2028 NAEP Science Assessment Framework Update: Feedback from Public Comment

Overview

The current update of the 2028 NAEP Science Assessment Framework underway is the first conducted under the Board's revised policy for Assessment Framework Development. During the May 2022 quarterly meeting, the Board unanimously adopted a charge to the Steering and Development Panels, describing key issues and initial guidance for the framework update.

The Board charge was developed following review and discussion of feedback gathered during an initial call for public comment and commissioned papers from science education experts on whether and how the current NAEP Science Framework (last updated in 2005) should be changed (additional information was included in the November 2021 and March 2022 Board meeting materials). In comparison with framework updates conducted under the previous Board policy, broader input was gathered at the beginning of the process to update the 2028 NAEP Science Framework.

In accordance with the new policy, the Board conducted an open call for panelist nominations during summer 2022, with support from Widmeyer/Finn Partners, the Board's Science Framework Strategic Communications contractor. Extensive and targeted outreach was conducted to hundreds of stakeholder groups and individuals from education, policy, industry, assessment, research and other science-related areas, in order to ensure representation of diverse backgrounds and perspectives on science education and assessment. The recruitment resulted in 120 applications.

Another change from previous framework updates is that the panelists were tasked with focusing primarily on developing a substantive outline of the framework (what is to be assessed and how). Project staff/consultants are taking the lead on filling in some of the background and rationale for the assessment framework recommendations, which will then be reviewed and edited by panel members. In previous framework updates, panelists spent considerable time drafting and revising the narrative text. The substantive outline was the basis for public comment, with limited narrative text serving as a working draft of the framework. Conducting public comment on a working draft of the framework earlier in the process was intended to make it easier to incorporate substantive feedback, compared to waiting until there is a near final document. One consequence of this approach, however, was that there were certain details the panel did not yet have a chance to discuss or fully address in the initial working draft.

Public Comment

NAEP is an important tool for education and policy leaders, and the frameworks determine what is measured by each NAEP assessment. The purpose of the formal public comment period is to disseminate information about the framework recommendations to a wide range of stakeholders with multiple perspectives and to provide the opportunity for submitting feedback.

It is critical that feedback on NAEP assessment frameworks is solicited from a diverse group of stakeholders while protecting the integrity of the process and the role of both panelists (recommendations) and Board members (approval). Board staff made several changes to how public comment was conducted in comparison with other recent updates to NAEP assessment frameworks, largely based on recommendations from the Board's strategic communications contractor for the science framework, Widmeyer/FINN Partners.

Public comment took place between March 13 – April 17. In advance of public comment opening, the Governing Board conducted outreach to over 700 individuals and organizations to notify them about the upcoming opportunities to provide feedback and to learn more by registering for one of the informational webinars. Information was disseminated through half a dozen email blasts; the Governing Board monthly newsletter that reaches over 12,000 individuals nationwide; and through 49 posts on the Board's Facebook, LinkedIn, and Twitter handles that resulted in a potential reach of more than 493,000 people and 1,088 engagements (including likes, shares, and comments). Board members, staff, contractors, panelists, and advisory committee members also forwarded to their networks, including organizations of which they are members. Organizations who co-hosted webinars with the Board (see public comment summary) also disseminated information directly to their networks. Finally, a notice was posted in the Federal Register.

The project website (www.naepframeworkupdate.org) was used for the public comment process. A PDF of the framework working draft was posted at 12:00 a.m. ET on March 13, along with a one-page information sheet (about NAEP, the Governing Board, and the framework revision process), and a structured form to submit feedback. The structured form represented a change from previous collections of public comment for NAEP assessment frameworks in which respondents were instructed to send an email with their feedback and/or upload comments on a word document in tracked changes. When public comment is completely open-ended, it is generally the case that few respondents comment on each issue or theme and it can be difficult to interpret whether silence means the respondent did not have an opinion or whether it did not occur to them to comment on a particular issue or question that they would have responded to if they had been prompted. The structured form also was intended to make it easier to summarize the feedback received with less need for interpretation, as well as to ensure that feedback would be as comprehensive and useful as possible. The last question asked respondents to include any other feedback or comments they had to capture additional input and avoid constraining the responses. The instructions for submitting feedback and questions that were included on the feedback form can be found at the beginning of the working draft.

A single informational slide deck was prepared for use in eight webinars and five in-person presentations conducted between late March and mid-April. For each presentation, Board staff Sharyn Rosenberg presented information about NAEP, the Governing Board, and the process of updating the framework; rotating Development Panel members presented a high-level overview of the framework recommendations; and audience members had the opportunity to ask questions. Attendees were notified upfront that the purpose of the presentations was to provide information about the process and recommendations and encourage the submission of feedback through the

official form. Following each webinar, registrants and attendees received an email thanking them for their interest in the NAEP Science Assessment Framework and directing them to the project website for the official submission of feedback.

Excluding project staff and panelists, approximately 625 people registered for one of the webinars or in-person presentations on the framework recommendations; approximately 300 people attended one of the webinars or in-person conference presentations. The webinar registrants included: representatives of Departments of Education in 42 out of 50 states; teachers, school staff, and district staff; policymakers; researchers and professors in science and science education; assessment specialists; curriculum specialists; business representatives; parents; and other members of the general public.

The feedback form was removed from the project website at midnight on April 17, and a total of 29 responses were received from the public. A summary of the feedback, along with the raw comments by question, are included in this attachment.

Concurrent with public comment, Board staff asked NCES to review the working draft framework from an operational perspective, recognizing that some of the information needed to implement the framework recommendations is not yet available at this preliminary stage of the process and will require additional input. A memo from NCES Acting Associate Commissioner Daniel McGrath is also included in this attachment.

Next Steps

The Development Panel met virtually on May 2 to discuss initial plans for addressing the feedback received from NCES and the public. In accordance with the Board policy, the Assessment Development Committee (ADC) will exercise its responsibility to monitor framework development activities and provide direction to the framework panels by reviewing the feedback received and discussing potential policy guidance to provide to the Development Panel as they work to respond to the comments. **During the May plenary session, the Panel Leadership Team will present a brief overview of the framework recommendations and preliminary plans to respond to the feedback received. ADC Chair Patrick Kelly will share ADC recommendations for policy guidance and will moderate Board member discussion.**

The Development Panel will continue to work on framework revisions in small groups and will meet in-person in Washington, DC on June 5-6 to engage in large group discussions, followed by a webinar with the full Steering Panel on July 10. A revised draft of the framework will be prepared by mid-July for inclusion in the August quarterly Board meeting materials. Additional revisions will be made to address any outstanding issues and based on discussions at the August Board meeting, with the intention to provide a final document for Board action at the November 2023 quarterly Board meeting.

Additional Background on the Science Framework Panels and Project

The current Board policy charges the ADC with recommending a slate of panelists for approval by the Executive Committee. The process and criteria for assembling a slate of Steering and Development Panel members that balanced and optimized many different factors was discussed during the August 2022 ADC meeting, and Board staff and contractors provided support to ADC to finalize their recommendation of panelists to put forward to the Executive Committee in late August 2022. The Board evaluated applications with the goal of constructing a balanced panel of stakeholders with diverse perspectives on issues relevant to the Board charge.

The following factors were prioritized in constructing a balanced panel: individuals specifically nominated to represent a national organization, given the critical need to engage various constituencies; panelist role; experience and expertise overall and the specific sub-content areas covered by the framework; demographic characteristics, including race, gender, and geography; previous experience with and stance on the Next Generation Science Standards (NGSS), including both NGSS developers and critics, and practitioners in states that have adopted NGSS standards, NGSS-alike standards, and non-NGSS standards; and diverse perspectives on issues relevant to the Board charge. The Executive Committee met by webinar on August 29 and unanimously approved the proposed slate of panelists and alternates put forward by ADC. All 30 invited panelists agreed to participate on the Development and/or Steering Panels.

The role of the Steering Panel is to formulate high-level guidance about the state of the field and how to implement the Board charge; the role of the Development Panel is to develop the content of the framework and specifications documents. The Development Panel engages in detailed deliberations about how issues outlined in the Board charge and Steering Panel discussions should be reflected in a recommended framework. Board policy specifies that the Steering Panel should include 30 members, of which 20 members continue as the Development Panel.

In July 2022, the Board awarded contract number 91995922C0001 to WestEd (as the result of a competitive bidding process) to carry out the process of recommending updates to the current NAEP Science Assessment Framework. The Project Management Team consists of Mark Loveland, Taunya Nesin, Steve Schneider, Marianne Perie, and Megan Schneider. As project director, Mark Loveland provides day-to-day leadership, guidance, and liaising with the Governing Board. Project Director, Mark Loveland, and Science Content Lead, Taunya Nesin, have oversight for all programmatic activities. Steve Schneider serves as a senior advisor to project activities. A panel leadership team of four work with WestEd and Board staff to plan meetings, facilitate panel discussions, and represent the panel's work to the Governing Board. Together, they and Dr. Nesin are leading the Steering and Development Panel activities, and Dr. Nesin also coordinates the Educator Advisory Committee (EAC). Measurement Lead, Dr. Perie, coordinates the Technical Advisory Committee (TAC). Ms. Schneider serves as Project Manager, documenting all project activities. In addition to the project leaders, the broader project team includes additional science subject matter experts, members of the science measurement team, project coordinators, and research assistants. Additional information about the project team and participants in the framework update can be found at: www.naepframeworkupdate.org.

The Board policy does not include any explicit guidance on the panel leadership structure, but previous NAEP framework panels have typically had a chair or two co-chairs. Board staff proposed, and ADC agreed, that the 2028 NAEP Science Framework Panels would not have a single individual designated as chair; instead, four members of the Development Panel serve as a panel leadership team. The rationale for this change is to ensure that a variety of backgrounds and diverse views be represented in the panel leadership; achieving balance on multiple factors is much more difficult when a single individual is designated as the panel leader. Members of the panel leadership team share responsibility for facilitating panel meetings, working towards panel consensus, and presenting to the Board. The four members of the panel leadership team are: Aneesha Badrinarayan, Jenny Christian, Nancy Hopkins-Evans, and Joseph Krajcik. Their biographies are included in this attachment.

Development of Recommendations to Update the 2028 NAEP Science Assessment Framework

On October 17-18, all 30 members of the Steering Panel met (in Washington, DC with a few panelists participating virtually via Zoom) to begin the process of recommending updates to the framework. ADC Chair Patrick Kelly delivered the Board charge, and Assistant Director for Assessment Development Sharyn Rosenberg provided other parameters and guidance in accordance with Board policies and the NAEP legislation. NCES Item Development Lead Nadia McLaughlin presented information about the current NAEP science assessment. WestEd staff presented background information and facilitated the meeting, which included several opportunities for panelists to discuss substantive issues both in small groups and as a full group. Panelists generated several initial recommendations and identified areas for further discussion and resolution by the Development Panel. A summary of the initial recommendations from the Steering Panel was presented to the Board during the November 2022 quarterly meeting.

All 20 members of the Development Panel met in person in Washington, DC on December 12-13, and January 26-27. In addition, several virtual panel meetings took place between November 2022 and March 2023. Panel members worked in small groups between meetings to generate content for individual sections of the framework, which was then discussed and deliberated by the larger group. Members of the Technical Advisory Committee and Educator Advisory Committee took turns attending panel meetings and listening and contributing to the discussions. Key takeaways from the TAC and EAC meetings were communicated back to the panel. Panelists worked to finalize recommendations to put forth during the formal public comment period. Prior to the opening of the formal public comment period, framework recommendations were shared with the Board during the March 2023 quarterly Board meeting.

2028 NAEP Science Assessment Framework Panel Leadership Team

Aneesha

Aneesha Badrinarayan, Panel Leadership Team
Director of State Performance Assessment Initiatives
Learning Policy Institute

Aneesha Badrinarayan leads projects related to state performance assessments. 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 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. Her portfolio includes leading several multi-state teams of leaders and experts to redefine "alignment" in the era of new state standards; developing criteria for innovative large-scale and classroom assessments; providing professional learning and strategic guidance for state leaders; and conducting analyses of state, local, and expert efforts to design and implement performance assessments and systems of assessment in science.

Badrinarayan earned a M.S. in Neuroscience at the University of Michigan, where she served as a research fellow for the National Institute of Mental Health, and a B.A. in biology from Cornell University.

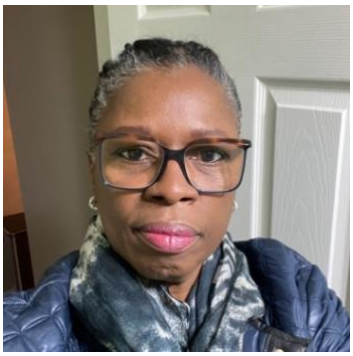
Jenny

Ms. Jenny Christian, Panel Leadership Team
STEM Director of Science and Wellness
Council of the Great City Schools District Representative
Dallas Independent School District

Jenny Christian is the STEM Director of Science & Wellness in Dallas Independent School District. Dallas ISD comprises 384 square miles and encompasses 16 cities, including Dallas. The district is the second-largest public school district in the state, and the 14th-largest district in the nation. The school district serves approximately 160,000 students in pre-kindergarten through the 12th grade, in 227 schools, employing nearly 20,000 dedicated professionals.

Raised on the border of Mexico, Jenny has served in multiple teacher and administrative roles in seven school districts, over the past 27 years. She has her master's degree in Aerospace Studies from the Odegard School of Aerospace Sciences. She has served as a Space Science Consultant at Brooks City-Base in San Antonio, and as a Flight Director for the Challenger Space Center. She has also contributed as an active panel member on STEM education advisory councils for NASA's Network of States, the Girl Scouts, and the National Urban Wellness Coalition Steering Committee.

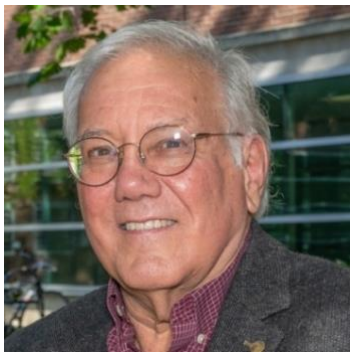
2028 NAEP Science Assessment Framework Update Panel Leadership Team



Nancy

Dr. Nancy Hopkins-Evans, Panel Leadership Team
Associate Director for Program Impact BSCS
Science Learning

Nancy Hopkins-Evans is the Associate Director for Program Impact at BSCS Science Learning. As a former college chemistry professor, she understands and cares deeply about students having exceptional learning experiences in science that leverage their communities and cultures while building conceptual understanding as they figure out science ideas instead of learning about science through memorization of facts and theories. She has worked in large and small school systems developing and implementing curriculum, professional learning and assessment aligned to state standards, the common core state standards, and the Next Generation Science Standards. She presents at conferences and leads professional learning for teachers, principals, directors, and superintendents focused on experiences and activities that support effective teaching and learning for ALL students particularly those from under-served and under-estimated communities. She recently served on a National Academies of Sciences, Engineering and Medicine committee to develop the consensus study report entitled, Call to Action for Science Education, Building Opportunity for the Future. She holds degrees in chemistry from Chestnut Hill College and Villanova University and earned a Ph.D. in biological chemistry from the University of Michigan.



Joe

Prof. Joseph Krajcik, Panel Leadership Team
Lappan-Phillips Professor of Science Education
Michigan State University College of Education

Joseph Krajcik serves as director of the CREATE for STEM Institute and is the Lappan-Phillips Professor of Science Education and a University Distinguished Professor at Michigan State University. Throughout his career, Joe has collaborated with colleagues and science teachers to design and test project-based learning environments to improve teaching practices and to research student learning and engagement. Joe has also investigated the design of formative assessment to promote student learning and recently, he has explored the use of machine scoring to assess open-ended assessment tasks. Joe served as president NARST from which he received the Distinguished Contributions to Science Education Through Research Award in 2010. He served as lead writer for developing Physical Science Standards for the NGSS and the lead writer for the Physical Science Design team for the Framework for K-12 Science Education. In 2020, Joe was elected to the National Academy of Education and received the prestigious McGraw Prize for Innovation in Pre-K-12 Education and in 2021, the International Society for Design and Development in Education Prize for Excellence in Educational Design. He has published over 100 peer reviewed manuscripts and his book on Project-based Learning is in its fifth edition.

IES Institute of Education Sciences

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

To: Lesley Muldoon
Executive Director
National Assessment Governing Board

From: Daniel McGrath
Delegated Authority of Associate Commissioner
National Center for Education Statistics
Institute of Education Sciences
U.S. Department of Education

Date: April 28, 2023

Re: NCES Review of the Working Draft Science Framework for the 2028 National Assessment of Educational Progress (Version Released for Public Comment, March 2023)

This memorandum summarizes the technical review of the working draft Science Framework for the 2028 National Assessment of Educational Progress (NAEP) by the National Center for Education Statistics (NCES). NCES appreciates the opportunity to review the framework and commends the work of the Development Panel to update the current science framework and to ground it in contemporary research and practice.

Following the adoption of this framework by the National Assessment Governing Board, NCES will operationalize the new framework as specified by the Education Sciences Reform Act (P.L. 107-279). To this end, NCES staff and contractors have reviewed the draft framework carefully and have identified areas that pose technical or operational challenges, or that require clarification to facilitate our work to implement the 2028 NAEP science framework. NCES recognizes that the draft that was issued for public comment is an early working draft.

We have organized our comments around four themes:

- Design feasibility
- Cost implications
- Trend implications
- Clarifications needed to support framework implementation

Design Feasibility

Expanded scope of content to be assessed. Compared to the current NAEP science framework, the draft framework is broader and reflects an expansion of content expectations compared to the current science framework. Within the Disciplinary Concepts (DC) dimension, there are many more concept statements in the draft framework than there are content statements in the current framework, as shown in Table 1, below. This expansion reflects additional and elaborated content and, by extension, an expectation of deeper and more sophisticated understanding of science disciplinary concepts. For example, in fourth grade Physical Science, the Motion and Forces subgrouping includes demonstrating knowledge of "rate of change" (P4.5) and "net zero forces" (P4.6), which are potentially challenging and complex concepts that require advanced cognitive understanding for fourth grade students. In Earth Science, the subgrouping Earth's Place in Space has little alignment with the current framework at grade 12; E12.1-E12.3 in the draft framework are more advanced and approach different topics than the previous framework. As another example, dynamic equilibrium (P12.9) is an advanced topic for high school students. Moreover, there are more statements in Physical Science than in Life Science and Earth & Space Sciences, and a number have multiple parts and would be assessed by more than one item. The emphasis on Physical Science also has implications for the difficulty of the assessment since not all grade 12 students have an opportunity to take a course in physics.

Table 1. Scope of Disciplinary Concepts in Current and Draft Frameworks

	Current NAEP Framework (Number of Statements)			**Draft 2028 Framework** (Number of Statements)		
	Grade 4	**Grade 8**	**Grade 12**	**Grade 4**	**Grade 8**	**Grade 12**
Physical Science	15	16	23	20	30	30
Life Science	7	12	13	15	28	26
Earth & Space Sciences	11	15	13	13	20	17
Total	33	43	49	48	78	73

In addition to greater breadth and depth in the Disciplinary Concepts, the draft framework calls for assessing eight Science and Engineering Practices (SEPs), twice the number of practices in the current framework. Further, the draft framework includes an extensive set of expectations for students in carrying out these practices. There are also seven Crosscutting Concepts (CCCs) to consider as part of test design and development, a dimension that was not explicitly required in the current NAEP science framework. Table 2 shows the total number of statements included in the Science and Engineering Practices and Crosscutting Concepts in the draft framework. Though not every statement would be assessed in an operational assessment, this illustrates the breadth of the practices and concepts articulated in the framework.

Table 2. Number of SEP and CCC Statements in the Draft Framework

	Draft 2028 Framework		
	Grade 4	**Grade 8**	**Grade 12**
Science and Engineering Practices (8 practices)	69	70	61
Crosscutting Concepts (7 concepts)	17	31	28

The broad definition and scope of the NAEP science domain overall has implications for the assessment design required to cover the domain of science achievement as defined in the framework. The current NAEP science assessment design is based on a total of 13 blocks at grade 4, 14 blocks at grade 8, and 16 blocks at grade 12 to reliably report achievement for each of the three subscales (Physical Science, Life Science, and Earth & Space Sciences).[1] From a psychometric perspective, science achievement could still be reliably reported using the same number of blocks (and items), but from a content perspective the science domain, as defined in the draft framework, would not be addressed as completely with the same number of blocks. This threatens the construct validity of the NAEP science assessment and the inferences that can be made. A larger assessment would be necessary to more adequately cover the framework; the implication of expanding the size of the assessment is discussed further in the section below about costs.

Lack of definition at the lower end of performance. The framework has expanded the scope of content and has done so primarily at the higher end of science skills and knowledge. An assessment designed for this framework is likely to make the NAEP science assessment difficult, particularly when the draft framework does not provide enough description of lower-level skills and knowledge to support reliable measurement at the lower end of the NAEP achievement scale. To properly describe the science achievement of all students and allow for equitable measurement, the assessment needs to include items that are appropriate for students across the performance distribution. However, the greater depth and challenging nature of the content in the framework will make it difficult to develop enough easier items to reliably measure achievement of students at the lower end of the performance distribution. About one-third of students nationally (at grades 4, 8 and 12) scored below Basic on the 2019 NAEP science assessment and the assessment needs to be able to assess these students.

Constraints associated with multi-dimensional items. The draft NAEP science framework requires assessment items that draw on three dimensions of science learning, with the stipulation that every item assess at least two dimensions and ideally three. It also requires that items be designed around "compelling phenomena and/or problems based in real-world

[1] The number of items in a block varies by grade and whether a block comprises discrete items or is a block composed of a Hands-on Task (HOT) or Interactive Computer Task (ICT). A "discrete item" block has 15-19 items, depending on the grade, while HOTs have 4-8 items and ICTs have 6-12 items.

contexts" so that students can demonstrate "knowledge in use." These requirements will necessarily require a heavy reliance on items organized in sets in which multiple items are associated with a common stimulus (in NAEP parlance this includes multi-part items and item sets) or, again, in NAEP parlance, Scenario-based Tasks (SBTs).

NAEP has a history of including in its assessments items that assess multiple dimensions and that are based in real-world contexts involving compelling phenomena. However, in the interest of improving measurement efficiency, minimizing burden on students, and maintaining practical development and data collection costs, NAEP typically employs stand-alone discrete items along with the more elaborate multi-dimensional, set-based items. The latter are typically used to supplement NAEP assessments, rather than constituting the core, because they tend to have increased complexity and language load, resulting in more difficult items and reduced measurement efficiency and, consequently, increased program costs. These concerns are elaborated further below.

> ***Complexity and language load.*** Item stimuli will need to present enough information to provide the appropriate context to support items in which students engage in more than one dimension. This goes hand in hand with greater complexity and language load, as illustrated by the example items provided in the draft framework. It would be difficult for stand-alone, discrete items to meet the requirements established in the draft framework (i.e., multi-dimensional items and science phenomena and problem-based contexts) and be feasible for a large-scale assessment. The draft framework includes two examples of stand-alone items (Example item 1 and Example item 2). However, these example items would not be appropriate for NAEP given the time it would take for students to process and respond and the limited measurement information a single item provides. Increased complexity and language load will also make it challenging to determine whether the items are measuring reading comprehension rather than science achievement. Further, we anticipate that increased reading and language loads may result in more difficult and complex items, particularly given the greater depth and more challenging content discussed above.
>
> ***Measurement efficiency.*** We know from experience that items of this nature often take longer for students to process and respond to, so it is reasonable to conclude that fewer items could be assembled in a 30-minute block, constraining the amount of measurement information we can get from a single block and reducing measurement efficiency. For example, the interactive computer tasks (ICTs) in the current NAEP assessment, are a useful point of comparison. Each ICT includes 6-12 items and is presented in a 30-minute block, compared to 15-19 items in a 30-minute block composed of discrete items (which may include stand-alone items or item sets). Reducing the amount of measurement information per block consequently decreases the measurement information we get for a student. If fewer items can be assembled into a single block the assessment will need to include more blocks to cover the same

amount of content and, necessarily, a larger student sample. Larger student samples will increase program costs to administer the science assessment, which is elaborated further in the "Cost Implications" section below. The number of items per subscale at the test form level (i.e., two 30-minute blocks) is an important factor for reliable performance estimation. Including fewer items in each test form will reduce the reliability of estimates of student achievement. Also, items in sets have a narrower content focus, given the shared stimulus, which can limit the diversity of content in a block and for an individual student, which can contribute to "sparseness" and reduce reliability of the estimates.

It is imperative that the 2028 NAEP science framework and forthcoming item specifications provide explicit guidance on how to implement the goals articulated in the framework. The framework needs to include examples of multi-dimensional items in which students are engaged in sensemaking and problem solving in real-world contexts that do not include a high language load and high level of cognitive complexity. A diverse set of example items (measuring different DCs, SEPs and CCCs), annotated to explain how students engage with the dimensions of science assessed, will provide essential guidance to item developers. Examples should also include two-dimensional items that measure a disciplinary concept and a crosscutting concept. Example items also need to include scoring rubrics. It is particularly important that example items include those that are accessible to students at the lower end of the performance distribution.

The draft framework calls for an even distribution of response time for the three disciplinary concepts. With the heavy reliance on set-based items (item sets or SBTs), the time spent on each disciplinary concept is unlikely to directly translate to the amount of measurement information collected for each disciplinary concept. We expect that there will be a large range of response times for different items that will be driven by the context provided and the dimensions in which students are engaged rather than by the specific item type. Rather than specifying the distribution of assessment items across disciplinary concepts in terms of student response time, NCES suggests specifying this in terms of the percentage of items.

Cost Implications

The requirements outlined in the draft framework have significant and unsustainable cost implications for the NAEP program.

First, as previously discussed, authentic and engaging phenomena or problem-based scenarios that require students to address problems, make sense of phenomena and make informed decisions will require multi-part items, item sets and SBTs, resulting in fewer items that can be assembled into a single block and consequently requiring more blocks to support the framework's measurement and content targets. The more blocks required to sufficiently measure the framework will result in a larger student sample size and increased costs to administer the assessment.

