2028 NAEP Science Assessment Framework Update: Revised Draft Following Public Comment

July, 2023

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 took the lead on filling in some of the background and rationale for the assessment framework recommendations, which was 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 have a chance to discuss or fully address in the initial working draft.

Revised Framework Draft Following Public Comment and May Board Meeting Discussion

Shortly before the May quarterly Board meeting, the Development Panel met virtually on May 2 to discuss initial plans for addressing the feedback received from NCES and the public. During the May Board meeting, the Panel Leadership Team presented an update to the full Board on initial plans to revise the draft framework following public comment, including: 1) prioritization rules for cutting down the volume of content to be assessed (due to feasibility and cost concerns); 2) an initial draft complexity framework to illustrate how assessment items can be developed to measure students across a wide range of performance; and 3) additional guidance for creating discrete and multi-part items that still reflect the multidimensional nature of the intended construct. The Board discussed the need to balance feasibility and cost concerns with validity and quality, identifying the point at which the intended construct can be adequately measured at a reasonable cost. The Board also acknowledged the likely need to start new trendlines to implement the proposed framework, given the substantive changes to the construct of science achievement and the small number of existing items that can likely be carried forward to a new assessment.

ADC put forward the following proposed policy guidance to provide to the panel, which received consensus from the full Board during the May plenary discussion:

- **Significantly cut down on the volume of content to reduce the number of items and students needed without jeopardizing the measurement of the intended construct**
- **Provide detailed guidance, examples, and evidence that the framework can support items at the lower end of the scale**
- **It is not necessary to assume that the majority of existing items can be carried forward, even if this ends up leading to a break in trend**

The Development Panel continued to work on framework revisions in small groups and met in-person in Washington, DC, on June 5-6 to engage in large group discussions. ADC Vice Chair Christine Cunningham attended the June panel meeting to provide and clarify the Board policy guidance. The Steering Panel met virtually on July 10 to discuss and provide input on a revised draft of the framework in advance of it being included in these Board materials.

The revised framework draft (attached) reflects several important changes and additions from the public comment version that was released on March 13, including:

- The narrative text has been expanded and made more consistent and coherent, including edits for clarity (General)
- The description of the Board's commitment to equity has been revised following discussions at the March quarterly Board meeting, and review/revision by the Assessment Development Committee and the Executive Committee (Executive Summary/Introduction)

- The number of disciplinary concept statements has been reduced from 199 to 128, with a more even distribution of statements across the three domains (for comparison purposes, the 2009-2024 NAEP Science Framework contains 125 content statements) (Chapter 2)
- The eight science and engineering practices have been grouped into four pairs of practices with the requirement that at least 10 percent of items at each grade level represent each of the four groupings (Chapter 2)
- The text describing how the practices and crosscutting concepts can be implemented has clarified that there is no expectation for most of the sub-statements under each practice and crosscutting concept to be covered by the assessment (Chapter 2)
- The framework has clarified that two-dimensional items should include a disciplinary concept and a practice, and that crosscutting concepts will represent the third dimension when feasible. Guidance has been added indicating which crosscutting concepts pair most naturally with each of the four groupings of practices. (Chapter 2)
- Edits and clarifications have been made to the disciplinary concept statements, science and engineering practices, and crosscutting concepts (Chapter 2)
- Guidance about the mathematics required has been added to the disciplinary concept statements and science and engineering practices, where applicable (Note: additional math guidance will be included in the accompanying *Assessment and Item Specifications*) (Chapter 2)
- A draft complexity framework has been added to describe how the assessment can support items across the full range of student performance (Chapter 3)
- Guidance for assessment design has been expanded, including item types and formats; digital tools; and phenomena and contexts (Chapter 3)
- A wider variety of sample items has been added, along with annotations describing how key requirements of the framework are reflected, to illustrate a variety of item types, different complexity levels, and other prominent features of assessment design (Chapter 3 and Appendix B)
- Guidance for subject-specific contextual variables has been added to the description of reporting (Chapter 4)
- Draft achievement level descriptions (ALDs) have been developed (Appendix A)

An iterative review and revision process is currently underway, including to address comments from NCES. The purpose of the ongoing reviews is to ensure that the final version of the framework recommended to the Board for action in November has successfully addressed any outstanding concerns about implementation.

During the upcoming discussion at the August Board meeting, the Panel Leadership Team will briefly describe the key elements of the framework with an emphasis on revisions made following public comment. Patrick Kelly will moderate Board member questions and discussion, including determining whether any additional policy guidance from the Board should be provided to the Development Panel as they further revise the document for planned Board action in November.

Next Steps

Shortly after the August Board meeting, the Development Panel will meet virtually on August 11 to discuss any outstanding revisions to the framework (including policy guidance from the Board, if applicable) and the accompanying *Assessment and Item Specifications* document which provides additional guidance to NCES on how to operationalize the framework. For example, there is ongoing work in the following areas:

- Determining whether some of the details currently in the framework should be moved to the *Assessment and Item Specifications*
- Streamlining the Executive Summary/Introduction and Chapter 1 (including updating the summary table of changes since the previous framework), once the remaining chapters are finalized
- Adding more sample items to illustrate: additional disciplinary concepts, science and engineering practices, and crosscutting concepts; scenario-based tasks; and additional engineering problem contexts
- Adding scoring guides for constructed response sample items
- Continuing to refine sample items, the complexity framework, and achievement level descriptions
- Copyediting and additional formatting of text, tables, and images
- Adding more information about the use of tools

A joint virtual meeting of ADC and the Committee on Standards, Design and Methodology (COSDAM) is expected to take place this fall to review an initial draft of the Assessment and Item Specifications. There are also several rounds of NCES review of this document (which is intended primarily for them) underway between July and October, with the goal of finalizing the document in advance of the November quarterly Board meeting for action in conjunction with the framework. The Development Panel has a final virtual meeting scheduled for October 3 to discuss any remaining issues or questions as the documents are finalized for Board action.

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; both states that have adopted the Next Generation Science Standards (NGSS) and those that have other science 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.

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, were included in the May Board materials.

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 was not yet available at this preliminary stage of the process and required additional input. A memo from NCES Acting Associate Commissioner Daniel McGrath was also included in the May Board materials.

2028 NAEP Science Framework Panel Leadership Team

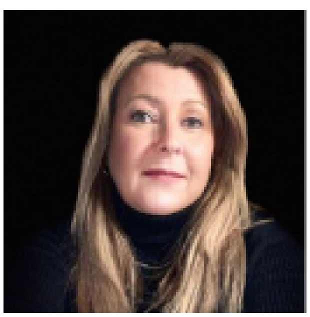
Aneesha

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

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.

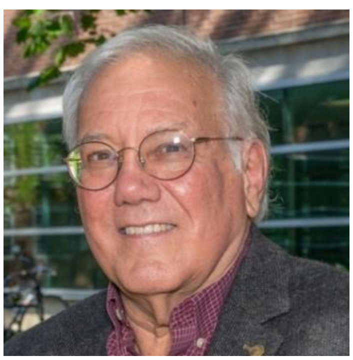
Raised on the border of Mexico, Jenny has served in multiple teacher and administrative roles in seven school districts, over the past 27 years. She has her master's degree in Aerospace Studies from the Odegard School of Aerospace Sciences. She has served as a Space Science Consultant at Brooks City-Base 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 Framework Panel Leadership Team

Nancy

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 Learning. As a former college chemistry professor, she understands and cares about students having exceptional learning experiences in science that leverage 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 principals, directors, and superintendents focused on experiences and activities that support effective teaching and learning for ALL students particularly those from served and under-estimated communities. She recently served on a National 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

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.

2028 NAEP Science Assessment Framework

Draft #2

Submitted for National Assessment Governing Board review
July 17, 2023
Contract # 91995922C0001

WHAT IS NAEP?

The National Assessment of Educational Progress (NAEP) is a continuing and nationally representative measure of trends in academic achievement of U.S. elementary and secondary students in various subjects. For nearly four decades, NAEP assessments have been conducted periodically in reading, mathematics, science, writing, U.S. history, civics, geography, and other subjects. By collecting and reporting information on student performance at the national, state, and local levels, NAEP is an integral part of our nation's evaluation of the condition and progress of education.

THE 2023–2024 NATIONAL ASSESSMENT GOVERNING BOARD

The National Assessment Governing Board was created by Congress to formulate policy for NAEP. Among the Governing Board's responsibilities are developing objectives and test specifications and designing the assessment methodology for NAEP. [Will need to UPDATE list in October 2023].

For further information, contact:

National Assessment Governing Board

800 N. Capitol St. NW Suite 825

Washington, DC 20002–4233

www.nagb.org

November, 2023

The 2028 NAEP Science Framework was developed for the National Assessment Governing Board by WestEd under contract 91995922C0001.

This document includes descriptions of science and engineering practices, crosscutting concepts, and disciplinary core ideas from *A Framework for K-12 Science Education: Practices, Crosscutting Concepts,* and excerpts from *Next Generation Science Standards: For States, By States* and associated Appendices, with permission granted by the National Academies Press, Washington, D.C.

Table of Contents

Table of Exhibits

EXECUTIVE SUMMARY

What is NAEP?

The National Assessment of Educational Progress (NAEP), often called The Nation's Report Card, is the largest nationally representative and continuing assessment of what students in public and private schools in the United States know and are able to do in various subjects. Since 1969, NAEP has been a common measure of student achievement across the country in science, mathematics, reading, and seven other subjects. NAEP results enable comparisons of what sampled students know and are able to do among states and jurisdictions, among various demographic groups, and over time. By law and by design, NAEP does not produce results for individual students or schools.

In 1988, Congress created the National Assessment Governing Board (Governing Board) as an independent, nonpartisan organization responsible for setting policy for NAEP. The 26 members of the Governing Board include governors, state legislators, state and local school officials, educators, researchers, business representatives, and members of the general public who are appointed by the U.S. Secretary of Education. The development of the NAEP assessment, administration, scoring, and reporting are carried out by the National Center for Education Statistics (NCES), located within the U.S. Department of Education's Institute of Education Sciences (IES).

As the ongoing national indicator of the academic achievement of U.S. students, NAEP regularly collects information on representative samples of students and periodically reports on student achievement in reading, mathematics, writing, science, and other subject areas. NAEP assessments are administered to students in grades 4, 8, and 12 at the national level and sometimes also for states and districts that volunteer to participate at the state level or in the Trial Urban District Assessment (TUDA) program.

NAEP scores are always reported at the aggregate level, not for individual students or schools. (By law, NAEP cannot report results for individual students.) For science, NAEP results are reported at the national and state levels and for urban school systems that volunteer for the Trial Urban District Assessment component of NAEP.

The NAEP Authorization Act of 2002 (NAEP, P.L. 107-279) is the governing statute of NAEP. This law stipulates that NCES develops and administers NAEP and reports NAEP results. Under the law, the Governing Board's responsibilities include determining the assessment schedule, developing the assessment frameworks that provide the blueprints for the content and design of the assessments, and setting the achievement levels.

By law, NAEP assessments shall not evaluate personal beliefs or publicly disclose personally identifiable information, and NAEP assessment items shall be secular, neutral, nonideological and free from racial, cultural, gender, or regional bias. Although broad implications for academic subject matter may be inferred from the assessment, NAEP does not specify how any subject area should be taught; nor does it prescribe a particular curricular approach to teaching any subject.

The NAEP program is strongly committed to equity and advances this goal through the design, administration, and reporting of assessments that strive to be inclusive and accessible for all participating students. NAEP assessments align with current educational measurement standards[1] for fair and unbiased assessments. Through contextual questionnaires, NAEP gathers and reports data that may enhance understanding of factors related to differential student achievement.

NAEP data can be used as a tool for researchers and policy makers by providing reliable information on student achievement in reading, mathematics, science, and other subjects at the national and state levels, as well as for a set of large urban districts. The NAEP website (http://nces.ed.gov/nationsreportcard) provides subject-matter achievement results (as both scale scores and achievement levels) for various subgroups; results of surveys taken by students, teachers, and school leaders to provide information on context factors such as school facilities and teaching methods, as well as the history of state and district participation, publicly released assessment questions and scoring guides. The website also contains user-friendly data analysis software to enable access to all aspects of NAEP data, perform significance tests, and create customized graphic displays of NAEP results.

Frameworks and Specifications Documents

The development of a new or improved NAEP assessment begins with the creation of a framework that describes the subject matter to be assessed for students in grades 4, 8, and 12 and the assessment questions to be asked as well as the assessment's design and administration. In accordance with Governing Board policy, a framework focuses on "important, measurable indicators of student achievement to inform the nation about what students know and are able to do without endorsing or advocating a particular instructional approach."

Each framework is accompanied by an item specifications document that serves as the "assessment blueprint" with additional information about item development. Unlike frameworks that are intended for a general audience, specifications documents are intended for a more technical audience, including NCES and the contractors who will develop the assessment items.

The broad-based process used in the development of the frameworks and specifications documents means that current thinking and research are reflected in the descriptions of what students should know and be able to do in a given subject. Therefore, these documents are

[1] American Educational Research Association, American Psychological Association, and National Council of Measurement in Education, 2014; International Test Commission, 2019; IRA/NCTE Joint Task Force on Assessment, 2010.

frequently used as resources and models for the development of state assessments.

The NAEP Science Framework

The 2028 NAEP Science Framework answers the broad question: What science knowledge, skills, and practices are to be assessed on NAEP at grades 4, 8, and 12? As an assessment framework it does not cover all relevant content for each grade level; some concepts, practices, and activities in school science are not suitable to be assessed in a short, on-demand assessment, although they may well be important components of a school curriculum. This document also does not attempt to answer the question: How should science be taught? That is a state or local decision and not suitable to be specified by NAEP.

The NAEP Science Assessment has been administered on a digital platform since 2019. Each student who is randomly selected for the NAEP assessment receives a sample of questions that take one hour to complete. At the beginning of the assessment session, students interact with a tutorial that presents all the information needed to take the assessment on the digital platform. Assessment items include both selected response and constructed response formats and represent a variety of item types. When students finish answering assessment questions, they participate in a survey that is administered on the same platform as the assessment, answering both general questions and science-related questions (e.g., about participation in science activities in and out of school). Data from these questionnaires, along with surveys completed by the participating students' science teachers and school administrators, provide valuable context to student achievement in science.

Following is a chapter-by-chapter overview of the 2028 NAEP Science Framework.

Chapter 1 - Overview begins with a brief history of NAEP science assessments and explains the rationale for the current update. It then describes the framework development process and key features of the 2028 NAEP Science Framework, which are briefly summarized below.

- The design of the new framework is based on widely accepted common state science standards and assessments. However, it is intended to inform development of an assessment, not to advocate for a particular approach to instruction or to represent the entire range of science content and skills.
- The breadth of the science principles represented in the source materials made it necessary to focus on the foundational and pervasive knowledge within each discipline and to pare down the science content to be assessed.
- The framework is based on scientific knowledge and processes derived from tested explanations and supported by accumulated empirical evidence. Explanations of natural phenomena that rely on nonscientific views are not reflected in the framework.
- Science content is presented in detailed, grade-specific charts that also allow the reader to see the progression in complexity of ideas across grades.
- Every attempt has been made to be free of error in describing the science content. The

language used strives to be accurate but not so technical as to make the framework inaccessible to a wide audience.

- The focus is on students' conceptual understanding, that is, their knowledge and use of science facts, concepts, principles, laws, and theories. Students' abilities to engage in some components of scientific inquiry and engineering design are also reflected in the framework.
- Questions on the achievement assessment and survey are non-ideological and do not concern family beliefs or attitudes.
- A variety of assessment formats are recommended, including the use of item sets and scenario-based tasks in addition to standalone items.

Chapter 1 concludes with a discussion of the challenges involved in using the framework to develop a complete assessment, given several significant challenges.

Chapter 2 - Dimensions of Science Achievement describes the science disciplinary concepts, science and engineering practices, and crosscutting concepts that operationalize the meaning of science achievement that the NAEP Science Assessment is designed to measure. The chapter introduction also explains the relationship between the science content chosen for this assessment, and two other assessment documents in wide use by the great majority of states: *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas* (NRC, 2012) and the *Next Generation Science Standards: For States, By States* (NRC Vol 1&2, 2013). Chapter 2 also describes in detail the following three dimensions of science content selected for the 2028 assessment.

- *Disciplinary Concepts* are well-tested theories and explanations developed by scientists organized into three major disciplinary groupings: Physical Science, Life Science, and Earth and Space Sciences.
- *Science and Engineering Practices* are ways of working to develop scientific explanations of phenomena or design engineering solutions to problems.
- *Crosscutting Concepts* are used across all science disciplines to provide scientists and engineers (and thus also students) with tools for applying their knowledge of science to new phenomena or problems.

Chapter 3 - Assessment Design explains how items are to be constructed to measure students' abilities to combine the disciplinary concepts, practices, and crosscutting concepts described in Chapter 2, to make sense of phenomena and design solutions to problems.

An assessment designed to measure science achievement as defined in this framework requires students to reason about compelling phenomena and/or meaningful problems. No item will assess rote content or procedural knowledge, and all items will require some reasoning with the targeted dimensions. Every item will be multidimensional. Each item will include at least one disciplinary concept and one science and engineering practice. Some items will also include at least one crosscutting concept. Discrete items are single, standalone items. Multi-part items have multiple components that are scored together producing a single score. Item sets contain multiple questions that are linked to a single phenomenon or problem, but with several items that are scored separately. Scenario-based tasks consist of multiple items built around a common problem/situation. These items typically include a storyline with interactive or static (but complex) components.

The distribution of items by content area should be approximately equal across Physical Science, Life Science, and Earth and Space Sciences at all grades. With respect to science and engineering practices and crosscutting concepts, at all grades, the emphasis should be on meaningful representation rather than a strictly equal distribution.

Chapter 4 - Reporting describes the methods that NCES will use to report results using both scale scores and the following achievement levels:

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

When sample size allows, results will be reported at the state and TUDA district level for various subgroups of students. Chapter 4 describes the interactive software tool on the NAEP website that allows users to compare student achievement results according to demographic factors such as gender, race/ethnicity, and geographic region. In addition to assessment scores, the website also reports survey data from students, teachers, and school surveys, providing data on context variables related to science achievement.

As the nation's only ongoing survey of students' educational progress, and due to its rigorous design, NAEP has become an increasingly important resource for obtaining information on what students know and are able to do. NAEP reports (known as The Nation's Report Card) are widely disseminated through the media to compare student performance in a given subject across states, within the subject over time, and among groups of students within the same grade. This major revision of the 2028 National Assessment of Educational Progress (NAEP) Science Assessment Framework is intended to ensure that NAEP results continue to serve the nation.

NAEP SCIENCE PROJECT STAFF AND COMMITTEES

The Science Framework for the National Assessment of Educational Progress is the result of extraordinary effort and commitment by hundreds of individuals across the country, including some of the nation's leading scientists, science educators, policy makers, and assessment experts. Every attempt has been made to be accurate in the description and representation of the science content. Existing framework development policies and procedures ensure periodic reviews and revisions, if necessary, to reflect advancements in scientific knowledge. Because this is a public endeavor, the Governing Board (www.nagb.org) welcomes suggestions for future versions of the framework.

GOVERNING BOARD STAFF

STEERING AND DEVELOPMENT PANEL MEMBERS

WESTED STAFF

TECHNICAL ADVISORY COMMITTEE MEMBERS

EDUCATOR ADVISORY COMMITTEE MEMBERS

CONTRIBUTING GROUPS AND INDIVIDUALS

CHAPTER ONE: Overview

1A. A Brief History of NAEP Science

Data on science achievement has been a part of NAEP since assessments began in 1969. Science achievement results were reported ten times between 1969 and 1999. In 2004, work began on a major revision of the NAEP science assessment to take into account the rapidly changing nature of science and science education, as well as advances in assessment methodologies. In 2007 and 2008 Pilot tests of the new assessment were conducted, including hands-on performance tasks and interactive computer tasks. The result was a new science framework fielded in 2009. The 2009 NAEP Science Framework was re-published by the Governing Board with minimal changes as the 2019 NAEP Science Assessment Framework (NAGB, 2019). The same framework has been the basis for NAEP science assessments in 2011, 2015, and 2019. In this document it will be referred to as the 2009-2024 NAEP Science Framework since it will also provide the basis for the 2024 assessment.

1B. Process for Developing the 2028 NAEP Science Assessment

During late summer and fall 2021, the Governing Board conducted a review of the current science framework to determine whether and how it should be updated for assessments in 2028 and beyond. In accordance with Board policy, the review included an open comment period and commissioned papers and discussions with science educators and experts. Based on this review and other relevant research, at its May 2022 quarterly Board meeting the Governing Board determined that the NAEP Science Assessment Framework needed to be updated.

According to the Governing Board policy on Assessment Framework Development for NAEP, new and updated frameworks are to be developed by a Steering Panel consisting of educators, state and local school administrators, policymakers, researchers and technical experts, assessment specialists, and other content experts and users of assessment data.

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. Board policy specifies that the Steering Panel should include 30 members, of which 20 members extend their service as members of a Development Panel. The role of the Development Panel is to follow the decisions of the larger group as it works with Governing Board staff to develop the framework and specifications documents.

The Governing Board conducted an open call for panelist nominations from mid-June through mid-July 2022. Extensive and targeted outreach was conducted to hundreds of stakeholder groups and individuals representing education, policy, industry, assessment, research, and other science-related areas. The Board evaluated applications to serve on the panels with the goal of constructing a balanced panel of stakeholders. The following factors were prioritized: a) individuals specifically nominated to represent a national organization, given the critical need to engage various constituencies; b) panelist role; c) experience and expertise overall and in the specific sub–content areas covered by the framework; d) demographic characteristics, including race, gender, and geography; e) 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 f) different perspectives on issues relevant to the Board charge. Thirty individuals were invited to serve on the Steering and Development Panels, and all who were invited agreed to participate.

The Governing Board unanimously adopted the following charge to the Steering and Development Panels that would subsequently be convened to develop an updated science framework:

- NAEP must account for greater convergence in state science standards but cannot endorse the standards of any particular state or group of states.
- NAEP should remain forward-looking and consider what students should know and be able to do in science to be successful in college and careers.
- Updates should consider whether the definition of student achievement in science needs to incorporate relevant aspects of the 2014 NAEP Technology and Engineering Literacy (TEL) Framework[2].
- Updates to the NAEP Science Assessment Framework should prioritize relevance, utility, and validity over the need to maintain trend lines but continuing the trend lines is desirable if possible.
- Updates should balance the emphasis on content and practices to ensure that the measurement of skills does not occur in isolation from content knowledge.
- Updates should be bound by considerations of feasibility, including technical issues (i.e., ensuring that the framework can be operationalized), cost (e.g., accounting for scenario-based tasks being more expensive than other item types) and the NAEP legislation (including but not limited to the requirements for NAEP to be nonsectarian).
- Updates should support the development of assessment items reflective of students who have a wide range of knowledge and skills in science.

In July 2022, the Board awarded a contract to WestEd through a competitive bidding process to convene the panelists, conduct meetings, and assist in creating the new framework and item specification document. Additional assistance was provided by a Technical Advisory Committee (a group of six measurement experts who provided feedback on technical issues) and an Educator Advisory Committee (a group of eight science educators who provided feedback on

[2] The National Assessment of Educational Progress in Technology and Engineering Literacy (NAEP TEL) is a computer-administered assessment that measures problem-solving abilities related to design and systems, the use of digital tools for collecting and communicating information, and students' understanding of issues related to technology and society. The NAEP TEL was administered to a nationally representative sample of 20,500 8th grade students in 2014 and again to 15,400 8th grade students in 2018.

issues particularly relevant to practitioners).

The panelists were tasked with developing a substantive outline of the framework that would invite public comment at an earlier stage of the process compared with prior science frameworks to allow ample time to address substantive feedback. Public comment was solicited from March - April 2023, and the framework has been revised in response to the feedback received. The Panel Leadership Team presented updates and engaged in discussion with the Governing Board at every quarterly meeting beginning in November 2022. Board action on the framework and specifications is anticipated for the November 2023 quarterly meeting. [NOTE: This will be revised as needed when the Board approves the Framework.]

1C. The Changing Construct of Science Achievement

Although NAEP has measured science achievement since its inception in 1969, the definition of science achievement has changed considerably over the decades. A major purpose of the new science framework is to anticipate how K-12 science achievement should be defined—that is, how the construct of science achievement is to be operationalized for assessment—in 2028 and beyond. As stated in the Governing Board policy, "The framework shall determine the extent of the domain and the scope of the construct to be measured for each grade level in a NAEP assessment."

The definitions of science achievement in the early NAEP assessments emphasized knowledge of scientific concepts and theories, and the interpretation of natural phenomena based on that knowledge. Gradually, capabilities of scientific inquiry, such as observation, inference, and experimentation came to be recognized as valuable and measurable, and more recently the ability to apply science principles in understanding and improving technologies. These changes and others were reflected in two influential documents developed in the 1990s: *National Science Education Standards* (National Research Council [NRC], 1996) and *Benchmarks for Scientific Literacy* (American Association for the Advancement of Science [AAAS], 1993). These documents influenced the development of state science standards during the early 2000s, which were mandated by a federal law, passed with overwhelming bipartisan support in 2001, that came to be known as No Child Left Behind (NCLB). Consequently, the 2009-2024 NAEP Science Framework drew heavily from these two documents.

The 2009-2024 NAEP Science Framework called for students at all three grade levels to be assessed on their understanding of concepts and theories in physical science, life science, and Earth and space sciences, and for items to be constructed in which they were to demonstrate their knowledge through four types of science practices: identifying science principles, using science principles, using science inquiry, and using technological design. The percentages of item types in the 2009-2024 framework varied across the grade levels. For example, recognizing that Earth and Space Sciences was rarely taught at the high school level, a larger percentage of items in Earth and Space Sciences was called for at the eighth-grade level (40%) versus at the high school level (25%). And in recognition of the fact that few students learned about technology and engineering at any grade, only 10% of items were to have students apply science through technological design.

Since the 2019 NAEP Science Framework was developed for implementation in 2009, the practice of science education in the nation's schools has undergone yet another fundamental change, guided largely by the release of a seminal document developed by a blue-ribbon panel of scientists, engineers, educators, and researchers under the auspices of the National Research Council (NRC) of the National Academy of Sciences. *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Concepts* (NRC, 2012) provides a sound, evidence-based foundation for assessment standards by drawing on current scientific research—including research on the ways students learn science effectively—and identifies the science all K–12 students should know and be able to do. The NRC Framework includes several innovations, leading to a new definition of the construct of science achievement that includes:

- Updates of disciplinary core ideas in Physical Science, Life Science, and Earth and Space Sciences as the endpoints of instruction at grades 2, 5, 8, and 12.
- The introduction of crosscutting concepts that apply to nearly all fields of science and engineering.
- Identification of specific and measurable practices common to science and engineering in place of the more amorphous inquiry skills.
- A call for the teaching and assessment of science to integrate all three dimensions of science—disciplinary core ideas, crosscutting concepts, and practices of science and engineering—to make sense of natural phenomena, and to solve challenging problems in real-world contexts.
- Recognition of the interrelation among science, engineering, society, and the environment.

Over the past decade a great majority of states have patterned their standards and assessments on the essential ideas and specific definitions in *A Framework for K-12 Science Education.* Although full adoption of the new ways of teaching have been slow to take effect (Banilower et al., 2018) the development of new curriculum materials and professional development programs aligned with the new standards continue to move the nation in a common direction.

