Approved by both organizations' Boards of Directors in 2018, they were publicly released soon afterwards. Despite NSTA suspending its relationship with CAEP in 2019, it is clear from recent meeting presentations that they have been found by and are being employed by science teacher preparation programs. Over the long-term, as programs seek state-level accreditation, the use of standards prepared and promoted by national organizations, aligned to *AFK12SE*, strengthen their autonomy to seek innovative ways to prepare new teachers of science. NSTA has also invested considerable resources in providing learning materials for preservice teachers of science, currently counting more than 1200 such students nationwide using these materials directly in 2021-22 alone.

The present time is also a critical juncture for science teacher preparation. Many university-based programs have indicated substantial declines in enrollments, at a time when many senior teachers are retiring from the profession. Mining Title II data indicates a steady decline in the number of new secondary science licenses granted, down approximately 20% since 2013 across science content areas, with an uneven distribution across the country. Some studies have indicated that more than half of new teachers leave the field within six years. The impact of the COVID on education represents a substantial disruption to the profession, as high numbers of teachers plan to leave the classroom in the next few years. Thus, there is a consistent but not even need for new teachers of science while the supply of qualified teachers of science declines.

Even as schools emerge from the pandemic, it is likely that by 2028, the average teacher of science will be younger, less experienced, and less likely to stay in the field. At the same time, science teaching methods courses continue to emphasize student-centered 3-dimensional science instruction, with *NGSS*-based lesson planning a key feature. These new teachers are also more likely to work collaboratively if opportunities exist and are much more connected as digital natives through a variety of hand-held platforms. This creates opportunities for professional learning in their careers.

Professional Learning

One of the factors known to support and sustain teachers of science is high-quality professional learning. While school districts may provide professional development, this is often generic to specific issues of the district and not necessarily content specific. When it comes to science, the most valued professional learning opportunities are those that are science-specific, emphasizing pedagogical content knowledge, and are collaborative in nature. In addition, many funding sources, such as NSF, require that professional learning be more than one-shot opportunities, supported by research that shows effective professional learning is sustained over time. Professional learning opportunities are central to helping teachers adopt new approaches, but access to these new approaches has been uneven across different regions, and across time, putting some constraints on the penetration of *AFK12SE*-informed innovations in the classroom. This is a confounding factor that would need to be accounted in the 2028 NAEP Science Framework.

Like many content areas, science professional learning has been provided through face-to-face conferences at the state, regional, and national level. These opportunities, however, have been in decline for several years, in large part due to increasing costs, diminishing professional learning funds, and a reluctance by many teachers to be away from their classes over absences potential negative impact on student test scores. The COVID pandemic forced a complete shutdown of conference opportunities, are only now slow returning to operation. The general availability of substitute teachers insufficient to cover teacher illnesses, let alone allowing teachers to attend conferences, has only magnified the impact of the last two years on teacher professional learning. In the short term, many organizations turned towards virtual offerings, which while offering high quality content, lack the general spontaneity of conferences and are constrained by Internet bandwidth. NSTA has responded with providing virtual conference offerings, as well as thematic webinars on multicultural science teaching and social justice and science teaching. It is quite evident, however, that “zoom saturation” has impacted participant satisfaction. Making recordings available for future viewing, however, has been supportive of teachers shifting schedules. NSTA has continued to innovate in professional learning, through district- and school-partnership programs, reaching blocks of teachers at once, and has recently introduced online professional learning units. All of these adaptations aside, there is a not yet solid information on the magnitude of professional learning needs unmet at the current time, and thus this aspect of “recovery” is not yet fully defined. Supporting and retaining teachers of science in the run-up to the next NAEP will require a comprehensive plan for addressing these needs, one which NSTA is positioned to support.

Equity

The issues of equity were raised by student performance over the last to administrations of NAEP, such that there was either flat performance lines or widening gaps between groups based on normal distinctions in diversity. The social protests of 2020 and the ongoing pandemic have made issues of equity quite prominent, and the complexity of potential interacting variables compound understanding equity and its role in student NAEP performance. For students, there remain equity of access issues to high-quality instructional materials and learning experiences, particularly those informed by or modeled after *AFK12SE*. There are also access equity issues for those students to have diverse, highly qualified teachers prepared in preparation programs that adhere to national standards for science teacher preparation (as described above). Those same teachers need consistent professional learning experiences that match their students evolving needs. And for both students and teachers, there remain equity access to contemporary technologies for data collection and analysis, such as data loggers and probeware. Online applications, databases, and remote learning platforms, heavily dependent on Internet bandwidth and wireless access are not readily available in all areas, particularly in rural areas or in areas where even basic computing technology is unavailable. Without such access, remote learning options and access to advanced science classes in domains where qualified teachers are thin on the ground will remain a persistent issue, and existing differences in NAEP performance will continue to grow.

Despite these difficulties, there is much promise that has come from understanding the social and neurological processes underpinning learning, including science. From a practitioner standpoint, the Universal Design for Learning (UDL) framework outlines the variety of instructional approaches that can support inclusive learning environments for all students, and to the extent that UDL frameworks are mindfully employed in both instruction and assessment, the greater the pressure on NAEP assessments to adjust in turn and provide more equitable results. NSTA has developed professional learning opportunities and instructional frameworks build around Sensemaking, which requires equity of student voice and opportunity across all aspects. NSTA has offered a series of multi-part webinars on multicultural science education and social justice in the science classroom, with strong participation synchronously and asynchronously. They have outlined the multiple variables, conditions, and circumstances that make directly addressing equity in learning such a complex issue. The bottom line is that there is no one, or one sequence of events, presented in a linear or sequential fashion, that would instantly produce equitable learning environments. Instead, the accumulation of small effects, consistently applied and measured by the cumulative effects, is most likely to result in an emergent pattern of equitable learning. Thus, the NAEP science framework must either anticipate either differences in opportunity and access from the start, or work from a position of emergent patterns, if it is to have value in directing policy and resources.

Questions 2 & 3

2. What would be lost if comparisons cannot be made to results from prior assessments?
3. What would be lost if potential changes to the framework are limited by the need to maintain trend?

Because the 2009 NAEP Science Framework has already been employed several times, there is an established trend line, based on prior models of teaching and learning science. There are factorial analyses that would have to be reconsidered in any future use of this framework, given the different conditions and expectations faced by teachers of science, some of which enhance and others hinder student learning. Policies based on results from continuing trend from the 2009 NAEP science framework would have little practical value, since they would be based on already obsolete curricula. As described above, the situation is much more complex and influenced not just by new curricula, but changes in teacher preparation, uneven distribution of resources for teacher professional learning in science, and the slow evolution of state assessments in science to keep pace with curricular innovations. Maintaining trend with the existing science framework would deprive policymakers and educators of a richer understanding of not just the impact of curricular innovations, but a deeper appreciation of existing inequities in the resource distribution to support these innovations and enhance student learning opportunities in science.

In conclusion, it is not sufficient to note the multiple divergences between the 2009 NAEP Science Framework and *AFK12SE* and *NGSS* in building an argument for substantial changes in the 2028 Science Framework, but also the philosophical differences between the two and the flexible and variable implementation at the classroom level of *NGSS*. There are multiple variables, interactions of variables, positive and negative feedback loops, and emergent patterns performance that would be undetectable by maintaining the existing NAEP Science Framework. These same divergences create the opportunity for devising an assessment framework much more likely to be informed by these variables and inform more effective and meaningful educational policies.