Second, it is more resource-intensive to develop scenario-based, multi-dimensional items, including item writing and developing scoring rubrics, pretesting and item revisions, and iterative reviews. We can also expect greater attrition based on pretesting and pilot data for these types of items, which means that we will need to pilot at least two items for every one needed for the operational assessment and possibly more. Attrition can occur more frequently for multi-part and set-based items (including SBTs), where one part of an item or one item in a set may be unsuccessful in the pilot and must be revised or dropped from the item pool, affecting other items in the set (e.g., impacting the storyline and scaffolding in an SBT) so they potentially also need to be revised or dropped. Also, it can be difficult to hit the right balance between creating sufficient contexts for students without compromising measurement, given the demands of phenomena and problem-based scenarios, resulting in more items not surviving to an operational assessment. Depending on the complexity of the scoring rubrics, scoring pilot and operational items could be resource-intensive, requiring large training sets and more time to conduct scoring.

Third, though NCES has not yet conducted a trend item pool alignment study at this stage of framework development, given the differences between the current and draft frameworks, we anticipate that relatively few items in the trend pool will align to the draft framework (compared to recent instances of new or updated NAEP frameworks). This means that we will need a larger development effort to fully implement the draft science framework in 2028.

Trend Implications

It is important that the Governing Board understand that maintenance of trend under this framework will be very difficult and probably not possible from a construct and technical perspective. Given the magnitude of shift in the construct from the current framework to the draft framework, maintaining trends in NAEP science achievement is unlikely from a construct validity standpoint. Further, measuring trends generally requires that about two-thirds of the blocks from one administration are carried over to the next. However, as described above, NCES anticipates that relatively few items in the trend pool will align to the draft framework so it is likely that there will be few trend blocks from the current science assessment that can be included in the 2028 science assessment. Also, shifts in the expected distribution of items across the three disciplinary concepts, although relatively modest, will make it more difficult to maintain trend from previous assessments since the emphasis has changed.

Clarifications Needed to Support Framework Implementation

NCES recognizes that the draft framework that was issued for public comment was an early draft and that the Development Panel is continuing to develop and elaborate aspects of the framework. With that in mind, NCES has listed below areas that require clarification and refinement to ensure sufficient alignment with NAEP's large-scale assessment methodology

and for NCES to understand what is required to operationalize the framework.

Reflecting a wide range of learners. NCES appreciates that the framework articulates the importance of ensuring that the NAEP science assessment is responsive and relevant to the diversity of the student populations that take NAEP. Nevertheless, some guidance provided will be challenging to implement in a group-level assessment like NAEP. The framework acknowledges the constraints inherent in a large-scale assessment, but the definitions and principles included suggest expectations that may not be feasible for NAEP. For example, the framework says that *"*...assessment scenarios/contexts should be relevant and specific to a range of specific communities rather than trying to represent an average. *The expectation is that all students could see themselves and their peers represented in some phenomena/problems, not in every phenomenon/problem*" (italics added). NCES agrees that the science assessment should reflect a broad range of communities. One challenge is that NAEP is very large assessment (in terms of the amount of assessment material) but an individual student will see only a small portion of the item pool. It is important that the framework acknowledges that and considers representation across the item pool, not for an individual student. Also, NAEP is administered at a national level across many unique communities and so it will be important to have further guidance on how to select phenomena/problem situations without introducing unintended biases. The framework also provides specific features of culturally relevant scenarios/contexts, many that can be achieved and are consistent with how NCES approaches item development for other NAEP subjects. NCES would need more practical guidance, however, on how to achieve an item pool with items that "reflect real, specific phenomena and problems particular communities care about."

Item types. The framework includes little in the way of specific digital item types that could be used in the science assessment. There is a list of possible item types on page 53 (in the section about balancing response types), but NCES expects a framework to provide a description of the selected-response and constructed-response item types that could be included and how they could be employed in the service of measurement, along with examples. Relatedly, the draft framework includes a section (page 52) called "Item Types" which defines terms that refer to item formats in NAEP assessments-- discrete items, multi-part items, item sets and Scenario-based tasks (SBTs)--rather than different item types in the way "item types" is used elsewhere in the framework and traditionally in the field of educational measurement.

Purpose and components of SBTs. The framework calls for the use of SBTs to focus on "interdisciplinary activities, authentic modeling, extended investigation design, and evaluation of design solutions." NCES needs more clarity on what is envisioned by the Development Panel. For example, should SBTs be used to have students conduct virtual experiments? What interdisciplinary activities are envisioned? How would an SBT support evaluating a design solution? Are there particular disciplinary concepts, practices, or crosscutting concepts that would be most suitable to be assessed using the SBT format? How are SBTs different from

item sets? There also appears to be a misconception about what types of items are included in SBTs. The framework says that SBTs will not include selected-response and constructed-response items. In other NAEP assessments, and in the ICTs in the current NAEP science assessment, SBTs do include selected-response and constructed-response items. Some of these "items" are more like interactions rather than traditional item types, but these too would be classified as either selected-response and constructed-response, depending. NCES needs to understand if there is a requirement for particular types of interactions or response modes in an SBT.

Performance expectations. Illustrative performance expectations that demonstrate how the three dimensions of the framework are to be combined to create multidimensional performances for evaluation, similar to what is provided in the current framework, will be essential for operationalizing the framework. This will help define student evidence statements that concretely demonstrate how we expect students to manifest their skills and knowledge for evaluation. Critically, this will also help to inform how scoring rubrics should be approached and also how scale-based reporting will address multiple dimensions.

Mathematics content expectations. The draft framework does not specify the level of mathematics content expected at each grade (as in current framework). It is essential that item developers know what is expected regarding, for example, computations, data representations, and measurement units. This is particularly important because the draft framework includes the practice "Using Mathematics and Computational Thinking". The framework does specify the level of mathematics content expected in the science and engineering practice "Analyzing and Interpreting Data" and may be a suitable example to apply to other areas of the framework.

Digital tools and other interactive features to support measurement. The framework does not include guidance on digital tools that should be made available to students (e.g., graphing tools, digital scales, embedded spreadsheets). Considering the role that digital tools and other interactive features could play in carrying out investigations, organizing and presenting data, and addressing design problems, for example, the framework should provide guidance on what these could include and how they could be employed in the service of measuring science achievement.

Claims. NCES needs clarification about the role of the "claims" (page 5 of the draft framework) for operationalizing the framework. It is not clear if the assessment should be structured around addressing the three claims or if there is another role for the claims.

Inclusion of disciplinary concept clarifications and boundaries. Provision of clarifications and boundaries for the disciplinary concept statements is immensely helpful for item writers but may be more appropriately positioned in the assessment and item specifications document. Also, the approach should be more consistently applied across the three disciplinary concept groupings.

Example items. It is essential that the framework include high-quality example items that reflect the demands of the framework, include scoring rubrics, reflect a range of item types, and are feasible and appropriate for a large-scale assessment. Example items should include those assessing a range of disciplinary concepts, science and engineering practices, and crosscutting concepts, include a variety of contexts, problems, and phenomena and reflect considerations for culturally responsive items. They should also be annotated to explain how students engage with the dimensions of science assessed and how the examples illustrate framework requirements.

Terminology. The draft framework includes several terms that require further clarification. Examples are listed below.

- Page 37 says: "The NAEP Science Assessment will ask students to engage these abilities as part of achieving a successful response to **multidimensional tasks**." Please define "multidimensional tasks."
- Page 37-38 says: "**Appropriate technology** refers to using the simplest level of technology that can achieve the intended purpose in each location, using fewer natural resources, emitting less pollution, and costing less. Appropriate technologies are often small-scale and make use of expertise available in the local community." The definition of "appropriate technology" is unclear, particularly in relation to an assessment.
- Page 53 uses some terms in reference to response types that need to be defined: "**digital tool-based and object-based constructed responses**," "**limited option responses**," "**discourse responses**" and "**collaboration responses**."
- The framework introduces the term "**sensemaking**" on page 5 and makes several additional references to the term as it relates "knowledge in use" (page 51), but it remains unclear how this term is defined as it relates to an assessment, especially in contrast to "sensemaking" when students are learning in a classroom setting.

Cognitive Complexity Framework. The draft framework refers to a forthcoming "cognitive complexity framework that will be applied to NAEP item development" (page 58). The expectations articulated in such a framework will have significant implications for NCES's implementation of the framework and potentially the feasibility of doing so. NCES will provide comments about the cognitive complexity framework when it is available.

Content Clarifications. Descriptions of the three dimensions of the framework require some clarifications to support NCES's implementation of the framework. Clarity is needed related to: (1) the progression of ideas and/or statements across grades; (2) distinctions between ideas (statements) within grades and expectations for what is to be assessed; (3) distinctions among crosscutting concepts; and (4) what is intended by a statement or terms used in a statement. Some examples of these issues follow.

- Across grades:
    - It is not clear how some of the SEPs across the grade spans. For example, S4.1 and S8.2 are very similar to each other and S4.13 and S8.11 are also very similar to each other.
- Within grade:
    - Disciplinary concepts P8.12 and P8.14 both mention an object's change in motion as a result of the sum of forces acting upon it. How should these concepts be distinguished when considering item alignment?
    - Disciplinary concept L12.11 appears to overlap with L12.9, which covers the role of cellular respiration in supporting life processes.
    - Disciplinary concept P12.9 describes dynamic equilibrium, an advanced topic for high school students. Clarification as to the scope and depth that is to be assessed is needed.
- Distinctions among crosscutting concepts:
    - It is not clear how the crosscutting concepts will be distinct enough to align items to them and measure the crosscutting concepts separately. For example, concepts 4 and 5 both involve systems. Concepts 1 and 2 both involve patterns and relationships.
- Unclear statements/terminology:
    - In disciplinary concept P4.1, further definition of "a small set of pieces" and the types of objects meant by "a great variety of objects" is needed given that at this grade level the atomic level is out of scope.
    - Disciplinary concept L12.7, says "the sugar molecules this formed…"; is this in reference to photosynthesis?

Reporting. Chapter 4 of the draft framework is less developed than the other chapters and NCES understands that the Development Panel is still working on it. Nevertheless, we provide feedback below on information that should be included in a future draft and that we will likely need to comment on when it is available.

The draft framework calls for labeling achievement results for the three disciplinary concept scales as Sensemaking in Physical Science, Sensemaking in Life Science, and Sensemaking in Earth & Space Sciences, providing the rationale that the results reflect the “integration of the three dimensions.” Given that not all items will assess three dimensions, this could inadvertently lead to misrepresentation of what the NAEP results mean.

The draft framework does not include guidance on the subject-specific contextual variables that should be collected from students, teachers, and schools and used to provide context for the NAEP achievement results. When this information is available NCES will provide feedback on their feasibility.

Thank you for the opportunity to review the draft framework and provide feedback on issues that pertain to NCES’ operationalization of the framework.

Summary of Public Comments Received on the Proposed 2028 National Assessment of Educational Progress (NAEP) Science Assessment Framework[1]

[1] This summary was produced by Dr. Danielle Tyree of Safal Partners under subcontract to WestEd as part of Governing Board contract 91995922C0001 to WestEd, 2028 NAEP Science Framework Update.

Introduction

The public comment period for the working draft of the 2028 NAEP Science Assessment Framework was held from March 13 – April 17, 2023. There were several informational webinars and conference presentations held between March 20 and April 11 to provide an overview of the framework recommendations and encourage participants to submit feedback through the official form on the project website (www.naepframeworkupdate.org).

Event Type	Organization/Cohost	Date	Stakeholder Group	Registrations/ Participants
Conference	Council of State Science Supervisors	3/20	State administrators	42
Webinar	National Assessment Governing Board	3/21	General public	81 / 36
Conference	National Science Education Leadership Association	3/22	School leaders	19
Conference	Education Leaders of Color	3/23	Education leaders and policymakers	18
Conference	National Science Teaching Association	3/25	Teachers	5
Webinar	International Technology and Engineering Educators Association	3/28	Technology/ Engineering educators and researchers	62 / 21
Webinar	Council of Chief State School Officers	3/30	State and business leaders	34 / 13
Webinar	Council of Chief State School Officers and Council of State Science Supervisors	3/30	State administrators	78 / 37
Webinar	Board on Science Education at the National Academies	3/31	Science and science education researchers	119 / 35
Webinar	The School Superintendents' Association	4/4	School district leaders	63 / 28
Webinar	National Assessment Governing Board	4/11	General public	81 / 38
Webinar	National Science Teaching Association	4/11	Teachers	21 / 10

Board members and staff, contractors, panelists, consultants, and others internal to the project were not included in these counts.

Feedback was collected using a structured form on the project website. Feedback was received from 20 individuals and nine organization representatives.

Name	Title and Organization
Individuals	
Alicia Alonzo	Associate Professor, Michigan State University
Lauren Brodsky	Assessment Lead, Learning Design Group, Lawrence Hall of Science, UC Berkeley
Michele Dischino	Professor, Technology and Engineering Education, Central Connecticut State University
Barbara Dunham	Teacher, Birmingham City Schools
Jake Foster	Founder, STEM Learning Design
Brian Gong	Senior Associate, Center for Assessment
Jeremy Haack	Science Educational Specialist, Maryland State Department of Education
Mary Headrick	Middle School Science Specialist, Alabama Math, Science, and Technology Initiative- University of Alabama in Huntsville (AMSTI-UAH)
Howard Lyon	Parent
Spencer Martin	Science Curriculum Instructional Coach, Kansas City, Kansas Public Schools
Peter Mecca	Adjunct Professor of Biology, University of Maryland Global Campus
Linda Morell	
Tony Perry	Postdoctoral Associate, Massachusetts Institute of Technology
K. Renae Pullen	Science Specialist, Caddo Parish Public Schools
Matthew Richard	Teacher, Olathe School District USD 233
Aracelis Janelle Scharon	9-12 Science Teacher, Bloom High School
Ted Willard	Senior Subject Matter Expert in Science, Discovery Education
Anonymous Respondent	Science Standards Specialist
Anonymous Respondent	Supervisor of Secondary Science, Public School
Anonymous Respondent	
Organizations	
ACT	
Cognia	
The Education Trust	
Maryland State Department of Education	
Michigan Department of Education	
National Association of Scholars	
Sunnyside Unified School District	
Virginia Department of Education	
Wisconsin Department of Public Instruction	

Feedback Areas

The feedback form included questions in the following areas:

Overall Representation of Concepts and Skills

- Relevance of science concepts
- Representation of what students should know and be able to do in science
- Proposed concepts or skills to add
- Concepts or skills not necessary to include
- Organization of framework
- Sample items
- Glossary terms

The Three Dimensions of Science

- Disciplinary concepts
    - Style of statements
    - Representation of concept statements
    - High/low priority subtopics or concepts
- Science and Engineering Practices
    - Representation of practices
    - High/low priority practices
    - Lowest priority practices
    - Technology and engineering concepts
- Crosscutting Concepts
    - Representation of crosscutting concepts
    - High/low priority crosscutting concepts

Assessment Design

- Dimensionality of assessment
- Balance of disciplinary concept domains and item types

Reporting

- Priority topics for contextual questionnaires
- Prioritization of implementing recommended changes versus maintaining trendlines

Other

(The exact wording of all questions can be found in the Appendix.)

Overall Representation of Concepts and Skills – Chapter 1

A majority of respondents indicated their agreement with the following two statements:

I find the NAEP Science Framework to effectively highlight the latest and most relevant science concepts that students should know and be able to do (24 out of 27 respondents agreed).

The NAEP Science Framework is a good representation of what students should know and be able to do in science (23 out of 27 respondents agreed).

Respondents indicated that the following concepts and skills should be added:

- *Physical Science: Energy (grade 8); Matter and Interactions (grade 12); Circuit electricity, solution chemistry, mass/energy and waves*
- *Life Science: Marine and other aquatic ecosystems*
- *Earth and Space Sciences: Clarifying details and statements included in ESS1 and ESS2 of the Next Generation Science Standards*
- *Facets of the nature of science and scientific literacy*
- *How practices or crosscutting concepts will be prioritized in two-dimensional items*
- *Geographic diversity and potential bias associated with phenomena*

Specific suggestions can be found in the Appendix on page 23.

Respondents indicated that it was not necessary to include the following concepts or skills, or that additional clarification was needed:

- *The domain defined by the three dimensions is too large to cover in an assessment*
- *Crosscutting concepts are especially difficult to assess well*
- *The practices take the scientific design process and divide them into parts, but NAEP will need to figure out whether to assess in pieces or as larger combinations*
- *The balance of assessment items is specified only for the disciplinary concepts and not for practices or crosscutting concepts*
- *Statements related to ethical and social responsibility should be reviewed carefully to ensure they do not measure personal beliefs or opinions that are prohibited by law*
- *The description of the Board's commitment to equity in chapter 1 should include additional information on cultural relevance and responsiveness to avoid the impression that this is a traditional approach to bias and sensitivity concerns*
- *The description of phenomena and problems needs additional clarification*
- *Information on cognitive complexity is too generic to be of practical use*
- *Inclusion of specific disciplinary concepts at grade 4, and "what if questions" and "tinkering" as specific practice statements; as well as exclusion of "planning and/or conducting investigations" as a grade 4 practice*
- *Content that goes beyond the grade-level expectations of the NRC Framework will produce an equity issue for students who are not exposed to those ideas and score lower*
- *Exclusion of certain ideas regarding the universe and Earth's history suggests that NAEP is bowing to political pressure*
- *There is too much emphasis on disciplinary boundaries*
- *The opening of chapter 4 (p. 71) makes a re-titling of the three subscales sound like a minor labeling issue, but that framing masks the fundamental shift that this framework is aiming for*
- *Several pieces of the disciplinary concepts appear to exceed what is required by the disciplinary core ideas in the Next Generation Science Standards*
- *There is incorrect information included in disciplinary concepts L12.12 and L12.16.*

Specific suggestions can be found in the Appendix on page 26.

Organization of Framework

A majority of respondents agreed that the organization of the framework is useful to understand the content and context of what students should know and be able to do in science. Suggestions for improvement included the following:

- *A multidimensional framework should use a measurement model and reporting that highlights different expertise rather than traditional content categories*
- *Include some discussion of what three-dimensional items look like (well-discussed in chapter 3) in chapter 2 in advance of the disciplinary concept statements. Doing this and including a sample item would provide helpful context for considering the dimensions laid out in chapter 2.*
- *Use a landscape orientation rather than portrait to allow for more comparison of grouped ideas and tracking across grades. Physical lines in the tables (such as those included for practices) are helpful.*
- *Consider matched numbering of ideas across grades, such as by using an additional categorization to keep associated ideas together – or adding an additional code to specify subareas*
- *Remove boundary and clarification statements from the body of the framework and instead include in the Assessment and Item Specifications*
- *The framework does not allow for enough challenging content for students who are scientifically talented*

Specific suggestions can be found in the Appendix on page 29.

Sample Items

A majority of respondents indicated that the sample items are a useful representation of what the NAEP Science Framework should measure. The following suggestions were made related to examples or sources for items and tasks to consider including in the framework:

- *Seaweed is not a good context for the sample items because it currently has a negative connotation with its overabundance in southern waters and resulting problems*
- *Look across websites of states that have assessments aligned to the NRC Framework*
- *Consider items from the Next Generation Science Assessment project (https://ngss-assessment.portal.concord.org/middle-school)*
- *Annotate sample items to illustrate the points intended*
- *Include sample items for using computational models to illustrate the boundaries*
- *Be clear about how sample items will use digital tools*
- *Include images or visual examples of phenomena to illustrate what they are*
- *Include examples of scenario-based tasks similar to this example for grade 4 NAEP mathematics: https://npd.naep.ed.gov/totw/2022/app/en/main.html?subject=Math4*
- *Consider the Stanford Assessment Project*
- *Rubrics and sample papers should be included, along with additional items for all disciplinary areas and grade levels*
- *Include examples of contexts and scenarios that are culturally relevant and based on current topics*

- *Consider Science Assessment Item Collaborative Materials (https://csaa.wested.org/spotlight/science-assessment-item-collaborative/)*
- *Consider what states will be assessing by 2028*
- *Consider examples from Inner Orbit; language from STEM Teaching Tools*

Specific suggestions can be found in the Appendix on page 33.

Glossary Terms

The following recommendations were made for terms to include in a framework glossary:

- *Acronyms used frequently such as NAEP, TEL, and NRC*
- *Sources used other than the NRC Framework*
- *Terms used in specific disciplinary concept statements*
- *Phenomena*

Specific suggestions can be found in the Appendix on page 37.

The Three Dimensions of Science - Chapter 2

The feedback form included questions about the style, representation, and prioritization of the disciplinary concepts, science and engineering practices, and crosscutting concepts.

Disciplinary Concepts

Style of Statements

Several respondents were unsure what this question was asking and did not perceive there to be a difference in the style of the concept statements across the three domains. Those who did perceive differences commented most often on the use of the clarification and boundary statements. Specific feedback included:

- *The most helpful features are parenthetical statements to provide examples; boundary statements phrased in the format beginning, "Students are not expected to"; and clarification statements phrased in the format beginning, "Emphasis is on"*
- *Clarification statements in Life Science are particularly helpful; Life Science also includes examples, which are helpful*
- *Clarification and boundaries for Earth and Space Sciences strike the best balance (Physical Science includes only boundaries, not clarifications, and Life Science sometimes includes too much information); they are also stylistically easier to understand*
- *The style for Physical Science is preferable because it is more concise*
- *Clarification statements are important because they narrow down possible interpretations of concept statements for assessment*
- *A possible alternative is to provide only concept statements in the main body of the framework and include an expanded version of the concept statements with boundaries and clarifications in an appendix.*
- *The need for boundary and clarification statements may indicate that the concept (idea) statements are not clear.*
- *Some content statements are not specific enough to guide decisions for assessment developers.*

Specific suggestions can be found in the Appendix on page 38.

Representation of Concept Statements - *Grade 4*

Feedback on the extent to which the concept statements represent what respondents believe should be assessed by NAEP included the following comments, grouped by grade:

- *Many disciplinary concept statements are based on the NRC framework that has informed most state standards and are generally appropriate for NAEP to assess*
- *The concept statements represent a relatively basic level of knowledge that will be necessary as a foundation for future learning to build upon*
- *Given limited science instruction time historically, Grade 4 targets may be overly aspirational*
- *Specific examples provided of concepts unlikely to be encountered by grade 4 include aspects of Physical Science (P4.11); Life Science (L4.5); and Earth and Space Sciences (E4.1, E4.3, E4.4)*
- *There are some statements that express extremely similar content (e.g., P4.6, P4.7, P4.8)*
- *The following information should be added: general understanding of climate (E4.12), gases when addressing the phases of matter*

Representation of Concept Statements - *Grade 8*

- *Many disciplinary concept statements are based on the NRC framework that has informed most state standards and are generally appropriate for NAEP to assess*
- *Some aspects, particularly at Grade 8, seem overly aspirational and may be better suited for Grade 12*
- *There are some statements that repeat each other (e.g., P8.3 and P8.5) or express extremely similar content (e.g., L8.10, L8.15, and L8.16)*
- *Some statements are very long and may need to be scaled back or broken into multiple pieces*
- *The number and content of some statements related to climate change and evolution may need to be reduced based on the extent to which states include these in their standards*
- *In Physical Science light should be more explicitly called out in several of the statements (or clarifications) relative to wave concepts*
- *For Earth Sciences, the current set of grade 8 concept statements includes at least as much content as the NGSS, and teachers do not have time to cover all of the NGSS*
- *It would be worth including some phenomena at only one grade level instead of two or three (e.g., E8.1, E8.2, E8.4)*
- *It is problematic that there is a large disparity in the number of disciplinary concepts in each domain (PS = 30, LS=28, ES=20)*
- *Specific examples provided of concepts unlikely to be encountered by grade 8 include: Physical Science (PS8.3, PS8.5, PS8.24); Life Science (LS8.8); and Earth and Space Sciences (ES8.3, ES8.5, E8.7, E8.18)*
- *The following concepts are missing: ESS1.A, ESS2.C3, ESS2.C.4, ESS2.D.3*

Representation of Concept Statements - *Grade 12*

- *Many disciplinary concept statements are based on the NRC framework that has informed most state standards and are generally appropriate for NAEP to assess*
- *There are some statements that express extremely similar content, are very focused on definitions as opposed to larger conceptual understanding, or are very long*

- *Information about how teachers are managing the NGSS performance expectations may help prioritize what should be included on the NAEP assessment*
- *It is problematic that there is a large disparity in the number of disciplinary concepts in each domain (PS =30, LS=26, ES=17)*
- *Specific examples provided of concepts unlikely to be encountered by grade 12 in Physical Science include: PS 12.2, 12.4, 12.7, 12.11, 12.13, 12.15, 12.24*
- *There are multiple components to E12.10 that should be separated*
- *The following concepts are missing: PS2.C; ESS 1.A.1, 1.A.3, 1.A.4, 1.C.1, 1.C.2, 2.A.1, 2.A.2, 2.A.3, 2.D.2, 2.D.3, 3.D.2; meiosis and protein synthesis; in Physical Science Matter and Interactions: types of interactions based on electron configurations, and explaining how the bulk properties of materials are determined by small scale structures*

Specific suggestions can be found in the Appendix beginning on page 41.

High Priority Concepts

While some respondents would not recommend assigning a higher priority to certain subtopics or concepts, other respondents had specific subtopics or concepts they would recommend emphasizing, including:

- *Prioritize subtopics or concepts that 1) are most likely to have a direct impact on students' current and future lives; 2) most clearly intersect with the crosscutting concepts (e.g., P8.9 might have higher priority than P8.2); 3) foster the habits of mind that foster science (e.g., curiosity, empiricism, formulation of questions about nature, how to get nature to answer them objectively); 4) students will use to make informed decisions while a member of a democratic society, and in an age of disinformation; 5) relate to cause and effect; and 6) relate to conservation of and interactions involving the transfer of matter, energy, momentum, and electric charge*
- *For Physical Science, P12.4, P12.5, P12.6, P12.12, P12.19, P12.28, and P12.29*
- *At the high school level, the Physical Sciences and Life Sciences are generally high priorities because of college and career readiness expectations, course content specificity and availability, and STEM career focus areas. At the middle and elementary levels, more time is spent on the Earth and Space concepts in preparation for high school curricula.*
- *If prioritizations are made, they should be documented in the Assessment and Item Specifications*

Specific suggestions can be found in the Appendix on page 52.

Low Priority Concepts

While some respondents would not recommend assigning a lower priority to certain subtopics or concepts, other respondents had specific subtopics or concepts they believed to be less important to assess, including:

- *Concepts that rely on seeing or hearing might generate bias for blind or deaf students*
- *P12.9 seems to be an odd topic to include while ignoring things more frequently taught. A lot of the wave material in PS4 seems in the weeds relative to other things that might be more resonant. A causal understanding of electric, magnetic, and gravitational fields is likely beyond the ability level of most Grade 12 students.*

- *The panel should review the prevalence and scope of standards for climate change and evolution across the states to determine the appropriate level of inclusion/assessment on NAEP*
- *Earth's Systems, Matter and Its Properties, subatomic particles, and plate tectonics*
- *Anthropogenic climate change*

Specific suggestions can be found in the Appendix on page 54.