1D. Science Achievement in the 2028 NAEP Science Framework

The Steering and Development Panels used *A Framework for K-12 Science Education* (NRC, 2012) as a foundational resource in responding to the Board charge and in making recommendations for updating the assessment construct for NAEP Science. Consistent with that document they defined the construct that the new framework will measure as follows:

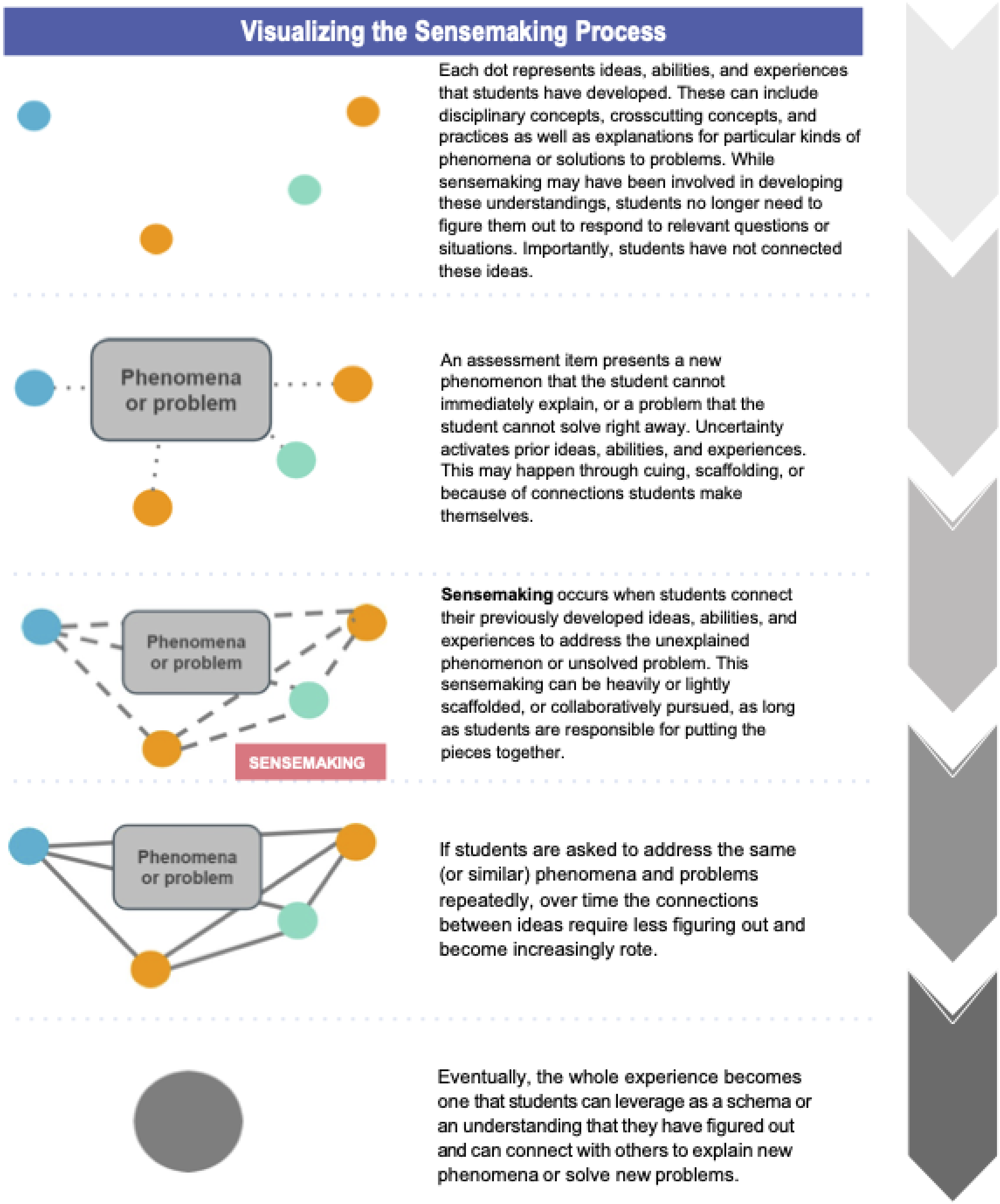
> ***Science achievement*** *is the ability to use relevant disciplinary concepts (Physical Science, Life Science, Earth and Space Sciences), science and engineering practices, and crosscutting concepts to identify and address problems, make sense of phenomena, and evaluate information to make informed decisions.*

The Panels also identified the following claims that they wanted the new framework to be able to support. Students are able to:

- Reason scientifically using disciplinary concepts in combination with science and engineering practices and crosscutting concepts.
- Address problems in the natural and designed world.
- Make sense of phenomena in the natural and designed world.
- Evaluate information and make decisions.

Built into the proposed construct is the idea of sensemaking, an essential aspect of all test items on the 2028 NAEP Science Assessment. In contrast to items that measure a student's ability to recall rote knowledge, NAEP science assessment items will require students to actively apply disciplinary concepts, science and engineering practices, and crosscutting concepts to figure out a phenomenon or address a real-world problem. The ability to apply disciplinary concepts using practices and crosscutting concepts is an intrinsic feature of sensemaking. The role of phenomena and problems in sensemaking is illustrated in Exhibit 1.1 and is described in more detail in Chapter 3.

Exhibit 1.1. Visualizing the Sensemaking Process[3]



[3] Adapted from Achieve (2019b). The Task Annotation Project in Science: Sense-making. Retrieved from https://issuu.com/achieveinc/docs/sense-making_02142019__7_

1E. Opportunity to Learn and an Expansive Understanding of Contextual Variables

NAEP testing is not intended to assess students or even individual schools, rather it is intended to evaluate the state of science learning across a region or a state, or large urban school districts that participate in the Trial Urban District Assessment (TUDA). Performance depends on students' opportunities to learn science both within and outside of school and on multiple social factors that condition those opportunities and that learning. This framework defines what should be measured about science learning by student testing and seeks also to place those results in the context of what can be gleaned about the students, their learning opportunities and their personal contexts from non-test questions that are asked of them, their teachers, and their school administrators. These factors are called contextual variables by NAEP. While NAEP student results are reported by categories such as gender, race, students classified as English language learners, or socioeconomic status, the differences across these categories are highly interrelated, and these categories alone do not tell us the full story of what the results are saying about the state of science learning in our nation's schools.

Opportunity to learn is generally understood to refer to inputs and processes that shape student achievement, including the school conditions, time and spaces devoted to science learning, teacher knowledge and beliefs about science learning, and the curriculum, instruction, and resources to which students have access. When opportunity to learn was first used as a concept, Carroll (1963, 1989) emphasized the time allowed for learning. For the past 50 years, the concept of opportunity to learn has continued to evolve, as have efforts to measure in-school opportunities to learn, with the majority of scholars focusing on the classroom as the unit of analysis and instruction as central. Research, for example, has documented the negative effects on achievement of policies and practices that are often found in schools serving children who live in poverty or have special needs, including an inadequate supply of science teachers with strong knowledge and skills, a tendency to offer few advanced science courses, and a common practice of tracking these students disproportionately into low-level courses that restrict their learning opportunities (e.g., Fergusan et al., 2007; Kohlhaas et al., 2010), all of which can be understood as instructional resources that shape what students learn.

In recent years, there has been significant research on science learning and on the conditions and contexts that affect it. Two NRC reports have summarized much of what is known in this domain, namely *Science and Engineering for grades 6-12* (NASEM, 2019) and *Science and Engineering in Preschool through Elementary grades* (Davis & Stephens, 2022). These reports have noted the historical tendencies to view science learning, particularly at the high school level, as in service to the production of scientists and engineers, and thus intended for a select group of students. They and other studies, including *A Framework for K-12 Science Education* (NRC, 2012) have argued that today a strong science education is needed for all students as preparation for life and community membership in the world of today, where many personal and community decisions require everyone to be able to interpret and apply scientific ideas and practices in the context of their daily lives. The NAEP Science Assessment, along with contextual information about the experiences of the participating students, is intended to measure how well that need is being met. Contextual information is critical to interpreting its results. Priorities for science-specific contextual variables are included in Chapter 4.

1F. Challenges of Developing a NAEP Assessment

Once a framework is completed and approved by the Governing Board, the next step is for the National Center for Education Statistics to develop the assessment items. Here we discuss three major challenges: (1) measurement constraints and the nature of the items included on the assessment, (2) time and resource constraints and how much can be assessed in NAEP, and (3) the timeline for the framework and the difficulty of developing a 10-year framework with the rapid explosion of knowledge in the Information Age. Although these challenges apply to assessments in all subjects, they are especially challenging in the area of science, due to the rapidly changing nature of the subject and wide diversity of potential topics. Each of these challenges is discussed below.

Measurement Constraints

Like any large-scale assessment in education, the workplace, or clinical practice, NAEP is constrained in what it can measure. This has implications for the proper interpretation of NAEP Science Assessment results. The NAEP Science Framework is an assessment framework, not a curriculum framework. Although the two are clearly interrelated, each has a different purpose and a different set of underlying assumptions. A curriculum framework is designed to inform instruction, to guide what is taught, and often, to guide how it is taught. It represents a very wide universe of learning outcomes from which teachers pick and choose what and how they teach. An assessment framework is a subset of the achievement universe from which assessment developers must choose to develop sets of items that can be assessed within time and resource constraints. Hence, the science content to be assessed by NAEP has been identified as disciplinary concepts that are central to the physical, life, and Earth and space sciences. As a result, some important outcomes of science education that are difficult and time consuming to measure (such as habits of mind, sustained inquiry, and collaborative research), but valued by scientists, science educators, and the business community, will be only partially represented in the framework and in the NAEP Science Assessment. Moreover, the wide range of science standards in the guiding national documents that could be incorporated into the framework had to be reduced in number to allow some in-depth probing of fundamental science content. As a result, the framework and the specifications represent a careful distillation that is not a complete representation of the original universe of achievement outcomes desirable for science education.

Assessment experts on the Development Panel and staff of the Governing Board also considered feasibility when drafting recommendations. For example, hands-on performance tasks, which were called for in the 2009-2024 NAEP Science Framework, have been eliminated and replaced by scenario-based tasks due to concerns about cost and feasibility of implementation.

Time and Resource Constraints

What NAEP can assess is limited by time and resources. Like most standardized assessments, NAEP is an “on-demand” assessment. It ascertains what students know and are able to do in a limited amount of time (60 minutes) and with limited access to resources (e.g., reference materials, feedback from peers and teachers, opportunities for reflection and revision). State standards, however, contain goals that require extended time (days, weeks, or months).

Like other on-demand assessments, NAEP cannot be used to draw conclusions about student achievement with respect to the full range of goals of science education. States, districts, schools, and teachers can supplement NAEP and other standardized assessments to assess the full range of science education standards. In addition to describing the content and format of an examination, assessment frameworks like this one signal to the public and to teachers the elements of a subject that are important. The absence of extended inquiry in NAEP, however, is not intended to signal its relative importance in the curriculum. Indeed, because of the significance of practices of science and engineering in science education, the framework promotes as much consideration of students’ abilities to use these practices as can be accomplished within the time and resources available for assessment.

Balancing Current and Future Standards and Curricula

The framework attempts to strike a balance between what can reasonably be predicted about future school science and what students are likely to encounter in their curriculum and instruction now and in the near future. It is a significant challenge to write a framework for the future. Cutting edge science research creates new knowledge and investigative practices at the intersection of disciplinary boundaries. For example, research on human and natural systems has generated new understanding about environmental science that is closely linked to knowledge generated in Physical Science, Life Science, and Earth and Space Sciences. Although the framework is organized into these traditional areas, features of current science research are woven throughout. Another example of burgeoning knowledge is the rapid development of technology, such as the transformation of our energy infrastructure from fossil fuels to renewable resources, and new developments in artificial intelligence.

The framework is intended to be both forward looking (in terms of the science content that will be of central importance in the future) and reflective (in terms of current school science). Because it is impossible to predict with certainty the shape of school science, the choices made for this framework should be revisited in response to future developments in school science.

A summary of the changes in the 2028 NAEP Science Framework, compared with the 2009-2024 framework, is shown in Exhibit 1.2. For more details on the three dimensions of science, see Chapter 2. For examples of items that combine these dimensions, see Chapter 3.

Exhibit 1.2. Summary of Changes in the 2028 NAEP Science Assessment

Topic	Change	Rationale
NAEP Science Construct	The framework defines the construct of science achievement and explains how this construct is operationalized using the three dimensions of science. This clearly defined construct helps to ensure that the assessment is measuring what it intends to measure (i.e., construct validity) by outlining exactly what is included and not included, helping to ensure that items can capture this construct and not elements outside of this construct.	Precisely defined constructs help to ensure that an assessment measures the construct it intends to measure rather than aspects not part of that construct, which creates construct-irrelevant variance. Without a precisely defined construct, it is hard to know whether items and other design features work toward measuring the intended construct or whether they might, in fact, be measuring something else.
Three Dimensions of Science	NAEP Science "content" has been redefined as any knowledge and reasoning skills that students need to know and be able to do on the NAEP Science Assessment. The content now includes updated and renamed science content statements (now disciplinary concepts) and science practices (now science and engineering practices), along with the addition of crosscutting concepts. These are now referred to collectively as the "three dimensions of science."	Prior NAEP Science frameworks organized what students should know and be able to do into two buckets: science content and science practices. Based on research presented in the NRC Framework, it is recommended that the science content covered on the NAEP Science Assessment now consist of science disciplinary concepts, science and engineering practices, and crosscutting concepts.

Topic	Change	Rationale
	Disciplinary Concepts are well tested theories and explanations developed by scientists organized into three major groupings: Physical Science; Life Science; and Earth and Space Sciences.	While the science ideas are still organized into three broad disciplinary groupings, NAEP science content statements have been renamed NAEP disciplinary concepts and updated to 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.
	NAEP Crosscutting Concepts have been added to the NAEP Science "content" and are defined as concepts used across all science disciplines that provide scientists and engineers and thus also students tools for asking productive questions and organizing their thinking.	With the introduction of the NAEP Crosscutting Concepts, based on findings reported in research on science learning, the updated definition of science achievement now describes the need for an assessment that can provide evidence about what students know and are able to do with all three dimensions of science.
	NAEP Science and Engineering Practices describe the skills and knowledge necessary to develop scientific explanations of phenomena and to design engineering solutions to problems.	NAEP science practices have been renamed "science and engineering practices" and updated to 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.

Topic	Change	Rationale
Technology and Engineering	Technology and engineering concepts that are relevant to science achievement have been integrated into the updated science and engineering practices.	The addition of technology and engineering concepts to NAEP Science 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. The framework incorporates concepts that represent the overlap between the NRC Framework and the NAEP Framework for TEL.
Assessment Design	The framework calls for students to use the three dimensions of science. Assessment items should require students to bring the three dimensions of science together to engage with the item. Items, item sets, and scenario-based tasks should be three dimensional whenever possible. No item will be one dimensional.	With the updated definition of science achievement, and the incorporation of the three dimensions of science, the assessment design should reflect the need for students to address all three dimensions in their demonstration of what they know and are able to do in science.
	The framework provides expanded recommendations and guidance on the following: • use of diverse tasks, phenomena, and contexts for items • considerations for language complexity • elimination of concept maps and replacement of hands-on tasks (HOTs) with scenario-based tasks	NAEP assessment items should be reflective of students who have a wide range of knowledge and skills. Feasibility is also a consideration, and HOTs are costly to administer and require additional personnel for implementation. Scenario-based tasks can address the same content as HOTs, but with easier administration and implementation.

Topic	Change	Rationale
	The framework calls for an even distribution of items across the three disciplines (Physical Science, Life Science, Earth and Space Sciences) and item types (selected response and constructed response) across grades 4, 8, and 12.	Prior NAEP science assessment frameworks called for differing distribution levels (higher percentage of Earth and Space Sciences at grade 8, lower percentage of Earth and Space Sciences at grade 12) based on NAEP data regarding students' course-taking patterns. 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.
	The framework includes a complexity framework in Chapter 3. The purpose of the complexity framework is to inform item development as to ensure that items are accessible to a wide range of learners.	The complexity framework will be applied to NAEP item development to reflect how complexity specifically scales within and across multidimensional science items. This, in part, guides the development of multidimensional items that assess the full range of student performance.
Reporting Results	Sub scale reporting categories in Physical Science, Life Science, and Earth and Space Sciences have "Sensemaking in" added to each.	With the updated definition of science achievement, and the incorporation of the three dimensions of science, the reporting of results for NAEP Science should reflect the emphasis on student scientific sensemaking.
	Recommendations for science-specific contextual variables have been updated and prioritized.	NAEP contextual variable survey items should be reflective of the changing nature of science instruction on opportunities for students to learn science.

Detailed lists of the disciplinary concepts, science and engineering practices, and crosscutting concepts to be assessed in 2028 and beyond are the focus of Chapter 2. Explanations of how these dimensions are to be combined to create assessment items are included with examples in Chapter 3. The processes for scoring, analyzing, interpreting, and reporting on NAEP Science achievement and contextual variables are summarized in Chapter 4.

CHAPTER TWO: Dimensions of Science Achievement

The NAEP Science Steering and Development Panels defined the construct that the 2028 NAEP Science Framework will measure as follows:

> ***Science achievement** is the ability to use relevant disciplinary concepts, science and engineering practices, and crosscutting concepts to identify and address problems, make sense of phenomena, and evaluate information to make informed decisions.*

The dimensions of science achievement are the disciplinary concepts, science and engineering practices, and crosscutting concepts. These dimensions are defined as follows:

- **Disciplinary Concepts** are well-tested theories and explanations developed by scientists organized into three major disciplinary groupings: Physical Science, Life Science, and Earth and Space Sciences.
- **Science and Engineering Practices** are ways of working to develop scientific explanations of phenomena or design engineering solutions to problems.
- **Crosscutting Concepts** are ideas that are used across all science disciplines and provide scientists and engineers and thus also students with tools for applying their knowledge of science to new phenomena or problems.

The use of multiple dimensions to make sense of phenomena is the essence of authentic science achievement.

2A. NAEP Science Disciplinary Concepts

Science knowledge continues to develop yearly, but it is neither possible nor desirable to try to teach and assess more science for K–12 students every year. Rather, it is important to select the core ideas from each disciplinary area, framed in an up-to-date way, to describe a base of knowledge for use that will allow students to become adults who can continue to learn scientific ideas throughout their lives, and build the interest and engagement with science that will allow those who choose to do so to enter careers that require going well beyond this base. The core ideas assessed as the NAEP Disciplinary Concepts below follow those recommended by the panel of distinguished scientists and educators who developed the NRC document *A Framework for K–12 Science Education*.

These concepts progress across the years of K–12 education. The exhibits below delineate the progressions across the grades 4, 8, and 12. Similar concepts, presented at a growing level of sophistication, are grouped in rows of the tables. Some concepts have no entry at grade 4 because their development is expected to begin later in the sequences of learning used by most schools across the U.S. Some disciplinary concepts include a clarification and/or a boundary statement. **Clarification statements** enhance disciplinary concepts by explaining the emphasis, giving examples, or providing a specific point of detail. **Boundary statements** tell the item writer what the item should not cover in relation to the disciplinary concept. [NOTE: When the framework is finalized, the clarification and boundary statements will be moved to only appear in the

Assessment and Item Specifications document.]

Since NAEP testing occurs midyear, the grade levels are defined by what a student would know and be able to do by the middle of that school year. The framework organizes the multiple disciplines of science into three major groupings and several subgroupings:

Exhibit 2.1. NAEP Science Discipline Groups and Subgroups

Physical Science	**Life Science**	**Earth and Space Sciences**
Matter and Its Properties • Properties of Matter • Structure of Matter • Phases of Matter and Atomic Substructure • Chemical Processes • Nuclear Processes Motion and Forces • Forces on an Object • Forces between Objects • Types of Forces Energy • Energy Flow and Transfer • Kinetic and Potential Energy • Thermal and Radiant Energy • Energy Conservation Waves and Their Role as Carriers of Information • Wave Patterns • Sound Waves • Electromagnetic Waves	From Molecules to Organisms: Structures and Processes • Structure and Function of Living Things • Reproduction • Matter and Energy in Organisms Ecosystems: Interactions, Energy, and Dynamics • Interdependent Relationships • Cycles of Matter and Energy Transfer • Ecosystem Dynamics, Functioning, and Resilience Heredity: Inheritance and Variation of Traits • Inheritance • Variation Biological Evolution: Unity and Diversity • Evidence of Common Ancestry and Diversity • Mechanisms of Change	Universe, Solar System, and Earth • Patterns of motion of space objects • Solar System • Formation of the universe Earth's Systems • Plate tectonics, patterns on the surface of the Earth • Earth's history • Water cycling, weathering, and erosion • Weather and Climate Earth and Human Activity • Natural Resources • Natural Hazards • Human Impacts on Earth Systems • Climate Change

These groupings should provide a coherent organization of the ideas to be tested.

The NAEP Science Assessment Framework will assess disciplinary concepts that are:

- useful in understanding the world and informing decisions in everyday life;
- central to the discipline;
- likely to endure after instruction;
- able to be measured meaningfully with items that engage students in sensemaking about a variety of phenomena and finding solutions to problems;
- critical to measure and monitor to understand large-scale trends in students' science learning; and
- included in most state standards.

The focus in the selection process was on the central principles of each discipline. The selected big ideas represent foundational and pervasive knowledge, key points of scientific theories, and underpinnings upon which complex understanding is built. A primary consideration was the grade-level appropriateness and accuracy within grade level of concept statements, based on the great majority of state standards. Once key concepts were identified within subtopics, the progression of ideas and performances, informed by available research, was tracked through grades 4, 8, and 12. A deliberate attempt was made to limit the breadth of science concepts to be assessed so that some important topics could be measured in-depth.

As an organizational tool in the Disciplinary Concept tables below, each disciplinary concept is preceded by a specific code in bold (e.g., "L12.10) Within a code, the letter denotes broad content area ("P" for Physical Science, "L" for Life Science, and "E" for Earth and Space Sciences); the number before the period denotes grade level (grade 4, 8, or 12); and the number following the period denotes the concept's order of appearance within a given content area and grade. For example, L12.10 denotes that this is the tenth concept to appear in the grade 12 section of the Life Science disciplinary concepts. Because the numbering within each content area and grade is sequential, code numbers do not necessarily indicate any relationships across grades.

Disciplinary Concepts in Physical Science

Matter and Its Properties

- Properties of Matter
- Structure of Matter
- Phases of Matter and Atomic Substructure
- Chemical Processes
- Nuclear Processes

Motion and Forces

- Forces on an Object
- Forces between Objects
- Types of Forces

Energy

- Energy Flow and Transfer
- Kinetic and Potential Energy
- Thermal and Radiant Energy
- Energy Conservation

Waves and Their Role as Carriers of Information

- Wave Patterns
- Sound Waves
- Electromagnetic Waves

NAEP Physical Science Disciplinary Concepts		
Grade 4	**Grade 8**	**Grade 12**
Big Idea: Matter and Its Properties How can the great variety of substances and processes of change in matter be explained?		
Sub-Topic: Properties of Matter		
P4.1: Different types of matter (materials) have different properties. Each material can be classified using a number of its properties. Materials with different properties are needed for different uses. [**Clarification:** Students should recognize the variety of solid and liquid matter and be able to interpret evidence about properties but are not expected to remember properties of materials.] [**Boundary**: Students should not be expected to recall the description of a substance.] [**Boundary**: Students are not expected to recognize gasses as matter at this grade level. At this grade level, students are not expected to develop understanding of the structure of matter.]	P8.1: Each pure substance can be identified by its characteristic properties. [**Boundary**: Students are not expected to remember characteristic properties of various substances.]	

Sub-Topic: Structure of Matter		
	P8.2: All substances are made from atoms. There are over 100 different types of atoms, which combine with one another in various ways. Atoms form molecules or extended structures. [**Boundary:** Students are not expected to recall the symbols of atoms.]	P12.1: All matter is made of atoms that contain protons that are positively charged and neutrons that have no electric charge in the nucleus and electrons that have negative charge that surround the nucleus. Neutral atoms can lose electrons to become positively charged ions or gain electrons to become negatively charged ions. [**Boundary:** Students are not expected to recall the number of protons, electrons, and neutrons with any given atom.] P12.2: Electrical attractions and repulsions between positively-charged nuclei and negatively-charged electrons explain both the structure of isolated atoms and the forces between two or more nearby atoms that cause them to form molecules, compounds, and extended materials (i.e., the formation of chemical bonds). **[Boundary:** Full understanding of atomic stability is beyond grade level, all that is included here is that the positive nucleus attracts the electron cloud, and that electrical forces between nearby or overlapping atoms can form stable

		structures such as molecules and extended materials.]
Sub-Topic: Phases of Matter and Atomic Substructure		
P4.2: Many materials can be solid and liquid depending on temperature. [**Boundary:** Gas is not assessed at this grade level.]	P8.3: In any state -- gas, liquid, or solid -- the temperature influences the motion of atoms and molecules. In solids the atoms are close together, held in place relative to each other by forces between them, and move only with small vibrations about those positions. In liquids, the atoms or molecules are close together but are moving around relative to one another. The atoms and molecules that make up gas are relatively far apart and move around freely. [**Clarification**: Stress is on how particulate models of these three phases of matter can account for phenomena and properties of matter.] [**Boundary:** The energetics of phase changes or calculation of gas laws are not included at this grade level.]	P12.3: In gasses or liquids, the motion of atoms or molecules leads to collisions between them. Such collisions are necessary for chemical processes to occur. Higher rates of collisions occur at higher temperatures because atoms are typically moving faster, and at higher pressure in a gas because the atoms are closer together.
Sub-Topic: Chemical Processes		
	P8.4: In a chemical reaction, the atoms of	P12.4: A stable molecule has less energy

	the reacting substances are regrouped in characteristic ways into new substances with different properties. Atoms only rearrange. As such the amount of matter does not change. [**Clarification:** The emphasis in this statement is on recognizing when a chemical process has occurred. A further consequence of the statement is that the total mass (or weight) of matter present does not change, which can be used to make inferences, for example that invisible gasses have left an open system.] **[Boundary**: Mass and weight are not distinguished at this level, and students are not expected to remember atomic masses.]	than the same set of atoms at rest far apart. Any process that results in a new set of molecules must start with some energy input that allows a break-up or the initial molecule or molecules to begin the process. Often this energy comes from the kinetic energy of colliding molecules. P12.5 In some chemical reactions, energy is released as higher kinetic energy of motions of the products compared to that of the reactants. P12.6: The total number of atoms of each type does not change in any chemical process; that is, atoms are conserved in all such processes. Knowing that atoms are conserved during chemical processes, together with knowledge of the characteristic chemical properties of each element, allows individuals to describe and predict chemical reactions. [**Clarification:** Students should know basic trends in the periodic table, like as you move left to right, you transition from metals, to gasses, to nonreactive gasses.] [**Boundary**: Students are not expected to know details of the electron structures of various elements or memorize the periodic table columns that show similar chemical properties.]

		[**Boundary**: In balancing chemical reactions, students should only be expected to work with small molecules.]
Sub-Topic: Nuclear Processes		
		P12.7: Nuclear processes, including fusion, fission, and radioactive decays of unstable nuclei, involve release or absorption of energy. The total number of neutrons plus protons does not change in any nuclear process. [**Clarification:** Fusion includes nuclear fusion processes in the sun that release energy that eventually leaves the sun's surface as radiation and particle flows.]
Big Idea: Motion and Forces How can motion be described? What makes the motion of an object change?		
Sub-Topic: Forces on an Object		
P4.3: Unequal forces acting on an object can change its motion or forces can balance against other forces to hold the object in place.	P8.5: The change in motion of an object is determined by the sum of the forces acting on it; if the net force on the object is zero, it will remain at rest or continue	P12.8 The motion of an object changes if and only if the sum of the forces acting on it is non-zero. [**Clarification**: quantitative treatment of

[**Clarification:** A force acts on a single object and is due to the effect of another object that may or may not be touching it. Each force has both a strength and a direction.] [**Boundary**: The changes in motion introduced at this level are limited to obvious visible examples such as starting, stopping, or bouncing. Balanced forces on an object moving in a straight line at a steady speed are not introduced.] [**Boundary**: Forces acting at a distance are treated only for visible cases, such as between the Earth and a weight hanging on a spring or a magnet picking up a small object.]	moving in a straight line with the same speed and direction as before. [**Clarification:** Examples of static non collinear force situations could include a ladder leaning against a wall or an object suspended by multiple springs or rubber bands.] [**Clarification**: Forces are treated only qualitatively.] P8.6: The greater the mass of the object, the greater the force needed to achieve the same change in motion. For any given object, a larger net force causes a larger change in motion. [**Clarification:** Forces are treated only qualitatively.]	aligned forces is included but non collinear forces are treated only qualitatively.]
Sub-Topic: Forces between Objects		
	P8.7: For any pair of interacting objects, the force exerted by the first object on the second object is equal in strength to the force that the second object exerts on the first but in the opposite direction.	P12.9: Momentum is always conserved whether within a system or between two different systems. This is a consequence of the fact that the forces between any two interacting objects are equal and opposite and thus result in equal and opposite changes in momentum. [**Boundary:** Students are expected to apply the concept of momentum and

		changes of momentum qualitatively only, except when all forces and motions are collinear. Momentum is defined non-relativistically, i.e., p=mv]
Sub-Topic: Types of Forces		
P4.4: Objects exert forces on each other when they are touching or colliding with each other. [**Boundary:** At this grade level, gravity is not discussed.]	P8.8: Electric and magnetic forces between two objects can pull them together or push them apart. The magnitude depends on the magnitude of the charges, currents, or magnetic strengths involved and on the distances between the interacting objects. P8.9: The gravitational forces between any two objects with mass will pull them toward each other. The gravitational force between any two masses is very small except when one or both of the objects have large mass—e.g., Earth and the sun.	P12.10: Forces between objects at a distance are explained by fields (gravitational, electric, and magnetic) permeating space that can transfer energy and momentum through space. Any object with mass is a source of a gravitational field which exerts an attractive force on any other mass. The strength of the pair of forces between any pair of masses is proportional to the product of their masses and depends on the distance between the two centers of mass. P12.11: Attraction and repulsion and magnetic effects between electric charges (their electromagnetic interactions) at the atomic scale explain the structure, properties, and atomic scale processes of matter and forces between surfaces in contact.

Big Idea: Energy Why do we care about keeping track of energy? Why are so many different phenomena associated with energy?		
Sub-Topic: Energy Flow and Transfer		
P4.5: Energy can move from place to place by the motion of objects or by sound, light, heat, or electricity. [**Clarification**: Students can recognize that turning a switch allows electricity (whatever that may be) to provide energy to a light bulb or a toaster far from the power plant or solar panel that "makes" electricity.] [**Boundary**: At this grade level, students are not expected to know the nature of electrical currents.] P4.6 When objects collide, the forces between them can transfer energy from one object to the other. Typically, a sound is produced, showing that some energy has been transferred to the air.	P8.10: When two objects interact, each one exerts a force on the other that can cause energy to be transferred from one object to the other. P8.11: Electric currents are generated in multiple ways using a variety of energy transfers to produce them. We use that energy to produce the movement of machines, heat, and/or light. All the energy so "used" is eventually transferred to the surrounding environment as thermal energy. [**Clarification**: Stress is on tracking energy flows into, out of and within systems in everyday processes.]	P12.12: When two objects interacting through a field change relative position, the energy stored in the field is changed.
Sub-Topic: Kinetic and Potential Energy		
P4.7: Objects in motion have energy. The	P8.12: The energy of motion of particles	P12.13: Energy is a quantitative property

faster a given object is moving, the more energy it has.	or waves is called kinetic energy; for massive objects it is proportional to the mass of the moving object and grows with the square of its speed. [**Boundary**: Includes calculating ½ mv^2 for given values of mass *m* and speed *v*, but does not include rearranging the formula or solving for *v*.] P8.13: Any system of objects contains energy because of the gravitational, electric, and magnetic interactions between the objects. This energy is called potential energy. The amount depends on the relative positions of objects. [**Clarification:** Stress is on macroscopic objects and their mass, charge, and magnetic properties.]	of any system. The amount of energy available for processes in that system depends on the motion and interactions of matter and radiation within that system. The availability of energy limits what can occur in any system. [**Clarification**: Assessment includes manifestations of energy at the microscopic scale modeled as either the motion of particles or radiation, or energy stored in fields. Emphasis is on qualitative association of directly detectable macroscopic manifestations of energy with microscopic scale underlying processes.] [**Boundary:** The quantum model of radiation as a flux of particles is not introduced.]
Sub-Topic: Thermal and Radiant Energy		
P4.8: Heat and light from the sun are major sources of energy on Earth.	P8.14: The energy associated with random movements of atoms and molecules is called thermal energy. In all matter, the atoms are moving. The more thermal energy, the more the motion of atoms. The term heat is used only for energy transferred between two objects or systems at different temperatures.	P12.14: When sunlight is absorbed at Earth's surface it is eventually re-radiated as infrared radiation that transfers heat into the atmosphere. The average temperature of the atmosphere is determined by how long the energy stays in the system until it is reradiated into space from the top of the atmosphere.

	[**Clarification:** Heat transfer can be by convection (matter flow), conduction, or radiation. Motion patterns are different for solids, liquids, or gasses. Changes of phase, whether solid->liquid or liquid ->gas, require added energy to occur, but take place with no change in temperature.] [**Boundary:** Relative motion of subatomic particles is not introduced.] P8.15: Two systems at the same temperature could have different total energy; the relationship between the temperature and the total energy of a system depends on the types, states, and amounts of matter present. [**Clarification:** It takes a different amount of energy to heat the same weight or volume of different substances by the same amount.] [**Boundary:** Assessment includes qualitative not quantitative application of this idea.]	[**Clarification:** Various gasses, known as greenhouse gasses, present in the atmosphere absorb and re-radiate infrared radiation so its path from Earth's surface to space is a series of many short steps that depends on the concentrations of such gasses.]
Sub-Topic: Energy Conservation		

	P8.16: Any object absorbs energy from, or loses energy to, the air or other matter it is touching depending on whether it is colder or hotter than the surrounding matter. Energy is spontaneously transferred out of hotter regions or objects and into colder ones. [**Clarification:** Thermal energy transfers through particle collisions or emission or absorption of infrared radiation are both included.] [**Boundary**: Infrared radiation is assessed only in cases where it can be felt as noticeable warming.]	P12.15: Energy cannot be created or destroyed, but it can be transferred from one place to another and between systems. [**Boundary**: Quantitative application of conservation of energy is limited to simple physical cases (e.g., a freely falling mass, a swinging pendulum, a mass bouncing on a spring). In such cases all needed formulae for energy are provided.] P12.16: Although energy cannot be destroyed, it can be converted to a less useful form, becoming thermal energy in the surrounding environment.
Big Idea: Waves and Their Role as Carriers of Information How can information be encoded, and sent over long distances and decoded? What physical phenomena do we use to do this?		
Sub-Topic: Wave Patterns		
P4.9: Waves are regular patterns of motion in matter (e.g., waves can be made in water by disturbing the surface).	P8.17: Waves of the same type can differ in amplitude and wavelength and multiple waves traveling together can add to give complex patterns that can be used to encode information. Waves of the same	P12.17: The speed of a wave depends on the type of wave and on properties of the medium through which it is passing. [**Clarification**: Emphasis on sound and

29

	type traveling in different directions can pass through one another and emerge unchanged. [**Clarification**: Examples of information carrying waves could include sound, light, and other electromagnetic waves such as AM/FM radio, WiFi, and Bluetooth.] [**Boundary**: The assessment at this grade level is qualitative only; it can be based on examples such as the fact that two different sounds can pass a location in different directions without getting mixed up.]	light waves.] P12.18: Information can be transmitted by continuous waves or as digital pulses and can be stored in digital form (e.g., a picture stored as the values of an array of pixels).
Sub-Topic: Sound Waves		
P4.10: Sound can make matter vibrate, and vibrating matter can make a sound. [**Boundary:** Assessment includes qualitative ideas about vibration only]	P8.18: A sound wave needs a medium through which it is transmitted. The medium can be solid, liquid, or gas. [**Clarification:** Emphasis is on phenomena involving sound transmission.] [Boundary: Assessment should include details of sound wave forms.]	
Sub-Topic: Electromagnetic Waves		

P4.11: Some materials allow light to pass through them, others allow only some light through, and others reflect or absorb all the light that reaches them and cast a dark shadow on any surface beyond them, where the light cannot reach. An object can be seen only when light produced by it or reflected from its surfaces enters the eyes. [**Clarification:** Emphasis is on phenomena involving light beams, light sources, mirrors, and shadows.**]** [**Boundary**: Facts or concepts about the speed of light are not assessed.]	P8.19: When light shines on an object, it is reflected, absorbed, or transmitted through the object, depending on the object's material and the frequency (color) of the light. A wave model of light is useful for explaining brightness, color, and the frequency-dependent bending of light at a surface between media. However, because light can travel through space, it cannot be a matter wave, like sound or water waves.	P12.19: Many seemingly unrelated phenomena, from x-rays to radio waves, are electromagnetic waves like light but have very different wavelengths and frequencies. Electromagnetic waves are produced by patterns of motion of charges or magnets. The wave is a pattern of changing electric and magnetic fields.

Disciplinary Concepts in Life Science

From Molecules to Organisms: Structures and Processes

- Structure and Function of Living Things
- Reproduction
- Matter and Energy in Organisms

Ecosystems: Interactions, Energy, and Dynamics

- Interdependent Relationships
- Cycles of Matter and Energy Transfer
- Ecosystem Dynamics, Functioning, and Resilience

Heredity: Inheritance and Variation of Traits

- Inheritance
- Variation

Biological Evolution: Unity and Diversity

- Evidence of Common Ancestry and Diversity
- Mechanisms of Change

NAEP Life Science Disciplinary Concepts		
Grade 4	**Grade 8**	**Grade 12**
Big Idea: From Molecules to Organisms: Structures and Processes How do organisms live, grow, respond to their environment, and reproduce?		
Subtopic: Structure and Function of Living Things		
L4.1: Plants and animals have both internal and external structures that serve central functions necessary for life -- growth, survival, behavior, and reproduction. [**Clarification:** Examples of structures could include thorns, stems, roots, colored petals, heart, stomach, lung, brain, and skin.] [**Boundary:** Assessment is limited to macroscopic structures within plant and animal systems.]	L8.1: For both single cells and multiple cellular organisms, special structures within cells are responsible for particular functions. [**Clarification:** Emphasis is on the cell functioning as a whole system and the primary role of identified parts of the cell, specifically the nucleus, chloroplasts, mitochondria, cell membrane, and cell wall.] [**Boundary:** Assessment of organelle structure/function relationships is limited to the cell wall and cell membrane. Assessment of the function of the other organelles is limited to their relationship to the whole cell. Assessment does not include the biochemical function of cells or cell parts.]	L12.1: Systems of specialized cells within organisms help them perform the essential functions of life, which involve chemical reactions that take place between different types of molecules. [**Clarification:** Assessment includes one or more of the following types of molecules: water, proteins, carbohydrates, lipids, or nucleic acids. Emphasis is on the idea that all life functions, including growth and reproduction, are a result of chemical reactions.] [**Boundary:** Assessment does not include identification of specific cell or tissue types, whole-body systems, specific protein structures and functions, or the specific biochemistry of protein

	L8.2: In multicellular organisms, the body is a system of multiple interacting subsystems that are groups of cells that work together to form tissues and organs that are specialized for particular body functions. [**Clarification:** Emphasis is on the conceptual understanding that cells form tissues and tissues form organs specialized for particular body functions. Examples could include the interaction of subsystems within a system and the normal functioning of those systems.] [**Boundary:** Assessment does not include the mechanism of one body system independent of others. Assessment is limited to the circulatory, excretory, digestive, respiratory, muscular, and nervous systems.]	synthesis. Assessment does not include the names, steps, or specific processes involved in chemical reactions.] L12.2: Multicellular organisms have a hierarchical structural organization, in which its systems support functions necessary for the organism's survival and reproduction. Each system is made up of numerous parts and is itself a component of the next level. [**Clarification:** Emphasis is on functions at the organism system level such as nutrient uptake, water delivery, and organism movement in response to neural stimuli.] [**Boundary:** Assessment does not include interactions and functions at the molecular or chemical reaction levels.] L12.3: Feedback mechanisms maintain a living system's internal conditions within certain limits. Feedback mechanisms discourage change by means of negative feedback or proceed with changes through a system of positive feedback. [**Clarification**: Emphasis is on the concept rather than cellular mechanisms of homeostasis. Examples of investigations could include heart rate response to exercise, stomata response

		to moisture and temperature, and root development in response to water levels.] [**Boundary:** Assessment does not include the cellular processes involved in the feedback mechanism.]
Sub-topic: Reproduction		
L4.2: Reproduction is essential to the continued existence of every kind of organism. Plants and animals have distinct and diverse life cycles. [**Boundary:** Assessment of plant life cycles is limited to those of flowering plants. Assessment does not include details of human reproduction.]	L8.3: Organisms reproduce, using a variety of structures and processes (both sexual and asexual) and transfer their genetic information to their offspring. [**Clarification:** Emphasis is on using models such as Punnett squares, diagrams, and simulations to describe the cause-and-effect relationship of gene transmission from parent(s) to offspring and resulting genetic variation. Examples of processes could include characteristic behaviors in animals. Examples of structures could include structures in plants that attract animals.]	L12.4: In most multicellular organisms, an organism begins as a single cell (a fertilized egg), and then divides successively to produce many cells. Mitosis is the process that allows all cells to divide after a period of growth. This process starts with a parent cell copying its genetic material and passing identical genetic material to both cells that result from the division (the daughter cells). [**Clarification**: Emphasis is on both the development of a fertilized egg and the role of cell division in growth and repair of multicellular organisms.] [**Boundary:** Assessment does not include specific gene control mechanisms or knowing the steps of mitosis from memory.]
Sub-topic: Matter and Energy in Organisms		

L4.3: All animals need food, water, and air in order to live and grow. They obtain their food from their surroundings – from plants or from other animals. Plants need air, water, minerals (in the soil), and light to live and grow. [**Clarification**: Examples could include that animals need to take in food, but plants do not, the different kinds of food needed by different types of animals, the requirement of plants to have light, and that all living things need water.]	L8.4: Photosynthesizers, (including plants, algae, and many microorganisms) use the energy from light to make sugars (food) from carbon dioxide from the atmosphere and water through the process of photosynthesis, which also releases oxygen into the atmosphere. [**Clarification**: Emphasis is on tracing movement of matter and flow of energy. Assessment could pick plants, algae, or microorganisms.] [**Boundary**: Assessment does not include the biochemical mechanisms of photosynthesis.] L8.5: Within individual organisms, food moves through a series of chemical reactions in which it is broken down and rearranged to form new molecules, to support growth, or to release energy. [**Clarification**: Emphasis is on describing that molecules are broken apart and put back together and that in this process, energy is released.] [**Boundary**: Assessment does not include details of the chemical reactions for photosynthesis or respiration.]	L12.5: The process of photosynthesis converts light energy to stored chemical energy by converting carbon dioxide plus water into sugars plus released oxygen. [**Clarification**: Emphasis is on illustrating inputs and outputs of matter and the transfer and transformation of energy in photosynthesis by plants and other photosynthesizing organisms. Examples of models could include diagrams, chemical equations, and conceptual models.] [**Boundary**: Assessment does not include specific biochemical steps.] L12.6: The process of cellular respiration is a chemical process in which the bonds of food molecules and oxygen molecules are broken, and new compounds are formed that can transport energy. [**Clarification**: Emphasis is on the conceptual understanding of the inputs and outputs of the process of cellular respiration.] [**Boundary**: Assessment should not include identification of macromolecules or the steps or specific processes involved in chemical reactions.] L12.7: As a result of photosynthesis and

		cellular respiration, energy is transferred from one system of interacting molecules to another. Matter and energy are conserved in each change. This is true of all biological systems, from individual cells to ecosystems. [**Clarification**: Emphasis is on the matter and energy transfer between photosynthesis and cellular respiration.] [**Boundary**: Assessment should not include specific matter or energy transfer steps]
Big Idea: Ecosystems: Interactions, Energy, and Dynamics How and why do organisms interact with their environment and what are the effects of these interactions?		
Sub-Topic: Interdependent Relationships		
L4.4: Most animals can move from place to place on their own, but plants cannot, and often rely on animals for pollination or to move their seeds around. Different plants survive better in different settings because they have varied needs for water, minerals, and sunlight.	L8.6: In any ecosystem, organisms and populations with similar requirements for food, water, oxygen, or other resources may compete with each other for limited resources, access to which consequently constrains their growth and reproduction. L8.7: Predatory interactions may reduce the number of organisms or eliminate whole populations of organisms. Mutually beneficial interactions, in contrast, may	L12.8: Ecosystems have carrying capacities, which are limits to the numbers of organisms and populations they can support. Organisms would have the capacity to produce populations of great size were it not for the fact that environments and resources are finite. This fundamental tension affects the abundance (number of individuals) of species in any given ecosystem.

	become so interdependent that each organism requires the other for survival. Although the species involved in these competitive, predatory, and mutually beneficial interactions vary across ecosystems, the patterns of interactions of organisms with their environments, both living and nonliving, are shared.	[**Clarification**: These limits result from such factors as the availability of living and nonliving resources and from such challenges as predation, competition, and disease.]
Sub-Topic: Cycles of Matter and Energy Transfer		
L4.5: Organisms obtain the materials they need to grow and survive from the environment. Many of these materials come from organisms and are used again by other organisms.	L8.8: Food webs are models that demonstrate how matter and energy are transferred between producers, consumers, and decomposers as the three groups interact within an ecosystem. Transfers of matter into and out of the physical environment occur at every level. Decomposers recycle nutrients from dead plant or animal matter back to the soil in terrestrial environments or to the water in aquatic environments. The atoms that make up the organisms in an ecosystem are cycled repeatedly between the living and nonliving parts of the ecosystem. [**Clarification**: Emphasis is on the conservation of matter and flow of energy into and out of various ecosystems.]	L12.9: Photosynthesis and cellular respiration (including anaerobic processes) provide most of the energy for life processes. L12.10: Plants or algae form the lowest level of the food web. At each link upward in a food web, only a small fraction of the matter consumed at the lower level is transferred upward to produce growth and release energy in cellular respiration at the higher level. [**Clarification**: The cycle of matter and energy transfer can be between all producers, consumers, and decomposers.]

	[Boundary: Assessment should not include identification of microscopic organisms.]	
Sub-Topic: Ecosystem Dynamics, Functioning, and Resilience		
L4.6: When the environment changes in ways that affect a place's physical characteristics (such as geography, effects of fire), temperature, precipitation, or availability of resources, some organisms survive and reproduce, some move to new locations, some move into the transformed environment, and some die.	L8.9: Ecosystems are dynamic in nature; their characteristics can vary over time. Disruptions to any physical or biological component of an ecosystem can lead to shifts in all its populations, therefore helping or hurting the health of the ecosystem, including its biodiversity. [**Clarification**: Disruptions may include introduction or removal of species, natural or human-induced disturbances. Examples of ways ecosystem health could be measured include ecosystem services (cleaning water, air, cycling of nutrients) or continuity of food webs.] L8.10: Changes in biodiversity can influence the resources and ecosystem services that humans rely on. [**Clarification:** Biodiversity includes genetic variation within a species in addition to species variation in different habitats and ecosystem types (e.g., forests, grasslands, wetlands). Examples of humans' resources that can be	L12.11: A complex set of interactions within an ecosystem can keep its numbers and types of organisms relatively constant over long periods of time under stable conditions. Extreme fluctuations in conditions or the size of any population, however, can challenge the functioning of ecosystems in terms of resources and habitat availability. L12.12: Changes induced by human activity in the environment — such as habitat destruction, pollution, introduction of invasive species, overexploitation, and climate change—can disrupt an ecosystem, reduce biodiversity, and threaten the survival of some species. [**Clarification:** Examples of human activities could include overpopulation, overexploitation, habitat destruction, pollution, introduction of invasive species, climate change, urbanization, the building of dams, and dissemination of invasive species.]

	influenced by changes in biodiversity include food, energy, and medicine. Examples of ecosystem services that humans rely on could include water purification, nutrient recycling, prevention of soil erosion, and pollination.]	L12.13: Humans depend on the living world for the resources and other benefits provided by biodiversity. Changes in biodiversity can influence resources and ecosystem services that humans rely on. [**Clarification**: Examples of changes in biodiversity that can influence resources could include food, energy resources, and medicines. Examples of ecosystem services that humans rely on could include water purification, nutrient recycling, prevention of soil erosion, and pollination.]
Big Idea: Heredity: Inheritance and Variation of Traits How are the characteristics of one generation passed to the next? How can individuals of the same species and even siblings have different characteristics?		
Sub-Topic: Inheritance		
L4.7: Many characteristics of organisms are inherited from their parents. These inherited characteristics may result in variations in how they look and function. Other characteristics result from individuals' interactions with the environment. Many characteristics involve	L8.11: Genes are located in the chromosomes of cells, with each chromosome pair containing two variants of each of many distinct genes. Each distinct gene chiefly controls the production of specific proteins, which in turn affects the traits of the individual.	L12.14: Each chromosome consists of a single very long DNA molecule, and each gene on the chromosome is a particular region of that DNA. Genes contain the instructions to code for the formation of proteins that determine traits. Not all DNA codes for a protein; some segments of DNA are involved in regulatory or

both. [**Clarification**: Examples of the environment affecting a trait could include that normally tall plants grown with insufficient water are stunted and that a pet dog that is given too much food and little exercise may become overweight.] [**Boundary:** Assessment does not include genetic mechanisms of inheritance and prediction of traits.]	[**Clarification**: Emphasis is on conceptual understanding that changes in genes may result in making different proteins.] [**Boundary:** Assessment does not include specific changes at the molecular level or mechanisms for protein synthesis.]	structural functions, and some have no currently known function. [**Clarification**: Emphasis is on the molecular aspect of DNA and its broad range of functions.] [**Boundary:** Assessment does not include the phases of meiosis or the biochemical mechanism of specific steps in the DNA to protein process.]
Sub-Topic: Variation		
	L8.12: In sexually reproducing organisms, each parent contributes half of the genes acquired (at random) by the offspring. Individuals have two of each chromosome and hence two alleles of each gene, one acquired from each parent. These versions may be identical or may differ from each other. Variations of inherited traits between parent and offspring arise from the subset of chromosomes (and therefore genes) inherited. [**Clarification:** Emphasis is on conceptual understanding that changes in genetic material may result in making different proteins.]	L12.15: In sexual reproduction, chromosomes can sometimes swap sections during the process of meiosis, thereby creating new genetic combinations and thus more genetic variation. Although DNA replication is tightly regulated and remarkably accurate, errors do occur and result in mutations, which are also a source of genetic variation. Environmental factors can also cause mutations in genes, and viable mutations are inherited. [**Boundary**: Assessment does not include the steps of meiosis.] L12.16: Environmental factors affect

	[**Boundary:** Assessment does not include specific changes at the molecular level, mechanisms for protein synthesis, or specific types of mutations.] L8.13: In addition to variations that arise from sexual reproduction, genetic information can be altered because of mutations. Although it is rare, mutations may result in changes to the structure and function of proteins. Some changes are beneficial, others harmful, and some neutral to the organism.	expression of heritable traits and hence affect the probability of occurrences of traits in a population. [**Clarification**: Emphasis is on describing the probability of traits as it relates to genetic and environmental factors in the expression of traits.] [**Boundary:** Assessment does not include Hardy-Weinberg calculations, the phases of meiosis, or the biochemical mechanism of specific steps in the process.]
Big Idea: Biological Evolution: Unity and Diversity How can there be so many similarities among organisms yet so many different kinds of plants, animals, and microorganisms? How does biodiversity affect humans?		
Sub-topic: Evidence of Common Ancestry and Diversity		
L4.8: Some kinds of plants and animals that once lived on Earth are no longer found anywhere. Fossils can provide evidence about these types of organisms that lived long ago and about the nature of their environments. [**Clarification:** Examples of evidence could include type, size, and distributions of fossil organisms. Examples of fossils	L8.14: The collection of fossils and their placement in chronological order (the fossil record) documents the existence, diversity, extinction, and change of many life-forms throughout the history of life on Earth. Similarities and differences in gross anatomical appearance or in embryological development, between organisms living today and between them	L12.17: Genetic information provides evidence of evolution. DNA sequences vary among species, but there are many overlaps. Such information is also derivable from the similarities and differences in amino acid sequences and from anatomical and embryological evidence. [**Clarification:** Emphasis is on a

and environments could include marine fossils found on dry land, tropical plant fossils found in Arctic areas, and fossils of extinct organisms.] [**Boundary:** Assessment does not include identification of specific fossils or present plants and animals. Assessment is limited to major fossil types and relative ages.]	and organisms in the fossil record, enable the reconstruction of evolutionary history and inference of lines of evolutionary descent. [**Clarification:** Emphasis is on patterns of changes in the level of complexity of anatomical structures in organisms and the chronological order of fossil appearance in the rock layers.] [Boundary: Assessment does not include the names of individual species or geological eras in the fossil record.]	conceptual understanding of the role each line of evidence has relating to common ancestry and biological evolution. Examples of evidence could include similarities in DNA sequences, anatomical structures, and order of appearance of structures in embryological development.] [**Boundary:** Assessment does not require familiarity with the details of specific technologies or organisms.]
Sub-topic: Mechanisms of Change		
L4.9: Species change over time. Sometimes the differences in characteristics between individuals of the same species provide advantages in surviving, finding mates, and reproducing. This can be especially true when a habitat changes. [**Clarification:** Examples could include that plants that have larger thorns than other plants may be less likely to be eaten by predators and that animals that have better camouflage coloration than other animals do may be more likely to survive and therefore more likely to produce	L8.15: Adaptation by natural selection acting over generations is one important process by which species change over time in response to changes in environmental conditions. Heritable traits that support successful survival and reproduction in the new environment become more common; those that do not become less common. Thus, the distribution of traits in a population changes. This can also be done artificially by humans selectively breeding for a desired trait in other organisms. [**Clarification:** Emphasis is on a change	L12.18: Evolution by natural selection results from the interaction of four factors: (a) the potential for a species to increase in number, (b) the genetic variation of individuals in a species due to mutation and sexual reproduction, (c) competition for an environment's limited supply of the resources that individuals need in order to survive and reproduce, and (d) the ensuing proliferation of those organisms that are better able to survive and reproduce in that environment, passing on those traits to offspring. Fitness, as measured by survival and reproduction

offspring. Examples of environmental changes could include changes in land characteristics, water distribution, temperature, food, and other organisms.]	in environmental conditions leading to an advantage of heritable traits for survival and reproduction, and thus a change in traits over time. Other examples include the influence of humans on genetic outcomes in other organisms through artificial selection, such as genetic modification, animal husbandry, or gene therapy. Examples could include breeds of dogs, horses, and cattle or development of modern maize from teosinte.] [**Boundary:** Assessment requires evidence of heritability but not precise gene-to-protein mechanism. Assessment does not include Hardy-Weinberg calculations.]	rates, may be altered if changes in the physical environment, whether naturally occurring or human induced, take place. [**Clarification:** Emphasis is on the influence each of the four factors has on the number of organisms, behaviors, morphology, or physiology in terms of ability to compete for limited resources and subsequent survival of individuals and adaptation of species. Examples could include mathematical models such as simple distribution graphs and proportional reasoning. For changes in the environment, the emphasis is on determining relationships for how changes to the environment such as deforestation, fishing, application of fertilizers, drought, flood, and the rate of change of the environment affect distribution or disappearance of traits in species.] [**Boundary:** Assessment does not include other mechanisms of evolution, such as genetic drift, gene flow through migration, and coevolution.]