Science and Engineering Practices

Representation of Practices

Feedback on the extent to which the science and engineering practices represent what respondents believe should be assessed by NAEP included the following comments:

- *The practices align to the NRC Framework and are appropriate for NAEP*
- *It may be difficult to accurately and fairly assess aspects of tradeoffs that involve aesthetics, values, morality, etc. on a large-scale science assessment*
- *Some of the learning expectations/statements are written very broadly, whereas they need to be focused on operationalized analysis/application (e.g., S4.49, S4.50, S8.70)*
- *Consider a closer alignment with the ITEEA's STEL Practices by including some softer skills such as collaboration, creativity, and attention to ethics*
- *There are some inappropriate practices for elementary students.*
- *There are misalignments in grade-level practices and alignment to the NRC framework and supporting documents (e.g., Line 362: Asking Questions & Defining Problems S4.1, S4.2, S4.4, S4.7, S4.9, S4.10, S8.4, S8.7, S8.8, S8.10, and S12.8; and Line 375: Developing & Using Models S4.14, S4.15; Line 390: Planning & Carrying Out Investigations S4.21)*

Specific suggestions can be found in the Appendix on page 56.

High Priority Science and Engineering Practices

Many respondents did not recommend assigning priority to certain science and engineering practices and considered all practices to critical for sense-making. Specific suggestions for high priority practices included:

- *Explanation, Modeling, and Argumentation should be prioritized because they are most directly involved in making sense of phenomena*
- *Engaging in Argument from Evidence; Obtaining, Evaluating, and Communicating Information; Asking Questions; and Analyzing and Interpreting Data should be prioritized*
- *Use the Instructional Leadership for Science Practices (ILSP) if there is a need to group practices: Investigating Practices, Sensemaking Practices and Critiquing Practices*
- *Practices 1, 3, 4, 6, 7, 8 should be prioritized; learners should know, understand, and apply accepted ways new information is developed and applied*

Specific suggestions can be found in the Appendix on page 62.

Low Priority Science and Engineering Practices

While some respondents would not recommend assigning a lower priority to certain science and engineering practices, other respondents had specific practices they believed to be less important to assess, including:

- *Some of the science and engineering practices may be difficult to authentically measure (e.g., Practice 1: Asking Questions and Defining Problems and Practice 8: Obtaining, Evaluating, and Communicating Information). Many of those skills are more suited to being meaningfully applied and assessed in a research project or larger performance tasks.*
- *Planning and Carrying Out Investigations can be assessed in a limited way with technology-enhanced items, but it is important to consider what is essential for assessment versus instruction*
- *It is important to ensure that the final proposed learning expectations/statements for the practices will be assessable with the item types available*
- *It will be important to decide what the frequency of AID and OECI will be on the assessment in comparison to other practices and what that prioritization of a practice over another or the difference in the number of practice-specific questions implies*
- *Engaging in tinkering to improve a design is unnecessary to assess*
- *Using Mathematics and Computational thinking could have less attention than the others*

Specific suggestions can be found in the Appendix on page 64.

Technology and Engineering Concepts

A majority of respondents agreed that the framework adequately reflects technology and engineering concepts and noted that it draws on the work that the NRC Framework did to identify the technology and engineering concepts that are needed by all students at the various grade bands. Specific comments included:

- *The degree to which students are actually exposed to using these standards in class may be less than desirable to meet these goals*
- *The tie to the TEL framework might seem like a stretch*
- *It may be helpful to separate the science practices from the engineering practices*
- *Some relevant aspects of technology and engineering are absent in the current framework, including the history of technology and its impacts and influence on society*
- *Connections to technology are vague or absent; engineering appeared to be prioritized*

Specific suggestions can be found in the Appendix on page 67.

Crosscutting Concepts

Representation of Crosscutting Concepts

Feedback on the extent to which the crosscutting concepts represent what respondents believe should be assessed by NAEP included the following comments:

- *The crosscutting concepts represent what should be assessed*
- *The crosscutting concepts should be clustered into four large clusters with the practices, perhaps as Data Interpretation, Scientific Inquiry and Investigation, Scientific Models and Arguments, and Engineering and Design Thinking, which would be similar to international frameworks such as PISA and TIMSS*
- *Some of the specific learning expectations/statements are written very broadly and feel quite conceptual, which may make them difficult to assess*

- *There are misalignments to the NRC framework and supporting documents (Appendix E, Appendix F, Appendix G, and the DCI/SEP/CCC descriptors found in the three foundation boxes associated with the NGSS Performance Expectations (Volume 1, The Standards, Next Generation Science Standards For States, By States)*
- *Reporting should include more information about crosscutting concepts*

Specific suggestions can be found in the Appendix on page 69.

High Priority Crosscutting Concepts

While many respondents would not recommend assigning a higher priority to certain crosscutting concepts, other respondents had specific crosscutting concepts they believed to be more important to assess, including:

- *If there is a need to group the crosscutting concepts in some way, we the NRC Framework could be used to make three groups – Patterns, Cause and Effect, and Systems*
- *Assessment of crosscutting concepts should consider the expected frequency of pairings with disciplinary concepts and practices (the Assessment and Item Specifications should contain more information about what is expected)*
- *Patterns are frequently taught and therefore would be fair to assess*
- *Mechanisms and Explanation: Cause and Effect; and Systems and System Models/Systems Thinking are the most important CCCs to assess*
- *Math*
- *Crosscutting concepts 1, 2, 5, and 7 because these present learners with opportunities that are easier to become attentive and engaged*
- *Patterns; Cause and Effect; Flow of Energy and the Cycling of Matter; Structure and Function are the most accessible for a wide range of students and their science experiences*
- *Because the Nature of Science is such a critical part of scientific literacy, it should not be left out of this exam (For example, students should know what a theory actually is and what the process of science looks like in order to make informed evaluations of media (mis)information)*
- *Systems; Structure and Function; Matter and Energy; Stability and Change; Patterns*

Specific suggestions can be found in the Appendix on page 73.

Low Priority Crosscutting Concepts

While many respondents would not recommend assigning a lower priority to certain crosscutting concepts, other respondents had specific crosscutting concepts they believed to be less important to assess, including:

- *Within the learning expectations/statements for some of the crosscutting concepts, some statements are more foundational recognitions or understandings that may be difficult to assess, e.g., C4.8 (To understand a phenomenon it often helps to develop a model of the system in which it occurs)*
- *Conservation, Flows, and Cycles: Tracking Energy and Matter is less important to assess because it is difficult to differentiate it from the Physical Science disciplinary concepts involving Matter and Energy*
- *Structure and Function seems to be a subset of Models; Cause and Effect; and Systems*

Specific suggestions can be found in the Appendix on page 75.

Overview of Assessment Design - Chapter 3

Dimensionality of Assessment

A majority of respondents agreed that the assessment should be three-dimensional whenever possible and individual items should be at least two dimensional. Support for multidimensional items was based on alignment with current approaches to science standards and instruction. Concerns raised about this approach included:

- *This requirement is unnecessarily restrictive and potentially unfair to some students, including those who are unfamiliar with 3D assessments*
- *Sometimes a one-dimensional question (measuring only a disciplinary concept) is appropriate depending on the complexity of the concept, particularly in 4th grade*
- *All items should be at least two dimensional, but no item should include only a disciplinary concept and a crosscutting concept (practices should be the required dimension)*
- *Multidimensional items may make it difficult to adequately measure the skills and knowledge of lower-performing students*

Specific suggestions can be found in the Appendix on page 76.

Balance of Disciplinary Concept Domains and Item Types

A majority of respondents agreed with the balance proposed for the disciplinary concept domains and item types at all 3 grades. Support for an even distribution of items by disciplinary concept domain was generally based on the areas having equal importance in science education. Concerns raised with this approach included:

- *It is important for the framework to distinguish between a total testing administration (all forms for a grade across all students) and an individual student's experience (single form)*
- *The distribution of items should include additional requirements for the practices and crosscutting concepts to ensure coverage of them as well*
- *For grade 4, the following distribution is recommended: 50% for Physical Science (how things work), 25% for Life Science (how living beings work), and 25% for earth and ecological studies (how the above can work together)*
- *The distribution of engineering/technology related items and nature of science related items should be specified*
- *Removing the hands-on-tasks may send the wrong message about science instructional practice*

Specific suggestions can be found in the Appendix beginning on page 79.

Reporting Results of the NAEP Science Assessment - Chapter 4

Priority Topics for Contextual Questionnaires

Responses to this question focused on various aspects of students' opportunity to learn, the nature of science instruction, as well as science learning outside the classroom. Recommendations included:

- *Amount of science instruction time, particularly at the elementary level*
- *Types of equipment and activities that students can access as part of that instruction*
- *Science activities outside the classroom*
- *Exposure to 3D assessment items*
- *Nature of science instruction*
- *Curriculum sequences and course taking patterns in middle and high school*
- *Opportunities to engage in phenomena-based science learning, investigations or engineering design, and multidimensional learning*

Specific suggestions can be found in the Appendix on page 85.

Prioritization of Implementing Recommended Changes Versus Maintaining Trendlines

A majority of respondents indicated that it was more important to implement the recommended changes to align with the way that states are currently teaching science than to restrict changes in order to prioritize comparisons with previous NAEP results. Feedback included the following comments:

- *The need to reflect current thinking in science education greatly outweighs the need to compare to earlier NAEP assessments*
- *The incorporation of the vision of science achievement reflected in the NRC Framework represents at least as large as the one from the 1996 NAEP Science Framework to the 2009 NAEP Science Framework; implementation of this framework could be diluted if forced to maintain trend*
- *The recommended changes are vital to reflect the needs and expectations for K-12 science education going forward; to not implement these changes for the sake of maintaining trend lines would only undermine the intention of NAEP (to assess what students know and are able to do) and foster misinterpretation and misunderstanding of the abilities of K-12 students in science*
- *NAEP is commonly referred to as the gold standard of large-scale assessments, but if what the NAEP science assessment measures is different from what is prioritized and measured by states it could erode NAEP's credibility and reduce the utility of the NAEP science assessment*
- *It is more important to align the assessment with what most states are teaching; the retention of the three disciplinary concept domains may still allow comparisons with previous results*
- *This framework represents a substantial shift in NAEP which is necessary and well timed; trying to maintain trend lines could lead to a situation where the shifts are not implemented as envisioned, leading to misinterpretation of the intent of the revised framework*
- *It is more important to measure what is important to measure in student understanding as opposed to avoiding a discontinuity, especially if avoidance results in the status quo*
- *Trying to match this new framework to the previous framework would not create a valid trendline; amongst these competing priorities, we have not seen a clear winner*

- *The previous results do not seem to be moving in a particular direction and are not that relevant, especially when science is not assessed on a more regular basis*
- *The proposed framework is much improved, so comparing future results to past results should not be prioritized*
- *Creating a test that better represents students' abilities to engage in all three dimensions is more important than comparing to past NAEP data*
- *It is unimportant to implement the recommended changes at this time because of the COVID pandemic; the ability to see NAEP results from past years and compare them to current/future ones should be imperative*
- *Given the investments made to transform science education, aligning assessments with the NRC Framework should be prioritized; there is a communications problem with comparisons over time*
- *Due to COVID it is important to maintain the trend line; the addition of clarification statements does not warrant a break in trend*
- *Change is needed, especially given the 3D science learning practices that have been in place for a decade*

Specific suggestions can be found in the Appendix on page 87.

Additional Comments

Respondents were given an open-ended field to include any additional comments. Many comments were received in a wide variety of areas, including the following topics:

- *Disciplinary concept statements, science and engineering practices, and crosscutting concepts*
- *Sample items*
- *Cognitive complexity*
- *Item types*
- *Equity (including achievement gaps, accessibility, students with disabilities, and bias)*
- *Phenomena and contexts*
- *Reporting*

More detail on the open-ended comments can be found in the Appendix on page 90.

Appendix – Public Comments by Question

Overall Representation of Concepts and Skills

1. **I find the NAEP Science Framework working draft to effectively highlight the latest and most relevant science concepts that students should know and be able to do.**

Resp 1: Disagree

Resp 2: Agree

Resp 3: Agree

Resp 4:

Resp 5: Agree

Resp 6: Agree

Resp 7: Agree

Resp 8: Agree

Resp 9: Agree

Resp 10: Disagree

Resp 11:

Resp 12: Agree

Resp 13: Agree

Resp 14: Agree

Resp 15: Agree

Resp 16: Agree

Resp 17: Agree

Resp 18: Agree

Resp 19: Disagree

Resp 20: Agree

Resp 21: Agree

Resp 22: Agree

Resp 23: Agree

Resp 24: Agree

Resp 25: Agree

Resp 26: Agree

Resp 27: Agree

Resp 28: Agree

Resp 29: Agree

2. **The NAEP Science Framework working draft is a good representation of what students should know and be able to do in science.**

Resp 1: Disagree

Resp 2: Agree

Resp 3: Agree

Resp 4:

Resp 5: Agree

Resp 6: Agree

Resp 7: Agree

Resp 8: Disagree

Resp 9: Agree

Resp 10: Disagree

Resp 11:

Resp 12: Agree

Resp 13: Agree

Resp 14: Agree

Resp 15: Agree

Resp 16: Agree

Resp 17: Agree

Resp 18: Agree

Resp 19: Disagree

Resp 20: Agree

Resp 21: Agree

Resp 22: Agree

Resp 23: Agree

Resp 24: Agree

Resp 25: Agree

Resp 26: Agree

Resp 27: Agree

Resp 28: Agree

Resp 29: Agree

3. **Are there missing concepts or skills that you believe the NAEP Science Framework should consider?**

Resp 1: Yes

Resp 2: No

Resp 3: Yes

Resp 4:

Resp 5: No

Resp 6: No

Resp 7: Yes

Resp 8: No

Resp 9: No

Resp 10: Yes

Resp 11:

Resp 12: Yes

Resp 13: No

Resp 14: Yes

Resp 15: No

Resp 16: No

Resp 17: No

Resp 18: No

Resp 19: Yes

Resp 20: No

Resp 21: No

Resp 22: No

Resp 23: No

Resp 24: No

Resp 25: No

Resp 26: No

Resp 27: Yes

Resp 28: Yes

Resp 29: Yes

4. **What specifically should be included?**

Resp 1:

Resp 2:

Resp 3: The idea that energy is given of or taken in during a chemical reaction is not really captured in middle school and it is an important idea that is both appropriate and necessary for that grade level.

The idea that energy is given of or taken in during a change of state is not really captured in middle school and it is an important idea that is both appropriate and necessary for that grade level.

Resp 4:

Resp 5:

Resp 6:

Resp 7: There are a couple of concepts we would like to be included in the MATTER AND INTERACTIONS Grade 12: 1) Types of interactions based on electron configurations 2) hot the bulk properties of materials are based on their small scale structure

Resp 8:

Resp 9:

Resp 10: Please see comments below for grades 8 and 12.

Resp 11:

Resp 12: PS: circuit electricity, solution chemistry including acids and bases, mass/energy equivalence, waves other than light and sound such as seismic and water;

LS: life science is overly focused on terrestrial ecosystems, marine and other aquatic ecosystems should be included;

ES: clarifying details and statements are missing fromESS1 and ESS2 from NGSS DCIs as mentioned in additional comments (such as our sun as star, absolute time scales in relation to Earth history and processes, surface processes).

Resp 13:

Resp 14: A fuller treatment of evolutionary biology and thought. A fuller treatment of the philosophical roots of modern science. Some indication of how this framework intends to foster creativity, and to match science instruction to students' own passions and talents.

Resp 15:

Resp 16:

Resp 17:

Resp 18:

Resp 19: As further detailed below, key facets of the nature of science and scientific literacy appear to be missing.

Resp 20:

Resp 21:

Resp 22:

Resp 23:

Resp 24:

Resp 25:

Resp 26:

Resp 27: Missing content within different levels (details provided provided below).

Resp 28: 521 - 523 describes the components of a multi-part item for the assessment. 533 shows the distribution of the concept by grade level. It is not clear from any of the items for the section on multi-dimensional items or the distribution of items the frequency of 3-dimensional component combinations. Since at least 2 items must be included which of the SEPs or CCs will be prioritized and what will be the frequency? Is this determined by grade level or will this be across grade levels with increased difficulty because of the CC and SEP alignment with math and ELA.

565-568 describes the approach to sensemaking based on the phenomena that will be used in the items. How will NAEP ensure that the phenomena selected are not geographically privileging certain students because of proximity to the phenomena based on lived experience? 644 - 664 does not address the impact of the results of any bias that may result or limitations of the phenomena students are being measured regarding learning. Unless the data will be disaggregated by item, this approach doesn't address the variation in results that may result because of the chosen item.

Resp 29: 8.P4U2.5 Develop a solution to increase efficiency when transferring

energy from one source to another.

is faster the same as efficient? maybe this part needs to be clarified.

5. **Is there anything included in the NAEP Science Framework working draft with which you disagree?**

Resp 1: Yes

Resp 2: No

Resp 3:

Resp 4:

Resp 5: Yes

Resp 6: Yes

Resp 7: No

Resp 8: Yes

Resp 9: Yes

Resp 10: Yes

Resp 11:

Resp 12: Yes

Resp 13: No

Resp 14: Yes

Resp 15: No

Resp 16: No

Resp 17: Yes

Resp 18: No

Resp 19: Yes

Resp 20: No

Resp 21: No

Resp 22: No

Resp 23: Yes

Resp 24: No

Resp 25: No

Resp 26: No

Resp 27: Yes

Resp 28: No

Resp 29: No

6. **What do you disagree with and why?**

Resp 1: Use of the Next Generation Science Framework in several operational state science assessments and surfaced serious challenges. 1) The domain defined by the DCIs, SEPs, and CCCs is too large to be "covered" in an assessment--even a highly matrix-sampled one, so the NAEP Framework will need to prioritize explicitly what it values. 2) The possible domain is too large for any student to learn well. That should be reflected in any ALDs. 3) The SEP take an experimental design process (largely) and break it into parts. NAEP will need to decide whether to leave assessment items/tasks in parts or try to assemble into larger, meaningful chunks that carry a context. 4) The SEP are a subset of valuable science skills; see, for example, additional skills in the AAAS framework that included more than experimental design for scientific evidence. 5) I haven't seen anyone who was able to put the CCC into an assessment framework well--are the CCC a type of content or a type of reasoning skill? If used in the NAEP Framework, what each CCC means and how they fit into the construct should be concretely defined. For example, do the CCC ever interact with each other?

Resp 2:

Resp 3:

Resp 4:

Resp 5: I am concerned that the Framework specifies only a distribution in terms of disciplinary concepts (a one-dimensional view of science achievement), while the rest of the Framework does a really nice job of reflecting the three-dimensional view of science achievement reflected in the NRC Framework. Since the assessment is measuring multi-dimensional science achievement, then guidance about the balance of practices and cross-cutting concepts seems equally important. It would be particularly helpful to specify the balance of science practices and engineering practices in the assessment.

Resp 6: 1. Some of the learning expectations/statements around ethical and social aspects of science, engineering, and technology may or may not be appropriate and need to be reviewed, to ensure that personal belief and opinion would not become entangled in assessment responses. This is important as noted in line 26 that "By law, NAEP assessments shall not evaluate personal beliefs." The discussion and learning opportunities around these topics are often more appropriate for the classroom and not suited for large-scale assessment.

2. The introduction of equity in Chapter 1 (briefly included in the paragraphs across lines 26-41) does not put forth the admirable dedication to equity considerations that is found in Chapter 3. We suggest it would be better to more completely discuss the move to culturally relevant and responsive assessment up front in Chapter 1. That may also help address or clarify the statement in lines 27-28 that "NAEP assessment items shall be secular, neutral, and nonideological and free from racial, cultural, gender, or regional bias," which (1) sounds like the traditional approach to bias and sensitivity concerns and does not signal a shift, and (2) is problematic with the use of the word "neutral" given the intent to acknowledge and incorporate culture and diversity.

3. While we agree completely with centering all items and tasks on phenomena and problems, some of the language in lines 459-501 lacks precision and needs to be addressed. The document first states phenomena and problems are necessary because students must have something to make sense of (lines 459-465) but then later states they are important because they provide access points for students (lines 492-495). It is really the former as to why we must set all the questions against phenomena and problems for a good assessment of the constructs and the proposed assessment claims. The rest of the paragraph starting with lines 492-495 better draws out aspects of the latter point; crafting high quality items is important to make sure students can have access points and show what they know relative to the constructs being tested – though equally important is the choice of phenomena and problems, ensuring they are grade/developmentally appropriate, relevant, and explainable with the level of disciplinary content, science and engineering practices, and crosscutting concepts expected of students in that grade.

4. The information on cognitive complexity (lines 724-729) is too general and generic to be of any practical use. That fact is magnified with nothing to exemplify this in the sample items. Only those familiar with emerging cognitive complexity frameworks for multi-dimensional science assessment (e.g., Achieve's A Framework to Evaluate Cognitive Complexity in Science Assessments) will be able to infer what sort of framework is to be used. Please at least cite the framework/source you will be basing this on.

Resp 7:

Resp 8: Yes, language related to programming, and vague terminology not represented in the NRC Framework, DCI/SEP/CCC descriptors, or found in appendices E, F or G. See detailed feedback.

Resp 9: I disagree with the inclusion of DCI: E4.3 (that is a 5th grade concept in most states), "what if" questions as an SEP, and "tinkering" as an SEP. I am strongly against the inclusion of tinkering (under constructing explanations and designing solution) as an assessment expectation and do not understand the evidence that supports its inclusion in our Nation's assessment framework when it is not a part of the NRC Framework and it is not a required disciplinary practice in most states. I think it is inappropriate to conflate tinkering and engineering. Why does elementary science get this treatment?

Moreover, I do not understand the exclusion of "investigation" in the SEPs for elementary. "Plan and/or conduct investigation" is no where in the elementary SEPs.

Resp 10: Given my time-constraints for this review, my focus was on the chosen content (disciplinary core ideas) for grades 8 and 12. The identified issues are listed below in the sections for grades 8 and 12.

Regarding the “ideas” that are beyond the grade-band, the 2012 Framework is central to so many of our state standards, and therefore we have an equity issue with those “ideas” that are beyond the grade-band. Students who have not gone beyond the Framework will score lower than those who have. This will skew the resulting stats and reporting.

In addition, I am concerned with the design of the test if the items in past tests were not three-dimensional. Trends will not be available given the differences in the new test as compared to past tests.

It was obvious in the Earth and Space Science “ideas” that some 2012 Framework elements were folded together, possibly for concision. However, by doing so, the important role of each of those elements/ideas is diminished. In addition, it appears from what was omitted from Earth and Space Science regarding the universe and Earth’s history, that the NAEP team is bowing to political pressure.

Resp 11:

Resp 12: The understanding of fields at grade 8 seems too aspirational to be fair to assess, Le Chatelier’s principle is overly aspirational as it is rarely taught; there are additional details in our other comments.

Resp 13:

Resp 14: I think there is too much emphasis on disciplinary boxes, and too little emphasis on allowing bright students to go beyond the narrow disciplinary boundaries. In some instances, the treatment is dumbed down. Paradoxically, in others, the treatment assumes more than students can be expected to master. The entire focus seems to be to lock-step all students through a set framework. This framework would not foster curiosity, creativity, or differences in inherent talent that students may present. Some will be gifted at math and logic, others will be gifted at seeing patterns and making unusual comparisons. As it is, this framework is a grab-bag of concepts that students will be forced to march through in lock-step.

Resp 15:

Resp 16:

Resp 17: While the opening of the Report Scale Scores section (p. 71) makes a re-titling of the 3 subscales sound like a minor labeling issue, that framing masks the fundamental shift that this framework is aiming for. The notion that the NAEP will no longer be a straight content assessment should be clearly articulated (and celebrated). That has significant implications for science curriculum and instruction, and the public's understanding of what matters for science literacy and proficiency in today's world. (It also provide a justification if it is no longer to continue trend lines.) In order to successfully make the case for the proposed transition, however, the reporting has to be able to better account for the other two dimensions (particularly science & engineering practices, in my opinion) other than just saying 'they are embedded in the items.' Otherwise people will likely continue to view the updated NAEP as just a content assessment.

Resp 18:

Resp 19: As further detailed below, the reporting categories are insufficient to spur discussion that is as rich as it could be.

Resp 20:

Resp 21:

Resp 22:

Resp 23: Several pieces of the Discipline Concepts seem to go beyond what the DCIs in the NGSS outline in the grade-level bands. E.g. PS8.3. My concern is the going beyond the grade bands as outlined with the standards will not provide metrics that are as meaningful in the assessment.

Resp 24:

Resp 25:

Resp 26:

Resp 27: Error through exclusion of material in L12.12 in that archaebacteria are not included and cannot fall into the category of algae. L12.16 All cells and organisms have the same genetic content. This is incorrect in gametes, in RBCs. This is an opportunity to reference most somatic cells vs ALL cells.

Resp 28:

Resp 29:

7. **The organization of the framework is useful to understand the content and context of what students should know and be able to do in science.**

Resp 1: Disagree

Resp 2: Agree

Resp 3: Agree

Resp 4:

Resp 5: Agree

Resp 6: Disagree

Resp 7: Agree

Resp 8: Agree

Resp 9: Agree

Resp 10: Agree

Resp 11:

Resp 12: Disagree

Resp 13: Agree

Resp 14: Disagree

Resp 15: Agree

Resp 16: Agree

Resp 17: Agree

Resp 18: Agree

Resp 19: Agree

Resp 20: Agree

Resp 21: Agree

Resp 22: Agree

Resp 23: Agree

Resp 24: Agree

Resp 25: Agree

Resp 26: Agree

Resp 27: Agree

Resp 28: Agree

Resp 29: Agree

8. **What suggestions do you have to improve the organization of information?**

Resp 1: If NAEP is going to continue to advocate for multidimensional science frameworks, consider a measurement model and reporting that highlights different expertise and not just traditional content categories.

Resp 2:

Resp 3:

Resp 4:

Resp 5:

Resp 6: First, we suggest including some of the discussion of what three-dimensional items operationally look like (well discussed in Chapter 3) into Chapter 2 before the learning expectations/statements. Doing this, perhaps via a sample item, would provide needed context for how to read and consider the three different parts (dimensions) of the expected content by showing how they need to be integrated together. As the framework stands currently, all the learning expectations/statements are segmented in their presentation, and that makes it hard to think about or evaluate the expectations we are setting for students in a holistic, integrated way aimed at sensemaking.

Second, we recommend using a landscape orientation instead of portrait to allow for more comparison of grouped ideas and the tracking of those ideas across grades. In some places the groupings are a little clearer than others, and where there are actual physical lines included (e.g., as attempted in the practices) that is helpful. We definitely suggest going back to the disciplinary content and crosscutting concepts and adding lines there too. [Adding matched numbering of ideas across the grades would help even more. We understand the traditional approach has been to sequentially number the statements but perhaps you could use an additional categorization to keep associated ideas together, e.g., PS4.5.1, PS4.5.2, etc., and associate the schema across all three grades.]

Third, we suggest you take out the boundaries and clarifications for the disciplinary content statements here and instead supply them with the statements in an expanded appendix, so that within the main body of the document, the idea groupings (with at least physical lines) are easier to view as groupings.

Resp 7:

Resp 8:

Resp 9:

Resp 10:

Resp 11:

Resp 12: CODING SYSTEM -- We would suggest having an intermediate classification in the content progressions so that is easier to see which concepts align with only a coding system and not needing the full text. For example, the motion and forces subsection of physical science could add the code MF after the grade. So, P4.5 would become P4.MF.5 (or 1 and renumber by subsection) and P8.11 would become P8.MF.11 so that the relation across grade bands would be clear. This would hold true for LS and ESS domains as well. This also aligns well with the NRC Designing Assessments for the Next Generation Science Standards that suggest that assessments of the NGSS or similar standards should place students along the continuum of a learning progression; therefore, using a coding system that allows for easier tracking of student performances across grade bands in the progression would be preferable. There may need to have subsections that are added below this level based on the added box grids in the statement lists. Adding this will help analysis, development, and school systems try to interpret and use the NAEP results to improve student learning. This will help to clarify boundaries at the grade levels and how the conceptual understanding changes from elementary through middle school and culminating in high school understandings.

Resp 13:

Resp 14: As I have said in earlier contexts, this is a potpourri of science concepts that are thrown together as what presumes to be a comprehensive science curriculum. I expect it will hold back the scientifically talented, and will foster an ill-founded concept of what science is on the rest.

Resp 15:

Resp 16:

Resp 17:

Resp 18:

Resp 19:

Resp 20:

Resp 21:

Resp 22:

Resp 23:

Resp 24:

Resp 25:

Resp 26:

Resp 27:

Resp 28:

Resp 29:

9. **The sample items are a useful representation of what the NAEP Science Framework is measuring.**

Resp 1:

Resp 2: Disagree

Resp 3:

Resp 4:

Resp 5: Agree

Resp 6: Disagree

Resp 7: Agree

Resp 8: Agree

Resp 9: Agree

Resp 10: Agree

Resp 11:

Resp 12: Disagree

Resp 13: Agree

Resp 14: Agree

Resp 15: Agree

Resp 16: Agree

Resp 17: Disagree

Resp 18: Agree

Resp 19: Agree

Resp 20: Agree

Resp 21: Agree

Resp 22: Agree

Resp 23: Agree

Resp 24: Agree

Resp 25: Agree

Resp 26: Agree

Resp 27: Agree

Resp 28: Agree

Resp 29: Agree

10. **What other examples or sources of available items or tasks would be helpful to illustrate the three dimensions of science?**

Resp 1:

Resp 2: While I think the questions do a good job of making students apply the three dimensions of science, I think the particular context selected (i.e., presence of seaweed) may not be the best choice. Most of the current news stories regarding seaweed focus on its overabundance in southern Atlantic waters and resulting problems, rather than the value of seaweed in the ecosystem. I think that better sources for sample items might come from problems related to relevant societal issues such as climate

change, sustainable energy sources, medical challenges, etc. I think the overall structure of the sample item is good, but I think the context could be improved to make it more timely, relevant and interesting to students.

Resp 3:

Resp 4:

Resp 5:

Resp 6: We would suggest looking across state websites for sample or released items from state assessments that are aligned to multi-dimensional (NRC Framework-based) standards. We would expect many states would be willing to agree to allow you to use some as exemplars.

Items from the Next Generation Science Assessment project (https://ngss-assessment.portal.concord.org/middle-school) may also provide some examples that are useful.

For all samples we strongly recommend annotation to assure you illustrate the points intended.

Resp 7: It would be useful to have more questions in which students use computational models. The boundaries of the assessments are unclear as to whether they are reading a graph, using an equation, deriving an equation or something else.

Resp 8: Some of the expectation statements include terms such as “use digital tools” (S12.33). It is unclear how this would be assessed based on the samples provided. BL input - The paper-pencil version of the example provided did not highlight features of the newly updated framework as effectively as an online interactive example would. I would like to see examples of scenario based tasks using an interactive online demonstration such as this example for 4th grade math: https://npd.naep.ed.gov/totw/2022/app/en/main.html?subject=Math4

Resp 9: I think a call out image or visual example of a phenomenon would be helpful. Many educators and education partners do not know what phenomena are. Having a more detailed understanding of what they are could be helpful.

Resp 10: The Stanford Assessment Project is a good example.

Resp 11:

Resp 12: Generally, we agree. However, there were not enough of them to fully answer the question. What was there was appropriate, but not enough of the domain was represented. It is difficult to tell what is being measured without representation of how responses are scored. Rubrics/sample papers would help. It would also be useful to have examples of grade 4/12 AND Physical and Earth Science items, especially if the level of detail in those domains is similar to this draft version.

It would be useful to also have examples of acceptable culturally relevant scenarios/contexts, as called for in lines 665-680.

Resp 13: If possible, scenario/problem - based exercises on current topics: water (toxic spills, uses in agriculture, industry, etc.); climate change on crops, food production, biodiversity, ethnobotany; land use in urban, suburban & rural areas; air quality. Identify and include issues "in the news."

Resp 14: Only three dimensions to science?

Resp 15:

Resp 16: Examples of tasks for grades 4, 8, and 12 (with similar 3D components) to illustrate the progression of items/tasks

Resp 17: The example items do not explicate for me how the multi-dimensional approach is carried out in practice. They do not discuss, for example, how each item specifically draws upon each of the associated dimensions. For any particular multiple-dimension item, a student could 'fail' that item because they did not have a grasp of any one of the dimensions the items draws upon. To use example item 1 to illustrate (p. 60-61), which relates to L8.11 and L8.16 and practice of developing and using models, a student might provide the wrong answer because they can't decipher or use the model (the web), or because they don't understand one of the two concepts (8.11 about dependencies or 8.16 about representing transfer of matter and energy). How will an analysis of the student's response account for why they got it wrong? It seems that a 'right answer' can be interpreted as an full understanding of all three elements. But a 'wrong answer' will not elucidate which of the three the student understood, or more specifically, which they did not understand. Ideally, wrong answer choices could be designed to elicit those distinctions, but there is no discussion of that in the text or in association with the sample items. (On a separate but related point, why specify dimensions of the practices in the practices section [pp.38-50] but then only refer to the overall practice in the item?) Sample Item #1 is the simplest of the sample items; they only get more complex from there and seemingly less able to speak to what elements of the multiple dimensions students have successfully (or not) applied to their answer. I don't see how reporting will be able to make a claim about disciplinary concept domain understanding without being able to account for the impact of the other two dimensions which are assumed as necessary to complete the items. A second reason that this is hard to envision in practice is that the items are all very text heavy, particularly the longer multi-part scenario examples. A student is as likely to trip up on the language, mis-interpret language somehow, or not have enough language proficiency to successfully complete these items. It would be great if there were example items that were more visual in nature, or even interactive, that did not rely so much on substantive and dense text.

Resp 18: Science Assessment Item Collaborative Materials (https://csaa.wested.org/spotlight/science-assessment-item-collaborative/)

Resp 19: What examples are coming from states working toward innovative assessment design? The examples here look like what states have already been doing for several years. It would be great to have more innovative analyses of student science learning in 2028.

Resp 20: Science Assessment Item Collaborative Materials (https://csaa.wested.org/spotlight/science-assessment-item-collaborative/)

Resp 21: N/A

Resp 22: consider examples from Inner Orbit; language from STEM Teaching Tools

Resp 23: Example 2 & example 3 seem especially headed in the direction of multi-component tasks to measure the framework.

Resp 24:

Resp 25:

Resp 26:

Resp 27: Sample animations or simulations would help.

Resp 28: Sample responses would be helpful to understand how the items are intended to be scored. Can a CEDS and OECI sample item be shared? The EAFE examples seemed very simplistic and borderline CEDS.

Resp 29:

11. **I have recommendations for terms to include in a framework glossary to aid understanding of the framework.**

Resp 1:

Resp 2: No

Resp 3: No

Resp 4:

Resp 5: No

Resp 6: Yes

Resp 7: No

Resp 8: Yes

Resp 9: No

Resp 10: No

Resp 11:

Resp 12: Yes

Resp 13: No

Resp 14: No

Resp 15: No

Resp 16: Yes

Resp 17: No

Resp 18: No

Resp 19: No

Resp 20: No

Resp 21: No

Resp 22: No

Resp 23: No

Resp 24: No

Resp 25: No

Resp 26: No

Resp 27: No

Resp 28: No

Resp 29: No

12. Which terms or phrases should be included?

Resp 1:

Resp 2:

Resp 3:

Resp 4:

Resp 5:

Resp 6: It would be helpful to include the acronyms (e.g., NAEP, NRC, TEL) used frequently in the document for easy reference for readers less familiar with them.

Resp 7:

Resp 8: While the NRC Framework is referenced, in my review I found that the NAEP framework relied heavily on other resources. This needs to be clear in both the glossary but also in the references and narratives describing the NAEP framework. These resources should include Appendix E, Appendix F, Appendix G, and the DCI/SEP/CCC descriptors found in the three foundation boxes associated with the NGSS Performance Expectations (Volume 1, The Standards, Next Generation Science Standards For States, By States).

Resp 9:

Resp 10:

Resp 11:

Resp 12: We are unsure whether this refers to science content terminology, assessment terminology, or some other list of appropriate words to include in a glossary. Does the wording of the DCIs imply what terms are fair to be tested or assumed to be understood by students? For example, aerobic and anaerobic appear in the DCI list (see L12.9/11); however, the terms biotic and abiotic do not see L8.11). Does this imply that no items will require knowing the terms biotic/abiotic, but some items might use aerobic/anaerobic?

Resp 13:

Resp 14:

Resp 15:

Resp 16: phenomenon

Resp 17:

Resp 18:

Resp 19:

Resp 20:

Resp 21:

Resp 22:

Resp 23:

Resp 24:

Resp 25:

Resp 26:

Resp 27:

Resp 28:

Resp 29:

About the Content

Disciplinary Concepts

13. **The physical sciences, life sciences, and Earth and space sciences concept statements currently are written in different styles. Which style do you prefer, and why?**

Resp 1:

Resp 2:

Resp 3: First off, the way this question is written led me to believe that the physical science concepts were written in one way, the life science concepts in second way, and the earth and space science concepts were written in a third way. After studying the text for a good bit, I determined that this was not the intent of the question.

I didn't notice a difference in the main statements, but I did notice differences in the clarifications and boundary statements. I liked the following three features:

1. Inclusion of parenthetical statements in the main statements to provide examples.

2. Boundary statements that were phrased in the format that began "Students are not expected to..."

3. Clarification statements that were phrased in the format that began "Emphasis is on..."

Resp 4: [I focused on Earth Science only] I appreciate that the statements highlight the way we want students to approach problems in Earth science, with unifying conceptual frameworks. However, I am concerned that, like the NGSS, many of the Earth science content statements are not specific enough to guide decisions for assessment developers (or for instruction, as NAEP ends up influencing instructional decisions). The NAEP assessments are supposed to be phenomena-driven and about student sensemaking, but many of the statements don't specify the class of phenomena students should be expected to reason about, leaving it open or providing a broad list of example phenomena. This leaves the choice of the phenomenon to the task developer, and also leaves them with no specifics about what students should be expected to know about the processes or mechanisms relating to any particular phenomenon. With the NGSS, I've seen this lead to items that a. require knowledge that students didn't have the opportunity to learn (because their curriculum focused on a different phenomenon from that list) b. don't require any specific knowledge so instead provide all the information in a text-heavy item that can't realistically require very deep sensemaking, or c. don't go into any specific processes/mechanisms at all and end up being very superficial. I would really strongly recommend further specifying the statements to more concretely define what a student should know and be able to do - for each statement it should be clear what the observation or class of phenomenon to explain/make predictions about is, and at what level of detail the student should know the mechanism.

Resp 5: The life science concept statements seem most useful to me for a couple of reasons:

1) The Boundaries consistently refer to the assessment and what it includes. (Statements in physical and Earth and space sciences seem to vary in terms of what is being described.)

2) I found the Clarifications in the life science concept statements to be particularly helpful. Physical sciences didn't include these Clarifications, and Earth and space sciences Clarifications were less consistent about what was being described. In general, I found the "Examples of" or "Emphasis is on…" Clarifications (more prevalent in the life sciences concept statements) to be the most useful.

Even though the Clarification statements add quite a bit to the length of the Framework, I think they are worth including because they provide information that is essential to item developers and reviewers, as well as those who are "consumers" of NAEP Science (and the NAEP Science Framework). Where concept statements could be interpreted in different ways, the Clarification statements narrow down the possible interpretations. This is especially important for the item development and review process, where

different interpretations of a given element of the Framework can add significant time and expense to the process. (It seems much more efficient, and more consistent with the Framework's intentions, to anticipate where there could be different interpretations and provide clarity upfront, rather than relying on item writers and reviewers to come to consensus about what might have been intended while items are being developed.)

Resp 6: Assuming this was in reference to the annotations (i.e., boundaries, clarifications), we believe the style in Earth and space sciences seemed to strike the best balance. In physical sciences only boundaries were provided, and in life sciences there were a very large number of annotations (boundaries and clarifications) where perhaps not everything needed explanation.

As an alternative, however, we'd like to again put forth the suggestion as we did in an earlier question about framework organization that in the main body of the framework you only include the concept statements and then in an appendix at the end, you can provide an expanded version of the concept statements with as many boundaries and/or clarifications for the statements as needed.

Resp 7: It was hard to tell the differences between styles in the actual DCI's were written. I appreciated the boundary and clarification statements. I liked how the questions in the biology sections (I assume the questions in italics) all relate to why a student would care about the material. The earth science questions were nice as they were directly related to processes on Earth, but the student buy in was not as obvious. The physical science questions were helpful, but did not relate to students.

Resp 8: I do not have style preference but do think it is important to keep the Clarification and Boundary statements.

Resp 9: Life science seems easiest to read.

Resp 10: The addition of the clarification statement and the boundary will assist those crafting the assessment items and with the grade-level descriptors. However, if such detail is needed, then the concept (idea) statements are not clear.

Resp 11:

Resp 12: Every DCI should include a boundary statement and most should include a clarification. Because this is an assessment framework, the boundaries of the assessment should be known in advance, like the PE of NGSS.

Resp 13: I'm a little confused by this question. If you're referring to concepts, clarification, and boundary, I welcome all three because it guides instruction and indirectly curriculum development.

Resp 14: I frankly can't see any difference. All comprise a laundry list of topics that are sometimes incoherent. 4th graders shouldn't be expected to grasp the difference between liquid, solid and gaseous materials?

Resp 15: I did not notice a style change

Resp 16: I prefer the physical science style as it is more concise than the earth/space science style and much less wordy than the life science style.

Resp 17: I'm not seeing obvious differences in styles in the concepts statements across the three sciences. The are all organized by grade (column) with a concept statement and sometimes a clarification and/or boundary.

Resp 18: We are not really sure what the difference in the styles is other than there are no clarification statements in PS. We prefer the LS and ESS which have both clarification and boundary statements.

We also prefer clarification and boundary statements where appropriate. They will ensure we agree on the expected rigor.

Resp 19: They do not seem significantly different enough to make a difference.

Resp 20: I am not really sure what the difference in the styles is other than there are no clarification statements in PS. I prefer the LS and ESS which have both clarification and boundary statements.

Resp 21: The life science has examples of the concepts whereas the physical science do not list examples. The examples tend to assist individuals with understanding the total concept.

Resp 22: Include both clarification and boundary statements to improve what is actually assessed versus what is beyond assessment limits

Resp 23: I don't think that I have an opinion on this

Resp 24:

Resp 25: No comment

Resp 26:

Resp 27: The inclusion of clarification and boundary statements make the content more visible to teachers and test developers; however, these statements may also serve as strict guardrails in terms of question development and use. This a careful dance when rolling out an assessment framework. We want students to develop conceptual understanding through engaging with phenomena, asking questions, etc. however, the framework provided such specificity in the clarifications (at times and inconsistently) that it will send an unintended message that content should be taught to the test.

Resp 28: Earth and space science concept statements are stylistically easier to understand along with the clarification and boundary. The Life Science concept statements are broader and require more interpretation to understand as written what is implied without looking at the performance indicators that NGSS provides to understand what students need to demonstrate as 3-D learning.

Resp 29:

14. **For grade 4, do the science concept statements represent what you believe should be assessed by NAEP? Why or why not?**

Resp 1:

Resp 2:

Resp 3: yes

Resp 4: [Earth Science] I think this is a strong basis for what should be assessed at grade 4, but I think it would be important to fully specify the set of observations/phenomena that could be considered in task development (more about this in my comments above). As a guiding principle for making those decisions for grade 4, I would focus on students making explanations and predictions for commonly observable phenomena that rely on a simple (one or two component) mechanism.

For Earth's Place in Space, E4.2 nicely includes the mechanism of Earth's rotation. I would specify that the set range of observations to explain could include: apparent motion of Sun and moon. And along with E4.3, students could explain the pattern of day and night. E4.1 content seems fine, but may be difficult to make a multidimensional sensemaking task with it.

For Earth's Systems, E4.4 is a nice mechanism for explaining the shaping of rock features, the cutting of channels, and alluvial fans. I think the inclusion of energy and gravity in the clarification goes too far for grade 4. E4.5 Seems fine in terms of having students use maps. The clarification makes it sound like students might also be expected to bring in knowledge of all the different features and events, which doesn't seem as appropriate. E4.6 has a lot of overlap with E8.7, not much is added for grade 8 (or 12). I think it may be a more appropriate fit for grade 8 (or 12), students at grade 4 can reason about these things, but deep time is even more abstract and the idea of having evidence for the history of Earth is probably less relevant. E4.7 can help explain the erosion and deposition phenomena of E4.4, but then also includes some more declarative content (in the concept statement and in the clarification) that doesn't add much. E4.8 is clear and a nice stopping point before grade 8.

Earth and Human Activity as a whole is less specified and leaves more for the task developers to decide, which can result in either really shallow tasks or tasks that may expect understanding of phenomena and mechanisms that may not be familiar or appropriate for 4th grade. It is hard to say without providing some more detail about what specific knowledge is expected of students. More about this in my comments about the style in which the concept statements are written.

Resp 5:

Resp 6: Overall we believe the science disciplinary concept statements for grade 4 represent what should be assessed by NAEP. A large portion of the statements reflect what is contained in the NRC A Framework for K–12 Science Education; knowing that most states in the nation have science standards based on that framework, the disciplinary concept statements presented in the new NAEP Science Framework are generally appropriate.

While not as prevalent in grade 4 as in grades 8 or 12, we note there are some statements that express extremely similar content (e.g., P4.6, P4.7, P4.8). The writing team should look carefully to make sure ideas are distinguished so that each concept statement is clear and distinct.

Resp 7:

Resp 8: No, specifically:

Physical Science

P4.11 “the earth exerts a gravitational pull on other objects near or far as well as those on its surface” 3-PS2-1 Clarification Statement includes: Assessment is limited to gravity being addressed as a force that pulls objects down. 5-PS2-1 Clarification Statement includes: “Down” is a local description of the direction that points toward the center of the spherical Earth. Based on this, P4.11 extends beyond 4th grade when including the vague term “far” combined with “pulls objects down”. The P4.11 Boundary includes “gravitational phenomena”. Does far mean kite? Airplane? Satellite? Moon?

Life Science

L4.5 includes “minerals (in the soil)” which is mentioned in the K-12 Framework but not explicitly cited in the NGSS documents. 2-LS2-1. States “Plan and conduct an investigation to determine if plants need sunlight and water to grow.”

Earth and Space Science

E4.1 is mentioned in the K-12 Framework grade band endpoint (grade 2) but not explicitly cited in the NGSS documents.

E4.3 is associated with 5-ESS1-1 and ESS1A The sun is a star that appears larger and brighter than other stars because it is closer. Stars range greatly in their distance from Earth. This extends beyond the 4th grade expectation.

E4.4 clarification statement introduces the term deposition which is not reflected in the K-12 Framework or NGSS documents.

Why is a general understanding of climate missing from the K-4 assessment framework?

E4.12 – “Changes to Earth’s global average temperature” - the ESS3.D DCI is not associated with grades K-4.

Resp 9: Yes, for the most part. I have concerns about...

1. E4.3: The sun is a star that appears larger and brighter than other stars because it is closer. Stars range greatly in their size and distance from Earth.

- In most states (NGSS and Framework aligned states), this is concept would not be taught until 5th grade. It is incredibly unfair to assess student on a core idea that most of them will not learn until 5th grade. More over, the last time a 4th grader would have discussed the sun in general terms would have been in 1st grade. Students making sense of the patterns of the sun (including seasonal patterns) would be most appropriate for most of our 4th grade students in the country. For that reason, I think this science idea should be eliminated.

Resp 10: did not review

Resp 11:

Resp 12: Generally, yes. Some aspects, particularly at Grade 8 seem overly aspirational and may be better suited for Grade 12. Even at Grade 4, given the limited science instruction time historically (see NAEP research by Blank or ACT research), the Grade 4 targets may be overly aspirational. The apparent emphasis on plate tectonics (which cannot be observed directly by students) over observable surface processes, is not appropriate, since more concrete observable ESS processes are more likely to be both

relevant to students and students are more likely to have had instruction relating to observable surface interactions and phenomena, especially at Grade 12 (for example, in environmental science).

Resp 13: Yes. Those included are basics for building a solid foundation to more complex information and therefore a more complex conceptual understanding.

Resp 14: Again, this is a laundry list of concepts that 4th graders can supposedly understand, marked off by boundaries that seem arbitrary or incoherent. Case in point: Clarification of food, asserting that animals take in food, but plants do not? Misses the opportunity to tell students that animals and plants have a common conception of food.

Resp 15: Yes, They seem open-ended and to allow considerable leeway.

Resp 16: Yes. Since the grade 4 test would be given midyear, I believe it represents reasonable expectations for student learning up to the 4th grade. The concept statements also represent a relatively basic level of knowledge that will be necessary as a foundation for future learning to build upon.

Resp 17:

Resp 18: Yes, the concepts identified reflect the work that the NRC Framework did to identify the core concepts that are needed by all students at the various grade bands.

Resp 19: Yes, it seems you've followed the NRC Framework and NGSS, which makes sense in the current national context.

Resp 20: Yes, the concepts identified reflect the work that the NRC Framework did to identify the core concepts that are needed by all students at the various grade bands.

Resp 21: Yes it states what the internal and external structures have

Resp 22: yes; the concepts look appropropriate to me based on the DCI progressions

Resp 23: I can't speak well to elementary

Resp 24:

Resp 25: No comment

Resp 26:

Resp 27: Consider including gases when addressing the phases of matter in fourth grade.

Resp 28: I teach secondary science. My expertise is limited to 6-12 science teaching.

Resp 29: Yes, because it starts building the foundations for concepts.

15. **For grade 8, do the science concept statements represent what you believe should be assessed by NAEP? Why or why not?**

Resp 1:

Resp 2:

Resp 3: yes

Resp 4: [Earth Science] In addition to my comments about the style of the statements, I would also note that the current set of grade 8 concept statements includes at least as much content as the NGSS. I have worked closely with developers of three very widely used NGSS curricula, and all have found in working with teachers (through both developing and using their curricula) that teachers don't have time to teach all of the NGSS Earth Science standards in middle school. I worry that if NAEP ends up driving some instructional decisions, it would continue to put pressure on schools to teach more things in less depth, or to just not get to the last units of a curriculum which often include climate change. I don't know what all of your constraints are, but I think it would be worth including some phenomena at only one grade level instead of two or three, even if that means only hitting a certain level of sophistication of a given mechanism. There are many ways to cut back on the content, I could imagine using a guiding principle to make decisions for each grade level. For example, something like the transfer of energy within and between Earth's systems could be a principle for 8th grade, in which case the phenomena could include seasons, volcanic activity driven by energy in earth's interior, weathering driving by energy from the sun, rainshadow effect, global warming, etc. But regardless of how the decisions are made, less content could really help lead to deeper, more consistent instruction in middle schools.

I started going through the content statements initially to provide potential phenomena to specify. However it began to seem like just too much content so it didn't make sense to go through all of it. I am just including the notes below as examples, with some thoughts about relevance to students' lives and helping students to understand that Earth science and claims about the history of Earth are based on evidence as considerations when specifying phenomena.

For Earth's Place in Space, E8.1 builds nicely on E4.2. I would limit the patterns to seasons and the phases of the moon, as two familiar, observable phenomena. I don't think the middle school understanding of gravity is appropriate for these ideas, and I would not include tides or the different positions of sun/moon/stars at different times (as it is partly covered in grade 4, and less relevant than seasons or moon phases). E8.2 seems really declarative, and though gravity is also in the NGSS, it doesn't seem appropriate given the expectations for understanding of forces in grade 8. E8.3 seems fine, but may be difficult to make a multidimensional sensemaking task with it.

For Earth's Systems, E8.4 is either really huge or really abstract. It would be really helpful to tie these ideas to a concrete set of possible phenomena to explain, to guide assessment development and make sure the phenomena driving tasks are appropriate. I think I would limit the cycling of matter to rock and water. It seems like the transformation and conservation of rock materials could be used to explain things like volcanic activity and sedimentary rock. A water focus could be merged with E8.8 to explain things like the rainshadow effect or humidity/rain near water.

E8.5 is a nice foundation for E8.6, and might be nice to test together. E8.6 has a lot in the clarification. Again I think it would be good to name a set of phenomena that can be explained with plate motion, and to make the creation and destruction of seafloor material more clearly part of the concept statement and not just a clarification. In terms of phenomena to be explained, I would propose: seafloor ridges and trenches (as things that are very closely linked to the creation and destruction of plate material), and mountains, volcanoes, and earthquakes (familiar and easily observable). E8.7 is nice as a way for

students to start thinking about the evidence we have for the history of Earth, but the clarification goes too far.