Disciplinary Concepts in the Earth and Space Sciences

Universe, Solar System, and Earth

- Patterns of motion of space objects
- Solar System
- Formation of the universe

Earth's Systems

- Plate tectonics, patterns on the surface of the Earth
- Earth's history
- Water cycling, weathering, and erosion
- Weather and Climate

Earth and Human Activity

- Natural Resources
- Natural Hazards
- Human Impacts on Earth Systems
- Climate Change

NAEP Earth and Space Sciences Disciplinary Concepts		
Grade 4	**Grade 8**	**Grade 12**
Big Idea: Universe, Solar System, and Earth How do we explain Earth's relationship to objects in space?		
Sub-topic: Patterns of Motion of Space Objects		
E4.1: Many objects in the sky change position and are not always visible due to Earth's rotation. The patterns of motion of the sun and moon can be observed, measured, described, and predicted. [**Clarification**: Emphasis is on the rise/set patterns of the sun and moon, and the day/night cycle.] [**Boundary:** Assessment should focus on patterns of one space object at a time.]	E8.1: The orbits of Earth around the sun and of the moon around Earth, together with the rotation of Earth on an axis that runs from its north pole to its south pole, cause observable and measurable patterns that can be used to predict apparent motions of the sun and moon and occurrence of tides and seasonal changes through models. [**Clarification**: Emphasis is on model representation of patterns of seasons, moon phases, and eclipses. Examples of models can be physical (e.g., globes), graphical (e.g., charts), or conceptual (e.g., analogies).] [**Boundary**: Assessment is limited to the patterns of motion used to explain seasons, tides, and phases of the moon;	E12.1: Cyclical changes in the shape of Earth's orbit, together with changes in the orientation of the planet's axis of rotation— occurring from tens of years to hundreds of thousands of years—have altered the intensity and distribution of sunlight falling on Earth. This variation drives changes in Earth's climate patterns over time. [**Clarification**: Emphasis is on patterns of cyclical changes in global temperatures as indicated by either direct measurements or from proxies for Earth's climate such as fossils, ice, or sea-floor cores.] [**Boundary**: Items should not assess knowledge of any specific episode of time or change in paleoclimate.]

	assessment does not include the names or shape of specific constellations.]	
Sub-topic: Solar System		
E4.2: Some objects in the solar system can be seen with the naked eye, and some require tools that extend human perception. [**Clarification**: Emphasis is on the collection of data using Earth-based instruments, space-based telescopes, and spacecraft to determine similarities and differences among solar system objects.] [**Boundary:** Assessment does not include recalling facts about properties of the planets and other solar system bodies.]	E8.2: The solar system consists of the sun and a collection of objects, including planets, their moons, and asteroid belts in orbit around the sun. Gravitational interactions between the Sun and planets in the Solar system produce orbital patterns that can be observed and predicted. [**Boundary**: Assessment does not include Kepler's Laws or retrograde motion.]	E12.2: Orbiting objects can be described in terms of their elliptical paths around the sun as described by Kepler's laws. These orbits can change slightly due to gravitational effects from, or collisions with, other objects in the solar system. [**Clarification:** Emphasis is on Newtonian gravitational laws governing orbital motions and Keplerian laws for orbital period and shape, which apply to human-made satellites as well as planets and moons.] [**Boundary:** Newtonian gravitational laws should be limited to the force of attraction between two masses. Kepler's laws with respect to planetary motion are limited to the general proportionality of distance from the Sun and orbital period. Assessment does not include the mathematics of ellipses and square/cube ratios.]
Sub-topic: Formation of the Universe		

E4.3: We can observe objects in the sky such as the moon, sun, other planets, and other stars. The sun is a star that appears larger and brighter than other stars because it is closer. [**Boundary**: Assessment is limited to relative distances, not other factors affecting brightness.] E4.4: Unlike stars, the moon and other planets do not make their own light but reflect light from the sun so we can see them from Earth.	E8.3: The sun and its solar system are a small piece of a large group of stars called the Milky Way, which is only one of many such galaxies spread out in the Universe. Scientific instruments collect and provide information about space objects to understand how they formed, became distributed, and evolved. [**Clarification**: Emphasis is on the large range of scales (powers of ten) and the enormity and dynamic nature of the Universe.] [**Boundary**: Assessment does not include recall of numbers about scales or any details of particular structures.]	E12.3: The study of stars' light spectra and relative brightness is used to identify compositional elements of stars, their movements, and their distances from Earth. This is used to develop explanations of the formation, age, and change over time of the Universe. [**Clarification**: Emphasis is on how changes in those stellar spectra, such as red-shift and blue-shift, provide evidence for the expansion of the Universe since the Big Bang.] [**Boundary**: Assessment is limited to evidence provided by stellar spectra and not on cosmological features immediately following the Big Bang.]
Big Idea: Earth's Systems What are Earth's systems and how do they change?		
Sub-topic: Plate tectonics, patterns on the surface of the Earth		
E4.5: Locations of local, regional, and global surface features and phenomena reveal patterns on Earth's surface. [**Clarification**: Emphasis is on patterns including the locations of mountain	E8.4: The Earth consists of layers, including a solid, rigid outer layer divided into plates, which are always moving very slowly. Interactions between Earth's moving plates result in changes of	E12.4: The transfer of thermal energy from the Earth's interior, generated from radioactive decay, toward the surface, along with the gravitational movement of denser materials back toward the interior, drives the flow of matter inside the Earth.

ranges, earthquakes, and volcanoes.] [**Boundary**: Assessment should focus on observable features and phenomena on the surface of the Earth.]	physical features. [**Clarification**: Emphasis is on changes resulting from interactions that can be chemical and/or physical; emphasis is on changes due to long-term processes, such as plate motion, creating mountain ranges, ocean trenches, and shaping of coastlines.] [**Boundary**: Assessment focuses on what is causing physical features to form (plate motion) rather than recalling specific vocabulary terms for features or plate motion.]	This convection cycle moves Earth's plates and causes the patterns of physical features. [**Clarification**: Emphasis is on radioactive decay as the source of thermal energy and convection as the mechanism driving plate motion and the resulting patterns in features.] [**Boundary**: Assessments should be limited to the patterns in landforms and rock materials, not on recalling specific mineral formation or process that formed any specific rock type or landform.]
Sub-topic: Earth's History		
E4.6: Earth and life on Earth have changed over time. The occurrence and location of certain fossil types provide evidence for changes in environmental conditions and the development of life over time. [**Clarification:** Emphasis is on the relationship between the fossil and the rock it formed in, i.e., the rock is the same age as the fossil.] [**Boundary:** Assessment does not include processes of fossil formation, or	E8.5: The geologic time scale interpreted from fossils and the sequence of rock strata provides a way to reconstruct how and when major events in Earth's history in terms of relative time. [**Clarification:** Emphasis is on how evidence from rock strata is used to piece together major geological events in Earth's history.] [**Boundary:** Assessment does not include recalling the names of specific periods or epochs and events within them.]	E12.5: The decay of radioactive isotopes in minerals and rocks provides a measurement for dating rock formations and for providing evidence for Earth's formation and early history. [**Clarification:** Emphasis is on evidence that includes absolute ages obtained by radiometric dating of meteorites, moon rocks, and minerals.] [**Boundary**: Calculations of age dates should be limited to geometric progressions (doubling/halving, such as in half-lives).]

knowledge of specific fossil organisms. The expression of time is limited to sequences that represent relative time.]		
Sub-topic: Water Cycling, Weathering, and Erosion		
E4.7: Water is found in oceans, rivers, lakes, and air. The downhill movement of water drives the flow of water toward the ocean. [**Clarification**: Emphasis on water includes liquid (rain, river, ocean, clouds), solid (snow and ice), gas (vapor).] E4.8 Rocks on Earth's surface can be broken into pieces and moved by water, wind, and living organisms; this causes continual, observable changes to surface features. [**Clarification**: Emphasis is on phenomena highlighting observable short-term physical changes to features, such as a change in the course of a stream, the change in the shape of a beach, or a rock slide.]	E8.6: The movement of water within the water cycle is a function of phase changes - evaporation, condensation, freezing, and melting. E8.7: Water continually cycles within and among land, ocean, and atmosphere. Water's movements, both on the land and underground, are driven by gravity and change the land on and below Earth's surface. [**Clarification**: Emphasis is on the relationship between specific elements of the water cycle and their function in changing the shape of the land. Changes to the surface can include eroding of high regions, depositing sediment in low regions, formation of caves, or landslide movements of Earth materials.] [**Boundary**: Assessment should not require quantitative understanding of the phase changes, including latent heats of vaporization and fusion.]	E12.6: Interactions between the hydrosphere and the geosphere are influenced by water's unique properties, including its exceptional capacity to absorb, store, and release large amounts of thermal energy; expand upon freezing; dissolve and transport materials; separate different chemical elements, and change the properties of rocks. [**Clarification**: Emphasis is on stream transportation and deposition, surface weathering due to water expansion while freezing, ice sheets scouring the surface of the land beneath them, creating massive amounts of sediment; and chemical weathering of minerals in rocks.]

Sub-topic: Weather and Climate		
E4.9: Patterns in when and where weather conditions occur can be used to make predictions about the kind of weather that can be expected in a region. [**Clarification**: Emphasis is on the record of temperature and precipitation over time, among other measurements, that provides the basis for weather patterns.] [**Boundary**: Assessments should emphasize general seasonal patterns of temperature and precipitation and avoid specific conditions based on elevation or geographic location. Assessment does not include climate change.]	E8.8: Weather is influenced by interactions involving sunlight, the ocean, the atmosphere, ice, and landforms. Because the interactions are so complex, weather patterns in a given location can only be predicted through probabilities (likelihood to occur), and only for a short period of time into the future. [**Boundary**: Assessment should include the use of multiple forms of evidence but does not include recalling the names of atmospheric phenomena (e.g., cloud types or types of weather fronts).] E.8.9: Influences on the climate at a given place include latitude, altitude, local and regional geography, and oceanic and atmospheric flow patterns. [**Clarification**: Emphasis is on how multiple variables influence the climate in a given region. Focus is on only one region.] [**Boundary**: Assessment should include multiple pieces of data but does not include specific dynamics of the Coriolis effect.]	E12.7: The absorption, reflection, storage, and redistribution of visible and infrared energy from the Sun among the atmosphere, hydrosphere, and geosphere, and the reradiation of infrared energy into space, lead to the geographic and temporal patterns in Earth's climate. [**Clarification**: Emphasis is on the mechanisms by which ocean and atmospheric circulations exert a major influence on weather and climate on a global scale.] E12.8: Geological and historical evidence indicates changes in past climates are linked to alterations in the composition of atmosphere and variations in solar output or Earth's orbit. The time scales of these changes vary from sudden - few tens of years (e.g., large volcanic eruptions or changes in ocean circulation), to gradual - millions of years (e.g., movement of Earth's plates). [**Clarification**: Emphasis is on the varied time scales of the changes in climate throughout Earth's history.] [**Boundary**: Assessment of the results of

		changes in climate is limited to changes in surface temperatures, precipitation patterns, glacial ice volumes, sea levels, and biosphere distribution.]
Big Idea: Earth and Human Activity How do Earth's system processes and human activities affect each other?		
Sub-topic: Natural Resources		
E4.10: Humans depend on natural resources because all living things need water, air, and resources for food, transportation, and shelter, which influences where they live. [**Clarification**: Emphasis is on the various ways humans depend on natural resources.] [**Boundary**: Assessments should be limited to resources that are not specific to a geographic region and should not focus on knowledge of one specific resource.]	E8.10: Natural resources are distributed unevenly by biogeochemical processes around the planet as a result of Earth systems processes. Humans depend on the Earth's geosphere, hydrosphere, atmosphere, and biosphere for resources, both renewable and nonrenewable, within human life spans. [**Clarification**: Emphasis is on the interrelated nature between humans' dependence on natural resources, their distribution, and the way in which they are formed.] [**Boundary**: Assessments should not require knowledge of any specific biogeochemical process to explain how resources are formed.]	E12.9: Resource availability guides the development of human societies. All forms of energy production and resource extractions have associated economic, social, and environmental cost/benefit factors. [**Clarification**: Emphasis is on the relationship between the resources (e.g., sand, lumber, minerals) that humans use and their impacts.] [**Boundary**: Assessment should not require specialized knowledge of specific resources. Assessments should require students to consider multiple impacts from any particular resource.]

Sub-topic: Natural Hazards		
E4.11: Natural hazards are caused by natural processes. Depending on where one lives, some kinds of natural hazards are more likely than others. [**Clarification**: Emphasis is on patterns of observable phenomena in terms of where natural hazards occur.] [**Boundary**: Assessments do not include knowledge of processes that cause any specific hazard to occur.]	E8.11: Some natural hazards are typically preceded by observable phenomena, which provide a warning for their occurrence (e.g., volcanic eruptions and severe weather). Other hazards occur suddenly and often with very little or no advance warning (e.g., earthquakes and tornadoes). Data on the duration and frequency of the warning signs reveal patterns of natural hazards in a region, which can help forecast the locations and likelihoods of future events in order to minimize risks. [**Clarification**: Emphasis is on how humans use data collected on hazards over long time periods to minimize risk from future events.] [**Boundary**: Assessment should be limited to types of natural phenomena that negatively impact humans, and how the recognition of patterns of past events (landslides, high water marks, weather maps) can be used as a future predictor of such events.]	E12.10: Land use and city planning can affect the frequency and intensity of the impacts of some natural hazards; some have significantly altered the size and location of human populations. [**Clarification**: Emphasis is on how human activities (environmental, agriculture, infrastructure, etc.), have affected and been affected by the occurrences of natural hazards.]
Sub-topic: Human Impacts on Earth Systems		

E4.12: Human activities cause changes to the local areas where they live. Human choices can increase or decrease the positive and negative impacts on the land, water, and air. [**Clarification**: Emphasis is on the ways humans' actions can influence their impacts on the environment around them. Human activities should emphasize actions students can take at home, at school or in their local community.] [**Boundary**: An assessment should focus on a positive or a negative impact on the environment.]	E8.12: Human activities have significantly altered the biosphere, atmosphere, and geosphere, sometimes damaging or destroying ecosystems and causing the extinction of organisms. Human choices can minimize harm to other organisms and risks to the health of the regional environment. [**Clarification**: Emphasis is on the regional impacts, including examining how activities in one part of a region (e.g., upstream in a watershed) can impact another area of the same region (e.g., downstream in the watershed).] [**Boundary**: Data should be regional, not global, and focus on positive and/or negative effects to either air, water, OR land.]	E12.11: When the sources of an environmental problem are understood, applying engineering and design solutions, new technology, and other creative ideas can mitigate negative impacts on Earth's resources and global environment, while inaction on the problem could magnify the negative impacts. When the sources of such problems are not well understood, some actions could magnify the problems. [**Clarification**: Emphasis is on the sustainability of human societies and the biodiversity that supports them, which requires responsible monitoring and management of natural resources.] [**Boundary**: Assessment should address the relationship between no more than two global regions (continental scale).]
Sub-topic: Climate Change		
	E8.13: Human activities that release greenhouse gasses, such as production and combustion of fossil fuels, are major factors in the current rise in Earth's temperature. Monitoring the production and reducing the use of fossil fuels can slow the increase in global temperatures	E12.12: Current models predict that, although future regional climate changes will be complex and varied, average global temperatures will continue to rise. Changing the outcomes predicted by global climate models strongly depends on reduction of the amounts of human-generated

	as well as the effects of climate change. [**Clarification**: Emphasis is on the role that individual and industrial activities have on the rise of global temperatures and the various ways they affect life on Earth. Greenhouse gasses include methane and CO_2.] [**Boundary**: Assessment can include the analysis of data but should not require the analysis of climate models.]	greenhouse gasses added to the atmosphere each year, but are also influenced by uncertainties about behavioral, economic, and political factors and how they will impact potential solutions and their success. [**Clarification**: Emphasis is on the use of climate models to explain and predict how changes in human activities could impact the ocean, atmosphere, and biosphere, while also accounting for the uncertainty of human societal factors.] [**Boundary**: Assessment should be limited to the analysis of one climate model and its impacts.]

2B. NAEP Science and Engineering Practices

> *Engaging in the practices of science helps students understand how scientific knowledge develops; such direct involvement gives them an appreciation of the wide range of approaches that are used to investigate, model, and explain the world. Engaging in the practices of engineering likewise helps students understand the work of engineers, as well as the links between engineering and science. Participation in these practices also helps students form an understanding of the crosscutting concepts and disciplinary ideas of science and engineering; moreover, it makes students' knowledge more meaningful and embeds it more deeply into their worldview. (NRC 2012, p. 42)*

The ability to engage in the practices of science and engineering (along with the ability to use crosscutting concepts) allows students to apply their disciplinary science knowledge as they develop explanations of phenomena or design solutions to engineering problems. The 2028 NAEP Science Assessment will ask students to engage these abilities as part of achieving a successful response to multidimensional items.

The goal of science is the construction of theories that can provide explanatory accounts of features of the world. A theory becomes accepted when it has been shown to be superior to other explanations in the breadth of phenomena and conditions it accounts for and in its explanatory coherence and parsimony. Scientific explanations are explicit applications of theory to a specific situation or phenomenon, perhaps with the intermediary of a theory-based model for the system under study. The goal for students is to construct logically coherent explanations of phenomena that incorporate their current understanding of science, or a model that represents it, and are consistent with the available evidence.

In engineering, the goal is a designed solution to a problem rather than an explanation. The term *engineering* applies to any such design, whether it is for an object, a system, or a process. The domain of the problem can be any area of applied science or technology. The problem can arise from individual, community, or global needs or wants. The process of developing a design is iterative and systematic, as is the process of developing an explanation or a theory in science. Engineers' activities, however, have elements that are distinct from those of scientists. These elements include specifying constraints and criteria for desired qualities of the solution, developing a design plan, producing and testing models or prototypes, selecting among alternative design features to optimize the achievement of design criteria, and refining design ideas based on the performance of a prototype or simulation. Engineering design focuses on the appropriate use of technology. 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 practices of science and engineering occur in social contexts and can be used in ways that can benefit or harm individuals, communities, or the environment. As students use these practices, they should do so ethically and recognize the risks and harms that can be caused and have been caused by negligent use, as well as the benefits.

The ability to engage in the practices of science and engineering (along with the ability to use crosscutting concepts) allows students to apply their disciplinary science knowledge as they develop explanations of phenomena or design solutions to engineering problems. The NAEP Science Assessment will ask students to engage these abilities as part of achieving a successful response to multidimensional items. The elements of the practices listed in Exhibits 2.3 - 2.10 focus on aspects of engaging in the practices that can be assessed in large-scale science- and engineering-oriented assessments and do not include all aspects needed in instruction. The bulleted sub-statements in the charts under each science and engineering practice are individual elements that pull out aspects of each practice that might be assessed at this grade level but not every sub-statement needs to be assessed by NAEP.

The practices for the NAEP Science Assessment are as follows:

- **Asking Questions and Defining Problems**
- **Planning and Carrying Out Investigations**
- **Analyzing and Interpreting Data**
- **Using Mathematics and Computational Thinking**
- **Developing and Using Models**
- **Constructing Explanations and Designing Solutions**
- **Engaging in Argument from Evidence**
- **Obtaining, Evaluating, and Communicating Information**

Scientific work involves all these practices used in an iterative and recursive process to achieve an eventual explanation based on a well-tested model. It further requires honest reporting and critical review to be effective. Engineering work likewise involves an iterative and recursive process using all these practices to design, test, and redesign to achieve a successful problem solution. Individuals and organizations working within science and engineering also consider how their work contributes to ecological and social matters and how to optimize their work, products, and applications to benefit society and minimize harms, including consideration of unintended negative effects. Scientific theories and explanations are empirically based and subject to revision based on new or evolving evidence. The same applies to student-developed models, explanations, and engineering design solutions. Students should be able to reflect on what needs to be revised and whether additional evidence is required to improve the outcome or strengthen the claim.

For NAEP Science Assessment purposes, the science and engineering practices have been organized into four categories: Investigating, Analyzing, Explaining, and Evaluating (Exhibit 2.2). On any given assessment, at least 10% of the items must fall into each of the four categories.

Exhibit 2.2. NAEP Science and Engineering Practices

Investigating	Asking Questions and Defining Problems
	Planning and Carrying Out Investigations
Analyzing	Analyzing and Interpreting Data
	Using Mathematics and Computational Thinking
Explaining	Developing and Using Models
	Constructing Explanations and Designing Solutions
Evaluating	Engaging in Argument from Evidence
	Obtaining, Evaluating, and Communicating Information

Asking Questions and Defining Problems

Scientific questions arise in a variety of ways. They can be driven by curiosity about the world; inspired by the predictions of a model, theory, or findings from previous investigations; or driven by the need to solve a problem. Scientific questions are distinguished from other types of questions in that the answers lie in explanations supported by empirical evidence. Engineering design work also begins with asking questions to help define a problem to solve.

Many aspects of asking questions do not lend themselves to assessment. The aspects of questioning listed below are those that can reasonably be the practice element of a science or engineering assessment item.

Exhibit 2.3. Asking Questions and Defining Problems

Grade 4	**Grade 8**	**Grade 12**
Ask questions to inform an investigation or develop an explanation or model of phenomena.		

Grade 4	Grade 8	Grade 12
• Ask questions to help refine observations, develop interpretations of data, develop and/or evaluate models, or define an engineering problem. [**Boundary**: Statistical displays are limited to bar graphs and pictographs for categorical data, and line plots for measurement data (whole number measurements only).] • Ask "what if" questions about a system or phenomenon being observed that could be investigated empirically.	• Ask questions to clarify and/or refine an observation, model, or explanation of phenomena; or to clarify and/or refine an engineering problem. • Ask questions that can be answered with empirical evidence to investigate relationships between variables in a system model or in phenomena.	• Ask questions that arise from examining a model, an explanation, or a design plan to clarify and/or identify additional needed information or tests. • Ask investigable questions to determine relationships, including quantitative relationships, between independent and dependent variables in a model, and when appropriate frame a hypothesis about potential findings.
Ask questions as part of understanding, evaluating, and/or challenging the work of others.		
• Ask questions to clarify an argument or interpretation of a data set.	• Ask questions to clarify or respectfully challenge the evidence and/or the premise(s) of an argument or interpretation of a data set.	• Ask and/or evaluate questions that challenge the premise(s) of an argument, the interpretation of a data set, or the suitability of design considerations.
Define a design problem that addresses a need		

Grade 4	Grade 8	Grade 12
• Define a design problem to provide a solution for a situation people want to change that can be solved through the development of a new or improved object or tool.	• Define a design problem that considers relevant scientific principles and potential impacts on people and the natural environment that may limit possible solutions and can be solved through the development of an object, tool, process, or system that includes multiple criteria and constraints.	• Define a design problem that involves the development of a process or system with interacting components and criteria and constraints that may include social, technical, ethical, and/or environmental considerations.

Planning and Carrying Out Investigations

Scientific investigations may be undertaken to describe a phenomenon and to test a theory or model for how the world works. The purpose of engineering investigations might be to determine conditions under which the design solution needs to function, to find out how to fix or improve the functioning of a technological system, or to compare different solutions to see which best solves a problem. Whether students are doing science or engineering, it is always important for them to state the goal of an investigation, predict outcomes, and plan a course of action that will provide the best evidence to support their conclusions or design solutions. Students should design investigations that generate data to provide evidence to support claims they make about phenomena. Students should build engineering investigations that address the criteria and constraints.

Over time, students are expected to become more systematic and careful in their designing methods, including the selection of instruments and tools to analyze data. In laboratory experiments, students are expected to decide which variables should be treated as results or outputs, which should be treated as inputs and intentionally varied from trial to trial, and which should be controlled, or kept the same across trials. In the case of field observations, planning involves deciding how to collect different samples of data under different conditions, even though not all conditions are under the direct control of the investigator. In engineering investigations, students should become more adept at using tools, materials, and processes as well as taking into consideration constraints and criteria.

The NAEP Science Assessment will not require students to carry out experiments with physical equipment, but simulations or virtual laboratories could be made available for some items.

Exhibit 2.4. Planning and Carrying Out Investigations

Grade 4	**Grade 8**	**Grade 12**
Developing or revising an investigation plan		
• Plan an investigation to explore a scientific question or design problem taking into consideration appropriate variables and tests.	• Evaluate and/or revise an experimental design that can produce data to serve as the basis for evidence that meets the goals of the investigation or design problem.	• Plan an investigation that will produce data to serve as the basis for evidence as part of building and revising models, supporting explanations for phenomena, or testing solutions to problems. Consider possible confounding variables or effects and evaluate the investigation's design to ensure appropriate variables are controlled.
Selecting and evaluating appropriate tools for an investigation		
• Select methods and/or tools for collecting data.	• Select and evaluate tools to collect, record, and analyze data.	• Select and evaluate appropriate tools to collect, record, analyze, synthesize, and evaluate data.
Predicting expected outcomes		

• Make predictions about what would happen if a variable changes. • Predict the outcome of an experiment, or a design solution based on a model, a phenomenon, or on a design plan.	• Predict the change in a dependent variable when a change in an independent variable occurs. • Predict the outcome of an investigation or test of a design plan and provide arguments that support that prediction.	• Predict the direction and magnitude of change of a dependent variable for a change in the independent variable and provide rationale to support the prediction. • Predict the outcome of an investigation or test of a design plan and support that prediction with an argument including evidence from models, evidence from prior experiments, and/or the application of science knowledge to support the prediction.

Analyzing and Interpreting Data

Data must be organized, analyzed, and interpreted to serve as the evidence to support claims. In the data-rich world of today, this work has become a discipline called "data science." Students, like scientists and engineers, use a range of tools to display and analyze data and to identify patterns, sources of error, and degrees of certainty in sets of data. They organize and analyze data to test model-based predictions, to infer relationships and trends in a system, to provide evidence for claims and arguments, to support or refute hypotheses or explanations, or to compare different solutions to specific design criteria and determine which design best solves the problem within given constraints.

Exhibit 2.5. Analyzing and Interpreting Data

Grade 4	Grade 8	Grade 12
Displaying data to observe patterns and relationships		

• Represent data in tables and/or various graphical displays (e.g., bar graphs and pictographs) to provide information or visualize relationships that can help to explain phenomena or solve design problems.	• Construct, analyze, and/or interpret graphical displays of data and/or large data sets from an investigation (e.g., maps, charts, graphs, and/or tables) to identify relationships between variables (linear vs. nonlinear relationships, causal vs. correlational relationships, and temporal and spatial relationships).	• Construct, analyze and/or interpret representations of small and large data sets from an investigation using tools, technologies, and/or models (e.g., computational, mathematical), including statistical analysis (descriptive statistics) and probability.
Analyzing data to support or reject claims about phenomena or improve design solutions		
• Analyze data to determine whether it supports or refutes a claim about a phenomenon or design solution. • Analyze data from tests of two objects designed to solve the same problem to compare the strengths and weaknesses of how each performs.	• Analyze data to provide evidence to support or reject a model or explanation or use to improve a design solution.	• Analyze data to provide evidence to support or reject a model or explanation or use to optimize a design solution relative to criteria for success.
Evaluating the quality and adequacy of data		
	• Evaluate the limitations of the data for the intended use, considering factors such as quantity and	• Evaluate whether the data are sufficient in quantity, accuracy and reliability for the purpose intended and

	quality of the data, the tools used to obtain it, and its presentation.	suggest needed improvements.

Using Mathematics and Computational Thinking

Both science and engineering require mathematics and information technology. Students apply their understanding of mathematics in science and engineering contexts. It is also in these contexts that they are expected to manipulate quantities with physical units, not just pure numbers.

This practice links to student assessment of mathematics and ability to use computational tools, and the progression of expectations across grade levels is therefore closely aligned with the mathematics expected at each grade level. The item demands for students using this practice should not exceed NAEP mathematics assessment expectations at the same grade level. Items should not be purely a mathematical or computational item that can readily be completed without demonstrating any understanding of the disciplinary content of the item. Additional guidance about mathematics is provided in the NAEP Science Assessment Item Specifications.

Exhibit 2.6. Using Mathematics and Computational Thinking

Grade 4	Grade 8	Grade 12
Using Mathematics		
• Apply simple mathematical concepts and/or processes (addition, subtraction, measurement, and simple computation) to a scientific question or a design problem. [**Boundary**: Computations involve whole numbers only and do not include converting units of measure.]	• Apply mathematical concepts and/or processes (such as ratio, rate, percent, basic operations, and simple computations) to scientific questions and/or design problems. • Interpret and use quantities involving ratios based on two different types of units of measure (e.g., speed, density, and population density).	• Apply mathematical techniques (such as functions, statistical reasoning, and computational algorithms) to represent and solve scientific questions and/or design problems. • Interpret and apply ratios, rates, percentages, and unit conversions in the context of complicated measurement problems involving

64

		quantities with derived or compound units (such as mg/mL, kg/m^3, etc.).
Computational Thinking		
• Break a process into a series of steps.	• Use algorithms (a series of ordered steps) to solve a design problem. • Apply digital tools and/or mathematical concepts and arguments to test and compare proposed solutions to design problems.	• Apply or revise algorithms when analyzing data or designing, programming, testing, and revising scientific models, explanations, and design solutions. • Apply mathematical expressions, computer programs, algorithms, or simulations of a process or system to evaluate the model by comparing the outcomes with what is known about the phenomena or design problem.

Developing and Using Models

A practice of both science and engineering is to use and construct models as helpful tools for representing ideas and explanations. These tools include diagrams, drawings, physical replicas, mathematical representations, analogies, and computer simulations. Scientists use the term *model* for all these, whereas engineers may talk of a design plan for a diagrammatic representation of a system or a prototype for scaled physical replica. In science, models are used to develop questions and predictions and are repeatedly tested and revised until they can provide successful predictions for tests. They then form the basis of an explanation of the phenomenon of interest. They are likewise a key part of the process of engineering design and of troubleshooting to analyze and identify flaws in designed systems.

Students are expected to develop, test, critique, and apply models as a core feature of their science and engineering assessment. They use models to express, examine, and refine their thinking and support their arguments for a claimed explanation.

While the full cycle of developing a model takes too much time to be included as an assessment item the phrase "develop a model" is included in the elements described below to cover inclusion of items that ask students to carry out some part of the work of model development.