Resp 5:

Resp 6: Overall we believe the science disciplinary concept statements for grade 8 represent what should be assessed by NAEP. A large portion of the statements reflect what is contained in the NRC A Framework for K–12 Science Education; knowing that most states in the nation have science standards based on that framework, the disciplinary concept statements presented in the new NAEP Science Framework are generally appropriate.

However, we note there are some statements that repeat each other (e.g., P8.3 and P8.5) or express extremely similar content (e.g., L8.10, L8.15, and L8.16). The writing team should look carefully to see where ideas can be streamlined and/or distinguished so that each concept statement is clear and distinct. There are also some statements that seem quite focused on definitions as opposed to larger conceptual understanding, and we are uncertain, based on the statements alone, if these can be appropriately translated into application in assessment items (e.g., P8.21). Finally, some statements are quite lengthy and as a result the targeted construct seems to get lost; if all the content is needed, perhaps these can be broken apart into multiple statements (again being sure each statement is a unique/distinct construct).

In terms of domain-specific feedback, in physical sciences we would suggest that light be more explicitly called out in several of the statements (or clarifications) relative to wave concepts. In Earth and space sciences, while the ideas presented in the concept statements in Earth's Place in Space reflect disciplinary ideas in the NRC Framework, they seem overshadowed by the number of statements in the other two subgroupings. We also wonder if you have gathered data from reviewing all states' standards as to how inclusive states are relative to climate change concepts. While climate change concepts are included in the NRC Framework as well, we have some concerns about whether the number or content of those statements would need to be scaled back (especially at grade 12, maybe not quite as much here at grade 8). The same may be said for some evolution concepts in life sciences.

Resp 7:

Resp 8: Yes, however there is a duplicate concept, specifically:

Physical Science

P8.3 and P8.5 are identical concepts

Resp 9: Yes. These seem include the major concepts and science ideas students will have learned at the end of that particular grade band. Topics relating to climate, human activity, and evolution are included!

Resp 10: No. When there is a disparity in the number of ideas in each domain (PS = 30, LS=28, ES=20), there is a problem. When comparing what is in the 2012 Framework and what is proposed here, I quickly identified numerous physical science "ideas" that were above the grade-band (see the list below), and therefore based on research, beyond the reach of ALL students.

PS concerns: PS 8.3; 8.5; 8.24 (all above grade-band)

Missing: PS2.C: Stability and Instability of Physical Systems

LS concerns: LS 8.8 (oversteps content, includes three separate concepts), 8.15 (redundant with 16)

Missing:

LS1.D Information Processing

LS4.B.1: Natural selection leads to a predominance of certain traits in a population, and the suppression of others.

ES concerns: E8.3 (combines practices of science; equity issue if this not presented in a school curricula); ES8.5 (should be in grade 12)

Missing:

ESS1.A .1: Earth and its solar system are part of the Milky Way galaxy, which is one of many galaxies in the universe.

ESS2.C.3: Global movements of water and its changes in form are propelled by sunlight and gravity.

ESS2.C.4: Variations in density due to variations in temperature and salinity drive a global pattern of interconnected ocean currents. (only listed in the clarification statement in ES 8.9)

ESS2.D.3: The ocean exerts a major influence on weather and climate by absorbing energy from the sun, releasing it over time, and globally redistributing it through ocean currents.

Resp 11:

Resp 12: Generally, yes. Some aspects, particularly at Grade 8 seem overly aspirational and may be better suited for Grade 12. Even at Grade 4, given the limited science instruction time historically (see NAEP research by Blank or ACT research), the Grade 4 targets may be overly aspirational. The apparent emphasis on plate tectonics (which cannot be observed directly by students) over observable surface processes, is not appropriate, since more concrete observable ESS processes are more likely to be both relevant to students and students are more likely to have had instruction relating to observable surface interactions and phenomena, especially at Grade 12 (for example, in environmental science). Given the limitations of science instructional time, some of the Grade 4 targets may need to move to Grade 8.

Resp 13: Yes. Those included build on the foundation to more complex information and therefore a more complex conceptual understanding.

Resp 14: Again, there's nothing in the laundry list that I disagree with, just that it does not capture the essence of what science is, and in fact might foster a phony conception of science as an accumulation of factoids.

Resp 15: Yes,

Resp 16: Yes. The 8th grade concepts represent an appropriate level of understanding that indicates a deeper understanding and level of mastery when compared to the corresponding 4th grade statements.

Resp 17:

Resp 18: Yes, the concepts identified reflect the work that the NRC Framework did to identify the core concepts that are needed by all students at the various grade bands.

Resp 19: Why does "assessment … not include the mechanism of one body system independent of others"? Does making sense of some phenomena always require multiple body systems?

Resp 20: Yes, the concepts identified reflect the work that the NRC Framework did to identify the core concepts that are needed by all students at the various grade bands.

Resp 21: Yes it discuss and describes what living things are made of

Resp 22: yes, although E8.7 and E8.18 go beyond the 6-8 DCI progression

Resp 23: Yes. The content builds up into 12 grade and seems to be consistent with grade-appropriate content.

Resp 24:

Resp 25: No comment

Resp 26:

Resp 27: In P8.5 the boundary states that students need not know any details of the substructure. Does this mean that they do not need to know proton, neutron, and electron locations and roles, or does this mean a knowledge of these but no further breakdown of the construction of these subatomic particles (standard model).

Resp 28: Although the DCIs on their own make sense, we need to see what CCs and SEPs are being assessed with the DCI to fully answer this question. Student 3-D science learning requires cohesiveness and depth relative to the combination of the DCI with the CC and SEP.

Resp 29: yes. It increases in complexity and concepts are developed for deeper understanding.

16. **For grade 12, do the science concept statements represent what you believe should be assessed by NAEP? Why or why not?**

Resp 1:

Resp 2:

Resp 3: yes

Resp 4: [Earth Science] My overall comments in the section about style apply here again - I think it would help to fully specify the expectations of students. I am less familiar with the amount of content at grade 12 than I am at grade 8, though I do think it's worth getting some insight into how teachers are managing the NGSS PEs for Earth and Space Science. If they're not able to adequately treat all of them, it would be really wonderful to make decisions here about what to prioritize so that the NAEP assessments can be a step towards helping the problem.

Resp 5:

Resp 6: Overall we believe the science disciplinary concept statements for grade 12 represent what should be assessed by NAEP. A large portion of the statements reflect what is contained in the NRC A

Framework for K–12 Science Education; knowing that most states in the nation have science standards based on that framework, the disciplinary concept statements presented in the new NAEP Science Framework are generally appropriate.

However, similar to grade 8, there are some statements that express extremely similar content, or seem quite focused on definitions as opposed to larger conceptual understanding, or are quite lengthy.

In terms of domain-specific feedback, in Earth and space sciences, again the concept statements in Earth's Place in Space seem overshadowed by the number of statements in the other two subgroupings, and we wonder even more so than for grade 8 if you have gathered data from reviewing all states' standards as to how inclusive states are relative to climate change concepts. Lack of opportunity to learn is not the same as not having learned, and so if climate change is less represented and/or not taught, interpretation of the data could become difficult. It will be important to take this into account and consider if the aim is to compare student understanding across states, or to set an expectation about what content should be taught for all students. The same may be said for some evolution concepts in life sciences.

Resp 7: For the most part:

In physical science MATTER AND INTERACTIONS we would like to see a the concepts:

Types of interactions based on electron configurations (mainly in terms of bonding and relative strengths) which can then be applied to

Explaining how the bulk properties of materials are determined by small scale structures. This is somewhat addressed in P12.16, but seems to fit in the Matter section.

In MOTION AND FORCES:

P12.12: Could the sum of forces causes a mass to accelerate be called out as its own statement. We recognize that these concepts were derived in terms of momentum but calling it out on its own could draw attention the concepts of accelerations.

P12.14: The statement goes in depth when describing gravitational interactions, but can statements about charge interactions be added as this lays the groundwork for chemical interaction and can lead into EM wave communication. This concepts is also useful for explaining friction at the atomic level.

Resp 8: No, specifically :

Physical Science

Why is P12.7 so specific when L12.6 also addresses photosynthesis? Is there a reason it doesn't simply address carbon-cycling as described in the K-12 Framework?

Resp 9: Yes. These seem include the major concepts and science ideas students will have learned at the end of that particular grade band. Topics relating to climate and evolution are included!

Resp 10: No. When there is a disparity in the number of ideas in each domain (PS =30, LS=26; ES=17), there is a problem. When comparing what is in the 2012 Framework and what is proposed here, I quickly

identified numerous physical science "ideas" that were above the grade-band (see the list below), and therefore based on research, beyond the reach of ALL students.

PS concern: PS 12.2; 12.4; 12.7; 12.11; 12.13; 12.15; (12.11, 12.13, and 12.15 all overstep the grade level - why would ALL students need to know these "ideas?"), 12.24 (this should be in ES = ESS2.D)

Missing: PS2.C: Stability and Instability of Physical Systems

ES concerns: E12.10 includes multiple components from various parts of the 2012 Framework and should be separate entities given the importance of each piece.

Missing:

ESS1.A.1: The star called the sun is changing and will burn out over a lifespan of approximately 10 billion years.

ESS1.A.3: The Big Bang theory is supported by observations of distant galaxies receding from our own, of the measured composition of stars and non-stellar gases, and of the maps of spectra of the primordial radiation (cosmic microwave background) that still fills the universe.

ESS1.A.4: Other than the hydrogen and helium formed at the time of the Big Bang, nuclear fusion within stars produces all atomic nuclei lighter than and including iron, and the process releases electromagnetic energy. Heavier elements are produced when certain massive stars achieve a supernova stage and explode.

ESS1.C.1: Continental rocks, which can be older than 4 billion years, are generally much older than the rocks of the ocean floor, which are less than 200 million years old.

ESS1.C.2: Although active geologic processes, such as plate tectonics and erosion, have destroyed or altered most of the very early rock record on Earth, other objects in the solar system, such as lunar rocks, asteroids, and meteorites, have changed little over billions of years. Studying these objects can provide information about Earth's formation and early history.

ESS2.A.1: Earth's systems, being dynamic and interacting, cause feedback effects that can increase or decrease the original changes.

ESS2.A.2 (partial): Evidence from deep probes and seismic waves, reconstructions of historical changes in Earth's surface and its magnetic field, and an understanding of physical and chemical processes lead to a model of Earth with a hot but solid inner core, a liquid outer core, a solid mantle and crust.

ESS2.A.3 (partial): The geological record shows that changes to global and regional climate can be caused by interactions among changes in the sun's energy output or Earth's orbit, tectonic events, ocean circulation, volcanic activity, glaciers, vegetation, and human activities. These changes can occur on a variety of time scales from sudden (e.g., volcanic ash clouds) to intermediate (ice ages) to very long-term tectonic cycles.

ESS2.D.2: Gradual atmospheric changes were due to plants and other organisms that captured carbon dioxide and released oxygen. (Folded into E12.10)

ESS2.D.3: Changes in the atmosphere due to human activity have increased carbon dioxide concentrations and thus affect climate.

ESS3.D.2: Through computer simulations and other studies, important discoveries are still being made about how the ocean, the atmosphere, and the biosphere interact and are modified in response to human activities. (although this may be addressed within E12.14)

Resp 11:

Resp 12: Generally, yes. Some aspects, particularly at Grade 8 seem overly aspirational and may be better suited for Grade 12. Even at Grade 4, given the limited science instruction time historically (see NAEP research by Blank or ACT research), the Grade 4 targets may be overly aspirational. The apparent emphasis on plate tectonics (which cannot be observed directly by students) over observable surface processes, is not appropriate, since more concrete observable ESS processes are more likely to be both relevant to students and students are more likely to have had instruction relating to observable surface interactions and phenomena, especially at Grade 12 (for example, in environmental science). Again, because of the limitations of instructional time in Grades K-4, the continued sliding of some topics from Grade 4 to Grade 8 may cause some Grade 8 targets to slide upward to Grade 12.

Resp 13: Yes. Those included build on the foundation to more complex information and therefore a more complex conceptual understanding. One can see the horizontal articulation along with the vertical articulation to establish a coherent assessment.

Resp 14: Ditto my comments on the above question

Resp 15: Yes

Resp 16: Yes. The 12th grade concept statements are complex and represent detailed content and skills. Only students who choose an educational path with a wide variety of science content courses would be expected to perform well; however, it is this broad base of understanding that we should encourage to achieve the goal of developing scientifically literate citizens.

Resp 17:

Resp 18: Yes, the concepts identified reflect the work that the NRC Framework did to identify the core concepts that are needed by all students at the various grade bands.

Resp 19: Yes, it seems you've followed the NRC Framework and NGSS, which makes sense.

Resp 20: Yes, the concepts identified reflect the work that the NRC Framework did to identify the core concepts that are needed by all students at the various grade bands.

Resp 21: Yes is describes the functions of life and further discusses the chemical reaction

Resp 22: yes; the concepts look appropriate to me based on the DCI progressions

Resp 23: Yes. The content builds from 8th grade and seems to be consistent with grade-appropriate content.

Resp 24:

Resp 25: No comment

Resp 26:

Resp 27: Student in 12th grade should be able to answer questions concerning meiosis and protein synthesis. The knowledge of the periodic table in 12th seems to be lowered as compared to physical science and not a logical vertical progression of concepts.

Resp 28: Although the DCIs on their own make sense, we need to see what CCs and SEPs are being assessed with the DCI to fully answer this question. Student 3-D science learning requires cohesiveness and depth relative to the combination of the DCI with the CC and SEP.

Resp 29: yes. It increases in complexity and concepts are developed for deeper understanding.

17. **What are the science subtopics or concepts that you believe are the highest priority to be measured by NAEP? Why?**

Resp 1:

Resp 2:

Resp 3:

Resp 4:

Resp 5: I would think about emphasizing subtopics or concepts that are a) most likely to have a direct impact on students' current and future lives and b) most clearly intersect with the crosscutting concepts. For example, within the Matter and Its Properties subgrouping, P8.9 (related to flow of energy and matter) might have higher priority than P8.2 (related to properties of substances). Across subgroupings, for example, Energy subgrouping might be more important than the Motion and Forces subgrouping. Or the Earth and Human Activity subgrouping might be more important than Earth's Place in Space.

If such prioritizations are made, it would be important to document those, at least in the Assessment and Item Specifications document.

Resp 6: Overall the concept statements have fairly good alignment with the NRC Framework, and that framework already tried to distill down to the foundational disciplinary core ideas for students in grades K-12. Within the set of concepts presented, we definitely would not recommend any sort of hierarchy to prioritize or deprioritize certain concepts over others (after consideration of climate change and evolution is complete).

Resp 7: Note: This was completed by high school physical science teachers so biology concepts are not present, but are important.

P12.4 Explaining chemistry based on atomic interactions is a key idea that can be used to explain reactinos.

P12.5 Energy in reactions can be used to explain why certain reactions are more favorable and not spontaneous

P12.6 Knowing that matter is conserved can help to draw attention to energy needs and uses

P12.12 Forces and their effects on motion is something that all students can experience

P12.19 It is good to know that energy can not be just appear

P12.28

P12.29 Understanding electromagnetic waves serves as the basis for a lot of the technology that we use to observe our universe and communicate with one another

Resp 8: A subtopic or concept at or within the Physical Science, Life Science, or Earth and Space Science should not be prioritized over another.

Resp 9: I think they are all important. I do not have a priority, and I don't think their should be any priorities. I imagine trying to prioritize practices, concepts, and/or topics would be extremely limiting when developing multidimensional assessment items. I would prefer the focus be on equitable science assessment items.

Resp 10: Those that students will use to make informed decisions while a member of a democratic society, and in an age of disinformation.

Resp 11:

Resp 12: Conservation of and interactions involving the transfer of matter, energy, momentum, and electric charge which are foundational to understanding nearly all other ideas in science which leads to an understanding of observable interactions. These concepts should be measured in the context of and application to topics that have potential impact on everyday life for most people , which would include issues around climate change, resources, biodiversity, human safety (including natural and human-generated hazards), and disease.

Resp 13: energy, ecosystems, evolution, earth systems, earth and human activity

Resp 14: The highest priority should be fostering the habits of mind that foster science: curiosity, empiricism, formulation of questions about nature, how to get nature to answer them objectively. (This is NOT the 'scientific method.) Some attention to history of science would have been appreciated, including reasons why past scientists may have thought about nature very differently from how we think about it now, rather than simply dismissing them as 'wrong' compared to now, which is reflexively regarded as 'right.' Symbiosis, for example, is a large topic left entirely untaught. Similar to the narrow focus on Darwinian selection as the only logical model for evolution. Both those can only be critically taught if the historical context is well understood. Trouble is, of course, that it may not be possible to 'assess' these things when the whole object is assessment.

Resp 15: The actual data presented in real life (or in a fictional situation) should be accurate (or possible). Outside of this document, I have seen warming water supposedly heated by "room temperature" air that that is stated to have reached 30 degrees Celsius. 30 degrees Celsius is not "room temperature" air by any available definition.

Resp 16: Earth and Human Activity is a high priority concept since it has the potential to impact our students' choices related to the use of natural resources and the development of new technologies. Ecosystems are also high priority for similar reasons.

Resp 17:

Resp 18: Assigning priority to certain content standards over others is not something that we would advocate. The NRC Framework has already identified the core concepts that are needed by all students, further delineating them would not be advantageous.

Resp 19: It would be important to prioritize the earth and human activities standards, as they are potentially the most consequential for society.

Resp 20: Assigning priority to certain content standards over others is not something that we would advocate. The NRC Framework has already identified the core concepts that are needed by all students, further delineating them would not be advantageous.

Resp 21: Cause and effect because it describes how one reaction leads to another reaction of the phenomena

Resp 22: all

Resp 23: I am not sure that I have an opinion about this. It is difficult to standardize what is a priority to measure given the volume of possible directions it could go.

Resp 24:

Resp 25: No comment

Resp 26:

Resp 27: Conservation of Matter of Energy and energy flow. System interactions at all levels. The relationship of structure and function....cross cutting concepts.

Resp 28: At the high school level, the physical sciences and life sciences are generally high priorities because of college and career readiness expectations, course content specificity and availability, and STEM career focus areas. At the middle and elementary levels, we spend more time on the Earth and Space concepts in preparation for the high school curriculum.

EAFE and Scale and Function are high priorities along with physical sciences.

Resp 29:

18. **What are the science subtopics or concepts that you believe are the lowest priority to be measured by NAEP? Why?**

Resp 1:

Resp 2:

Resp 3:

Resp 4:

Resp 5:

Resp 6: As noted in a prior response, we believe that if not already done, the writing team needs to review the prevalence and scope of standards for climate change and evolution across the states to determine the appropriate level of inclusion/assessment on NAEP. Validity concerns can arise if the assessment is testing concepts that we know are consistently not being taught to large numbers of students.

Resp 7:

Resp 8: A subtopic or concept at or within the Physical Science, Life Science, or Earth and Space Science should not be prioritized over another.

Resp 9: I think they are all important. I do not have a priority, and I don't think their should be any priorities.

Resp 10: The "ideas" listed above that are above the grade-bands should be omitted. Equity....

Resp 11:

Resp 12: P12.9 – Le Chatelier's principle seems to be an odd topic to include while ignoring things more frequently taught and everyday life ramifications like colligative properties, solutions, or reaction rates that might help people to understand drug effectiveness weakening over time. A lot of the wave material in PS4 seems in the weeds relative to other things that might be more resonant such as how lenses work to improve vision, over P12.30. A causal understanding of electric, magnetic, and gravitational fields is likely beyond the ability level of most Grade 12 students. Specifics about phases of the moon which students frequently struggle with and do not have a real practical application. Knowing how the relative alignment of sun, earth, and moon affect tides is important; however, full, new, waxing, waning, etc. terminology probably is not that important. Similarly, although a general understanding of plate tectonics is required to understand the timing and occurrence of everyday phenomena such as earthquakes and volcanoes, a true causal understanding of plate tectonics, as implied in E12.4, the clarification for E12.5, and E12.7, is not critical to this understanding and likely not a focus of instruction in high school. See specific comments to specific standards in attached document.

Resp 13: matter, earth's place in space

Resp 14: Anthropogenic climate change. It's activism backed up by pseudoscientific claptrap.

Resp 15: I don't see a lowest priority. In every case the concepts should be factual or if there might be an undiscovered variable, I hope it is found and explained.

Resp 16: Earth's Systems and Matter & Its Properties would be lower priority in my opinion. Understanding subatomic particles and plate tectonics may be less important in the average American's daily lives and decision-making processes.

Resp 17:

Resp 18: There are not topics that are a lower priority, but rather some topics that might be disadvantageous to certain student groups, that might need special attention. Concepts that rely on seeing or hearing might generate bias for blind or deaf students.

Must be deliberate about contextualization of concepts so students can access the ideas.

Resp 19: They are all important. Putting one branch of science above another is not productive. The critical piece is scientific literacy, which is what this assessment should be focusing on at its core.

Resp 20: There are not topics that are a lower priority, but rather some topics that might be disadvantageous to certain student groups, that might need special attention. Concepts that rely on seeing or hearing might generate bias for blind or deaf students.

Resp 21: N/A

Resp 22: none

Resp 23: I am not sure that I have an opinion about this. It is difficult to standardize what is a priority to measure given the volume of possible directions it could go.

Resp 24:

Resp 25: No comment

Resp 26:

Resp 27: N/A

Resp 28: I am not sure what criteria are being used to decide the ranking.

Resp 29:

Science and Engineering Practices

19. **Across grades 4, 8, and 12, do the science and engineering practices represent what you believe should be assessed by NAEP? Why or why not?**

Resp 1:

Resp 2: I think they're good, but I might consider a closer alignment with the ITEEA's STEL Practices by including some softer skills such as collaboration, creativity and attention to ethics ... while perhaps harder to measure, these are arguably equally important skills for students to master.

Resp 3: The general scope of the practices seems appropriate, but the number of practices that are listed seemed unnecessary. I believe there were more statements of practices in the NAEP Framework than there are in NGSS. And based on my experience with NGSS, there are some of the elements of the Practices in NGSS that are not used very frequently.

Resp 4:

Resp 5: Yes - they reflect the science and engineering practices in the NRC Framework.

Resp 6: Overall we believe the science and engineering practices represent what should be assessed by NAEP and we support the inclusion of the same 8 practices as presented in the NRC Framework, which by extension are the same practices found in many state standards. However, we do notice that some of the learning expectations/statements are written very broadly, whereas they need to be focused on operationalized analysis/application (e.g., S4.49: Engage in tinkering to improve a design, S4.50: Try new technologies and generate strategies for improving existing ideas, or S8.70: Report investigation results and design considerations accurately and completely.).

Also, we are not sure why there is a general description/introduction provided for each practice, whereas this same information is not provided for the crosscutting concepts. Please either provide it for both dimensions, or provide it for neither.

Resp 7: Yes. They show a clear escalation of skills and outline necessary skills in a scientists toolbox.

Resp 8: No, while the practices are appropriate, there are misalignments in grade-level SEPs and alignment to the NRC framework and supporting documents (Appendix E, Appendix F, Appendix G, and the DCI/SEP/CCC descriptors found in the three foundation boxes associated with the NGSS Performance Expectations (Volume 1, The Standards, Next Generation Science Standards For States, By States). Specifically, these misalignments are outlined below:

Line 362

Asking Questions & Defining Problems

- S4.1 – "develop or refine models" is beyond the 4th grade SEP, as stated in Appendix F – Science and Engineering Practices introduction to the Grades 6-8 column "Asking questions and defining problems in 6–8 builds on K–5 experiences and progresses to specifying relationships between variables, and clarifying arguments and models"

- S4.2 – "compare various suggested models" is beyond the 4th grade SEP, as stated in Appendix F – Science and Engineering Practices introduction to the Grades 6-8 column "Asking questions and defining problems in 6–8 builds on K–5 experiences and progresses to specifying relationships between variables, and clarifying arguments and models"

- S4.4 – "models" should be removed from this statement as it is beyond the 4th grade SEP as stated in Appendix F – Science and Engineering Practices introduction to the Grades 6-8 column "Asking questions and defining problems in 6–8 builds on K–5 experiences and progresses to specifying relationships between variables, and clarifying arguments and models"

- S4.7 – "quantity" should be removed from this statement as it is beyond the 4th grade SEP as stated in Appendix F – Science and Engineering Practices introduction to the Grades 3-5 column criteria should be limited to qualitative relationships, "Asking questions and defining problems in 3–5 builds on K–2 experiences and progresses to specifying qualitative relationships."

- S4.9 – This statement closely aligns with the 3-5 engineering SEP descriptor (e.g. 3-5-ETS1-1). My concern is that the K-4 statements encompass ALL of the K-5 Asking Questions and Defining Problems statements. Although this SEP may be taught in grade 3 and 4, student that have this SEP in 5th grade would be at a disadvantage. Therefore, I suggest moving it to the grade 8 level.

- S4.10 – This statement is beyond the 4th grade SEP as the K-2 and 3-5 statements in Appendix F – Science and Engineering Practices do not include considering impacts on society (individuals and groups) or the environmental impacts. These area are addressed in the Grades 9-12 column in Appendix F – Science and Engineering Practices.

- S8.4 – remove the word "respectful" as the term is subject to varied interpretations.

- S8.7 – Why is the term "system" included here, as that is more closely aligned to the CCCs.

- S8.8 and S8.10 - These statements are beyond the 8th grade SEP as considering impacts on society (individuals and groups) or the environmental impacts are addressed in the Grades 9-12 column in Appendix F – Science and Engineering Practices.

- S12.8 – remove – this statement is redundant to S12.7

Line 375

Developing & Using Models

- S4.14 and S4.15 - General concern that ALL statements articulated in the Grades 3-5 column for Developing & Using Models in Appendix F – Science and Engineering Practices are represented in the 4th grade column. A student that does not demonstrate this SEP until the 5th grade would be progressing appropriately, but at a disadvantage on the assessment. I recommend at a minimum that S4.14 and S4.15 be removed, as they align most closely with the 3-5 engineering SEP descriptor (e.g. 3-5-ETS1-1).