Exhibit 2.7. Developing and Using Models

Grade 4	Grade 8	Grade 12
Developing and using models to explain phenomena or design a solution.		
• Develop, use, and/or revise a model to describe and explain a phenomenon or describe a design proposal. • Identify and describe how the parts of a model and the relationships between them represent a phenomenon.	• Develop, use, and/or revise a model to describe, explain, and/or predict phenomena by identifying relationships among parts and or quantities in a system, including both visible and invisible quantities. • Use a model to test ideas about phenomena in natural systems or proposed design solutions.	• Develop, use, and/or revise a model that includes mathematical relationships (including both visible and invisible quantities) to describe, explain, and/or predict phenomena or to test a proposed design solution.
Identifying and addressing limitations of models		
• Identify limitations of a model for a phenomenon in terms of what the model can or cannot yet explain.	• Evaluate limitations of a model for a phenomenon and propose revisions to address what the model cannot yet explain.	• Evaluate merits and limitations of two different models of the same proposed tool, process, mechanism, or system to select or revise a model that best fits the evidence or design criteria.

Constructing Explanations and Designing Solutions

While students at the K–12 level are not expected to develop new scientific theories, they are expected to apply scientific knowledge appropriate to their grade level to explain phenomena or to develop designs that offer a solution to a real-world problem. The process involves a cycle of all eight science and engineering practices, which together allow students to achieve this goal. The elements listed for this practice are those specific to completing the cycle to achieve and present the desired explanation or design.

Each suggested explanation is a claim that must be supported with an argument based on evidence (see the following practice). In science, the argument is most often model-based, and the evidence enters in the process of testing and revising the model. Likewise, engineering design proposals must be supported by tests of the design through prototypes or simulations of the proposed design.

Exhibit 2.8. Constructing Explanations and Designing Solutions

Grade 4	Grade 8	Grade 12
Data based explanations		
• Develop an evidence-based description or explanation supported by evidence and reasoning of a phenomenon or the action of a designed device.	• Construct or revise an explanation that uses a chain of cause and effect or evidence-based associations between factors to account for the qualitative or quantitative relationships between variables in a phenomenon.	• Construct or revise an explanation that uses a chain of cause and effect or evidence-based associations between factors to account for the qualitative or quantitative relationships between variables in a phenomenon.
Model-based explanations		
• Relate an explanation of a phenomenon to a model.	• Evaluate whether a model provides sufficient explanation of the phenomenon and how the model could be revised to	• Evaluate a model-based explanation or a design proposal using empirical evidence and the application of

	better explain the observations.	disciplinary concepts.
Designing and comparing solutions		
• Compare multiple possible solutions to a design problem based on how well each is likely to meet the criteria and constraints of the problem.	• Apply scientific ideas or principles to propose tests or tradeoffs needed to optimize a design.	• Evaluate, and/or refine a solution for a complex design problem, based on scientific knowledge, evidence, prioritized criteria, and trade-off considerations.

Engaging in Argument from Evidence

Evidence in science and engineering is based on the analysis of empirical data and its comparison with the predictions of a model or the goals and constraints of a design plan.

Scientists argue to critique or defend a model or explanation; engineers likewise argue to support the merits or critique flaws of a design. Students likewise are expected to argue, or critique proposed models, explanations, and designs—both their own and those of others—using evidence from multiple sources as part of the cycle of testing and improving them.

Evidence in science and engineering is based on the analysis of empirical data and its comparison with the predictions of a model or the goals and constraints of a design plan. The evidence that the students are expected to use in supporting or refuting an argument in an assessment context should be provided to them, possibly also with evidence that is not to be used.

Exhibit 2.9. Engaging in Argument from Evidence

Grade 4	Grade 8	Grade 12
Construct an argument to support or refute a model, explanation, or design solution		

• Construct and/or support an argument with evidence to support or reject a claim about a phenomenon or a design solution. • Make a claim about the merits of a design solution by citing relevant evidence about how it meets the criteria and constraints of the problem.	• Construct an argument with evidence and scientific reasoning to support or reject a proposed model, explanation, or design solution for a problem. • Identify evidence that could be used to refute a claim about a phenomenon.	• Construct an argument with evidence and scientific reasoning to support or reject a proposed model, explanation, or design solution for a problem.
Evaluate and/or improve an argument for an explanation, model, or design solution		
• Evaluate an argument based on the evidence or reasoning it includes.	• Revise an argument that supports or rejects a model, explanation, or design solution for a problem to address new evidence. • Compare and critique two arguments on the same question to analyze their fit with the evidence and/or whether they emphasize similar or different evidence and/or interpretations.	• Revise an argument to support or reject a model, explanation, or design solution for a problem to address new evidence or to address a counterclaim. • Compare and evaluate the arguments for two competing design solutions, based on design criteria, empirical evidence, and/or relevant factors such as economic, societal, environmental, or ethical considerations.

Obtaining, Evaluating, and Communicating Information

Reading, interpreting, evaluating, and producing scientific and technical texts, which can include both written and visual information along with data presentation and mathematical relationships, are fundamental practices of science and engineering, as is communicating clearly and persuasively using both verbal and visual resources.

Being a critical consumer of information about science and engineering requires the ability to read or view reports of scientific or technological advances or applications (whether found in the press, the internet, or social media) and to recognize the salient ideas; identify sources of error and methodological flaws; and distinguish observations from inferences, arguments from explanations, and claims from evidence. Scientists and engineers employ multiple sources to obtain information used to evaluate the merit and validity of claims, methods, and designs.

Evaluating information is a critical skill in the world today, where both information and misinformation (even deliberate disinformation) are widely available through digital sources.

Students need to know how to compare information from multiple sources and, where contradictions exist, to use reasonable criteria to determine the most reliable sources and to argue for the merits or unreliability of a source of information.

Communicating information, evidence, and ideas can be done in multiple ways: using tables, diagrams, graphs, models, interactive displays, and equations; speaking; writing; and discussing.

Almost every NAEP Science item will require students to use their skills in reading and interpreting text, combining that with graphic information to understand the item context, and to communicate their conclusions, so these aspects of this practice are not stressed in the table of elements of the practice to be specifically assessed.

Exhibit 2.10. Obtaining, Evaluating, and Communicating Information

Grade 4	**Grade 8**	**Grade 12**
Evaluating Information		
• Evaluate whether the information presented is evidence, an opinion, or a fictional story. • Evaluate whether the information presented in a text summarizing a graph or table of data accurately reflects the claim that could be made from	• Assess the credibility, accuracy, and possible bias of an article on a science topic (e.g., based on where it is found, the qualifications of the source, and/or the evidence given to make the claim.) • Evaluate information from two different	• Assess the credibility, accuracy, and possible bias of an article on a science topic (e.g., based on where it is found, the qualifications of the source, and/or the evidence given to make the claim. • Evaluate scientific and/or technical

the data.	sources to determine whether there are conflicts between them. ● Identify and critique standard flaws in science-related arguments, (e.g., poor assumptions, cause vs. correlation, faulty explanations, or overgeneralizations from limited data).	information from multiple sources, assessing the evidence used by and the information on qualifications and expertise of each source. ● Identify and critique standard flaws in science-related arguments (e.g., poor assumptions, cause vs. correlation, faulty explanations, or overgeneralizations from limited data).

2C. NAEP Science Crosscutting Concepts

> *Some important themes pervade science, mathematics, and technology and appear over and over again, whether we are looking at an ancient civilization, the human body, or a comet. They are ideas that transcend disciplinary boundaries and prove fruitful in explanation, in theory, in observation and in design. (American Association for the Advancement of Science, 1990, p. 123)*

> *These crosscutting concepts were selected for their value across the sciences and in engineering. These concepts help provide students with an organizational framework for connecting knowledge from the various disciplines into a coherent and scientifically based view of the world. (NRC Framework, 2012, p. 83)*

The idea that broad concepts common to nearly all fields of science and engineering should be included as an essential part of science education for all students was initially proposed in the seminal work *Science for All Americans* (AAAS, 1990), in which they were referred to as "themes." Later, the list of these concepts was refined and renamed "unifying concepts" in the *National Science Education Standards* (NRC, 1996), and further refined as "crosscutting concepts" in *A Framework for K–12 Science Educatio*n (NRC, 2012).

The crosscutting concepts used in the *2028 NAEP Science Assessment Framework* are based on those defined in the NRC 2012 document. Students who have learned to use these concepts in many contexts as tools for sensemaking (i.e., asking productive questions, examining a model under development, designing a solution, etc.) will be able to apply their knowledge of disciplinary concepts to explaining unfamiliar phenomena or solving challenging problems.

Crosscutting concepts are deeply linked to practices of science and engineering and are conceptual tools that guide effective and reflective practice.

If an item is three dimensional, it requires using the disciplinary concept, science and engineering practice, and crosscutting concept to answer the item. The item should elicit evidence that the student thought about that crosscutting concept (i.e., the student will demonstrate using the crosscutting concept to solve the item). For example, if the crosscutting concept is Scale, proportion, and quantity, at grade 8, then selecting or generating the correct response could require reasoning about the disciplinary concept in relation to how the observed function of a system may change with scale.

The seven crosscutting concepts for the NAEP Science Assessment are as follows:

- **Patterns**
- **Mechanisms and explanation: Cause and effect**
- **Scale, proportion, and quantity**
- **Systems and system models/systems thinking**
- **Conservation, flows, and cycles: Tracking energy and matter**
- **Relationships between structure and function**
- **Conditions for stability and change in systems**

Exhibits 2.11-2.17 describe each of the crosscutting concepts in detail. The bulleted sub-statements in the charts under each crosscutting concept are individual elements that pull out aspects of each crosscutting concept that might be assessed at this grade level but not every sub-statement needs to be assessed by NAEP.

Patterns

Patterns exist everywhere—in regularly occurring shapes or structures and in repeating events and relationships. Patterns are discernible in the symmetry of flowers and snowflakes, the cycling of the seasons, and the repeated base pairs of DNA. Noticing patterns is often a first step to organizing and asking scientific questions about why and how the patterns occur.

One major use of pattern recognition is in classification, which depends on careful observation of similarities and differences; objects can be classified into groups on the basis of similarities of visible or microscopic features or on the basis of similarities of function. Such classification is useful in codifying relationships and organizing a multitude of objects or processes into a limited number of groups. Patterns of similarity and difference and the resulting classifications may change, depending on the scale at which a phenomenon is being observed. For example, isotopes of a given element are different—they contain different numbers of neutrons—but from the perspective of chemistry they can be classified as equivalent because they have identical patterns of chemical interaction. Once patterns and variations have been noted, they lead to questions; scientists seek explanations for observed patterns and for the similarity and diversity within them. Engineers often look for and analyze patterns, too. For example, they may diagnose patterns of failure of a designed system under test in order to improve the design, or they may analyze patterns of daily and seasonal use of power to design a system that can meet the fluctuating needs.

The ways in which data are represented can facilitate pattern recognition and lead to the development of a mathematical representation, which can then be used as a tool in seeking an underlying explanation for what causes the pattern to occur. Biologists studying changes in population abundance of several different species in an ecosystem can notice the correlations between increases and decreases for different species by plotting all of them on the same graph and can eventually find a mathematical expression of the interdependencies and food web relationships that cause these patterns.

The human brain is remarkably adept at identifying patterns, and students progressively build upon this innate ability throughout their school experiences.

Exhibit 2.11. Patterns

Patterns: Observed patterns in nature guide organization and classification and prompt questions about relationships and causes underlying them.		
Grade 4	**Grade 8**	**Grade 12**
• Similarities and differences in patterns can be used to sort, classify, communicate, predict, and explain, with various representations (such as physical graphs or diagrams) to describe and analyze features of simple natural phenomena and designed products. [**Boundary**: Statistical displays are limited to bar graphs and pictographs for categorical data, and line plots for measurement data (whole number	• Patterns in data can be identified and represented using graphs, charts, and tables. Analyzing patterns can help identify cause and effect relationships and estimate probabilities of events.	• Patterns in data can be identified and represented using graphs, mathematical relationships, and statistical quantities . Analyzing correlated patterns can help identify cause and effect relationships and estimate probabilities of events, but correlation alone is not sufficient information to infer a causal relationship.

measurements only).]		

Mechanisms and Explanation: Cause and Effect

Cause and effect is often the next step in science, after a discovery of patterns or events that occur together with regularity. A search for the underlying cause of a phenomenon has sparked some of the most compelling and productive scientific investigations. Any tentative answer, or 'hypothesis,' that A causes B requires a model or mechanism for the chain of interactions that connects A and B. For example, the notion that diseases can be transmitted by a person's touch was initially treated with skepticism by the medical profession for lack of a plausible mechanism. Today infectious diseases are well understood as being transmitted by the passing of microscopic organisms (bacteria or viruses) between an infected person and another. A major activity of science is to uncover such causal connections, often with the hope that understanding the mechanisms will enable predictions and, in the case of infectious diseases, the design of preventive measures, treatments, and cures.

In engineering, the goal is to design a system to cause a desired effect, so cause-and-effect relationships are as much a part of engineering as of science. Indeed, the process of design is a good place to help students begin to think in terms of cause and effect, because they must understand the underlying causal relationships in order to devise and explain a design that can achieve a specified objective.

When students perform the practice of "Planning and Carrying Out Investigations," they often address cause and effect. At early ages, this involves "doing" something to the system of study and then watching to see what happens. At later ages, experiments are set up to test the sensitivity of the parameters involved, and this is accomplished by making a change (cause) to a single component of a system and examining, and often quantifying, the result (effect). Cause and effect is also closely associated with the practice of "Engaging in Argument from Evidence." In scientific practice, deducing the cause of an effect is often difficult, so multiple hypotheses may coexist. For example, though the occurrence (effect) of historical mass extinctions of organisms, such as the dinosaurs, is well established, the reason or reasons for the extinctions (cause) are still debated, and scientists develop and debate their arguments based on different forms of evidence. When students engage in scientific argumentation, it is often centered on identifying the causes of an effect.

Exhibit 2.12. Mechanisms and Explanation: Cause and Effect

Mechanisms and explanation: Cause and effect: Events have causes, sometimes simple, sometimes multifaceted. Deciphering causal relationships, and the mechanisms by which they are mediated, is a major activity of science and engineering.

Grade 4	Grade 8	Grade 12
• Cause-and-effect relationships are routinely identified, tested, and used to explain changes. • Events that occur together with regularity might have a cause-and-effect relationship or might have some other shared explanation.	• Relationships can be classified as causal or correlational, and correlation does not necessarily imply causation. • Cause-and-effect relationships may be used to predict phenomena in natural or designed systems. • Phenomena may have more than one cause, and some cause-and-effect relationships in systems can only be described using probability.	• Empirical evidence is required to differentiate between cause and correlation and make claims about specific causes and effects. • Cause-and-effect relationships can explain and predict complex natural and human-designed systems. Such explanations may require examining and modeling small scale mechanisms within the system.

Scale, Proportion, and Quantity

Scale, proportion, and quantity are important in both science and engineering. These are fundamental assessments of dimension that form the foundation of observations about nature. Before an analysis of function or process can be made (the how or why), it is necessary to identify the what. These concepts are the starting point for scientific understanding, whether it is of a total system or its individual components. Any student who has ever played the game "twenty questions" understands this inherently, asking questions such as, "Is it bigger than a bread box?" in order to first determine the object's size.

An understanding of scale involves not only understanding that systems and processes vary in size, time span, and energy, but also that different mechanisms operate at different scales. In engineering, "no structure could be conceived, much less constructed, without the engineer's precise sense of scale. At a basic level, in order to identify something as bigger or smaller than something else—and how much bigger or smaller—a student must appreciate the units used to measure it and develop a feel for quantity. Grades 4 and 8 items can include the units of miles, yards, feet, and inches for length, area, and volume as appropriate; pounds and ounces for weight; and fahrenheit for temperature. In grade 12, items should use metric units and temperatures in centigrade.

The ideas of ratio and proportionality as used in science can extend and challenge students' mathematical understanding of these concepts. To appreciate the relative magnitude of some properties or processes, it may be necessary to grasp the relationships among different types of quantities—for example, speed as the ratio of distance traveled to time taken, density as a ratio of mass to volume. This use of ratio is quite different from a ratio of numbers describing fractions of a pie. Recognition of such relationships among different quantities is a key step in forming mathematical models that interpret scientific data.

The crosscutting concept of Scale, proportion, and quantity figures prominently in the practices of Using mathematics and computational thinking and in Analyzing and interpreting data. This concept addresses taking measurements of structures and phenomena, and these fundamental observations are usually obtained, analyzed, and interpreted quantitatively. This crosscutting concept also figures prominently in the practice of Developing and using models.

Scale and proportion are often best understood using models. For example, the relative scales of objects in the solar system or of the components of an atom are difficult to comprehend mathematically (because the numbers involved are either so large or so small), but visual or conceptual models make them much more understandable (e.g., if the solar system were the size of a penny, the Milky Way galaxy would be the size of Texas).

Exhibit 2.13. Scale, Proportion, and Quantity

Scale, proportion, and quantity: In considering phenomena, it is critical to recognize what is relevant at different size, time, and energy scales and to recognize proportional relationships between different quantities as scales change.		
Grade 4	**Grade 8**	**Grade 12**
• Natural objects and/or observable phenomena exist from the very small to the immensely large or from very short to very long time periods.	• The observed function of natural and designed systems may change with scale. Phenomena that can be observed at one scale may not be observable at another scale. • Time, space, and energy phenomena can be observed at various scales using models to study systems. Proportional	• Explanations of phenomena observable at one scale may require models of the system or of processes at many-orders-of-magnitude-smaller scale (e.g., macroscale processes in matter require atomic level understanding of forces between and among atoms).

	relationships (e.g., speed as the ratio of distance traveled to time taken) among different types of quantities provide information about the magnitude of properties and processes.	● Algebraic thinking is used to examine models and scientific data and predict the effect of a change in one variable on another (e.g., linear growth vs. exponential growth).

Systems and System Models/Systems Thinking

Systems and system models are useful in science and engineering because the world is complex, so it is helpful to isolate a single system and construct a simplified model of it. To do this, scientists and engineers imagine an artificial boundary between the system in question and everything else. They then examine the system in detail while treating the effects of things outside the boundary as either forces acting on the system or flows of matter and energy across it—for example, the gravitational force due to Earth on a book lying on a table or the carbon dioxide expelled by an organism. Consideration of flows into and out of the system is a crucial element of system design. In the laboratory or even in field research, the extent to which a system under study can be physically isolated or external conditions controlled is an important element of the design of an investigation and interpretation of results. The properties and behavior of the whole system can be very different from those of any of its parts, and large systems may have emergent properties, such as the shape of a tree, that cannot be predicted in detail from knowledge about the components and their interactions.

Models can be valuable in predicting a system's behaviors or in diagnosing problems or failures in its functioning, regardless of what type of system is being examined. In a simple mechanical system, interactions among the parts are describable in terms of forces among them that cause changes in motion or physical stresses. In more complex systems, it is not always possible or useful to consider interactions at this detailed mechanical level, yet it is equally important to ask what interactions are occurring (e.g., predator-prey relationships in an ecosystem) and to recognize that they all involve transfers of energy, matter, and (in some cases) information among parts of the system. Any model of a system incorporates assumptions and approximations; the key is to be aware of what they are and how they affect the model's reliability and precision. Predictions may be reliable but not precise or, worse, precise but not reliable; the degree of reliability and precision needed depends on the use to which the model will be put.

Exhibit 2.14. Systems and System Models/Systems Thinking

Systems and system models/systems thinking: A system is an organized group of related objects or components: System models can be used for understanding and predicting the behavior of systems.		
Grade 4	**Grade 8**	**Grade 12**
● To explain or make predictions about a phenomenon it often helps to develop a model of a system of related parts, each of which plays some role in the phenomenon.	● A system model specifies the essential components and quantities involved in a phenomenon and the relationships or interactions between them. The model includes both material and conceptual aspects of the system, such as forces between objects or relationships between species. System models can help analyze and explain a phenomenon, and, after testing, to make predictions about the phenomenon. ● Systems may interact with other systems; they may have sub-systems and be a part of larger more complex systems. ● Engineers design systems to achieve particular functions or do specific items. An engineering design plan includes a system model.	● A system model is used to explain or simulate and predict phenomena that occur in the system. A system model defines a boundary for each system or subsystem, and delineates and, where relevant, quantifies all necessary parts of the system. The parts include both invisible features such as forces, or flows and transfers of energy or information. Such models may include equations that describe relationships between relevant quantities in the system. ● Engineered systems are designed to achieve particular functions. Such systems may be specific objects (e.g., a satellite) or involve large-scale networks of objects (e.g., a transportation

	Engineers also use system models to troubleshoot system failures.	system).

Conservation, Flows, and Cycles: Tracking Energy and Matter

Energy and matter are essential concepts in all disciplines of science and engineering, often in connection with systems. "The supply of energy and of each needed chemical element restricts a system's operation—for example, without inputs of energy (sunlight) and matter (carbon dioxide and water), a plant cannot grow. Hence, it is very informative to track the transfers of matter and energy within, into, or out of any system under study.

In many systems there also are cycles of various types. In some cases, the most readily observable cycling may be of matter—for example, water going back and forth between Earth's atmosphere and its surface and subsurface reservoirs. Any such cycle of matter also involves associated energy transfers at each stage, so to fully understand the water cycle, one must model not only how water moves between parts of the system but also the energy transfer mechanisms that are critical for that motion.

Consideration of energy and matter inputs, outputs, and flows or transfers within a system or process are equally important for engineering. A major goal in design is to maximize certain types of energy output while minimizing others, in order to minimize the energy inputs needed to achieve a desired item.

Exhibit 2.15. Conservation, Flows, and Cycles: Tracking Energy and Matter

Conservation, flows, and cycles: Tracking energy and matter: Tracking energy transfers and matter flows, into, out of, and within systems helps one understand their system's behavior.		
Grade 4	**Grade 8**	**Grade 12**
• To understand the function of a system, it is often useful to keep track of the flows and cycles of matter into, out of, and within the system. The only way that the total weight of matter in a system can change is by flow	• Matter is conserved because atoms are conserved in physical and chemical processes. • Energy manifests itself to our observation in multiple different ways, including in	• Flows of matter and transfers of energy into, out of, and within a system are analyzed and described using a system model. The amount of matter or energy in any system changes only by flow

of matter into or out of the system. **[Clarification**: In grade appropriate contexts (e.g., needs for a healthy organism) the emphasis is on students modeling a system, defining the boundary of the system, keeping track of what matter moves across that boundary as the system functions, and recognizing how such flows are part of that functioning as they engage in sensemaking.]	mechanical, thermal, electrical, and chemical processes. Energy can transfer between these different observed effects and between objects or systems. ● To analyze the function or behavior of a system it is often useful to track and model the energy transfers and matter flows. Within any natural or designed system, transfers of energy are needed to drive any motion or cycling of matter.	of matter or transfer of energy into or out of the system. ● Tracking of matter flows and energy transfers is useful because the availability of matter and/or energy within a system limits what can occur and regulates how the system functions.

Relationships Between Structure and Function

Structure and function are complementary properties. The shape and stability of structures of natural and designed objects are related to their function(s). The functioning of natural and built systems alike depends on the shapes and relationships of certain key parts as well as on the properties of the materials from which they are made. A sense of scale is necessary in order to know what properties and what aspects of shape or material are relevant at a particular magnitude or in investigating particular phenomena—that is, the selection of an appropriate scale depends on the question being asked. For example, the substructures of molecules are not particularly important in understanding the phenomenon of pressure, but they are relevant to understanding why the ratio between temperature and pressure at constant volume is different for different substances.

Similarly, understanding how a bicycle works is best addressed by examining the structures and their functions at the scale of, say, the frame, wheels, and pedals. However, building a lighter bicycle may require knowledge of the properties (such as rigidity and hardness) of the materials needed for specific parts of the bicycle. In that way, the builder can seek less dense materials with appropriate properties; this pursuit may lead in turn to an examination of the atomic-scale structure of candidate materials. As a result, new parts with the desired properties, possibly made of new materials, can be designed and fabricated.

Exhibit 2.16. Relationships Between Structure and Function

Relationships between structure and function: The way an object is shaped or structured determines many of its properties and functions.		
Grade 4	**Grade 8**	**Grade 12**
• Different materials have different substructures, which can influence how they behave (function). • Within any system, natural or designed, the structures of objects, their composition, influences the overall function of the system and its subsystems.	• Complex macroscopic and microscopic structures within systems can be visualized and modeled. These structures and their relationships influence how the system and its subsystems behave. • Structures can be designed to serve particular functions by taking into account properties of different materials and how materials can be shaped and used.	• The functions and properties of natural and designed objects and systems can be inferred from their overall structure, the way their components are shaped and interconnected, and the molecular substructures of various component materials. • Designing new systems or structures requires a detailed examination of the properties of different materials and intentional design of the shapes and structures of different components and of connections between and among components.

Conditions for Stability and Change in Systems

Stability and change are the primary concerns of many, if not most scientific and engineering endeavors. Stability denotes a condition in which some aspects of a system are unchanging, at least at the scale of observation. Stability means that a small disturbance will fade away—that is, the system will stay in, or return to, the stable condition. Such stability can take different forms, with the simplest being a static equilibrium, such as a ladder leaning on a wall. By contrast, a system with steady inflows and outflows (i.e., constant conditions) is said to be in dynamic equilibrium. For example, a dam may be at a constant level with steady quantities of water coming in and out. A repeating pattern of cyclic change (e.g., the moon orbiting Earth) can also be seen as a stable situation, even though it is clearly not static.

An understanding of dynamic equilibrium is crucial to understanding the major issues in any complex system—for example, population dynamics in an ecosystem or the relationship between the level of atmospheric carbon dioxide and Earth's average temperature. Dynamic equilibrium is an equally important concept for understanding the physical forces in matter. Stable matter is a system of atoms in dynamic equilibrium.

In designing systems for stable operation, the mechanisms of external controls and internal "feedback" loops are important design elements; feedback is important to understanding natural systems as well. A feedback loop is any mechanism in which a condition triggers some action that causes a change in that same condition, such as the temperature of a room triggering the thermostatic control that turns the room's heater on or off.

A system can be stable on a small time scale, but on a larger time scale it may be seen to be changing. For example, when looking at a living organism over the course of an hour or a day, it may maintain stability; over longer periods, the organism grows, ages, and eventually dies. For the development of larger systems, such as the variety of living species inhabiting Earth or the formation of a galaxy, the relevant time scales may be very long indeed; such processes occur over millions or even billions of years. Example systems that are appropriate for each grade can be found in the disciplinary concepts in Chapter 2, the sample items in Chapter 3, and the NAEP Assessment Item Specifications.

Exhibit 2.17. Conditions for Stability and Change in Systems

Conditions for stability and change in systems: For both designed and natural systems, conditions that affect stability and factors that control rates of change are critical elements to consider and understand.		
Grade 4	**Grade 8**	**Grade 12**
• Questions about what is changing in a phenomenon, what makes it change, and what keeps it from changing are useful ways to examine a phenomenon. Change in conditions can be described or predicted	• Stability or change over time in a system depends on external conditions as well as on relationships and conditions within the system. • Systems can appear stable on one time scale but viewed on a	• Rates of change are quantifiable and are important quantities to consider in modeling any system. • Feedback mechanisms within a system are important elements for explaining or

for a stable or ongoing situation (e.g., a growing plant, a healthy body). [**Clarification**: In grade appropriate contexts, the emphasis is on students considering what conditions are important for a system to function and how those conditions affect the functioning of a system, and making predictions about what happens when particular conditions change (e.g., needs for food and oxygen for a healthy growing organism; range of conditions in which that organism can live; a change in the slope of a ramp changing the time it takes for a given ball to roll down it).]	longer time scale are seen to be changing.	designing for either the stability or instability of the system. • Changes in a system can be caused by changes in other systems or in conditions affecting the system as well as by prior changes within the system. The scale of the effect is not always comparable to that of the cause but may be much larger or smaller.

Chapter 2 describes what students should be able to know and do with respect to the three dimensions of science achievement. Chapter 3 provides guidance on how these dimensions can be combined to create assessment items.

CHAPTER THREE: Assessment Design

The purpose of Chapter 3 is to describe how assessments can be designed to measure the dimensions and construct described in Chapters 1 and 2. Chapter 3 is organized into the following sections:

- 3A. Types of Items
- 3B. Distribution of Items
- 3C. Scientific Sensemaking in NAEP Science
- 3D. Features of Phenomena and Problems Used in Item Contexts
- 3E. Features of Multidimensional Items
- 3F. Assessing the Full Range of Student Performance
- 3G. Reflecting A Wide Range of Students
- 3H. Performance Expectations
- 3I. Digital Tools

3A. Types of Items

The essential element of any test is an ***item***, the basic scorable part of assessment. The NAEP Science Framework provides recommendations and guidelines for developing items for the 2028 NAEP Science Assessment for a broad audience. A technical specifications document that accompanies this framework will describe in greater detail how items are to be developed and used in the overall design of the 2028 NAEP Science Assessment. In brief, items will be constructed according to the following guidelines:

- The assessment will include a variety of item types, including selected response and constructed response formats, discrete and multipart stand-alone items, item sets, and scenario-based tasks.
- Each item will assess students' understanding in the context of a compelling phenomenon or problem.
- The performance required by each item will involve sensemaking about the phenomenon or problem. No item will assess rote content or procedural knowledge.
- A two-dimensional item will include a disciplinary concept and a science and engineering practice.
- A three-dimensional item will include all three dimensions: a disciplinary concept, a science and engineering practice, and a crosscutting concept.
- Items will be constructed with different levels of complexity to assess students with a wide range of knowledge and skills.
- The assessment will take advantage of several digital tools to clarify the context of each item, the requested performance, and mode of response.
- The assessment as a whole will be responsive to learners with rich and diverse cultural and linguistic backgrounds, identities, and learning environments.