Line 390

Planning & Carrying Out Investigations

- S4.21 – "failure points" is beyond 4th grade, especially since according to the 3-5 statements of Planning & Carrying Out Investigations in Appendix F – Science and Engineering Practices indicate that students "plan and conduct investigations collaboratively"

- S4.22 – "parameters" is vague and beyond 4th grade, especially since according to the 3-5 statements of Planning & Carrying Out Investigations in Appendix F – Science and Engineering Practices indicate that students "plan and conduct investigations collaboratively"

- S4.23 – already addressed in SEP Asking Questions

- S4.24 - This statement is beyond the 4th grade SEP as the K-2 and 3-5 statements in Appendix F – Science and Engineering Practices do not include considering impacts on society (individuals and groups) or the environmental impacts. These area are addressed in the Grades 9-12 column of Planning & Carrying Out Investigations in Appendix F – Science and Engineering Practices.

- S4.25 - This statement is beyond the 4th grade SEP as the K-2 and 3-5 statements in Appendix F – Science and Engineering Practices do not include students determining what tools are needed to gather data and what quantities need to be defined. These area are addressed in the Grades 6-8 column (what tools) and Grade 9-12 (how much data) of Planning & Carrying Out Investigations in Appendix F – Science and Engineering Practices.

- S4.30 – “make predictions” should be limited to own experiences, this statement extends predictions beyond changing variables (Grade 3-5 in Appendix F – Science and Engineering Practices) to predict overall outcomes and to use models (own or class-consensus) to make these predictions. General concern that ALL statements articulated in the Grades 3-5 column for Planning & Carrying Out Investigations in Appendix F – Science and Engineering Practices are represented in the 4th grade column. A student that does not demonstrate this SEP until the 5th grade would be progressing appropriately, but at a disadvantage on the assessment.

- S8.19 - This statement is beyond the 8th grade SEP as the 6-8 statements in Appendix F – Science and Engineering Practices do not include considering impacts on society (individuals and groups) or the environmental impacts. These area are addressed in the Grades 9-12 column of Planning & Carrying Out Investigations in Appendix F – Science and Engineering Practices.

-S12.19 – seems to already fall within the process outlined in S12.16

Line 401

Analyzing and Interpreting Data

-General concern that ALL statements articulated in the Grades 3-5 column for Analyzing and Interpreting Data in Appendix F – Science and Engineering Practices are represented in the 4th grade column. A student that does not demonstrate this SEP until the 5th grade would be progressing appropriately, but at a disadvantage on the assessment. Recommend at a minimum to remove S4.33 and S4.36 since they do not directly correlate to the language in the Grades 3-5 column.

-S8.29, S8.32, and S8.33 – recommend removal of these statements, as the language does not closely align with the statements in the Grades 6-8 column for Analyzing and Interpreting Data in Appendix F – Science and Engineering Practices

Line 409

Using Mathematics and Computational Thinking

S4.43 - This statement is vague and as presented does not align to the SEP of Using Mathematics and Computational Thinking.

S4.44 – This statement does not align SEP of Using Mathematics and Computational Thinking (seems more aligned to SEP Planning and Carrying Out Investigations)

- 4th Grade General Concern – While the statements S4.37, S4.38, S4.39, S4.40, S4.41, and S4.45 align directly with the statements in the Grades K-2 and 3-5 column for Using Mathematics and Computational Thinking in Appendix F – Science and Engineering Practices, there are zero SEP descriptors for Using Mathematics and Computational Thinking associated with K-4 NGSS PEs.

- S8.38 – This statement is very specific to ratios and types of units of measure. Why?

-S8.42 – This statement is more closely aligned with the SEPs Developing & Using Models and Analyzing & Interpreting Data. Recommend removing statement since it is already been covered by those SEPs.

- S12.32 – “programming” should be removed from this statement as it is a specific interpretation of computational tools.

-S12.33 – unclear how student would demonstrate proficiency of this statement on a standardized assessment unless access to digital tools are provided.

-S12.34 and S12.39 are identical statements.

Line 421

Constructing Explanations & Designing Solutions

-S4.50 – “try new technologies” is vague, unclear how student would demonstrate proficiency of this statement on a standardized assessment.

-S4.52 – “model” should be removed as it is already associated with the SEP Developing and Using Models Statements.

-S8.50 – statement should be removed as it is already associated with the SEP Developing and Using Models Statements.

-S8.51 – “critique” does not align with other statements in the Grades 6-8 column of Constructing Explanations & Designing Solutions in Appendix F – Science and Engineering Practices and is beyond the 8th grade expectation for this SEP.

-S12.46 and S12.47 - does not align with other statements in the Grades 9-12 column of Constructing Explanations & Designing Solutions in Appendix F – Science and Engineering Practices

Line 428

Engaging in Argument from Evidence

-S4.60 - does not align with other statements in the Grades K-2 and 3-5 column of Engaging in Argument from Evidence in Appendix F – Science and Engineering Practices and is beyond the 4th grade expectation for this SEP.

-S8.59 – aligns more closely with the Grades 9-12 column of Engaging in Argument from Evidence in Appendix F – Science and Engineering Practices and is beyond the 8th grade expectation for this SEP.

-S12.55 – duplicates portions of statements S12.51 and S12.53 with the addition of “jointly developed and agreed-upon design criteria” which more closely aligns to the SEP Constructing Explanations & Designing Solutions.

Line 446

Obtaining, Evaluating, & Communicating Information

General concern that ALL statements articulated in the Grades 3-5 column for Obtaining, Evaluating, & Communicating Information in Appendix F – Science and Engineering Practices are represented in the 4th grade column. A student that does not demonstrate this SEP until the 5th grade would be progressing appropriately, but at a disadvantage on the assessment. Recommend at a minimum to remove S4.67 since there is not a direct correlation to the language in the Grades 3-5 column, as well as removing the work “comprehend” from S4.63.

-S4.69 – eliminate as the expectation is already captured in S4.68

-S8.62 and S8.70 – eliminate as the expectation is already captured in S8.68

-S8.65 – eliminate as the expectation is already captured in S8.66

-S8.67 – eliminate as the expectation is already captured in S8.64 (assess the credibility)

-S8.69 – eliminate as the expectation is aligned to the SEP Engaging in Argument from Evidence

-S12.61 and S12.62 – eliminate as the expectation is already captured in S12.60

Resp 9: Mostly. There are some inappropriate practices in elementary. Example: I have no idea why tinkering is specifically called out in the NAEP Science Framework for elementary. Elementary students have the capacity to "Generate and compare

multiple solutions to a problem based on how well they meet the criteria and constraints of the design solution." Tinkering is not a clearly defined expectation or science and engineering practice in most states and it is not in the Framework. It's an activity that a lot of students don't have the opportunity or PRIVILEGE to do and should NOT be a practice assessed by NAEP.

I only noticed these inappropriate additions (not related to the Framework or science expectations found in most states) in elementary.

Resp 10:

Resp 11:

Resp 12: Generally, yes. Applying science to solving real world problems as method of looking at engineering and design thinking is appropriate. Also, evaluating those potential solutions as the second portion is also critical. However, it may be difficult to accurately and fairly assess aspects of tradeoffs that involve aesthetics, values, morality, etc. on a large scale science test.

Resp 13: Yes! These practices are essential if learners are to understand how both science and engineering are practiced by scientists and engineers. Understanding these processes will help learners engage in critical thinking, analytical thinking, and argumentation.

Resp 14: All the concepts are good. The question is how they foster creativity and talent. I don't think they do, or at least I see little evidence of attention paid.

Resp 15: Yes. There seems to be a good mix.

Resp 16: Yes. While there is a high degree of redundancy among the grade level practices, that consistency of focus is important. Simply progressing from a simple task to a more complex task that requires the same skills is appropriate for this type of assessment.

Resp 17:

Resp 18: Yes, they align to the recommendations put forth by the NRC Framework.

Resp 19: They match the Framework and NGSS, which will be important in aligning to what’s happening in states.

Resp 20: Yes, they align with the recommendations put forth by the NRC Framework.

Resp 21: Yes cause and effect show the relationship among various reactions

Resp 22: yes, based on the SEP progressions

Resp 23: Yes. The practices seem to represent a major portion of the habits of mind and practices of scientists and engineers that we want students to engage in and they seem to articulate well across grade levels.

Resp 24:

Resp 25: No comment

Resp 26:

Resp 27: Yes but the SEPs that have high level of student engagement and ownership will be difficult to assess unless through constructed response questions.

Resp 28: Yes from the standpoint of including the practice. What is missing is the connection to the DCI.

Resp 29:

20. **What are the science and engineering practices that you believe are the highest priority to be measured by NAEP? Why?**

Resp 1:

Resp 2: I find it very hard to prioritize them as I think they're all essential.

Resp 3: Clearly, all 8 practices should be addressed, but I think Explanation, Modeling, and Argumentation should receive more attention on average than the other 5 because they are most directly involved in making sense of phenomena.

Resp 4:

Resp 5:

Resp 6: All of the practices are important understanding and skills constructs without which students cannot demonstrate the ability to identify and address problems, make sense of phenomena, and evaluate information to make informed decisions (NAEP measurement claim, lines 176-180). However, there may be some differences in the amount of emphasis you place on some of the practices over others based on how much is assessable by NAEP's format (see next response).

Resp 7: Engaging in argument from evidence and Obtaining, Evaluating and Communicating Information: Information is everywhere and having the ability to discern what is useful and true will be a way for them to identify reliable sources.

Asking Questions: Students live in an age where they can type questions in computers, but knowing what questions to ask to further one's knowledge is often difficult.

Analyzing and interpreting data: While collecting data for one's own self is important, people examine the work and data collected far more frequently. This skill can then be used to evaluate claims and sources of information.

While

Resp 8: No one science and engineering practice should be prioritized over another, as they are all critical in sense-making.

Resp 9: I think they are all important. I do not have a priority, and I don't think their should be any priorities. I imagine trying to prioritize practices, concepts, and/or topics would be extremely limiting when developing multidimensional assessment items. I would prefer the focus be on equitable science assessment items.

Resp 10:

Resp 11:

Resp 12: Data interpretation and scientific argument from evidence. These two get at the heart of what is evidence and whether it is interpreted correctly. Scientific argument from evidence using a modified Toulmin framework of claim-evidence-reasoning allows the interweaving of the strands to justify claims which could be models, arguments, hypotheses, etc. The development and use of modeling is the basis for a specific type of argument that is the only way that certain fields in science carry out investigation, e.g. climate science, and is therefore also critical. Understanding a Control of Variables approach is also critical, as well as understanding that in many scientific disciplines it is impossible to use a Control of Variables approach. This will also allow the exploration of correlation vs. causality which is a desperately needed skill for both scientific literacy and college and career readiness. Inherently, there is a lot of overlap between the SEP and the CC which make it extremely difficult to disentangle and rank in importance.

Resp 13: 1, 3, 4, 6, 7, 8 Learners should know, understand, and apply accepted ways new information is developed and applied.

Resp 14: Training engineers to build things that don't fail. A deeper dive into what makes things (systems) fail, and how to avoid the traps.

Resp 15: Again, real data or plausible data should be used. If the results do not fit what students see, there should be eventually an explanation.

Resp 16: I believe Analyzing and Interpreting Data is high priority. These skills willl equip students to recognize legitimate claims when they can properly analyze available data.

Resp 17:

Resp 18: Assigning priority to certain practices over others is not something that we would advocate. If there is a need to group the practices in some way to better generate representation on a form of the NAEP assessment, we recommend using the Instructional Leadership for Science Practices (ILSP) model of grouping them into – Investigating Practices, Sensemaking Practices and Critiquing Practices.

Students should have ample opportunity to demonstrate full facility with an array of SEPs.

Resp 19: We do not see a prioritization being useful here. Again, like content, the key is scientific literacy. Within the list of practices and subskills there, many current assessments leave out students asking questions and students evaluating the validity of information. Those skills should not be minimalized.

Resp 20: Assigning priority to certain practices over others is not something that we would advocate. If there is a need to group the practices in some way to better generate representation on a form of the NAEP assessment, we recommend using the Instructional Leadership for Science Practices (ILSP) model of grouping them into – Investigating Practices, Sensemaking Practices and Critiquing Practices.

Resp 21: I think the ones represented are correct regarding science and engineering practices. They appear to have a specific alignment and order

Resp 22: constructing explanations and engaging in argument from evidence - to promote critical thinking and the use of empirical evidence to support claims

Resp 23: 1. Asking Questions and Defining Problems

2. Developing and Using Models

6. Constructing Explanations and Designing Solutions

7. Engaging in Argument from Evidence

These seem to be some of the most prominent practices that students might encounter in instruction and there could be ways to meaningfully assess them in a testing framework

Resp 24:

Resp 25: No comment

Resp 26:

Resp 27: evaluation, using evidence support or disprove claims/arguments, scientific literacy in reading and interpreting graphs and data

Resp 28: EAFE, CEDS, AID, OEFI, and DUM in order of rank. EAFE is more challenging for students but assessed on SAT as a gatekeeper for student career and college readiness. It is also a SEP that can be used to scaffold other SEPS while students are learning more about the limitations of models, collecting data, etc.

Resp 29:

21. **What are the science and engineering practices that you believe are the lowest priority to be measured by NAEP? Why?**

Resp 1:

Resp 2: See previous response.

Resp 3: I think Using Mathematics and Computational thinking could have less attention than the other 7.

Resp 4:

Resp 5:

Resp 6: Some of the science and engineering practices may be difficult to authentically or deeply measure with the NAEP Science Assessment. For example, in our experience, authoring questions for Practice 1, Asking Questions and Defining Problems, sometimes leads to contrived items and/or items that don't fit well within a task set, especially for machine-scored, selected-response type items. Many of those skills are more suited to being meaningfully applied and assessed in a research project or larger performance tasks.

Likewise, we often observe the same issues for Practice 8, Obtaining, Evaluating, and Communicating Information. On a large-scale standardized assessment, often the questions cannot be very meaningful and/or become reading comprehension rather than a good skill assessment in the context of a research project or performance tasks. For example, S4.68 (Communicate scientific and/or technical information, reasoning, and ideas in written formats and various forms of media as well as tables, diagrams, and charts.) may be much more difficult to appropriately assess as compared to S8.67 (Evaluate the reliability of information based on where it is found, and the qualifications of the source and the evidence given to make the claim.).

We also caution the writing team to be aware and examine (perhaps paired with assessment authoring experts) that the final proposed learning expectations/statements for the practices will be assessable with the item types available for the NAEP Science Assessment. You noted that hands-on tasks are no longer going to be included, and it will be quite important to consider if your capacity for simulations can equally meet assessment needs.

Resp 7:

Resp 8: No one science and engineering practice should be prioritized over another, as they are all critical in sense-making.

Resp 9: I think they are all important. I do not have a priority, and I don't think their should be any priorities. However...

S4.5: Ask "what if" questions about a system or phenomenon being observed, and predict reasonable outcomes based on patterns or expected cause and effect relationships.

S4.49: Engage in tinkering to improve a design. (I take great issue with this one.)

^ Why are these so important that it needs to be an SEP on NAEP? Who decided this? Which states have these (ask "what if" questions, Tinkering) specific practices as a part of their standards & expectations? Why does elementary science have these practices that aren't noted in the Framework or found in most states' standards? Why does NAEP choose to value these? Why is an explicit bullet about plan/conduct investigations missing?

Resp 10:

Resp 11:

Resp 12: Asking questions and defining problems are important skills; however, they are difficult to assess objectively. Because this is specifically an assessment framework, the difficulty in measuring these skills fairly and accurately may lead to lower importance in a blueprint or limit them to constructed response items that requires a justification for how and why the question is worded. Carrying out investigations is limited even with TEI and interactive units as an aspect of the assessment and not part of a curriculum.

Again the focus here is what claims about what students know and can do can be ASSESSED given the constraints, not that these aspects are not important to teach.

Resp 13: 2, 5, Need to have more collaboration and PD for both science and mathematics educators.

Resp 14: I'm not going to prioritize the eight practices. All are good. My question is over the seemingly arbitrary boundaries, clarifications, etc that seem to have the overall effect of dumbing things down for the very bright, and fostering a faux picture of science and engineering for the rest.

Resp 15: I don't see a lowest priority. Observation and maybe recording what is seen will eventually explain irregularities. (Some people have apparently survived serioud cold by slowing their heart -- someone noticed this, and developed a system of chilling a heart so it would stop for delicate surgery, and then, surgery complete, heating the blood and the heart started again. Make the measurements ... you'll find something! (Dr. Rob't White's career is exemplary)

Resp 16: None of the science and engineering practices should be low priority. They all represent important skills that students need to develop.

Resp 17:

Resp 18: Assigning priority to certain practices over others is not something that we would advocate. All the practices are needed for students to engage in science and all of them are interrelated in the process of understanding science.

Resp 19: None.

Resp 20: Assigning priority to certain practices over others is not something that we would advocate. All the practices are needed for students to engage in science and all of them are interrelated in the process of understanding science.

Resp 21: N/A

Resp 22: using computational and mathematical thinking - most difficult, in my opinion and impacted by student's mathematical background

Resp 23: 4. Analyzing and Interpreting Data

5. Using Mathematics and Computational Thinking

I think that standardized tests in science can often over-index on the mathematics sections of assessment and can often mask elegant conceptual understanding because of how the mathematics are presented.

3. Planning and Carrying Out Investigations

8. Obtaining, Evaluating, and Communicating Information

These seem to be difficult to meaningfully assess

Resp 24:

Resp 25: No comment

Resp 26:

Resp 27: N/A

Resp 28: AID and OECI are common baseline SEPs that are used to determine grade-level proficiency across science, math, and ELA. AID and OECI are less rigorous in comparison to the other SEPs. It will be important for NAEP to decide what the frequency of AID and OECI will be on the test in comparison to other SEPs and what that prioritization of a SEP over another or the difference in the number of SEP-specific questions implies.

Resp 29:

22. Do you believe the NAEP Science Framework adequately reflects technology and engineering and the level at which students need to know and understand these concepts today? Why or why not?

Resp 1:

Resp 2: I'm running out of time before the window for public comments expires, however I do think that there are some relevant aspects of technology and engineering that are absent in the current framework, including the history of technology as well as its impacts and influence on society. Perhaps this would be too much to add; however, since Technology Education is typically elective in high school (when students are likely best able to consider ethical and societal implications), maybe it's not unreasonable to consider adding something along these lines.

Resp 3: yes

Resp 4:

Resp 5: I appreciate that NAEP has adopted the 8 practices from the NRC Framework. For assessment development purposes, it might be helpful to more clearly separate out the science practices from the engineering practices. Otherwise, I worry that assessment of each practice could be more heavily weighted to science or engineering, depending on which seems easiest to assess. (This is particularly problematic without specifying the desired weighting of science and engineering practices.)

Resp 6: As related to integration and involvement with science, we generally would answer yes. The NRC Framework has also considered this to some extent, and what is reflected in the NAEP Science Framework seems well matched. There is certainly much more that students need to understand about technology and engineering, but we believe it may extend beyond the scope of the intent here, which is

science assessment. You have likewise acknowledged this tension in lines 100-102 (“Updates should consider whether the definition of student achievement in science needs to incorporate relevant aspects of the 2018 NAEP Technology and Engineering Literacy (TEL) Framework.”). It may be helpful to incorporate this consideration into Chapter 2 as well, and after final determination of the appropriate level of technology and engineering inclusion (including trying to consider where technology and engineering may be in 2028), make a final statement with rationale relative to that determination.

Resp 7: Yes, but as mentioned above, skills related to examining arguments with a critical eye are crucial. Asking questions can lead us to "what's next".

Resp 8: Yes, and it should be noted that upon review it appears that engineering was prioritized by the authors in this draft.

Resp 9: Technology? No, but I didn't get the impression that was the function of the NEAP Framework.

Engineering? Yes in the sense of expectations of the SEPs. I understand and appreciate that the engineering expectations are in the SEPs and not the SEPs and DCIs. I think this is more appropriate and equitable for most students across the US.

Resp 10:

Resp 11:

Resp 12: Generally, yes. However, the degree to which students are actually exposed to using these standards in class may be less than desirable to meet these goals.

Resp 13: It appears so. However, I'd like to examine ITEEA Standards to compare.

Resp 14: Again, no objection to the topics, the worry is that structuring them in seeming lockstep will do little to advance K-12 science and engineering instruction.

Resp 15: Yes. Again with the proviso that curriculum materials are accurate and possible.

Resp 16: Yes. I find the framework's technology and engineering levels adequate for an assessment whose intent is to measure overall science knowledge.

Resp 17:

Resp 18: Yes, the concepts identified reflect the work that the NRC Framework did to identify the technology and engineering concepts that are needed by all students at the various grade bands.

Resp 19: Because technology and engineering ideas are only represented in the practices, and not the DCIs or CCCs, it is more difficult to understand how they will be integrated into assessment items/clusters. Lines 476-484 note that tasks will include “problems” that require engineering practices/thinking to figure out, but more details be said about this integration. Connections to technology, on the other hand, are vague or absent.

Resp 20: Yes, the concepts identified reflect the work that the NRC Framework did to identify the technology and engineering concepts that are needed by all students at the various grade bands.

Resp 21: Yes correlations and relationships are a crucial part of technology and engineering

Resp 22: yes; identifying problems, designing solutions including criteria and constraint considerations

Resp 23: Unsure

Resp 24:

Resp 25: No comment

Resp 26:

Resp 27: It is referenced in the content but is not visible in the sample questions provided.

Resp 28: In the applied sense of a DCI - yes and with respect to the SEP - UMCT. The tie to the TEL framework might seem like a stretch.

Resp 29:

Crosscutting Concepts

23. **Across grades 4, 8, and 12, do the crosscutting concepts represent what you believe should be assessed by NAEP? Why or why not?**

Resp 1:

Resp 2:

Resp 3: Yes

Resp 4:

Resp 5: Yes - they reflect the crosscutting concepts in the NRC Framework.

Resp 6: Overall we believe the crosscutting concepts represent what should be assessed by NAEP and we support the inclusion of the same 7 crosscutting concepts as presented in the NRC Framework, which by extension are the same crosscutting concepts found in many state standards.

However, we do notice that some of the specific learning expectations/statements are written very broadly and feel quite conceptual, which may make them difficult to assess. We have an example of this in our response to the question “What are the crosscutting concepts that you believe are the lowest priority to be measured by NAEP?”

Resp 7: YES! The CCC's show how you can link concepts from one course to another and how ideas flow through multiple scientific disciplines. This is necessary to remove "science in silos" and encourage more interdisciplinary thinking.

Resp 8: No, while the practices are appropriate, there are misalignments in grade-level SEPs and alignment to the NRC framework and supporting documents (Appendix E, Appendix F, Appendix G, and the DCI/SEP/CCC descriptors found in the three foundation boxes associated with the NGSS Performance Expectations (Volume 1, The Standards, Next Generation Science Standards For States, By States). Specifically, these misalignments are outlined below:

Patterns

Grade 8 – "identify patterns in data and provide estimations of chance events" - why does this not include explaining rates of change (5-ESS1-2)?

C8.1 and C8.2 together could read as if the assessment would be limited to identifying patterns for "estimation of chance events" and to "estimate probabilities of events." "Events" is vague. I recommend removing these additional criteria, or revising to "natural phenomenon or designed systems"

C12.2 and C12.3 pull language from Appendix G – Crosscutting Concepts – but do not "match language in the NRC Framework for K-12 Science Education or the CCC descriptors associated with NGSS PEs.

Cause and Effect

None

Scale, Proportion, and Quantity

C4.5 – this should be limited to very short to very long time periods (e.g. 3-LS4-1) since the size reference is associated with 5th grade (e.g. 5-PS1-1)

C4.6 – this is not appropriate for 4th grade since it is associated with 5th grade (e.g. 5-PS1-2 and 5-PS1-3)

C8.8 – where did this come from? C8.11 already captures aspects of this and is aligned to the language in the CCC descriptors associated with NGSS PEs.

C8.10 – This statement is aligned to the Science and Engineering Practice of Analyzing and Interpreting Data, not the Crosscutting Concepts. In addition, algebraic expressions and equations are aligned to the language in the Common Core State Standards for Mathematics and language from Appendix G – Crosscutting Concepts – but do not match language in the NRC Framework for K-12 Science Education or the CCC descriptors associated with NGSS PEs. To be aligned to the SEP language the statement would also need to modified to "linear and nonlinear relationships".

Systems and System Models

C4.7, C4.8, and C4.9 lift language from Appendix G – Crosscutting Concepts – but do not do not match language in the NRC Framework for K-12 Science Education or the CCC descriptors associated with NGSS PEs. Grade 3 and 4 CCC descriptors state "A system can be described in terms of its components and their interactions."

C8.13 – "analyze and explain phenomena occurring in that system" is beyond the 8th grade progression from Appendix G – Crosscutting Concepts and beyond the CCC descriptors associated with Middle School NGSS PEs

C8.14 – "quantities" is beyond the 8th grade progression from Appendix G – Crosscutting Concepts and beyond the CCC descriptors associated with Middle School NGSS PEs

C8.15 – Is the intent of this statement to further clarify what is meant by "essential components" from C8.14? How does "such as representations of forces between objects or relationships between species" playout in assessment writing, as this doesn't read as an assessment boundary.

C8.16 – this statement and the inclusion of model limits is beyond the 8th grade progression from Appendix G – Crosscutting Concepts and beyond the CCC descriptors associated with Middle School NGSS PEs. Model limits are associated with CCC descriptors associated with high school standards (e.g. HS-PS2-2)

C8.17 - This statement is aligned to the Science and Engineering Practice of Developing and Using Models, not the Crosscutting Concepts.

C8.18, C8.19, and C8.20 – these statements do not align to language associated with the CCC descriptors and seem to be aligned with the Engineering, Technology, and the Application of Science DCIs.

C12.18 - This statement is aligned to the Science and Engineering Practice of Developing and Using Models, not the Crosscutting Concepts.

Energy and Matter

C4.10 and C4.11 – These should be limited to only energy, as conservation of matter is beyond 4th grade (ex: PS1.A DCI descriptor for 5-PS1-2).

C8.22 – "transfers of energy are needed to drive any motion or cycling of matter" is beyond 8th grade, see CCC descriptor associated with HS (e.g. HS-LS2-3)

Concept of conservation in closed system is absent.

C12.22 – "matter and/or energy within a system as a limiting factor" is aligned to DCIs (e.g. HS-PS3-1) not CCCs.

C12.23 – aligned to DCIs (e.g. HS-PS3-3, HS-PS3-4) not CCCs.

Structure and Function

C4.13 – aligned to DCIs (e.g. 1-LS1-1) not CCCs.

C4.14 – feels more closely aligned to the Science and Engineering Practice of Developing and Using Models as well as the CCC of Systems and System Models.

Stability and Change

C4.16 and C4.17 – 4th grade CCC should be limited to how slowly or rapidly things may change as aligned to 2nd grade (e.g. 2-ESS1-1, 2-ESS2-1). Considering conditions, develop models, and predictions are not applicable to this CCC for this grade band.

C8.30 - This statement is aligned to the Science and Engineering Practice of Developing and Using Models, not the Crosscutting Concepts. Also feels very similar to HS CCC descriptors (e.g. Hs-PS1-6)

C8.31 – Feedback loops/cycles are associated with HS DCIs (e.g. HS-LS1-3, HS-ESS2-2)

C12.27 - This statement is aligned to the Science and Engineering Practice of Asking Questions, not the Crosscutting Concepts.

Resp 9: Yes, but some seem to be very specific and related to specific science ideas and practices (ex. modeling). I wonder what that means for assessment development and students' ability to make sense of science assessment items.

I don't think it's a bad think. I quite appreciate it. I'm just left with some wonderings.

Resp 10:

Resp 11:

Resp 12: We are unsure that they can be disentangled from the SEP or DCI (energy and matter flows). Likely they should be clustered into four large clusters with SEP perhaps as Data Interpretation, Scientific Inquiry and Investigation, Scientific models and arguments, and engineering and design thinking which would be similar to international frameworks like PISA and TIMSS. Scale and proportion and system and system models are such large overarching ideas that they will be difficult to assess; nearly everything goes into these two buckets, and they overlap with everything so they are too big to adequately assess. However, it is critical that these CCCs are represented on the assessment as integrated with the other dimensions -- for example, understanding varying scales for position and time are foundational for understanding most of Earth Science, and much of the mention of scale has been taken out of or deemphasized in the Earth Science DCI statements.

Resp 13: Yes. They are not specific to a specific subject area. These show connections across content areas and can provide different perspectives for differing learners.

Resp 14: They're all important! Question is how students navigate through them, and a sorting mechanism that will allow the talented and bright to advance over the others.

Resp 15: Yes. Math is Math. Logic is how science works.

Resp 16: Yes. The crosscutting concepts should be assessed as they provide context and allow students to make cross-curricular connections and demonstrate skills that the other dimensions cannot assess.

Resp 17:

Resp 18: Yes, they align to the recommendations put forth by the NRC Framework.

Resp 19: The alignment to the NRC Framework and NGSS makes sense.

Resp 20: Yes, they align with the recommendations put forth by the NRC Framework.

Resp 21: Yes they are very descriptive

Resp 22: yes, based on the CCC progressions

Resp 23: Yes. These crosscutting concepts remain consistent but also build upon each other in grade bands. The 7 concepts seem to represent a good cross-section of what big ideas students need to engage in in scientific and engineering endeavors.

Resp 24:

Resp 25: No comment

Resp 26:

Resp 27: Yes and we like the table that supported the CCC. There is significant overlap in the table with SEP table (which is appropriate and very evident).

Resp 28: Yes. Although again, it would be better to see the DCI with the CC to address this question more specifically.

Resp 29:

24. **What are the crosscutting concepts that you believe are the highest priority to be measured by NAEP? Why?**

Resp 1:

Resp 2:

Resp 3: I think that the CCCs of Mechanisms and explanation: Cause and effect and Systems and system models/systems thinking are the most important CCCs to address

Resp 4:

Resp 5:

Resp 6: We believe all 7 crosscutting concepts are important as themes that pervade science and support students in explanation and sensemaking. We would not identify any particular crosscutting concepts that should have a lower priority, but we do wonder whether you might operationally see that some crosscutting concepts get addressed more often based on more organically pairing with a greater or lesser number of disciplinary concepts and practices. We recommend you take some time to explore this up front so you can be accurate and appropriate with blueprints, achievement level descriptors, etc.

Resp 7:

Resp 8: No one cross-cutting concept should be prioritized over another, as they are all critical in sense-making.

Resp 9: I think they are all important. I do not have a priority, and I don't think their should be any priorities. I imagine trying to prioritize practices, concepts, and/or topics would be extremely limiting when developing multidimensional assessment items. I would prefer the focus be on equitable science assessment items.

Resp 10:

Resp 11:

Resp 12: As stated before, these overlap so much with the SEP that they are difficult to distinguish – how is “patterns” not just an example of a type of data interpretation? How is cause and effect not just a subset of scientific argument? They are all important; however, what sort of Multidimensional IRT or

other complex model will be able to disentangle these so that they can generate a separate score. They should all be represented.

Resp 13: 1, 2, 5, 7 These present learners with opportunities that are easier to become attentive and engaged.

Resp 14: Again, prioritizing some of the seven over others seems a pointless exercise.

Resp 15: Probably math..

Resp 16: I believe all the crosscutting concepts are equally important.

Resp 17:

Resp 18: Assigning priority to certain crosscutting concepts over others is not something that we would advocate. If there is a need to group the crosscutting concepts in some way to better generate representation on a form of the NAEP assessment we recommend using the NRC Framework and making three groups – Patterns, Cause and Effect, and Systems.

Resp 19: A focus on these seven makes sense. What we do not see here is any connection to the nature of science, which typically gets lumped into the crosscutting concepts within the NGSS, even though it's a separate section of the NRC Framework. Because the Nature of Science is such a critical part of scientific literacy, it should not be left out of this exam. For example, students should know what a theory actually is and what the process of science looks like in order to make informed evaluations of media (mis)information.

Resp 20: Assigning priority to certain crosscutting concepts over others is not something that we would advocate. If there is a need to group the crosscutting concepts in some way to better generate representation on a form of the NAEP assessment we recommend using the NRC Framework and making three groups – Patterns, Cause and Effect, and Systems.

Resp 21: They are all represented in the order they should be

Resp 22: patterns; cause and effect; flow of energy and the cycling of matter; structure and function - most accessible for a wide range of students and their science experiences

Resp 23: Unsure

Resp 24:

Resp 25: No comment

Resp 26:

Resp 27: Systems, structure and function, matter and energy, stability and change, patterns

Resp 28: They are all relevant and honestly depend on what is expected to be taught at schools so that all students could have a fair chance at being assessed. So for example, patterns are a given that most science teachers will include and students can demonstrate their skills and learning.

Resp 29:

25. **What are the crosscutting concepts that you believe are the lowest priority to be measured by NAEP? Why?**

Resp 1:

Resp 2:

Resp 3: I think that the CCC of Conservation, Flows, and Cycles: Tracking Energy and Matter is less important than the other 7 to address, mostly because I think that it is extremely difficult to differentiate the content in these CCCs from the Physical Science DCIs involving Matter and Energy

Resp 4:

Resp 5:

Resp 6: As noted in our response to the previous question, we would not set a lower assessment priority for any particular crosscutting concepts, but we do wonder if you could see different emphases emerging based on some crosscutting concepts more organically pairing with a greater or lesser number of disciplinary concepts and practices. We also know that Crosscutting Concept 5, Conservation, Flows, and Cycles: Tracking Energy and Matter, holds significant overlap with disciplinary concepts, so you may need to be careful about the weighting/representation for those constructs.

Also, within the learning expectations/statements for some of the crosscutting concepts, some statements are more foundational recognitions or understandings that may be difficult to assess, e.g., C4.8 (To understand a phenomenon it often helps to develop a model of the system in which it occurs.). How would one ask a question to assess that on NAEP? It is more likely to be an understanding that was used by the assessment developers to build a task, and that task is going to present students a model or ask them to develop the model for a phenomenon. A developer would not write a question to ask, "What could you do to help understand this phenomenon?" and make the answer, "Build a model…." In light of this, as noted with some statements for practices (too broad/non-operationalized) and disciplinary concepts (too definitional/narrow), we recommend a careful review of all statements for their operational assessability specific to the testing structure and format (item types) of NAEP. Perhaps some learning expectations/statements should be flagged as understandings that questions will be built upon but are not able to be directly assessed. That may perhaps be too fine a line, but we offer it as something to consider.

Resp 7:

Resp 8: No one cross-cutting concept should be prioritized over another, as they are all critical in sense-making.

Resp 9: I think they are all important. I do not have a priority, and I don't think their should be any priorities.

Resp 10:

Resp 11:

Resp 12: Structure and function seems to be a subset of models, cause/effect, and systems. Again, this is more an issue of overlap rather than a lack of importance.

Resp 13: 3, 4, 6 These present learners with opportunities that may take longer to develop due to their nature.

Resp 14: Again, prioritizing some of the seven over others seems a pointless exercise.

Resp 15: Again these crosscutting concepts should be presented flawlessly. Perhaps there could be "What is Wrong with this Picture?" digressions to make a temporary detour into the impossible world.

Resp 16: I believe all the crosscutting concepts are equally important.

Resp 17:

Resp 18: Assigning priority to certain crosscutting concepts over others is not something that we would advocate. All the crosscutting concepts are needed for students to engage in science and all of them are interrelated in the process of understanding science.

Resp 19: None.

Resp 20: Assigning priority to certain crosscutting concepts over others is not something that we would advocate. All the crosscutting concepts are needed for students to engage in science and all of them are interrelated in the process of understanding science.

Resp 21: N/A

Resp 22: systems and systems models - addressed somewhat by modeling SEP

Resp 23: Unsure

Resp 24:

Resp 25: No comment

Resp 26:

Resp 27: N/A

Resp 28: I am not sure what criteria are being used to decide the ranking.

Resp 29:

Three-dimensional Assessment Design

26. **Do you agree with the description of the assessment design below? Why or why not?** The NAEP Science assessment should be three-dimensional whenever possible. Each item and each multi-part item should be at least two dimensional and three dimensional if appropriate. Item sets and scenario-based tasks should be three dimensional. No item will be one dimensional.

Resp 1:

Resp 2:

Resp 3: I do not agree with this statement as currently phrased. I think that every task should include a practice and then at least one of the other two dimensions. No item should be one dimensional.

An item that assess ONLY a DCI and a CCC is not appropriate.

Resp 4:

Resp 5: Yes. I think this captures the nature of the NRC Framework, while being realistic about what is possible. I think it is particularly important to specify that "no item will be one dimensional" because this has significant implications for how the scenario-based tasks will be interpreted. (Without the statement, I could imagine a set of items that each assess a single dimension being considered acceptable, when that is not consistent with the NRC Framework.)

However, the description seems incomplete. I think it would be important to also include in this description something about the centrality of phenomena or problems, such as the statement on lines 461-462 ("… all tasks are designed around compelling phenomena and/or problems").

Resp 6: Yes, we are in complete agreement. This is what is required to meet the assessment claims (lines 181-190) and aligns with the guidance on multi-dimensional science assessment practices from the NRC and the CCSSO SAIC (Science Assessment Item Collaborative).

Resp 7:

Resp 8: Yes, this approach aligns to both 3D instruction and the STEM Teaching Tools that guide writing 3D formative assessments.

Resp 9: Yes. For students to engage in sensemaking and have the best opportunity to demonstrate their understanding of science, it's important that their assessment questions aren't flat/are multidimensional. I appreciate tasks will be 3D.

Resp 10: Agree. Although there is an equity issue if students are unfamiliar with 3D assessments.

Resp 11:

Resp 12: Having all multidimensional items may make it difficult to clearly define what students near the lower end of the assessment framework will be able to do, especially at the fourth grade. Items that require students to identify a single data point, identify a controlled or manipulated variable, etc., allow all students to provide evidence as to what knowledge, skills, and abilities they possess. Additionally, having some of these items will strengthen claims about which aspect of a multidimensional task is most likely the KSA that the student struggled with. These should not make up a large percentage of the test; however, if the floor is set too high for assessment tasks, there may be a substantial portion of students that the assessment is not capable of placing along a progression of skills as suggested by the NRC Report on Assessing the Next Generation Science Standards. Admittedly, the substantial overlap between the Crosscutting concepts and science and engineering practices, may minimize this risk; however, this risk should be addressed. There is a risk that if you allow test developers to make one-

dimensional items, you may receive more than the limited number you would like to target. However, a few of these items may help gather important information. No items should be simply a DCI recall.

Resp 13: Yes. We should not encourage students to become "Fact Sheets" or Jeopardy players. Items should mentally engage learners that require them to interact and connect with the assessment.

Resp 14: This is like asking me whether I believe in motherhood and apple pie. All in favor of making science instruction as broad and as multi-dimensional as possible. Can this be done through siloing concepts and collections of factoids and passing it off as science? Doubtful, in my opinion.

Resp 15: Yes. If it's one dimensional has no function and goes nowhere and really has no useful meaning.

Resp 16: Yes. The test format should align with the 3-dimensional nature of our science standards.

Resp 17: In principle, I do agree with this. In practice, however, it is harder to envision. See my comments in the sample item question above. I do appreciate the use of phenomena and scenario based approaches.

Resp 18: Yes, the only way to assess three-dimensional standards is with three- or two-dimensional items that reflect the use of multiple dimensions. Rote memorization is not what these standards are seeking to assess and most memorization items are just one-dimensional.

Resp 19: Yes, this is a reasonable assessment design being followed by states currently.

Resp 20: Yes, the only way to assess three-dimensional standards is with three- or two-dimensional items that reflect the use of multiple dimensions. Rote memorization is not what these standards are seeking to assess and most memorization items are just one-dimensional.

Resp 21: Yes this provides needed details to fully understand the concept.

Resp 22: yes; this avoids rote memorization of either content, skills, and/or explanatory themes

Resp 23: Yes. While it can be challenging to assess three dimensions it is a useful goal that could contribute to helping shift instructional practices and goals. One dimensional questions are generally just around content knowledge or vocabulary and provide very little information around student's understanding of science.

Resp 24:

Resp 25: No. I do not agree with the statement because the description is unnecessarily restrictive and potentially unfair to some students. Why limit dimensions to two or three and not include one? A rationale is needed for this design choice. Items can be designed to measure one dimension, and some students [e.g., younger students (Grade 4) and older students who might understand the science but not be unable to express themselves well] could provide responses that more accurately reflect their science knowledge and/or ability. The description of the assessment design in its current form appears to be unnecessarily narrow and could potentially create biased items.

Resp 26:

Resp 27: Disagree. Sometimes a knowledge/one dimensional question is appropriate depending the complexity of the concept, particularly in 4th grade.

Resp 28: Yes. I agree.

Resp 29:

Balance of Assessment

27. **For Grade 4, do you believe there is an appropriate percentage of items for the various disciplinary concept domains and item types? Why or why not?**

Resp 1:

Resp 2:

Resp 3:

Resp 4:

Resp 5:

Resp 6: For all three grades, we would request that you please distinguish between a total testing administration (by which we mean all forms for a grade across all students) and an individual student's experience (single form). It may be acceptable for the student experience on a single form to have equal percentages of physical science, life science, and Earth and space science content, but cumulatively across all forms, we assume you will be testing in proportion to the final number of disciplinary content statements per domain, and that may end up being different in each grade. Exhibit 1 in Chapter 1 states that "Recommended distributions reflect shifts in expectations evident from reviews of state and national standards, policy documents from leading professional organizations, and expectations for science achievement on U.S. and international assessments." That statement led us to think of distributions in total (all forms), but as we read deeper into Chapter 3, we think you may have meant an individual form. Again, in both places, you need to be clear.

In regard to distribution by item type, we recognize that most NAEP assessments divide testing time evenly between selected response type items and constructed response type items. We do have some concern that the multi-dimensional constructs, with requirements for 2- and 3-dimensional items, may call this into question, however. Selected response type items are not always best suited to measuring particular multi-dimensional constructs and we suspect you may find you need greater proportions of some item types over others to accomplish the most valid measurement possible. These updated needs may require the percentage of selected response to constructed response to be different on the new assessment. As you review/update more sample items and tasks, we urge you to use that opportunity to critically interrogate the item types being used and the distribution that would result when pulled together in a test form.

Resp 7: Yes. Students need to understand how scientific concepts are interrelated and build upon one another. This distribution would encourage people to look at scientific concepts in each core discipline as something that can be used in another context creating a richer understanding.

Resp 8: Yes, we support the equal distribution of disciplinary concept domains (33.3% Physical Science, 33.3% Life Science, 33.3% Earth and Space Science) in each grade level (Grade 4, 8, and 12).

Resp 9: Yes. I think an even mix won't prioritize science topics.

Resp 10:

Resp 11:

Resp 12: Generally, the even balance is appropriate. Student performance in high school may reflect that all students are required to take Biology to graduate and fewer students take ESS classes, but that does not impact how the assessment is designed but might lead to a nuanced interpretation of the Grade 12 results.

Resp 13: Yes. One disciplinary concept domain is not more important or hold more value than the other two domains.

Resp 14: What could be fairer than 33.3%, 33.3% and 33.3%? Where did those numbers come from? Haven't a clue!

Resp 15: Perhaps 50% on the physical science portion (how things work) and maybe 25% on life science portion (how living beings work) and maybe 25% on earth and ecological studies (how the above can work together). I don't think average 4th graders know as much about how things work as they should

Resp 16: Yes. They are equally weighted.

Resp 17: Yes, equality across the three major domains makes sense.

Resp 18: Having each of the disciplines represent approximately 1/3 of the assessment makes perfect sense. This avoids privileging one discipline over others and signals the value of each discipline. Opportunities to assess more than a single traditional discipline ought to be allowable too. The traditional organization of science disciplines is not an accurate representation of either nature or its study.

Resp 19: A balance across disciplines and item types makes sense. What is not clear is the number of engineering/technology related items or nature of science related items. It would be important to ensure that a sufficient number of those items are in place as well.

We wonder about the wisdom of removing the science hands-on-tasks. It may send the wrong message about science instructional practice. Exhibit 1 suggests scenario-based tasks can assess the same content as hands-on tasks; while that may be true for content, it is not true for science and engineering practices or ways of thinking in science (crosscutting concepts). Therefore, with a goal of multi-dimensional tasks, not only content focused tasks, assessment power is lost by completely eliminating these tasks.

Resp 20: Having each of the disciplines represent approximately 1/3 of the assessment makes perfect sense. This avoids privileging one discipline over others and signals the value of each discipline. I would

also encourage creating item sets that integrate more than one domain to reinforce the integrated nature of science.

Resp 21: Yes

Resp 22: yes; all three domains are equally important at all grade levels and the balance of points between SRs and CRs promotes more opportunities for making claims about student science achievement

Resp 23: I am not sure that I have an opinion about this. It seems like a decent balance but I don't have strong opinions one way or another.

Resp 24:

Resp 25: The percentage of items is fine but insufficient given the assessment design. I suggest an additional requirement for inclusion. That is, the distribution of items should also include requirements for the other dimensions to ensure coverage of them as well. My concern is that (without the requirement) only one or two of the scientific practices (argumentation) and/or cross cutting concepts (patterns) will be used in the items design.

Resp 26:

Resp 27: Yes

Resp 28:

Resp 29:

28. **For Grade 8, do you believe there is an appropriate percentage of items for the various disciplinary concept domains and item types? Why or why not?**

Resp 1:

Resp 2:

Resp 3:

Resp 4:

Resp 5:

Resp 6: Please see our response under grade 4, as we have the same response for all grades.

Resp 7: Yes. Students need to understand how scientific concepts are interrelated and build upon one another. This distribution would encourage people to look at scientific concepts in each core discipline as something that can be used in another context creating a richer understanding.

Resp 8: Yes, we support the equal distribution of disciplinary concept domains (33.3% Physical Science, 33.3% Life Science, 33.3% Earth and Space Science) in each grade level (Grade 4, 8, and 12).

Resp 9: Yes. I think an even mix won't prioritize science topics.

Resp 10:

Resp 11:

Resp 12: Generally, the even balance is appropriate. Student performance in high school may reflect that all students are required to take Biology to graduate and fewer students take ESS classes, but that does not impact how the assessment is designed but might lead to a nuanced interpretation of the Grade 12 results. Many students may be exposed to more ESS in the grade band 5-8 than other bands, so this may require a nuanced interpretation of the results.

Resp 13: Yes. One disciplinary concept domain is not more important or hold more value than the other two domains.

Resp 14: What could be fairer than 33.3%, 33.3% and 33.3%? Where did those numbers come from? Haven't a clue!

Resp 15: Maybe 34% on physical science, 33% on life, and 33 on earth and space. Perhaps 1/3rd each.

Resp 16: Yes. They are equally weighted.

Resp 17: Yes, equality across the three major domains makes sense.

Resp 18: Having each of the disciplines represent approximately 1/3 of the assessment makes perfect sense. This avoids privileging one discipline over others and signals the value of each discipline. Opportunities to assess more than a single traditional discipline ought to be allowable too. The traditional organization of science disciplines is not an accurate representation of either nature or its study.

Resp 19: See answer for grade 4.

Resp 20: Having each of the disciplines represent approximately 1/3 of the assessment makes perfect sense. This avoids privileging one discipline over others and signals the value of each discipline. I would also encourage creating item sets that integrate more than one domain to reinforce the integrated nature of science.

Resp 21: Yes

Resp 22: yes; all three domains are equally important at all grade levels and the balance of points between SRs and CRs promotes more opportunities for making claims about student science achievement

Resp 23: I am not sure that I have an opinion about this. It seems like a decent balance but I don't have strong opinions one way or another.

Resp 24:

Resp 25: The percentage of items is fine but insufficient given the assessment design. I suggest an additional requirement for inclusion. That is, the distribution of items should also include requirements for the other dimensions to ensure coverage of them as well. My concern is that (without the

requirement) only one or two of the scientific practices (argumentation) and/or cross cutting concepts (patterns) will be used in the items design.

Resp 26:

Resp 27: Yes

Resp 28: I need to know the frequency of the SEPS and CCs tied to the DCIs to offer a response.

Resp 29:

29. **For Grade 12, do you believe there is an appropriate percentage of items for the various disciplinary concept domains and item types? Why or why not?**

Resp 1:

Resp 2:

Resp 3:

Resp 4:

Resp 5:

Resp 6: Please see our response under grade 4, as we have the same response for all grades.

Resp 7: Yes. Students need to understand how scientific concepts are interrelated and build upon one another. This distribution would encourage people to look at scientific concepts in each core discipline as something that can be used in another context creating a richer understanding.

Resp 8: Yes, we support the equal distribution of disciplinary concept domains (33.3% Physical Science, 33.3% Life Science, 33.3% Earth and Space Science) in each grade level (Grade 4, 8, and 12).

Resp 9: Yes. I do worry about earth and scape science only because I know that it is often devalued in some states on a high school level.

Resp 10: Absolutely! Earth and Space Science (ESS) has been disregarded over the years, and it is extremely important that students demonstrate proficiency in all the core ideas in ESS. Our future depends on how we as a society choose to interact with the Earth system. By levelizing ESS with PS and LS, this is a step in the right direction.

Resp 11:

Resp 12: Generally, the even balance is appropriate. Student performance in high school may reflect that all students are required to take Biology to graduate and fewer students take ESS classes, but that does not impact how the assessment is designed but might lead to a nuanced interpretation of the Grade 12 results.

Resp 13: Yes. One disciplinary concept domain is not more important or hold more value than the other two domains. It represents that all domains have relevance and importance across grade levels.

Resp 14: What could be fairer than 33.3%, 33.3% and 33.3%? Where did those numbers come from? Haven't a clue!

Resp 15: 25% on physical, 25 on life and 50 on earth and ecological studies.

By now they know pretty how things work, and need to fit in with other forms of life, and how they affect/afflict each other.

Resp 16: Yes. They are equally weighted.

Resp 17: Yes, equality across the three major domains makes sense.

Resp 18: Having each of the disciplines represent approximately 1/3 of the assessment makes perfect sense. This avoids privileging one discipline over others and signals the value of each discipline. Opportunities to assess more than a single traditional discipline ought to be allowable too. The traditional organization of science disciplines is not an accurate representation of either nature or its study.

Resp 19: See answer for grade 4.

Resp 20: Having each of the disciplines represent approximately 1/3 of the assessment makes perfect sense. This avoids privileging one discipline over others and signals the value of each discipline. I would also encourage creating item sets that integrate more than one domain to reinforce the integrated nature of science.

Resp 21: Yes

Resp 22: yes; all three domains are equally important at all grade levels and the balance of points between SRs and CRs promotes more opportunities for making claims about student science achievement

Resp 23: I am not sure that I have an opinion about this. It seems like a decent balance but I don't have strong opinions one way or another.

Resp 24:

Resp 25: The percentage of items is fine but insufficient given the assessment design. I suggest an additional requirement for inclusion. That is, the distribution of items should also include requirements for the other dimensions to ensure coverage of them as well. My concern is that (without the requirement) only one or two of the scientific practices (argumentation) and/or cross cutting concepts (patterns) will be used in the items design.

Resp 26:

Resp 27: Yes

Resp 28: I need to know the frequency of the SEPS and CCs tied to the DCIs to offer a response.

Resp 29:

Reporting Results – Contexts for Student Learning

30. **NAEP assessments are administered with contextual questionnaires that are used to interpret student achievement results. What are the science-specific topics or questions that you believe are the highest priority for contextualizing NAEP results? Why?**

Resp 1:

Resp 2:

Resp 3: I think the general sense of opportunities to learn is crucial, which involves many things, including the time devoted to science generally, whether specific topics are covered, what type of instruction students receive, what laboratory equipment students have access to, etc.

Resp 4:

Resp 5:

Resp 6: We would suggest the following as some topics that could be quite important in helping to promote meaningful interpretation of NAEP results:

• Student impression of whether the items are similar to or very different from the types of questions they are asked on tests in their science class – it would be helpful for this to be broken apart to address not only multi-dimensionality but also reading load, cultural responsiveness, etc.

• The nature of science instruction in the student's classroom (e.g., is it focused on facts, what kind of activities/tasks are done, are there driving questions that frame tasks, is what students are learning connected to the real world/real things/real events), from both the student and teacher perspectives

• The frequency/pattern of science instruction, not just in elementary school but middle and high school as well

• Curriculum sequences and course taking patterns in middle and high school

• If/the amount of climate change and evolution taught

• The kinds of activities students engage in outside of science class instruction (e.g., hiking/nature walks, science museums, clubs, camps, reading, art, coding, etc.)

Resp 7:

Resp 8: Administrators (these questions can highlight the types of supports needed to provide high-quality science instruction):

Are teachers provided instructional coaching in science?

Are K-8 students taught science in a self-contained model (one teacher all subjects) or a teaming model (a designated teacher provides science instruction)

PD available to administrator to support science instruction

PD available to administrator to evaluate science instruction

What do you prioritize when creating a master schedule of instruction?