NAEP assessments use a variety of item types to fully assess students' knowledge and skills. Different types of items are used for different purposes. These are best envisioned in two categories. Category one concerns the way test questions are arranged, either as short, separate items or in groups. These include:

A ***discrete item*** **(DI)** is a single, standalone item. Students need to be able to read the stimuli/prompt and answer the question in no more than a few minutes. Compared with other item types, discrete items allow for a large number of items to be included on the assessment, increasing the reliability of the assessment. Examples of discrete items are included below in section 3D and additional examples are included in Appendix B.

A ***multipart item*** **(MPI)** has a few parts that are dependent on each other. For example, a multipart item might ask students to make a choice or decision and follow up with another question to explain their reasoning. Multipart items take somewhat more time since than discrete items, but they can probe for deeper understanding than discrete items. Since multipart items are aimed at different aspects of a single performance, they generally receive a single score that may consist of multiple points. An example of a multipart item is included below in section 3D and additional examples are included in Appendix B.

An ***item set*** **(IS)** uses common stimulus material to ask a group of independent questions. Item sets make it possible to take advantage of efficiency by presenting rich and engaging stimulus material, then asking several questions to collect evidence through a number of different items. Since the items do not depend on each other, questions in an item set each receive a separate score. If an item is rejected as unreliable during testing, the other items in the set may be preserved. Examples of item sets are included in Appendix B.

A ***scenario-based task*** **(SBT)** is presented as an unfolding story, often with rich and engaging stimulus material such as images and video. Scenario-based tasks are often interactive, inviting students to respond to several short items and questions to engage their interest. However, the item does not have to be interactive to be a scenario-based task. Interactive and technology elements should be used when they make sense to serve sensemaking, not gratuitously. By interweaving context and items through a narrative arc, scenario-based tasks scaffold students' progress. With the inclusion of items at a range of complexity, SBTs have great potential to elevate what students know and are able to do. SBTs also make it possible to present more meaningful and compelling phenomena and problems, including those that require more background information. These could look like:

- A compelling phenomenon or intriguing problem that students engage in by manipulating a simulation or conducting an experiment in a virtual lab.
- A sequence of extended response items, with appropriate support, that ask students to authentically consider multiple data sources to construct a model or develop an argument.

- A request to develop a short research proposal to address a science research question or engineering challenge, with guidance for what to include as part of the proposal.

Due to a rich and engaging context and opportunity to ask several related questions, SBTs can surface more comprehensive information about certain aspects of student sensemaking (e.g., some SEPs and CCCs that are difficult to assess in discrete items) than other types of items.

The different item types can be used in combination to create an assessment that appropriately balances breadth and depth of content coverage, while also accounting for measuring a construct that requires time for students to process the phenomenon or context necessary for sensemaking. Discrete and multipart items can help ensure breadth of content coverage but require students to frequently switch between distinct phenomena and problems. Scenario-based tasks allow for in depth measurement of a single phenomenon or problem but generally take a lot of time relative to the information gathered. Although not a strict requirement, it is expected that item sets will play a prominent role in the implementation of this framework; groups of independent items that make use of some common phenomena and problems may provide the best balance of breadth and depth by creating opportunities to measure related but distinct content with independent items.

Within these item types, items will be either *selected response items* in which students choose a response from provided options, and *constructed response items*, in which the student responds by generating an original response using text, symbols, or other input. These are further divided into subcategories as follows.

Selected Response Items are closed-ended and finite, in terms of response options. Selected-response items typically ask students to select the correct answer from a list of options included in the item. Different types of selected response items that may be used on the 2028 NAEP Science Assessment are listed below.

- Single-selection multiple choice: In response to a prompt, students choose a single response from a set of (usually) four or more options.
- Multiple selection multiple choice: Students are prompted to choose two or more responses from a set of (usually) five or more options.
- Matching Table: Students mark their response to a list of statements in a table by marking each option as yes/no, true/false etc.
- Zone: Students respond to a prompt by marking or dragging a symbol into a different part of the answer space.
- In-line Choice: The student selects a single text option from a drop-down menu within a table or inline text.
- Grid: The student selects points on a grid to complete a task, such as creating lines and shapes, or plotting points.

Constructed Response Items are often interactive and typically ask students to write, or construct, the correct answer instead of selecting it. These are generally more challenging than selected response items because the alternative answers are not part of the item. Constructed response item types that may be used on the 2028 NAEP Science Assessment are listed below.

- Short Text: The student enters a word or short phrase into a box, or completes a sentence.
- Table Text: The student enters text into a table or chart.
- Extended Response: A prompt requires a written response that is several sentences long.
- Numeric Entry: The required response is a number or equation.

Varying the item types students engage with on the assessment is essential to balance complexity, time on task, and validity and reliability considerations. Discrete items, multipart items, item sets, and scenario based tasks may all use any combination of the item types described above.

3B. Distribution of Items

Balance by Disciplinary Concept Grouping

The distribution of items by discipline should be approximately equal across Physical Science, Life Science, and Earth and Space Sciences at all grades. With respect to crosscutting concepts and science and engineering practices, at all grades, the emphasis should be on meaningful representation rather than a strictly equal distribution. When an authentic query requires only an application of a practice to a disciplinary concept, a two-dimensional item is acceptable.

Exhibit 3.1. Distribution of Items by Disciplinary Concept Grouping and Grade

Percentage of items	Grade 4	Grade 8	Grade 12
Physical Science	33.3%	33.3%	33.3%
Life Science	33.3%	33.3%	33.3%
Earth and Space Sciences	33.3%	33.3%	33.3%

Balance by Response Type

The assessment will consist of about 65 percent selected response items and 35 percent constructed response items. Given, however, that items requiring constructed responses take a longer time to answer, it is anticipated that the amount of time students spend answering selected response items and constructed response items will be approximately equal.

Exhibit 3.2. Distribution of Items by Response Type

Percent of Items by Response Type	
Selected Response: 65% of items	Constructed response: 35% of items

Distribution of Science and Engineering Practices and Crosscutting Concepts

In doing science or engineering, the eight practices are used in an iterative and recursive cycle that often blurs the boundaries between them. For NAEP assessment purposes, the practices will be paired into four categories, labeled: Investigating, Analyzing, Explaining and Evaluating, as shown in Exhibit 3.3 below. These pairings put together practices most often used with a common purpose.

Exhibit 3.3. NAEP Science and Engineering Practices

Investigating	Asking Questions and Defining Problems
	Planning and Carrying Out Investigations
Analyzing	Analyzing and Interpreting Data
	Using Mathematics and Computational Thinking
Explaining	Developing and Using Models
	Constructing Explanations and Designing Solutions
Evaluating	Engaging in Argument from Evidence
	Obtaining, Evaluating, and Communicating Information

To ensure that a variety of the practices are used throughout the assessment, item developers are requested to use a minimum of 10% of the items at each grade level from each of the four groups of science and engineering practices, but may otherwise choose SEPs that work well within other item design considerations. Similarly, all seven crosscutting concepts should be used in items where appropriate for the item and grade level. More guidance around pairing crosscutting concepts with disciplinary concepts and science and engineering practices can be found in the NAEP Science Assessment and Item Specifications.

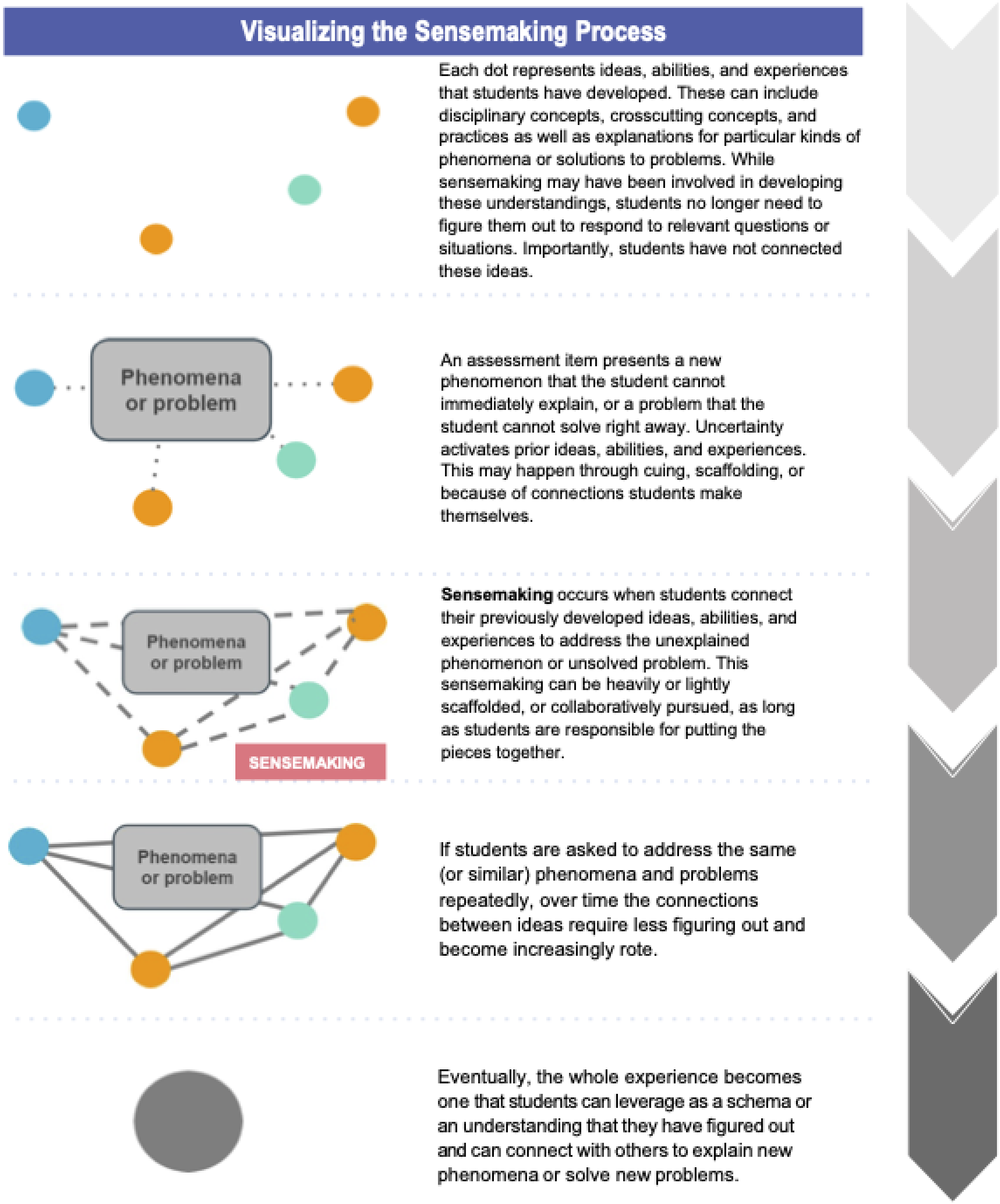
3C. Scientific Sensemaking

An essential aspect of all test items is that they will surface *sensemaking* (Exhibit 3.4). In contrast to items that measure a student's ability to recall rote knowledge, items that measure sensemaking require students to actively apply disciplinary concepts, science and engineering practices, and crosscutting concepts to figure out a phenomenon or address a real-world problem.

Items that require sensemaking enable students to demonstrate that they deeply understand and can apply the disciplinary ideas to figure out something in the world around them. It will therefore be necessary for all items to present either a phenomenon that invites explanation, or a problem that needs to be solved. For discrete items, this phenomenon may be a very simple observation or a single piece of data; for more complex items, the phenomenon- or problem-driven context may include more components for students to consider as part of scientific sensemaking.

The ability to apply disciplinary concepts using practices and crosscutting concepts is an intrinsic feature of sensemaking. The role of phenomena and problems in sensemaking is illustrated in Exhibit 3.4, and is described in the following section.

Exhibit 3.4. Visualizing the Sensemaking Process[4]



[4] Adapted from Achieve (2019b). The Task Annotation Project in Science: Sense-making. Retrieved from https://issuu.com/achieveinc/docs/sense-making_02142019__7_

3D. Features of Phenomena and Problems Used in Item Contexts

In this framework, an assessment designed to measure science achievement requires students to demonstrate scientific knowledge while engaging in the practices of science and engineering—that is, scientific sensemaking or problem solving using disciplinary concepts, science and engineering practices, and crosscutting concepts. To do so, all items are designed around compelling phenomena and/or problems. Without a phenomenon or problem at the center of an assessment item, there is nothing for students to make sense of, problem-solve about, or apply their knowledge to.

Phenomena are real-world events or processes that provide a setting for an item or set of items. They should be chosen to engage student attention and sensemaking that requires the targeted disciplinary concepts, practices, and crosscutting concepts for a satisfying explanation or effective solution. Problems are meaningful challenges that present a situation requiring a new or improved technology or processes. Where appropriate phenomenon and problem descriptions should include the impact, such as effects on people, animals, or the environment. To serve as the context of an item, phenomena and problems must require the application of a disciplinary concept identified in Chapter 2.

From the perspective of the student taking the assessment, they are answering questions about what, why, or how something occurs or what to do about a problem. If the phenomenon or problem is sufficiently compelling, the student will fully engage in the item and demonstrate their knowledge and skills.

Criteria for Selecting High-Quality Science Phenomenon and Problems

Phenomena and problems provide the context for all NAEP Science items. Some contexts will be short and simple; for example, they will have one or two sentences and one or two images. Other contexts will present more complex phenomena and problems or support a broader range of items. High quality phenomena and problems are important for science assessments because they provide access points for students, ensuring that all students can make their thinking visible, ensure assessments are accessible, and provide opportunities for all students to show what they know and are able to do. Following are criteria and guidelines for choosing high quality phenomena and problems.

High-quality items based on phenomena and/or problems: (a) position items to be compelling and motivating to students; (b) cue students toward the targeted dimensions they need to apply; (c) help students from a diversity of prior learning and lived experiences understand what they are being asked to do; and d) provide scaffolds for students to engage and demonstrate their understanding. In this way, high-quality phenomenon-/problem-based items are essential to truly surface what all students know and are able to do and to ensure that scores are trustworthy representations of students' knowledge and skills in science.

While the exact nature of contexts will depend on what disciplinary concept and practice are intended to be elicited, some common features of high-quality contexts for scientific sensemaking include the following:

- Focus on a specific, observable, and/or measurable event(s) that is relatable and motivating to students.
- An authentic question, puzzle, item, or other prompt that leads the student to use the targeted disciplinary concept and science and engineering practice (and crosscutting concept when appropriate) to explain the phenomenon or figure out a solution to the problem.
- Just the right amount of information about the phenomenon or problem that enables the student to engage their thinking, but not too much to be distracting.
- The context should be accurate and presented in an engaging way through text, images, video, or other means to engage student interest.
- The length of a phenomenon or problem description should scale with the scope of the assessment item. The context for a discrete item will be shorter than that for an item set or scenario-based task. The most important consideration is that the context is appropriate to measure the item-level targets.
- Require the appropriate level of conceptual understanding as described in Chapter 2, but not highly specific or technical levels of understanding beyond what students are expected to bring to the assessment.
- Avoid an additional cognitive burden by not asking students to hold a lot of contextual information in working memory or determine which pieces are relevant for each item.
- Do not give away the punchline. Avoid including information that students should have been expected to bring to the table. Leave space for students to demonstrate their understanding and not only their reading and logical reasoning skills.

Creating Contexts for Different Types of Items

The context for discrete selected response items should provide just enough information for the student to select the response that answers the question or challenge. For example, if the item is about data analysis, the context will need to provide data to analyze; if it is about making a claim from evidence, the context will need to provide evidence. In multiple choice questions, the answer choices themselves are also part of the information students use to understand and engage with the item, and should be designed accordingly.

Discrete constructed response items may ask students to engage more comprehensively in practices such as modeling, explaining, or arguing from evidence. Such items elicit a wide range of performances that allow for more expansive sensemaking than selected response items. Therefore, contexts for these items may provide more information. Like all items, the information should be only what is needed to engage with the item.

Multipart items, item sets, and scenario-based tasks will typically require more expansive contextual information to support a wider range of performances, and to compel student sensemaking throughout the set of items. This may begin with a compelling observation of a phenomenon, such as a volcanic eruption, or a meaningful problem, such as preventing a pandemic. Such contexts will often be richer, involve more text, images, and data than contexts for discrete items, and include multiple uncertainties that can be leveraged across many items. For such complex items, the context can be revealed one step at a time, providing just the amount of information needed to answer the next question or complete the next part of the item, so as not to burden the student with too much information to retain as they deploy their sensemaking abilities.

In coherent item sets, items may be presented in a particular order to help scaffold students through the set. In these items, like in multipart items, previous items will often provide students with additional context and contribute to schema development for the item. This should be accounted for intentionally.

Language Considerations in Contexts

Assessments that present phenomena and problems to enable sensemaking often require more language use (reading, writing) than do traditional assessments focused on recall and memorization. While this is necessary both to better engage learners and to elicit student sensemaking, attending to some specific considerations for language use can ensure that all learners can successfully engage with the assessment item. For example:

- Use only as many words as needed to convey a compelling and necessary context.
- Choose (and vary) narrative, expository, and scientific types of writing appropriate to the context.
- Use every-day language and active voice where possible.
- Analyze the reading level to ensure it is accessible to the vast majority of students.
- Use a variety of modalities to convey information, such as text, images, and video.
- Avoid using words that have different scientific and colloquial meanings.
- Use similar language conventions within and across disciplines.

Deciding What to Include and Exclude in the Context

The 2028 NAEP Science Framework is designed to enable students to demonstrate their conceptual understanding of disciplinary and crosscutting concepts, and to use science and engineering practices. However, students are generally ***not*** expected to know or recall specifics of a given phenomenon or a specialized topic. For example, students may be expected to understand how body systems work, but not the specific parts and functioning of the human digestive system. This means that in any given item, students will often need to be provided with additional contextual information for them to fully understand and access the question or perform the required item and apply their conceptual understanding appropriately. Chapter 2 of this framework identifies the information that students should bring to the table. All other details required for satisfactory responses would need to be provided in the context.

3E. Features of Multidimensional items

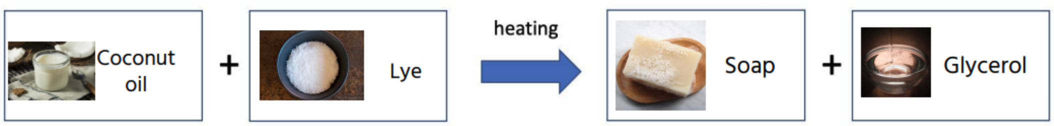
As described previously, measuring the construct described in this framework requires that each item requires students to bring together DCs, SEPs, and, when possible, CCCs to successfully address a question or accomplish a task. The following item is an example of a 3D discrete item.

Exhibit 3.5. Making Soap

Item ID: Making Soap (modified from NGSA)
Grade and discipline: 8th grade, PS
Item type: Single-Select multiple choice
Alignment: this item is a **3D item**, measuring parts of:

- DC: P8.4: In a chemical reaction, the atoms of the reacting substances are regrouped in characteristic ways into new substances with different properties. Atoms only rearrange. As such the amount of matter does not change.
- SEP: S8.13: Analyze data to provide evidence to support or reject a model or explanation or use to improve a design solution.
- CCC: C8.1: Patterns in data can be identified and represented using graphs, charts, and tables. Analyzing patterns can help identify cause and effect relationships and estimate probabilities of events.

One way to make soap is to heat a combination of coconut oil and lye.



The data table below shows properties of each substance in the process.

Sample	Mass (g)	Odor	Density (g/cm3)	Melting point (℃)
Coconut oil	100	Coconut	0.93	27
Lye	20	Odorless	2.13	318
Soap	115	Coconut	0.95	48
Glycerol	5	Odorless	1.26	17.8

Which data provide evidence that making soap involves a chemical reaction?

A. Coconut oil and soap both smell like coconut
B. The density of soap is different than the density of glycerol
C. The melting points of soap and glycerol are very different than the melting points of coconut oil and lye
D. The total mass of soap + glycerol is the same as the total mass of coconut oil + lye.

In this example, students have to apply their understanding of chemical reactions to analyzing data, while looking for patterns among the specific characteristic properties that will indicate that a chemical reaction has occurred.

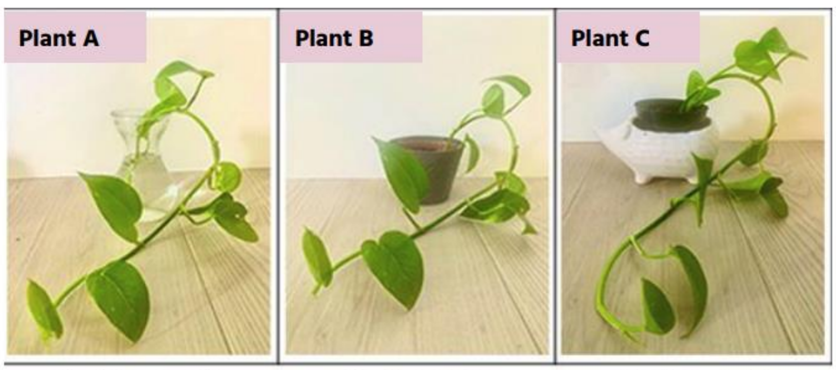
By contrast, the following example is a discrete grade 4 2D item.

Exhibit 3.6. Plant Growth

Item ID: Plant growth (adapted from NGSA)
Grade and discipline: 4th grade, LS
Item type: Multipart item, Single Select multiple choice and short constructed response
Alignment: this item is a **2D item**, measuring parts of:

- DC: L4.3: All animals need food, water, and air in order to live and grow. They obtain their food from their surroundings – from plants or from other animals. Plants need air, water, minerals (in the soil), and light to live and grow.
- SEP: S8:10: Predict the change in a dependent variable when a change in an independent variable occurs.

The plants below were recently placed in a classroom. They are all the same kind of plant, are on the same side of the room, near a window, and receive the same amount of light. Students decide to vary the water and soil conditions to figure out what plants need to grow.



Plant in vase with water and no soil	Plant in pot with soil and water	Plant in pot with soil and no water

Part A. Which plant do you expect will grow **<u>the least</u>** over one month?

a. Plant A
b. Plant B
c. Plant C

Part B. Provide one reason why you think that plant will grow less than the other plants in the investigation.

In this example, students have to apply their understanding of what plants need to grow to make a prediction, but there is no explicit use of a CCC required—the SEP and DC are sufficient to respond to this simple question.

Following are some of the questions that should guide development of multidimensional items:

- Is there an appropriate phenomenon or problem driving student thinking and responses?
- Does the item require students to demonstrate an understanding of at least one DC?
- Does the item require students to demonstrate their understanding of the DC through application of a SEP, and/or a CCC?
- Does the student need to engage in sensemaking to explain a phenomenon or solve a problem?
- Is the understanding appropriate to the grade-level being assessed?

Each discrete item and each multi-part item should be at least two-dimensional and three-dimensional if appropriate. Item sets and scenario-based tasks should include a minimum of one three-dimensional item. Each item will receive one score representing the integration of the dimensions measured by the item.

For additional sample items and annotations, please see Appendix B.

3F. Assessing the Full Range of Student Performance

It is important that NAEP provide a complete picture of student performance. Although there have been concerns that creating an assessment consisting largely of multidimensional items, item sets, and scenario-based items might prove too difficult for students who have not been provided the opportunity to develop proficiency in science, research from the learning sciences, including research on how students learn and develop three-dimensional science understanding, suggests otherwise (NRC, 2005, pp. 407-411; NRC, 2007; NASEM, 2017, pp. 5-14; NASEM, 2018, pp. 145-146). While traditional approaches to assessment often assume that rote understanding or simple procedural skills (e.g., definitions, facts, lab skills) are less cognitively complex and therefore more likely to be doable by students who are still developing their science understanding, this is not borne out in practice. Students do not learn by mastering one dimension at a time before integrating the dimensions, nor by memorizing content before applying it—they learn by using the dimensions together in increasingly sophisticated ways. Likewise, assessments intended to surface what students who have not yet mastered grade-level expectations know and can do may do so more effectively by varying the sophistication of multidimensional performances, rather than focusing on one-dimensional items.

Students at all grade levels and all performance levels can and do find success with multidimensional performances if students are presented with items and items that (a) use appropriately complex contexts, (b) sufficiently scaffold and support learners in engaging with the item, and (c) use the dimensions in appropriate combinations to right-size the complexity. These considerations are particularly important for multilingual learners and other students who may have conceptual understanding without having yet mastered vocabulary or rote facts and procedures. By focusing on multidimensional items that range in complexity, NAEP can better capture student thinking along progressions that mirror how student thinking develops.

The complexity framework that will be applied to NAEP item development will reflect how complexity specifically scales within and across multidimensional science items, including:

- the complexity of the phenomenon or problem context,
- the complexity of language, graphics, or mathematical elements,
- the complexity of the item stem, response mode, and response choices,
- the extent of sensemaking that is required of the student,
- the degree and nature of scaffolding and guidance provided, and
- the nature of the intersections of dimensions within items, including how each dimension contributes to the complexity of sensemaking in the item.

Complexity Framework

The proposed 2028 NAEP Science Framework is informed by the item complexity frameworks proposed by Achieve (2019a), Tekkumru-Kisa, Stein, and Schunn (2015), and WestEd, Center on Standards and Assessment Implementation, & Delaware Department of Education (2019). The purpose of the complexity framework is to inform item development as to ensure that items are accessible to a wide range of learners. The complexity framework considers two underlying contributors to complexity:

- The degree and nature of guidance provided to students. That is, how much direction or cueing are students given for what to consider and how to approach the item?
- The nature of reasoning required by students. That is, how sophisticated is the sensemaking required by students, and how does each dimension contribute to that sophistication in each item?

The complexity framework intentionally goes deeper than some traditional approaches to complexity (e.g., cognitive demand or content complexity approaches, such as Webb's Depth of Knowledge). By considering not only the overall complexity of each item, but how each dimension contributes to sensemaking, items can be designed more intentionally. For example, some items provide substantial scaffolding for engaging in the practice, with limited cueing for the disciplinary concepts, while other items engage in very simple sensemaking with disciplinary concepts while providing students the opportunity to more deeply engage with the SEP and/or CCC. In some items, the CCC can be used to reduce item complexity (e.g., by asking students to identify a pattern as a step toward figuring out the phenomenon) while in other items, the CCC expands complexity by asking students to consider a non-routine lens on a phenomenon or problem (e.g., asking students to examine a seemingly causal relationship that is actually correlational). These are important considerations for developing a balanced assessment that can intentionally surface a range of student thinking.

Exhibit 3.7. Complexity of Multi-Dimensional Items

	How does the DC contribute to the sophistication of sensemaking?	How does the SEP contribute to the sophistication of sensemaking?	How does the CCC contribute to the sophistication of sensemaking?	Overall
High	Students are given limited prompting about which DC to use. Students may leverage ideas from multiple DCs that are not closely related (within or across multiple disciplines). Students use DCs to address a significant uncertainty, with many possible alternative accounts.	Students are given limited prompting about which SEP to use, and how to engage in it. Students may use a series of SEP elements in a sequence of sophisticated thinking that expands the nature of sensemaking. Students use SEPs to navigate complex interactions among multiple components of phenomena and problems.	Students make decisions about which CCC to use to organize their approach to/reasoning within an item. Students explicitly use the CCCs to expand sensemaking. With limited prompting, students use CCCs to navigate phenomena and problems with significant uncertainty and many possible alternative accounts	Two or three dimensions are used to engage in a high degree of sensemaking. Students are given limited prompting about how to approach the item, requiring them to decide what understandings and practices to apply. Students address a high-degree of authentic uncertainty in the phenomenon or problem, navigating many possible (and valid) accounts.
Medium	Students are cued to use specific DCs to address the item. Students may leverage multiple components of a given DC together, OR demonstrate a	Students are cued to use specific SEPs and components of SEPs to address the item. Students use a single SEP component in	Students are cued to use a specific CCC component CCCs serve to focus student thinking within the item With guidance, students use the	Students are provided substantial cues for addressing the phenomenon or problem. They are prompted with specific DCs, SEPs, and CCCs, and provided

98

	sophisticated use of a single DC component. Students use DCs to address a moderate uncertainty, with limited alternative accounts.	support of authentic sensemaking, Students use SEPs to navigate simple interactions among components of phenomena and problems.	CCCs to navigate simple interactions among components of phenomena and problems with moderate uncertainty	guidance on how to use them. One dimension may be more heavily cued than others. Students address a moderate degree of uncertainty with limited possible accounts.
Low	Students are directed to use specific components of a DC to address the item. Students use limited DC components in routine or highly specific ways. Students engage in very simple application of the DC component to a phenomenon with a low degree of uncertainty.	Students are directed to use specific components of the SEP, using a well-defined set of actions or procedures. Students use the SEP as structure to make an idea visible, without using the SEP in service of significant sensemaking.	Students use given CCCs in service of very simple sensemaking, addressing phenomena and problems with limited uncertainty and limited alternative accounts.	Students use a well-defined set of actions to engage in the item and address the phenomenon or problem. They engage in simple application of DCs, SEPS, and CCCs, often involving one or two scaffolded steps. Students address a low degree of uncertainty with a single possible account.

For example, the following item illustrates a low complexity 12th grade item:

Exhibit 3.8. Permafrost, Version 1

Item ID: Permafrost melting (adapted from OSE)
Grade and discipline: 12th grade, PS
Item type: Selected Response, Drag and Drop (or Ordering)
Alignment: this item is a **3D item**, measuring parts of:

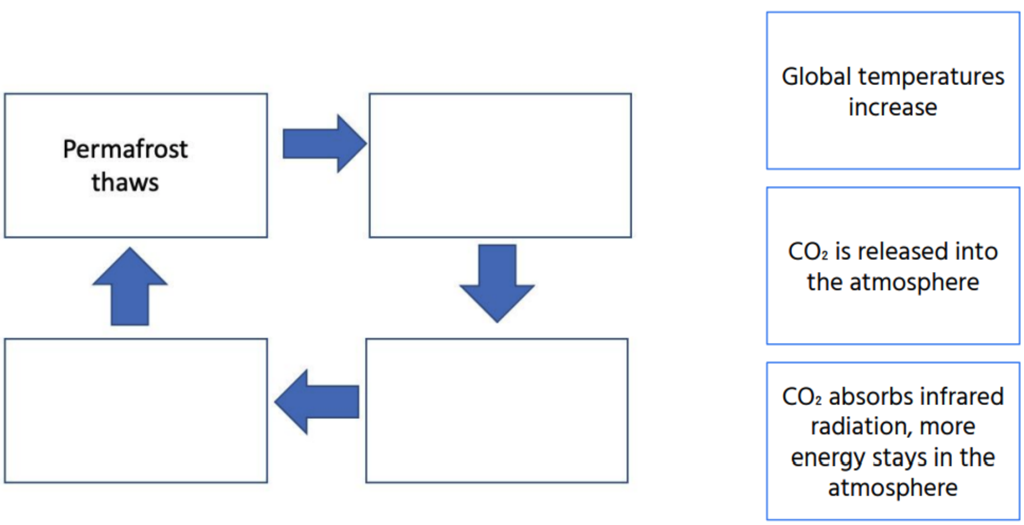
- DC: P12.14: When sunlight is absorbed at Earth's surface it is eventually re-radiated as infrared radiation that transfers heat into the atmosphere. The average temperature of the atmosphere is determined by how long the energy stays in the system until it is reradiated into space from the top of the atmosphere.
- SEP: S8.5: Develop, use, and/or revise a model to describe, explain, and/or predict phenomena by identifying relationships among parts and or quantities in a system, including both visible and invisible quantities.
- CCC: C12.14: Feedback mechanisms within a system are important elements for explaining or designing for either the stability or instability of the system.

Complexity: This item is an example of a low DC, low SEP complexity, low CCC complexity item.

Permafrost is a layer of soil and ice that is just below the surface in the Arctic. Historically, permafrost stayed frozen for many years at a time. However, in some areas permafrost is now melting, which can cause many changes to Earth's surface and living things.

One major concern is that permafrost contains carbon dioxide. The carbon dioxide is trapped when permafrost remains frozen, but is released as it melts.

Use the statements and arrows below to develop a model that shows the relationships between thawing permafrost and rising global temperatures.



This same item could be modified to be higher complexity by (1) requiring students to develop the model with significantly less support (Exhibit 3.9), and/or (2) asking students to consider implication and limitations of the model (Exhibit 3.10)

Exhibit 3.9. Permafrost, Version 2

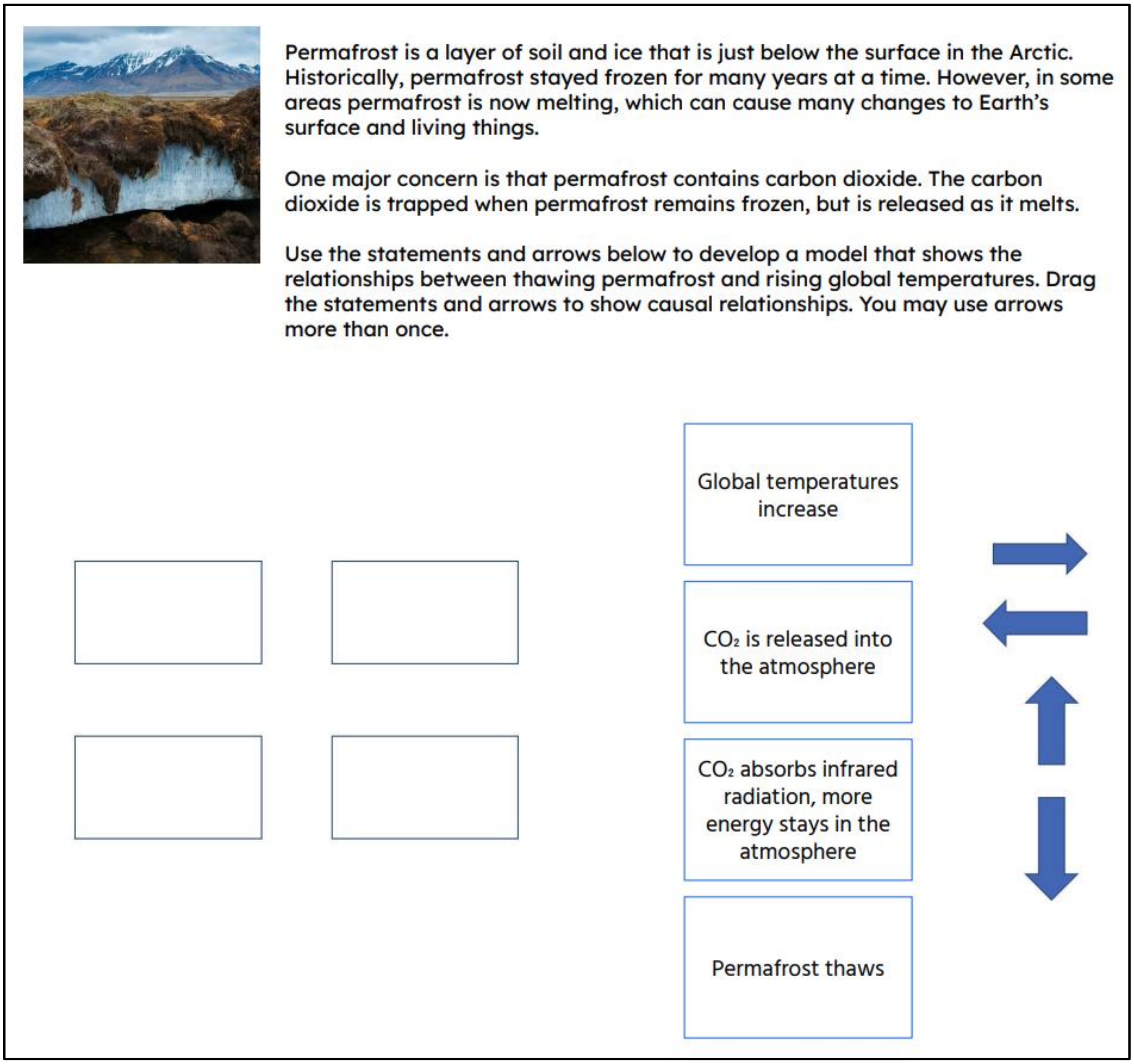
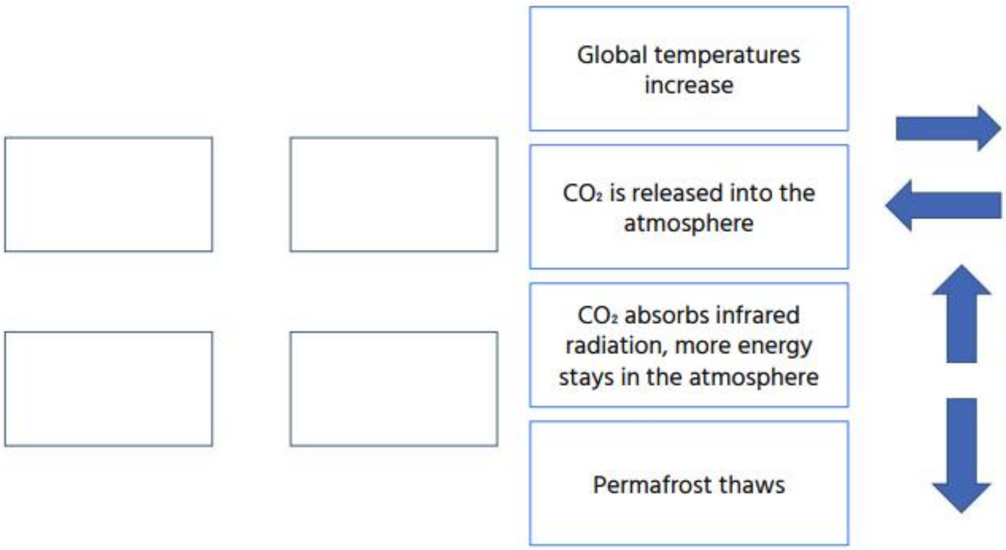


Exhibit 3.10. Permafrost, Version 3

Permafrost is a layer of soil and ice that is just below the surface in the Arctic. Historically, permafrost stayed frozen for many years at a time. However, in some areas permafrost is now melting, which can cause many changes to Earth's surface and living things.

One major concern is that permafrost contains carbon dioxide. The carbon dioxide is trapped when permafrost remains frozen, but is released as it melts.

Use the statements and arrows below to develop a model that shows the relationships between thawing permafrost and rising global temperatures. Drag the statements and arrows to show causal relationships. You may use arrows more than once.



Part B. What does your model predict about how the rate of melting of the permafrost might change over the next 50 years?

The rate of melting will:

A. Increase
B. Stay the same
C. Decrease

Part C. What is one limitation of your model?

For more examples of items at a range of complexity levels, including how a given item can be modified, please see Appendix B.

3G. Reflecting a Wide Range of Students

It is essential for the 2028 NAEP Science to be responsive and relevant to a wide diversity of students. Specifically, students taking the assessment should see themselves and their communities represented in the items across the assessment as a whole and the range of assets and funds of knowledge diverse learners bring to the table should be acknowledged as important elements of science achievement. Below are definitions and general principles for culturally relevant contexts for NAEP science followed by a list of particular features of these contexts.

General Principles and Definitions

- All students have culture, and when we think about diverse cultural representation, we mean to be inclusive of cultural and linguistic experiences across a range of geographies, cultural practices, disabilities, genders, and languages.
- Although some phenomena will be more relevant to some students than others, all students should be able to see themselves and their peers represented in some phenomena/problems included across the assessment. This framework does not suggest that every student be 'matched' with particular items, but rather that all learners should see a range of phenomena, geographies, and people represented such that the assessment, as a whole, is experienced as more culturally relevant.
- By varying the range of who the phenomena/problems are relevant to, we ensure that there is authentic relevance to multiple student groups and that all students are engaged in addressing scenarios/contexts and problems that are culturally novel.
- When contexts focus on legitimate interests of communities, it is more likely that all students will be engaged with the items, because they are inherently more compelling: someone really cares about this phenomenon or problem because it is having a real impact in the world. While traditional sensitivity reviews might flag all such contexts as problematic, a culturally relevant lens asks if the item elicits a productive affective response.
- Providing sufficient background information, including multiple modalities for conveying contexts and any additional information about a context for a phenomenon that is needed, will help reduce inadvertent issues of bias by ensuring all students have an opportunity to become familiar with a context. It should be noted that this kind of appropriate background information is essential in all items and can help ensure that student performance on the assessment is a trustworthy indicator of what they know and are able to do, not whether they were able to understand the task or were motivated to complete it.

Specific Features of Culturally Relevant Contexts and Assessment Design

- Contexts focus on real, specific phenomena and problems particular communities care about. Community interest can be determined through survey data (either available or

conducted), focus groups with diverse communities (available or conducted), news, impacts on lives and livelihoods, mission statements, engineering design solutions and community efforts, etc.
- Solutions and explanations presented to and generated by students through the assessment items improve people's lives and livelihoods.
- Item contexts consider geographic, demographic, and time-related factors to create enough distance between groups of students intended to be taking the assessment, and the phenomenon to limit any negative affective responses.
- Contexts include diverse representations of who is considered a scientist and engineer.
- Contexts position non-White people as (a) more than a stereotyped experience and (b) powerful doers and contributors to science and the broader world.
- Contexts do not include (or limit) gratuitous or superficial representation of diverse races, ethnicities, genders, etc.

In following example of a 2D grade 8 item, several features of culturally responsive items are included, such as (1) the use of native/home language as part of the item, (2) phenomenon that has deep value to a specific community (demonstrated by ongoing community activities and extensive documentation), and (3) the representation of non-traditional scientists, and whose knowledge is valued in science.

Exhibit 3.11. Limu Kohu

Item ID: Limu Kohu (adapted from University of Hawaii)
Grade and discipline: 8th grade, ESS
Item type: Multi-part, Single Select Multiple choice, short response constructed response.
Could part A be used as a stand-alone item? Yes.
Alignment: this item is a **2D item**, measuring parts of:

- DC: E8.12: Human activities have significantly altered the biosphere, atmosphere, and geosphere, sometimes damaging or destroying ecosystems and causing the extinction of organisms. Human choices can minimize harm to other organisms and risks to the health of the regional environment.
- SEP: S8.23: Identify evidence that could be used to refute a claim about a phenomenon.

Item-level claim (derived from targeted dimensions): Students can evaluate evidence about phenomena involving human impacts on the natural world, using their understanding of how human activities can significantly alter the biosphere.
Phenomenon/uncertainty: what human activities are contributing to the decline of limu kohu?
Complexity: This item is an example of a medium DC, low SEP complexity item.

Limu kohu is a type of seaweed that is native to the waters around Honolulu, Hawai'i, and is an important part of food systems as well as cultural and religious practices. Although Limu Kohu was easy to find for hundreds of years, Limu kohu populations around Honolulu have now been rapidly declining over the past 60 years.

Limu Kohu, found off of the coast of Honolulu.

Observations from Generations of Hawaiian Elders about Limu Kohu Growth and Harvesting
• Needs warm water and high salinity to grow • Grow and reproduces well on the edges of coral reefs • When limu kohu is trimmed, it regrows • When the base of the limu kohu is harvested, it cannot regrow

Part A. The following human activities are occurring in the area. Which is mostly likely to cause the **least** harm on limu kohu populations?

A. Industrial ways of harvesting limu picks all parts of the plant, including the base.
B. Many restaurants are using traditional harvesting practices that only harvest the top of the limu kohu.
C. Industry in the area is creating run-off that is changing the temperature and salinity of coastal regions.
D. Ships coming into the area have introduced invasive seaweed species that need the same resources as limu kohu.

Part B. Support your answer, using the information provided and your understanding of human impacts on the environment.

For additional examples, please see Appendix B.

3H. Performance Expectations

Although each student will answer only a subset of items, the full NAEP Science Assessment will measure student sensemaking in each of the disciplinary concepts in Chapter 2. The following guidance is provided to support item development, but is not intended to be prescriptive or limiting to item development.

An essential part of the item development process is to choose a *performance expectation*—something that the student can be expected to do to indicate they understand the targeted DC and can apply it via the cued associated practice (and crosscutting concept, when possible).

Following are examples that can be used to build items for grades 4, 8, and 12. Additional guidance for creating performance expectations is provided in the Assessment and Item Specifications.

Exhibit 3.12. Examples of Performance Expectations

Performance Expectation	DC	CCC	SEP	Rationale
Grade 4 Earth and Space Sciences Interpret patterns in sunrise/sunset data for a given location to explain seasonal differences in day length.	E4.1: Many objects in the sky change position and are not always visible due to Earth's rotation. The patterns of motion of the sun and moon can be observed, measured, described, and predicted.	C4.1: Similarities and differences in patterns can be used to sort, classify, communicate, predict, and explain, with various representations (such as physical graphs or diagrams) to describe and analyze features of simple natural phenomena and designed products.	S4.11: Predict the outcome of an experiment, or a design solution based on a model, a phenomenon, or on a design plan.	One of the first age appropriate CCCs for younger students to engage with is patterns. Sunrise/Sunset times have seasonal patterns to them that are caused by motion in the sun/earth system over the course of a year. This smaller idea (day length) is an important component to many larger ideas (seasonal temperature differences, light/temp cues for plant life cycles, etc.)
Grade 8 Physical Science Ask questions about the interactions between objects	P8.5: The change in motion of an object is determined by the sum of the forces acting on	C8.8: Systems may interact with other systems; they may have sub-systems and be a part of larger more	S8.2: Ask questions that can be answered with empirical evidence to investigate	Students can begin to ask questions to develop a qualitative understanding of forces at entry

to determine how changes in their motions is determined by the sum of the forces acting on each object.	it; if the net force on the object is zero, it will remain at rest or continue moving in a straight line with the same speed and direction as before.	complex systems.	relationships between variables in a system model or in phenomena.	points to making sense of phenomena related to interactions between objects. The sophistication of their questions grows as students progress toward mastery of complex material, providing opportunities to write items at all levels of difficulty and complexity.
Grade 12 Life Science Examine data on different types of grass that can be used in a design for a new public park. Take into account several factors when deciding on the type of grass that will have the smallest negative effect on the environment.	L12.12: Changes induced by human activity (anthropogenic change) in the environment — such as habitat destruction, pollution, introduction of invasive species, overexploitation, and climate change—can disrupt an ecosystem and threaten the survival of some species.	C12.16: Changes in systems depend on changes in other systems or conditions affecting the system as well as on changes within the system. . The scale of the effect is not always comparable to that of the change but may be much larger or smaller.	S12.19: Evaluate, and/or refine a solution for a complex design problem, based on scientific knowledge, evidence, prioritized criteria, and trade-off considerations.	By twelfth grade students are able to prioritize criteria and take into account information from several sources to decide how to solve an engineering problem in a way that minimizes the disruption of an ecosystem.

Once a performance expectation has been identified, it is then possible to find a realistic context for the item—a phenomenon that requires explanation, such as the changing length of day over the seasons, as in the 4th grade example, or a problem calling for an engineering design, such as the 12th grade example. The 8th grade expectation could be implemented through either a phenomenon or a problem. As the item developer gathers information and images about particular contexts, thinks about the level of complexity that is needed, and chooses an appropriate item type, the item starts to take shape.

Item scoring is straightforward for selected response items, which can be scored by machine. However, the large number of constructed response items require interpretation of open ended responses. A preliminary scoring guide is developed along with the first draft of an item, which is then refined after testing with a large number of students. Scores for constructed response items are not based solely on providing accurate descriptions of phenomena using appropriate vocabulary words or language skills. Rather, they are based on the logical application of the disciplinary concept to make meaning of a phenomenon, or to contribute to the solution of a problem. Scoring guides, which are developed using samples of student responses from the target age group, provide indicators to determine if the student correctly understands the disciplinary idea and is able to use the appropriate practice (and crosscutting concept, when appropriate). Scoring guides for some items allow for partial credit. Each item is given a single score, which is then combined with scores of other items to develop an overall score for sensemaking in Life Science, Physical Science, or Earth and Space Sciences.

3I. Digital Tools

The NAEP Science Assessment based on this framework will be administered via computer. Therefore, students will need a number of digital tools—and, at times, science-specific tools—to respond to the items. In a digitally based environment, for example, students will need to draw, highlight, and erase on the screen; to measure the dimensions of virtual objects; to plot data points on number lines; and to create and modify graphical representations. Additionally, the testing environment will need to provide computational tools equivalent to a four-function calculator at grade 4 and a scientific calculator at grades 8 and 12. Continuing a practice that has been in place for recent NAEP administrations, before the assessment, students complete a brief interactive tutorial designed to orient them to the digital tools they will use during the assessment. The 2019 tutorials for each grade level can be found on the Internet (Governing Board, 2019a, 2019b).

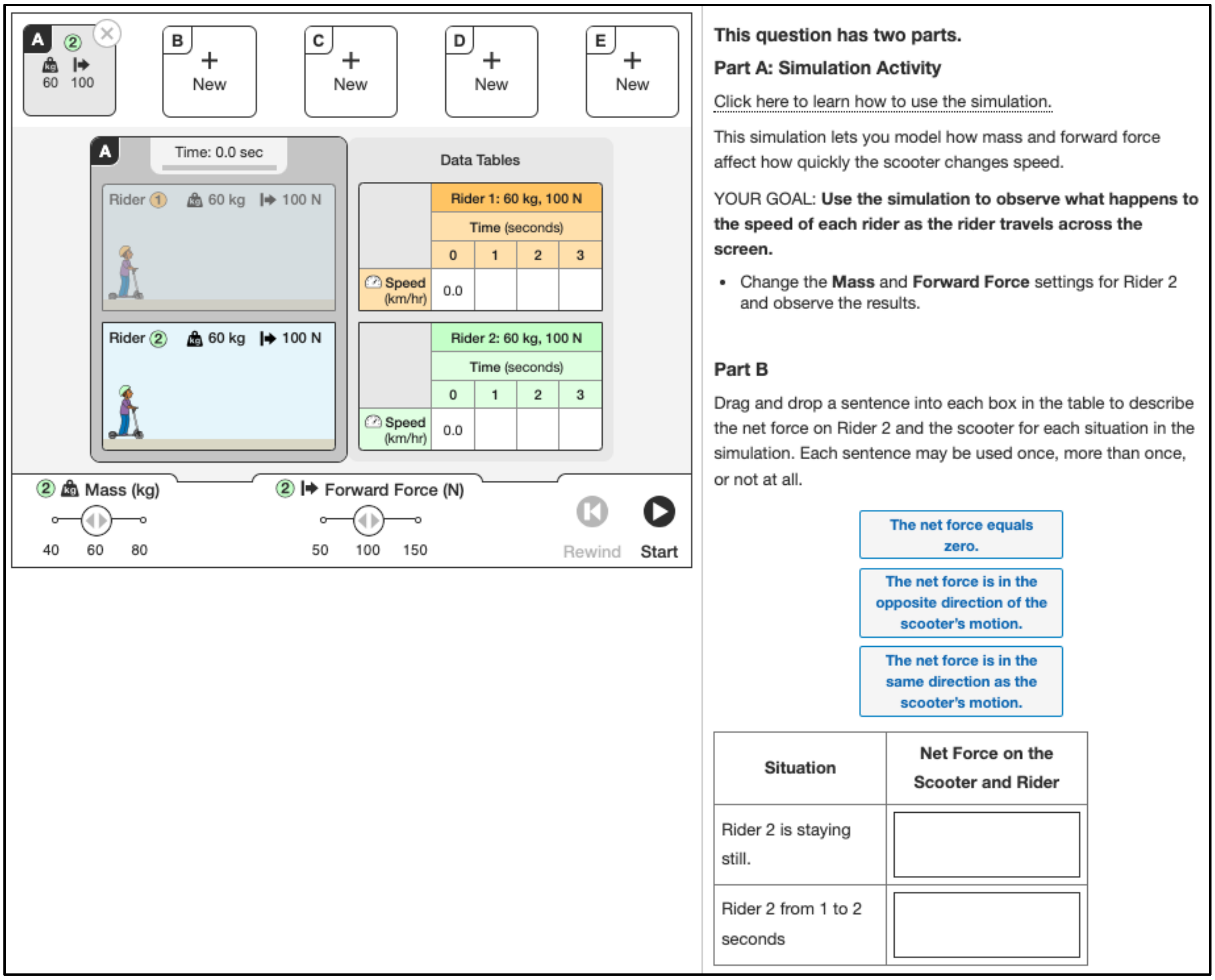
All digital NAEP assessments include system tools, which are always available and common across all NAEP assessments. There are also science and mathematics tools, which are specific to and only available for certain items on NAEP science assessments. The materials and accompanying items are carefully chosen to cause minimal disruption of the administration process and are typically only provided when relevant to solving the item.

The 2028 NAEP Science Assessment will include digital tools to support DCs, SEPs, and CCCs, presented previously. The illustrations in this framework are static screen shots to illustrate examples of these digital tools; however, the screen shots represent only a small subset of the many images, videos, and simulations students encounter during the assessment. Digital tools should be used when the item format offers advantages over other assessment modes. Examples include (but are not limited to) testing student scientific sensemaking related to the following situations:

- Using simulations and modeling tools for scientific phenomena that cannot easily be observed in real time, such as seeing things in slow motion (e.g., the motion of a wave) or at a higher speed (e.g., erosion caused by a river).
- Modeling scientific phenomena that are invisible to the naked eye (e.g., the movement of molecules in a gas).
- Working safely in lab-like simulations to collect and analyze data that would otherwise be disorderly in an assessment situation or hazardous (e.g., using dangerous chemicals).
- Situations that require several repetitions of an experiment while the student varies the parameters (e.g., rolling a ball down a slope while varying the mass, the angle of inclination, or the coefficient of friction of the surface).
- For manipulating objects, such as placing organisms into an ecosystem food web.

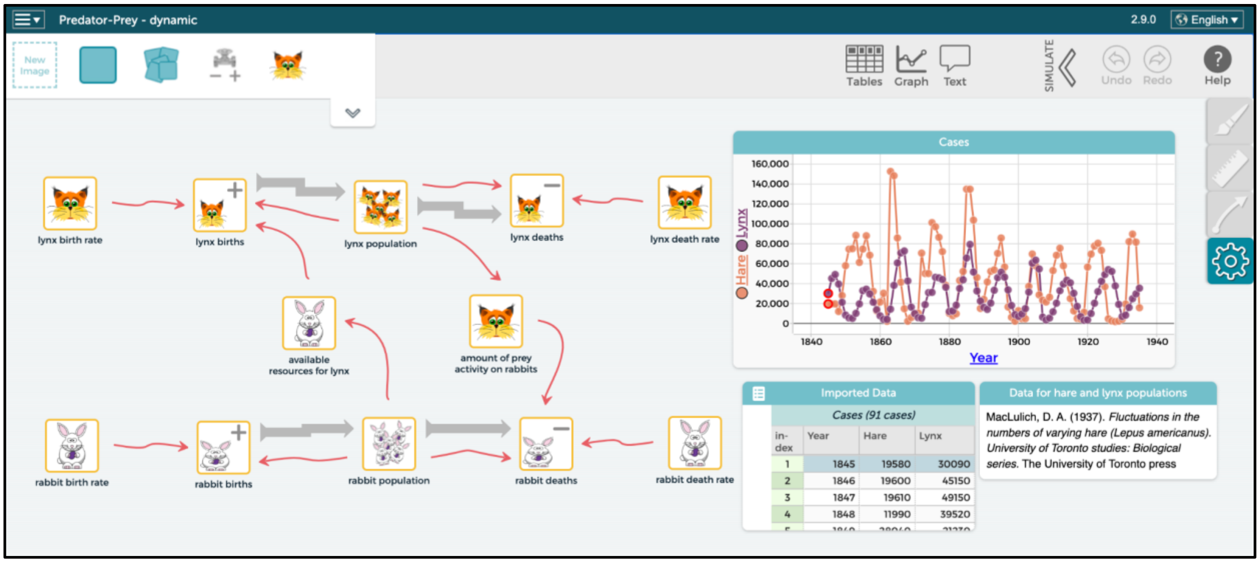
For example, the items in Exhibits 3.8, 3.9, and 3.10 illustrate use of the “drag and drop” digital tool. The following example (Exhibit 3.13) highlights how simulations might be used within NAEP Science. In this example, students use the simulation to better understand the forces acting within a phenomenon-based context.