Time spent on science instruction

Resp 9: Questions related to:

- frequency/time to learn science

- Opportunities to engage in phenomena-based science learning

- Opportunities to engage in investigations or engineering design

- Opportunities to engage in multidimensional learning, especially "doing science" the practices

- Questions about science learning out of school

- Feelings of identity, belonging, inclusion, and respectful participation in science

These questions could provide great insight.

Resp 10: Ch 3 - Overview of Assessment Design highlights key aspects of equitable assessments. If the assessment items adhere to the design principles listed, then contextualizing the results should not be an issue.

Resp 11:

Resp 12: Classroom instruction time, especially at the elementary level. Course titles that students have taken at the high school level. Access to hands on lab/field activities. Actual time spent doing and learning engineering and design thinking tasks in class. Access to virtual interactive activities (like PhET or virtual field trips).

Resp 13:

Resp 14: It would be richer if science wasn't siloed, but incorporated into broader systems of knowledge.

Resp 15: I think that truth-in-data should be the highest goal. The proof is in the pudding.

Resp 16: Science-specific topics should include the estimated time (per day) that the student currently spends learning about science. We are finding less time is being allowed for science in grades K-3 especially.

Resp 17:

Resp 18: For elementary, asking how much time is devoted to science instruction.

Another more general topic would be asking questions about three-dimensional instruction and the use of phenomena. Determining whether students are exposed to authentic phenomena and are given the opportunity to use the full potential of the three dimensions would help with score interpretation.

Resp 19: These questions are very important. It would be great to have them reported at the state level. Demographic information and opportunity to learn are particularly critical. It would also be nice to see more data around student identity. The comparative questions (28) appear to be culturally biased, as

being better than others and looking smart does not appeal universally across races/cultures/backgrounds. Questions like 29 seem to conflate classroom level learning desires with interest and grades, which might not be a valid analysis for all students. We would like to more directly know whether students like science and see themselves as science people. That could include learning beyond school, for example.

Resp 20: Asking questions about the instructional practices or instruction to see if students are getting three-dimensional instruction using phenomena.

Resp 21: The ones that are listed are adequate

Resp 22:

Resp 23: I read the paragraphs about this and am unclear about what that might look like. I would need to see some examples or more explantation to more accurately answer this question.

Resp 24:

Resp 25: Contexts should be varied to ensure that student understanding is measured accurately.

Resp 26: Success on achievement-related tasks such as the NAEP assessment is associated with domain-specific prior experiences, the expectancy of success, and subjective task value (e.g, Situated Expectancy-Value Theory; Eccles & Wigfield, 2020). As STEM education becomes more relevant to classroom practice (especially in elementary and middle schools), we should seek to understand the degree to which discipline-specific science, integrated STEM, and interdisciplinary learning experiences are associated with success on NAEP science for each of the grade levels. Using SEVT as a framework to guide this questionnaire would align with other NCES-led efforts (e.g., HSLS:09).

Resp 27: Time allocated to investigation, time allocated for student discourse and sensemaking, student favorite subcontent area of science, student response to importance of science in their lives or futures, student exposure to science outside of class (home, clubs, field trips, camps, etc).

Resp 28:

Resp 29:

31. **The extent of changes to a framework has implications for whether assessment results from the updated framework can be validly compared with results from the previous framework (i.e., continuing trend lines from 2009 to 2028 and beyond). To what extent do you consider it important to prioritize comparisons with previous NAEP science results? To what extent do you consider it important to implement the recommended changes even if they pose a significant risk to maintaining trend lines in NAEP science?**

Resp 1:

Resp 2: I think the value of improving the framework outweighs the value of comparisons to previous data.

Resp 3: I think the need to match the current thinking in science education greatly outweighs the need to be able to compare to earlier NAEP exams.

Resp 4:

Resp 5: The incorporation of the vision of science achievement reflected in the NRC Framework seems to represent a shift at least as large as the one from the 1996 NAEP Framework to the 2009 NAEP Framework. I am really pleased with the extent to which the draft 2028 NAEP Framework reflects the NRC Framework, but I worry that the implementation of the NAEP Framework could be diluted if the assessment is forced to maintain trend.

Resp 6: The recommended changes in the NAEP Science Framework, which require students to use multiple dimensions of science to identify and address problems, make sense of phenomena, and evaluate information to make informed decisions, are vital to reflect the needs and expectations for K-12 science education going forward, as clearly demonstrated by research, the NRC Framework, and the resulting changes to state science standards in the past decade. To not implement these changes for the sake of maintaining trend lines would only undermine the intention of NAEP (to assess what students know and are able to do) and foster misinterpretation and misunderstanding of the abilities of our K-12 students relative to science. Those understandings are seen by many as important in thinking about our nation's future in areas such as job/workforce, global competitiveness, environmental sustainability, etc. We understand that some stakeholder groups significantly value comparisons over time. We seriously question, however, given the magnitude in the shift of constructs in the new NAEP Science Framework, what utility such comparisons could possibly serve in terms of meaningful or accurate conclusions for the trend of student progress in science. Whether achievement on NAEP 2028 is high, low, or medium, it is going to be based on a very different set of constructs and application of knowledge than what was previously measured on the NAEP Science Assessment. The current and future assessments are not comparable in construct, a fact re-emphasized by the possibility of renaming the three disciplinary groupings for reporting (lines 881-885). The "story" that NAEP tells needs to be clear and, as emphasized in Chapter 4, not mislead users. States adopting new standards based on the NRC Framework have had to do standard setting for their new assessments, breaking the assessment's connection to their previous scale and results, as the only option for valid score interpretation and use.

Resp 7:

Resp 8: Brian Input - While the maintenance of NAEP trend is important, the NAEP science framework needs to be updated to allow a science assessment more consistent with the NRC framework. NAEP is commonly referred to as the gold standard of large-scale assessments, but if what the NAEP science assessment is measuring different from what is prioritized and measured by states, it is much more likely be misaligned with what is being taught within science classrooms across the country. A misalignment between NAEP assessment and state standards informing teacher instruction could lead to apples vs oranges results comparisons that could erode NAEP credibility and reduce the utility of the NAEP science assessment. It is more important that the NAEP Science Assessment Framework be updated to reflect research-based practices around science learning.

Resp 9: I don't care if there is a risk to maintaining trend lines. I value assessing students on 2D/3D science items.

Resp 10: Thank you for this question. The results from the new assessment if designed as suggested, can not be used in trend analysis. However, given the age of NGSS, we *should* be able to administer 3D assessment tasks.

Resp 11:

Resp 12: It is more important to align the assessment to what most states are teaching (which is currently over 90% based on NGSS/NRC Framework). The previous assessment was more similar to the NSES with science practices as a separate but equal portion rather than integrated in the current framework; however, the items clearly integrated the practices (with different names but similar skills) with content knowledge. Since recall questions were not allowed, there should be a way to compare the scales. The previous framework also had three strands with Life, Physical, Earth and the combined grid framework which is fairly comparable to the new framework.

Resp 13: Since there are many variables (different students, curriculum, instruction, assessment, cultural factors, economic factors, social factors, etc.) that contribute to learning, I question what information would be valuable. There are too many variables, some the same, some new, and some that evolved. What I see in this draft is the direction we should be going. However, it has implications for teacher professional development, along with changes in curriculum development and instruction at the local and state levels. The teachers must be supported.

Resp 14: I think it's more important to develop incipient scientists than having faux quantitative metrics that will let education bureaucrats congratulate themselves. Our current state of science education, which has been under the purview of bureaucrats for decades, is abysmal. This is unlikely to improve it.

Resp 15: If changes are made in the framework then comparisons to previous scores should be discouraged

Resp 16: I think it is less important to compare the new results to the previous assessment.

Resp 17: This framework represents a substantive shift in NAEP which is necessary and well timed. I think it is important to implement changes even if trend lines are broken. Trying to maintain trend lines could lead to a situation where the shifts are not implemented as envisioned, leading to mis-interpretation of the intent of the revised framework.

Resp 18: It feels more important to measure what is important in student understanding as opposed to avoiding a discontinuity especially if avoidance results in status quo. Though NAEP is not designed or intended to influence instruction, the fact is that it will. I do not advocate for NAEP to signal values in education that are outdated and harmful to many young people.

Resp 19: Trying to match this new framework to the previous framework would not create a valid trendline. One of the most valuable purposes of the NAEP Science is to see trendlines in scientific literacy. While the new framework will better align to current science learning research, it is less clear that it will provide better information about scientific literacy. Amongst these competing priorities, we have not seen a case for a clear winner.

Resp 20: The previous results seem to not be moving in any particular data, so they do not seem to be that relevant, especially when science is not assessed on a more regular basis. Would prefer to use the new framework and assess more often to create new data for comparison.

Resp 21: The comparison of data allows the viewers to determine trends regarding assessements

Resp 22: The 2028 NAEP is built on a very different - and much improved- framework so I would contend that comparing future to past results should not be prioritized

Resp 23: I am more interested in creating a test that can better represent students' abilities to engage in all three dimensions than in comparing to past NAEP data. I would lean more towards implementing the recommended changes.

Resp 24:

Resp 25: I believe it is quite unimportant to implement the recommended changes at this time because of the COVID pandemic. The ability to see NEAP results from past years and compare them to current/future ones should be imperative for IES. The effects of the COVID pandemic could be shown on the NEAP results so it is crucial for continuity among administrations to be maintained.

Resp 26: So many investments have been made to transform science education the US. We should prioritize aligning assessments with the NRC framework. This comes with a communications problem re: longitudinal interpretation.

Resp 27: Due to covid is important to maintain the trend line. The addition of the clarification statements does not warrant such a rescale.

Resp 28: The change is needed especially with the 3D science learning practices being in place for a decade.

Resp 29:

Additional Comments

Resp 1:

Resp 2: I apologize for leaving so many questions blank, but I ran out of time and answered only to those which I felt the most qualified to respond to. Thank you for the opportunity to comment.

Resp 3: I am happy to provide additional feedback or to clarify my comments if the Committee would find it useful

Resp 4:

Resp 5:

Resp 6: Equity: Lines 630-680

The framework writers have done an admirable job with this section. However, the section lacks specifics about student groups, and in lieu, perhaps a bullet with the following language can be added under line 655: Prior to the start of each development cycle, content developers and policymakers together analyze the latest available student demographic data and use this analysis to determine the communities to be represented on the assessment and the stakeholders to be involved in the planning, design, and development process.

On line 679, we recommend you strike “(or limit).” Let us commit fully to diversity, inclusion, and fairness.

Should more be called out specific to women/girls and minority groups, to honor the continued push to engage these groups more in science?

Sample Items: Lines 730-846

Sample items need to be annotated. Annotations are needed to help the reader understand how items meet the specifications and intent of the assessment. Lines 734-737 note “While these items are not all necessarily exemplary of all aspects of multidimensional items, they have been adapted to highlight specific important features and are intended to be a helpful guide to operationalizing this framework.” Without annotations, there is the risk for misunderstanding and erroneous assumptions about the quality or lack of quality of the assessment. For example, we identified some alignments we’d question (e.g., Line 743, Example 1, we would align to a different practice), but if the items were annotated with reasoning for alignments, context, item type, style, etc., there would be an explanation to understand the intention.

We’d also like to see the correct response and an explanation/analysis of the response provided for all items.

All items need to be reviewed for content accuracy and best assessment writing practices. For example on the graph under Line 781 for Example 3, the y-axis of the graph is not labeled, and per the key it represents two different quantities (number and percent). Two different quantities cannot share the same axis; the graph would need two axes labels and scales, on opposite sides, but that is not appropriate for grade 8. Also from an accessibility perspective there are too many lines and the colors are not appropriate for students who are color blind or have other vision impairments. The large number of lines also creates problems for brailling and alternative text. Despite the disclaimer at the start of the section, for credibility in the assessment and results, you must have solid, polished exemplars.

We also request more variety in the samples. The same “base” items are used to show a single item as well as different multi-part or task items. Recycling that same small set (with a static stimulus and fairly traditional item types) doesn’t highlight the variety of item types possible and the amount of innovation that could be present (lines 535-541). You should also vary the samples to show examples of the cultural responsiveness and relevance you outline. What does it look like to reflect what the text says? While the ecosystem disruption off a coastal city in these items is real and is a problem that particular communities care about, it’s an oft-used example. Bring more diversity to the samples to model what you will implement.

Scenario-Based Tasks: Lines 681-700

With the elimination of HOTs, we would like a little more information about how much/how often interactivity will center scenario-based tasks. Some science and engineering practices like Practice 3, Planning and Carrying Out Investigations, beg for simulations. Calling out the similarities/differences to the existing ICTs could help. We’d also encourage increased use of video and/or animation to support more authentic assessment, though care must be taken to include the appropriate considerations for accommodations and accessibility.

Resp 7:

Resp 8: Line 78 – "a national wake-up call" is a deficit lens.

Line 275 – "and elsewhere" is vague. When the NRC Framework for K-12 Science Education is explicitly referenced throughout, this feels like an unknown and could imply an unreferenced resource to which there is associated accountability.

BL Input – NCES and NAGB should take steps to ensure that items developed from the updated framework will be accessible for students on the lower-end of the score distribution and provide an accurate indication of those students skills, particularly in light of significant score drops for students scoring at the 25th and 10th percentiles from 2015 to 2019 not only for science, but also in all science content areas (physical science, life science, and earth and space science).

Resp 9: I appreciate the information included in the NAEP Framework.

Resp 10:

Resp 11: The Education Trust is an education advocacy organization committed to advancing policies and practices to dismantle the racial and economic barriers embedded in the American education system. A focus of that work has been in researching inequities(i) in the enrollment and persistence of Black and Latino students in advanced STEM courses.(ii) We identified how enrollment inequities are based on systemic biases and barriers in access, including the effect of assessments that too narrowly evaluate a student's knowledge and preparedness for advanced study. It is because of this work and our mission that we are grateful for the efforts of the National Assessment Governing Board to update the 2028 NAEP Science Assessment Framework to account for the diversity of learner "cultural and linguistic backgrounds, identities, and learning environments" (page 56, line 631).

Our work has confirmed that Black and Latino students and students from low-income backgrounds often aspire to go to college and are interested in STEM courses, but that they are often pushed out of the STEM pipeline by inequities in course availability, educator biases, and racialized tracking that begins as early as elementary school. We are optimistic about the clear intentions to reduce assessment bias by using accessible language, by ensuring items reflect the range of assets and funds of knowledge that students bring, and by grounding scenarios in real phenomena.

We believe that changes to language in the subsection titled, "Reflecting a Wide Range of Learners" will better reflect the intentions we outlined above. We suggest changing the following terms in lines 649 to 660:

Lines 649-650: CHANGE "students who do not have an average or majority" TO "student"

Lines 651-652: REMOVE "rather than trying to represent an average"

Line 658: CHANGE "culturally novel" TO "distinct from their experiences"

Line 660: CHANGE "average" TO "White normative" OR "normative"

We appreciate the intention to avoid scenarios and contexts that might reflect an averaged or colorblind experience, as stated in the first sentence of the excerpt above. However, students who may be stereotyped or excluded are not missing an "average or majority" experience. Our caution against this

framing is to avoid a frame that positions these students as deficient. Additionally, the "average or majority" experience that that has long been the basis for bias in standardized assessments could more accurately be described as the "White normative" experience. The "White normative" experience is neither reflective of an average nor a majority, particularly as student demographics continue to shift. It is clear that "culturally novel" refers to the novelty each student should find in a range of problems when they are reflective of authentic diversity. However, all other cultures become implicitly "novel" when discussed as one part of a broader effort to deviate from a standard of the White, normative experience.

Finally, we applaud the intentions of the Governing Board "to use surveys, focus groups, voices of community members in ways that ensure authentic relevance and diversity in contexts" and the specific naming of these approaches. We hope the Governing Board has considered how these approaches must intentionally and continually solicit difference to avoid a superficial capture of voice. We also hope the Governing Board is intentional in ensuring that the participants of these approaches reflect the diversity of student identities, including (but not limited to) racial and ethnic identity, economic background, sexual orientation, gender identity, family structure, housing status, and relationship to carceral systems.

(i) Patrick, K., Socol, A. R., and Morgan I. (2020, January 9). Inequities in advanced coursework. The Education Trust.

(ii) Patrick, K., Davis, J., and Socol, A. R. (2022, April 21). Why are Black and Latino students shut out of AP STEM courses?

Resp 12: Page 9, Lines 214/215: the reference to the NRC framework document needs to be italicized (see how it was done on page 5, lines 158/159.P4.8 and P8.15: It seemed like these 2 were related, however 4.8 talks about zero sum forces that "can" cause changes in motion, whereas in 8.15, it says that zero sum forces "will" cause changes in motion.The inconsistency bothered me. P4.16: the use of "beach" seems to come out of nowhere. In the original NGSS documentation, what is used is "a beach" rather than "the beach", which is slightly better.P4.19 (and optics in general) the terms that might be tested like transparent, translucent, and opaque should either be in the standard or in the clarification statement if students are expected to know them.P8.1 – over 100 different types of atoms is kind of disingenuous in that very few of the man-made ones other Americium and Plutonium have long enough half-lives to be useful or combine.If you are talking atoms then you are including isotopes and it should be "well over".However, this appears to imply elements and likely should use this term.

P8.3 & P8.5 are duplicates.Eliminate one.Page 11, There seemed to be a level of inconsistency here when going from gases (8.6) to liquids (8.7) to solids (8.8) with respect to the use of atoms, molecules, or both, and how the gas concept gives an example, but the liquid and solid concept statements do not. I would recommend using "atoms or molecules" in 8.7 and 8.8 (like in 8.6) and consider adding examples for clarity.P8.17 would prefer "extremely large mass" to "large mass". Compared to a person an elephant has a large mass but the gravitational force between a student and an elephant is negligible.P8.18 Seems aspirational.It was aspirational when included in NGSS .A causal relationship was level 3 of the LP for high school students and not likely to be mastered by the middle of 8th grade. Kaldaras et al(2021). Fields in general as a causal explanation of the phenomena are likely better placed in the grade 12 band. P8.16, on the other hand, explaining the relationships between the phenomena that are observable at the macroscopic level seems appropriate.P8.28 In addition to sound waves, mention of other waves such as seismic waves or water waves should be included in the statement, clarification, or boundary for clarity of what will be tested.P8.30 and P12.27 Digitized is not clear. Do students need to understand the

connection between digitized and wave pulses.P12.4 More explicit mention of crystalline, other “extended materials”, substances that are not molecular (see P12.4). Other concept statements refer only to “molecules”.Page 12, P12.8: “atoms and molecules” is used early on in the concept statement, but then at the end only “atoms” are used when talking about the effect of temperature changes on chemical reaction rates.I would recommend saying “atoms and molecules” at that second instance to match the earlier instance.

Page 12, P12.9: This concept statement talks about chemical equilibrium but does not use the term “equilibrium”. Is there a reason this term is not used? It could be introduced as a term because the definition is given. This parallels the Le Chatelier’s Principle NGSS standards that is also more aspirational than other standards and rarely, if ever, covered in high school chemistry classes below the AP level. Would suggest removing.I would argue that leaving out circuit electricity was one of the great errors of the NRC Framework. In the US, even students who have never lived in a detached house will draw one for the House-person-tree personality test. As a matter of scientific literacy and public safety,understanding simple wiring in a house circuit would seem to be at least as important, if not more important, than an understanding of electric and fields as the causal explanation of atomic phenomena. The cost from electrical fires and electrocutions in the home is billion dollars per year (out of characters, can supply mor

Resp 13: I commend everyone who is involved in this much needed activity. With the pandemic, I wished it would be an opportunity for educators and other stakeholders to really examine the education system in the US. I think the efforts by NAEP personnel is a great example of such.

Resp 14: I was not happy with the prompts or the structure of this means of commenting. The questions were shallow, meaningless, and structured to validate the exercise than to foster genuine critical input.

Resp 15: My daughter's 7th grade text had a graph of water changing phase over time ... without a heat-of-fusion plateau. The graph changed as ownership of the publisher changed. Wrong in '93, '94, '95, '97, heat of fusion plateau added for '99 ,2001 in partial fixes. bogus values 2011. 2012 - two state's editions, 2013. 2019 - 3 different editions, 2020 and 2021. I have ISBNs and books both! It's probably easier for a teacher to remedy a missing concept than to teach a given concept "against the book".

Resp 16: Lines 209-212: Include goal of scientifically literate citizens (not limited to those who choose science careers)

NAEP Physical Science-Matter & Its Properties: P8.3 is identical to P8.5

NAEP Life Science-Ecosystems: L4.7 and L4.8 are identical

NAEP Life Science-Ecosystems: Remove unnecessary wording (i.e. L8.14 "Similarly"; L12.15 "Moreover"

Resp 17:

Resp 18: As NAEP includes students with disabilities as a sub-group, it is important to consider how this assessment will be made accessible to those populations. This goes beyond just available tools but also to the design of the overall platform and items. The current examples would not pass some of the identified accessibility standards that are available. Another is considering how to assess DCI like P4.18 which disadvantages students with visual impairments.

Resp 19: • In reference to lines 882-883, it appears that the reporting categories will be "Sensemaking in _____ Science." In order to signal a focus on the full three dimensions of the NRC Framework for K-12 Science (and the NGSS and related standards), the categories need to be expanded beyond those current statements. From experience in Wisconsin schools, we have seen that reporting that is generic in relation to practices and crosscutting concepts tends to focus attention primarily on the area where it is not generic, in this case the content. We would strongly encourage using several practice-based categories at least, potentially in conjunction with broader sensemaking categories, which in our experience would be much more likely to move the field forward in their discussions about science learning.

• Cognitive complexity is not addressed in this document. We assume that the noted "Assessment and Item Specifications" document will address that alignment framework. In Wisconsin, after deliberating among different approaches, we landed on the revised DOK for science, largely to signal that we were not going to be using DOK 1 items, and wanted a mix of DOK 2 and 3 items - https://www.webbalign.org/dok-definitions-for-science. Other frameworks provide a more nuanced view of alignment to these three-dimensional standards and framework and might be considered as well—for example, this reference from Achieve: https://www.achieve.org/cognitive-complexity-science.

• Notably, this test design does not appear to be innovative in any significant way, unlike other NAEP work such as the Technology and Engineering Literacy Test. It would be great to have NAEP using some new item types or showcase learning structures to lead the way in what equitably assessing student understanding could look like. The framework, assessment overview, and samples provided in chapter 3 look like what many states are currently doing. NAEP has the potential to provide new strategies and ideas to states and districts on how three-dimensional science can be assessed, but this framework suggests that will not happen. We will largely receive similar information to our existing state test, and that is a missed opportunity.

Resp 20:

Resp 21: I think the document is well written and list all the appropriate data to assess students achievement.

Resp 22:

Resp 23: I think that this is a good direction and hope that the work and thinking continues in this way.

Resp 24: Please consider revising the use of "best" (such as in S8.40) to other descriptors such as "more appropriate" or "yield more accurate results."

Resp 25: @ Lines 524-526, "Importantly, disciplinary concepts, crosscutting concepts, and science and engineering practices do not receive separate scores." I believe additional rationale is needed for this statement to justify the potential loss of information that this can cause. Developmentally, a student's learning trajectories in the content area, practice, and crosscutting concept may evolve at different rates and in different ways. By requiring all items to measure 2 dimensions or more, interpretation of results could be confounded and obscure a student's true understanding and/or facility.

Resp 26:

Resp 27: No. Without student response it is difficult to determine the effectiveness of the prompt.

Resp 28: The data graphs are colored. This makes it difficult for students who are color-blind to determine the trend in the data. Accessibility is an issue. The use of different shapes for the points or line patterns would support different student needs.

Resp 29:

Respondents Table

Resp	First	Last	Stakeholder Group	State	Title	Organization
Resp1	Brian	Gong	Assessment specialist	NH	Senior Associate	Center for Assessment
Resp2	Michele	Dischino	Researcher/professor	CT	Professor, Technology and Engineering Education	Central Connecticut State University
Resp3	Ted	Willard	Other	MD	Senior Subject Matter Expert in Science	Discovery Education
Resp4	Lauren	Brodsky	Other	CA	Assessment Lead	Learning Design Group, Lawrence Hall of Science, UC Berkeley
Resp5	Alicia	Alonzo	Researcher/professor	MI	Associate Professor	Michigan State University
Resp6	Karen	Whisler	Assessment specialist	NH	Director Content Development	Cognia
Resp7	Matthew	Richard	K-12 teacher	KS	Teacher	Olathe School District USD 233
Resp8	Ann-Marie (those providing feedback are Leanne Weber and Brian Lloyd)	Mapes	State administrator	MI	State Assistant Administrator, Division of Educator, Student, and School Supports	Michigan Department of Education
Resp9	K. Renae	Pullen	School/district staff	LA	Science Specialist	Caddo Parish Public Schools
Resp10			Other	NJ	Science Standards Specialist	
Resp11	William	Rodick	Researcher/professor	DC	Spencer Scholar	The Education Trust

Resp12	Jay	Thomas	Assessment specialist	IA	Principal Content Specialist, Science	ACT, Inc.
Resp13	Peter	Mecca	Researcher/professor	VA	Adjunct Professor of Biology	University of Maryland Global Campus
Resp14	Scott	Turner	Researcher/professor	NY	Director, Intrusion of Diversity in the Sciences	National Association of Scholars
Resp15	Howard	Lyon	Parent	PA		
Resp16	Mary	Headrick	Other	AL	Middle School Science Specialist	AMSTI-UAH
Resp17	Jake	Foster	Other	MA	Founder	STEM Learning Design
Resp18	Tamara	Lewis	Assessment specialist	MD	NAEP State Coordinator	Maryland State Department of Education
Resp19	Kevin	Anderson	State administrator	WI	Science Education Consultant	Wisconsin Department of Public Instruction
Resp20	Jeremy	Haack	State administrator	MD	Science Educational Specialist	Maryland State Department of Education
Resp21	Barbara	Dunham	K-12 teacher	AL	Educator	Birmingham City Schools
Resp22			School/district staff	MD	Supervisor, Secondary Science	
Resp23	Spencer	Martin	School/district staff	KS	Science Curriculum Instructional Coach	Kansas City, KS Public Schools
Resp24			Researcher/professor	MO		
Resp25	Linda	Morell	Assessment specialist	CA		
Resp26	Tony	Perry	Researcher/professor	MA	Postdoctoral Associate	MIT

Resp27	Anne	Petersen	State administrator	VA	Science Coordinator/ Instructional Specialist II	Virginia Department of Education
Resp28	Aracelis Janelle	Scharon	K-12 teacher	IL	Science Teacher 9-12th Grade	Bloom High School
Resp29			School/district staff	AZ	Teacher Coach	Sunnyside Unified School District