Exhibit 3.13. Sample Simulation From a Multidimensional Item Set



Similarly, the example below (Exhibit 3.14) shows how digital tools might be used to allow students to construct dynamic models that they independently develop. This example leverages SageModeler, a free, open-source, web-based systems dynamics modeling tool commonly used in science education. This tool allows students to define variables, relationships, degree of influence, and to run models and collect data.

Exhibit 3.14. Sample Modeling Tool (SageModeler)



CHAPTER FOUR: Reporting Results of the NAEP Science Assessment

4A. NAEP Assessments and the Nation's Report Card

The NAEP Science Assessment provides the nation with a snapshot of what U.S. students know and are able to do in science. Results of the NAEP Science Assessment administrations are reported in terms of average scores for groups of students on the NAEP 0–300 scale and as percentages of students who attain each of the three achievement levels (*NAEP Basic*, *NAEP Proficient*, and *NAEP Advanced*). This is an assessment of overall achievement, not a tool for diagnosing the needs of individuals or groups of students. Reported scores are always at the aggregate level; by law, scores are not produced for individual schools or students. Results are reported for the nation as a whole, for regions of the nation, and sometimes for states and large districts that volunteer to participate. The NAEP results are published in an interactive report online as *The Nation's Report Card*.

The Nation's Report Card allows for examination of results by school characteristics (urban, suburban, rural; public and nonpublic) and other student characteristics (race/ethnicity, gender, English learner status, socioeconomic status, and disability status [i.e., supported by an Individualized Education Program]), as required by law. The NAEP Data Explorer is a publicly accessible tool that allows users to customize reports and to investigate specific aspects of student science achievement, such as performance by disciplinary area or by selected contextual variables. Also, reports of the results of survey questionnaires are produced each year on various topics (e.g., students' internet access and digital technology at home, instructional emphasis on science activities, confidence in science knowledge and skills, teachers' satisfaction, and views of school resources).

In 2002, NAEP initiated the Trial Urban District Assessments (TUDA) in five large urban school districts that are members of the Council of the Great City Schools (the Atlanta City, City of Chicago, Houston Independent, Los Angeles Unified, and New York City Public Schools districts). In 2003, additional large urban districts began to participate in these assessments, growing to a total of 27 districts by 2017. Sampled students in TUDA districts are assessed in the same subjects and use the same NAEP field materials as students selected as part of national or state samples. TUDA results are reported separately from the state in which the TUDA is located, but results are not reported for individual students or schools. With student performance results reported by district, participating TUDA districts can use results for evaluating their achievement trends and for comparative purposes.

4B. Reporting Scale Scores and Achievement Levels

NAEP reports average results on a scale of 0–300 in science. In the past, the average scores have also been reported on three disciplinary groups: Life Science, Physical Science, and Earth and Space Sciences. Reports from the new assessment will include average scores on the same three disciplinary groups, with an updated title for each to reflect the emphasis on student scientific sensemaking and problem solving on the assessment. Scale scores for the disciplinary groups will be reported using the following definitions of each reporting category:

- **Sensemaking in Physical Science**: The student reasons scientifically using disciplinary concepts in **physical science**, in combination with science and engineering practices, and crosscutting concepts.
- **Sensemaking in Life Science**: The student reasons scientifically using disciplinary concepts in **life science**, in combination with science and engineering practices, and crosscutting concepts.
- **Sensemaking in Earth and Space Sciences**: The student reasons scientifically using disciplinary concepts in **Earth and space sciences**, in combination with science and engineering practices, and crosscutting concepts.

NAEP will not, however, report on any of the three dimensions separately. That is, there will be no separate scores for students' knowledge of disciplinary concepts, practices of science and engineering, or crosscutting concepts. Given the goal to report on sensemaking in the three disciplinary groupings, all three dimensions are essential in surfacing and measuring students' abilities to apply their understanding of the disciplinary concepts to real-world contexts—the phenomena and problems that frame each item and group of items.

These definitions are intended to emphasize that a score for each disciplinary group reflects students' abilities to integrate the three dimensions of science: disciplinary concepts, science and engineering practices, and crosscutting concepts, and does not prioritize knowledge of the disciplinary concepts.

Since 1990, the Governing Board has used achievement levels for reporting results on NAEP assessments. Generic policy definitions for achievement at the *NAEP Basic*, *NAEP Proficient*, and *NAEP Advanced* levels describe in very general terms what students at each grade level should know and be able to do on the assessment (see Exhibit 4.1). Achievement level descriptions specific to the 2028 NAEP Science Framework are still under development and will be included in the final framework. These will be used to guide item development and initial stages of standard setting for the 2028 NAEP Science Assessment (if it is necessary to conduct a new standard setting).

Reporting on achievement levels is one way the Nation's Report Card helps the general public and policymakers interpret NAEP results. Results are reported as percentages of students within each achievement level range as well as the percentage of students at or above *NAEP Basic* and at or above *NAEP Proficient*. Students performing at or above the *NAEP Proficient* level on NAEP assessments demonstrate solid academic performance and competency over challenging subject matter. Following the first administration of the science assessment based on the updated framework, new Reporting ALDs will be created to specify certain skills in which students are likely to have demonstrated competency at each achievement level. Results for students not reaching the *NAEP Basic* achievement level are reported as below *NAEP Basic*. As noted, individual student performance cannot be reported based on NAEP results.

Note that the *NAEP Proficient* achievement level does not represent grade-level proficiency as determined by other assessment standards (e.g., state or district assessments), and there are significant differences between achievement in the context of NAEP as compared to the context of state-level annual tests. For one, teachers and students are not expected to have studied the NAEP framework or systematically aligned state standards or local curricula with it, nor are students expected to study for the assessment. Furthermore, the NAEP assessment is broader than a typical state grade-level test, for NAEP covers multiple years of study and does not focus on specific instructional units and school years. In addition, there is not a uniform definition of grade-level proficiency across states.

All achievement level setting activities for NAEP are performed in accordance with current best practices in standard setting and the Governing Board's *Developing Student Achievement Levels for the National Assessment of Educational Progress Policy Statement* (2018). The Governing Board policy does not extend to creating achievement level descriptions for performance below the *NAEP Basic* level.

Achievement level descriptions specific to the NAEP Science Framework were developed to elaborate on the generic definitions. Exhibit 4.1 presents the generic policy definitions. See Appendix A for the achievement level descriptions that illustrate how the policy definitions apply to NAEP Science for grades 4, 8, and 12.

Exhibit 4.1. Generic Achievement Level Policy Definitions for NAEP

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

4C. Contextual Variables

NAEP legislation requires reporting according to various student populations (see section 303[b][2][G]), including

- gender,
- race/ethnicity,
- eligibility for free/reduced-price lunch,
- students with disabilities, and
- English language learners.

NAEP users mistakenly may presume that the categories used to report data are related to causal explanations for observed differences (e.g., that gender predicts or explains performance differences or "achievement gaps"). However, scholars find that these differences reflect gaps in students' opportunities to learn. When results are interpreted in ways that emphasize achievement gaps without attending to opportunity gaps, score differences across subgroups of students can be misinterpreted as differences in student ability rather than as differences due to unequal educational opportunities.

The *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014) recommend that reports of group differences in assessment performance be accompanied by relevant contextual information, where possible, to both discourage erroneous interpretation and enable meaningful analysis of the differences. That standard reads as follows:

Reports of group differences in test performance should be accompanied by relevant contextual information, where possible, to enable meaningful interpretation of the differences. If appropriate contextual information is not available, users should be cautioned against misinterpretation. (Standard 13.6)

Contextual data about students, teachers, and schools are needed to fulfill the statutory requirement that NAEP include information, whenever feasible, that promotes meaningful interpretation of NAEP results. Contextual variables are selected to be of topical interest, timely, and directly related to academic achievement and current trends and issues in science. In the past, a range of information has been collected as part of NAEP.

4D. Science-Specific Contextual Variables

As noted in Chapter 1, research has informed an expanded view of the factors that shape opportunities to learn, including time, content and practices, instructional strategies (e.g., how students are grouped for learning; the scientific tasks they engage in; the opportunities students have to reason, model, and debate ideas), and instructional resources (e.g., human, material, and social resources that shape student access to science).

For example, research has demonstrated that what students learn is shaped by the availability of various science programs, curricula, extracurricular activities geared toward science, the percentage of teachers certified in science subjects, teacher years of experience, percentage of science teachers on an emergency license or vacancies/substitute teachers in the school, and number of teachers with science degrees, among other factors. Teachers' and administrators' beliefs about what science is, how one learns science, and who can learn science also affect student learning. What students learn is shaped by their sense of identity and agency. Students who see themselves, and who are seen by others, as capable scientific thinkers are more likely to participate in ways that further their learning; students who do not see themselves, and are not seen by others, as capable scientific thinkers are likely to be disengaged. Steele, Spencer, and Aronson (2002), for example, found that even passing reminders that a student is a member of one group or another—often, in this case, a group that is stereotyped as intellectually or academically inferior—can undermine student performance.

There are countless factors that shape what and when students learn. The NAEP Science student, teacher, and administrator surveys cannot possibly cover all such factors. Even though it would be helpful to ask students and teachers the same questions, this is also not possible given time constraints. Furthermore, questions about some factors may not be appropriate in the NAEP context. Given the constraints, not all topics can be addressed.

To support prioritization and ensure that NAEP results have appropriate context for interpretation, this framework sets the following topics to receive the greatest emphasis in the 2028 NAEP Science Assessment's contextual questionnaires (in order of priority).

- *Science content*. The 2028 NAEP Science Framework conceptualizes science content as disciplinary concepts, science and engineering practices, and crosscutting concepts. Therefore, contextual variables related to science content are expanded to include reference to NAEP Science and Engineering Practices and NAEP Crosscutting Concepts as well. Interpreting students' achievement requires a basic understanding of what science disciplinary concepts, science and engineering practices, and crosscutting concepts students have engaged with. Given variation across states in standards and frameworks, this information is crucial.
- *Teacher factors*. Research demonstrates that teacher quality is a critical in-school factor in predicting student achievement. This framework prioritizes the collection of data on teacher preparation and professional development, as well as teacher science knowledge for teaching, and for elementary students teacher confidence in teaching science topics.
- *Student science identity*. Research demonstrates that students' perceptions of their science identity directly relate to their learning. This framework prioritizes gathering information about students' science identities through questions that address student participation in activities such as discussion of phenomena, science ideas or evaluation of how a science problem or investigation is framed.
- *Instructional resources*. A range of resources influences instruction, including instructional leadership, additional instructional personnel, time, technology, curriculum, and materials. This framework prioritizes gathering information about school resources

that can inform the interpretation of results, including the time devoted to science teaching and learning in school, across current and prior grade levels, and the curricular and instructional materials at teachers' and students' disposal to support learning. In terms of technology, questionnaires will capture what technology is available to support science and engineering teaching and learning and how it is used.

- *Instructional organization and strategies*. Interpreting student achievement levels will also depend on understanding the instructional strategies used in science class, including collaborating in small-group work, engaging in science discussions, working hands-on and using grade appropriate measurement and data-analysis tools, and using a range of methods and tools to represent and model science phenomena and engineering design problems. This framework prioritizes gathering information both on the organization of classrooms and on the instructional routines and approaches that teachers use. It also includes what technologies and assessment approaches are used in instruction.

4E. Conclusion

As the Nation's Report Card, NAEP reports on student performance over time, presenting an analysis of national trends in students' science achievement. The 2028 NAEP Science Assessment is designed to assess the achievement of groups of students through robust and challenging assessments that are well aligned with current understanding of the three dimensions of science to be learned and that use technology in ways that maximize both student engagement and accessibility. The results of the assessment are informed by data on contextual variables that illuminate potential differences in opportunities to learn for students.

The ultimate goal of our nation's schools is to ensure that every student has access to learning high-quality science. NAEP plays an important role in providing a broad picture of students' knowledge and skills in science to the nation. NAEP scores, illuminated by relevant contextual information, can provide the public, families, students, and schools useful data on student performance that complements information provided by state tests that are more tightly aligned with specific state standards. As a view of present trends, it provides invaluable data to inform policy and practice in the future.

APPENDIX A: Achievement Level Descriptions

NAEP Grade 4 Science Achievement Level Descriptions		
NAEP Basic Level	**NAEP Proficient Level**	**NAEP Advanced Level**
Students should be able to demonstrate partial mastery and competency of the knowledge needed to understand science disciplinary concepts and the use of science practices and crosscutting concepts to evaluate the merits of the reasoning and interpretation of real-world situations. Students performing at this level should be able to perform tasks that focus on disciplinary concepts such as: • different types of matter (materials) have different properties, • balanced forces acting on an object keep the object in place, • water and light are needed for a plant's growth and survival, • the location of rocks and fossils can be used to establish Earth's history, • natural processes such as weathering change Earth's surface features, • humans can cause changes to the local areas where they live.	Students should be able to demonstrate solid academic performance and competency of the knowledge needed to understand relationships among closely related science disciplinary concepts and the use of science practices and crosscutting concepts to analyze and evaluate the merits of the reasoning and interpretation of real-world situations. Students performing at this level should be able to perform tasks that focus on disciplinary concepts such as: • matter (materials) can be classified based on its properties, • unequal forces acting on an object can change its motion, • varying amounts of water and light may affect a plant's growth and survival, • fossils can provide evidence for the nature of an environment where organisms lived long ago, • some short-term observable changes to Earth's surface features can be caused by wind or water,	Students should be able to demonstrate superior performance and competency of the knowledge needed to understand relationships among closely related science disciplinary concepts and the use of science practices and crosscutting concepts to analyze and evaluate the merits of explanations or predictions as applied to real world situations. Students performing at this level should be able to perform tasks that focus on disciplinary concepts such as: • matter (materials) with different properties have different uses, • two objects exert forces on each other even when the objects do not touch, • all organisms require food for growth and survival and that some animals obtain food from plants or from other animals, • the location of fossils within rock strata can be used to show the changes that occurred to Earth and

	• human activities can impact the land, water, and air.	life on Earth over time, • changes to Earth's surface features result from the action of water, wind, or living organisms, • human activities may have positive or negative impacts on the land, water, and air.
Students require a well-defined set of actions to be able to apply science practices and crosscutting concepts such as: • identifying testable and non-testable questions, • using simple models to explain a phenomenon, • selecting tools that are appropriate for the investigation of the flow of matter and energy through a system, • identifying an evidence-based argument about the conditions that lead to the changes of a system, • describing quantitative, measurable evidence needed to answer a scientific question, • using evidence to support the solution to a problem while considering the criteria that the solution should meet,	Students require substantial cues to be able to apply science practices and crosscutting concepts such as: • asking questions to refine observations about a system, • describing how the parts of a model represent the organization of a system, • describing observations or measurements that can be used as evidence to explain the flow of matter or energy through a system, • evaluating the merits of an evidence-based argument about the conditions that lead to the stability or the changes of a system, • organizing data sets to reveal patterns that can be used to answer a scientific question, • making a claim about the solution to a	Students require limited cueing to be able to apply science practices and crosscutting concepts such as: • asking questions to investigate cause-and-effect relationships about an observed phenomenon, • identifying the limitations of a model used to represent a phenomenon, • predicting the outcome of an experiment designed to explore how changes to the flow of matter or energy would affect a system, • comparing evidence-based arguments about the conditions that lead to the stability, or the changes of a system based on the evidence or the reasoning they include, • estimating or predict data points using

• using information from a variety of sources including written text, tables, diagrams, and/or charts to construct simple scientific explanation.	problem using evidence while considering criteria and constraints, • combining information in written text, tables, diagrams, and/or charts to describe patterns.	patterns in recorded data to further support the answer to a scientific question, • proposing a solution to a problem using evidence while considering how to prioritize criteria and constraints, • combining information in written text, tables, diagrams, and/or charts to describe cause-and-effect relationships.
In **physical science**, students should be able to integrate disciplinary concepts, science practices, and crosscutting concepts to engage in the *simple sensemaking* of phenomena when provided *a well-defined set of actions* to perform tasks such as: • using evidence to describe how temperature affects the physical state of a material, • proposing appropriate variables and tests when planning an investigation about the forces that objects exert on each other when they collide, • organizing data to reveal patterns related to the motion of an object and its energy, • asking questions to clarify the relationship between the parts of a	In **physical science**, students should be able to integrate disciplinary concepts, science practices, and crosscutting concepts to engage in the *moderate degree of sensemaking* of phenomena when provided *substantial cues* to perform tasks such as: • making a claim using data about how temperature affects the physical state of a material, • planning an investigation considering the variables to control and/or the number of trials to conduct to produce data to explore a scientific question about the forces that objects exert on each other when they are touching or colliding,	In **physical science**, students should be able to integrate disciplinary concepts, science practices, and crosscutting concepts to engage in the *high degree of sensemaking* of phenomena when provided *limited prompting* to perform tasks such as: • evaluating a claim based on the evidence or reasoning it includes about how temperature affects the physical state of a material, • making predictions using patterns in the data from an investigation about the forces that objects exert on each other when they collide, • analyzing patterns in data gathered for two different objects to explain the relationship between

simple model used to represent how an object can be seen only when light produced by the object or reflected from its surface enters the eyes.	• describing patterns in data to support the claim that the motion of an object is related to its energy, • asking questions to evaluate a model that represents how objects can be seen only when light produced by the object or reflected from its surface enters the eyes.	the motion of an object and its energy, • asking questions to identify the limitations of a model that represents how objects can be seen only when light reflected from its surface enters the eyes.
In **life science**, students should be able to integrate disciplinary concepts, science practices, and crosscutting concepts to engage in the *simple sensemaking* of phenomena when provided *a well-defined set of actions* to perform tasks such as: • communicating information about the life cycle of a plant, • use data to describe how organisms obtain the materials they need to grow and survive from the environment, • identifying evidence to support claims about how the environment can change the characteristics of an organism, • identifying data from tables or graphical displays to describe the relationship between the characteristics of organisms and their	In **life science**, students should be able to integrate disciplinary concepts, science practices, and crosscutting concepts to engage in the *moderate degree of sensemaking* of phenomena when provided *substantial cues* to perform tasks such as: • communicating information about the diverse life cycles of plants or animals, • analyzing data to explain that organisms obtain the materials they need to grow and survive from the environment, • making evidence-based claims about how the environment can change the characteristics of an organism, • representing data in tables or graphical displays to reveal patterns between the characteristics of organisms and their	In **life science**, students should be able to integrate disciplinary concepts, science practices, and crosscutting concepts to engage in the *high degree of sensemaking* of phenomena when provided *limited prompting* to perform tasks such as: • evaluating information from two or more sources to compare the life cycles of plants or animals, • evaluating data that can be used to support a claim about the types of materials that organisms obtain from the environment for their growth and survival, • evaluating the evidence to support claims about changes to the characteristics of an organism caused by the environment, • analyzing data to reveal cause-and-effect relationships

ability to survive, mate, or reproduce.	ability to survive, mate, or reproduce.	between the characteristics of organisms and their ability to survive, mate, or reproduce.
In **Earth and space sciences**, students should be able to integrate disciplinary concepts, science practices, and crosscutting concepts to engage in the *simple sensemaking* of phenomena when provided *a well-defined set of actions* to perform tasks such as: • using a simple model to explain why certain objects in the sky are not always visible from Earth, • identifying evidence to support arguments about how Earth and life on Earth has changed over time, • using patterns in data to describe the weather event that occurred in a region, • identify evidence to support a claim about how natural processes can cause hazards in some areas.	In **Earth and space sciences**, students should be able to integrate disciplinary concepts, science practices, and crosscutting concepts to engage in the *moderate degree of sensemaking* of phenomena when provided *substantial cues* to perform tasks such as: • using data to develop a model to represent why certain objects in the sky are not always visible due to Earth's rotation, • making arguments based on evidence about how Earth and life on Earth has changed over time, • using patterns in data to make predictions about the kind of weather expected in a region, • making an argument based on evidence for how natural processes can cause hazards in some areas but not others.	In **Earth and space sciences**, students should be able to integrate disciplinary concepts, science practices, and crosscutting concepts to engage in the *high degree of sensemaking* of phenomena when provided *limited prompting* to perform tasks such as: • revising a model that represents why certain objects in the sky are not always visible due to Earth's rotation, • evaluating multiple arguments based on the evidence or reasoning they include about how Earth and life on Earth has changed over time, • analyzing patterns in data to identify the conditions that would lead to a change in the weather in a region, • using evidence to support claims about possible cause-and-effect relationships between natural processes and hazards that occur in some areas.

NAEP Grade 8 Science Achievement Level Descriptions		
NAEP Basic Level	**NAEP Proficient Level**	**NAEP Advanced Level**
Students should be able to demonstrate partial mastery and competency of the knowledge needed to understand science disciplinary concepts and the use of science practices and crosscutting concepts to evaluate the merits of the reasoning and interpretation of real-world situations. Students performing at this level should be able to perform tasks that focus on disciplinary concepts such as: • temperature influences the motion of atoms and/or molecules in any state of matter, • when the net force acting on an object is zero, an object in motion will continue moving in a straight line with the same speed as before, • photosynthetic organisms use energy from light to change inorganic matter into food, • fossil records document the existence and extinction of many life-forms throughout Earth's history, • the movement of water within the water cycle is a function of phase changes,	Students should be able to demonstrate solid academic performance and competency of the knowledge needed to understand relationships among closely related science disciplinary concepts and the use of science practices and crosscutting concepts to analyze and evaluate the merits of the reasoning and interpretation of real-world situations. Students performing at this level should be able to perform tasks that focus on disciplinary concepts such as: • the relative distance between atoms and/or molecules in a sample of matter is different for solids, liquids, and gases, • the change in motion of an object is determined by the sum of the forces acting on it, • photosynthetic organisms use energy from light and an input of carbon dioxide and water to make sugars and release oxygen, • the evolutionary history of some organisms can be reconstructed based on the similarities and differences in gross anatomical features between organisms	Students should be able to demonstrate superior performance and competency of the knowledge needed to understand relationships among closely related science disciplinary concepts and the use of science practices and crosscutting concepts to analyze and evaluate the merits of explanations or predictions as applied to real world situations. Students performing at this level should be able to perform tasks that focus on disciplinary concepts such as: • in a sample of mater, the forces acting between the atoms and/or molecules influence their relative separation, • the change in motion of an object is determined by the sum of the forces acting on it and if the net force on the object is zero, it will remain at rest or continue moving in a straight line with the same speed and direction as before, • photosynthetic organisms use energy from light and an input of carbon dioxide and water releasing oxygen and making

• human activities have significantly altered the biosphere, atmosphere, and geosphere.	living today and organisms in the fossil records, • the movement of water on land and underground can cause changes to the land on Earth's surface, • human activities have altered Earth systems sometimes damaging or destroying ecosystems.	sugars that will undergo a series of chemical reactions to form new molecules used to support growth or to release energy, • fossil records document the existence, diversity, extinction, and changes to many life-forms based on the similarities and differences in gross anatomical features between organisms living today and organisms in the fossil records, • the movement of water on land and underground is driven by gravity and can change the land on and below Earth's surface, • human activities have altered Earth systems sometimes damaging or destroying ecosystems, but human choices can minimize harm to other organisms and risks to the health of the regional environment.
Students require a well-defined set of actions to be able to apply science practices and crosscutting concepts such as: • asking questions that arise from observations of phenomena to clarify	Students require substantial cues to be able to apply science practices and crosscutting concepts such as: • asking questions to verify patterns in the data gathered from an investigation of a phenomena,	Students require limited cueing to be able to apply science practices and crosscutting concepts such as: • asking questions to refine an explanation of cause-and-effect relationships in phenomena,

the evidence for an argument, • developing a simple model of a system to explain a phenomenon, • selecting and evaluating tools to collect data, • using evidence to support an argument about the stability or the changes that a system undergoes, • applying simple mathematical concepts (such as basic operations and simple computations) to scientific problems, • constructing graphical displays of data to identify relationships between variables, • describing a solution to a problem using scientific principles while considering prioritized criteria, • assessing the credibility of an article on a science topic based on the information it provides.	• developing a model to explain a phenomenon by identifying relationships among parts and or quantities in a system, • planning an experimental design to produce data that can be used as evidence for the flow of matter and energy through a system, • evaluating the merits of opposing arguments about the stability or the changes of a system, • applying simple mathematical concepts (such as ratios or proportional thinking) to scientific or engineering problems to address scale and quantity, • using graphical displays of data to identify linear vs. nonlinear relationships between variables, • evaluating a solution to a problem using scientific principles while considering criteria and constraints, • evaluating the information from two different sources about the stability and change in natural or designed systems to determine whether there are conflicts between them.	• revising a model to explain a phenomenon by identifying relationships among parts and or quantities in a system, • revising an experimental design to produce data that can be used as evidence for the flow of matter and energy through a system, • revising an argument about the stability or the changes that a system undergoes to address new evidence, • applying simple mathematical concepts (such as ratios, rates, or percent) to scientific or engineering problems to address scale and quantity, • using graphical displays of data to identify causal vs. correlational relationships between variables, • evaluating the merits of a solution to a problem using evidence while considering criteria and constraints, • identifying flaws in science-related arguments about the stability and change in natural or designed systems due to poor assumptions.
In **physical science**, students should be able to	In **physical science**, students should be able to	In **physical science**, students should be able to

integrate disciplinary concepts, science practices, and crosscutting concepts to engage in *simple sensemaking* of phenomena when provided *a well-defined set of actions* to perform tasks such as: • identifying evidence-based associations between temperature, the motion of atoms and/or molecules, and the physical state of a sample of matter, • planning an investigation to collect data that can serve as evidence for the fact that when the net force on an object is zero, an object at rest will remain at rest and a moving object will keep moving in a straight line at the same initial speed and direction, • constructing graphical displays of data to identify the relationship between kinetic energy and the mass of a moving object, • asking questions based on observations for how the material an object is made of influences the reflection or transmission of light shining on the object.	integrate disciplinary concepts, science practices, and crosscutting concepts to engage in the *moderate degree of sensemaking* of phenomena when provided *substantial cues* to perform tasks such as: • constructing an explanation that uses a chain of cause-and-effect associations for how temperature influences the motion and the relative separation between the atoms and/or molecules in a sample of matter, • evaluating an experimental design to produce data that can serve as evidence for the relationship between the sum of the forces acting on an object and the speed at which the object moves, • analyzing graphical displays of data and/or large data sets from an investigation to identify linear or nonlinear relationships between kinetic energy, the mass of a moving massive object, or its speed, • asking questions that can be answered using empirical evidence for how the material an object is made of influences whether light shining on the object is	integrate disciplinary concepts, science practices, and crosscutting concepts to engage in the *high degree of sensemaking* of phenomena when provided *limited prompting* to perform tasks such as: • revising an explanation that uses a chain of cause-and-effect associations for how temperature influences the motion and the relative separation between the atoms and/or molecules in a sample of matter, • revising an experimental design to produce data that can serve as evidence for the relationship between the sum of the forces acting on an object and the speed and direction in which the object moves, • evaluate and/or revise graphical displays of data to describe linear and nonlinear relationships between kinetic energy, the mass of a moving object, and its speed, • asking questions to clarify evidence for how the material an object is made of, or the frequency of the light, influences the reflection, absorption, or transmission of the light shining on the object.

	reflected or transmitted.	
In **life science**, students should be able to integrate disciplinary concepts, science practices, and crosscutting concepts to engage in *simple sensemaking* of phenomena when provided *a well-defined set of actions* to perform tasks such as: • assessing the credibility of a source that compares the structures and processes used by organisms for asexual reproduction based on its qualifications, • identifying data that can provide evidence to support a model that demonstrates how matter is transferred between, producers, consumers, and decomposers, • identifying evidence to support an argument about how variations in inherited traits between parent and offspring arise from the subset of genes inherited, • identifying evidence to support a claim about the changes a species undergoes over time in response to changes in environmental conditions.	In **life science**, students should be able to integrate disciplinary concepts, science practices, and crosscutting concepts to engage in the *moderate degree of sensemaking* of phenomena when provided *substantial cues* to perform tasks such as: • assessing the credibility and accuracy of the scientific information in a source that compares the structures and processes used by organisms for sexual or asexual reproduction, • analyzing data to provide evidence to support or reject a model that demonstrates how matter and energy are transferred between, producers, consumers, and decomposers, • constructing an argument using evidence to explain how variations in inherited traits between parent and offspring arise from the subset of genes inherited, • constructing an explanation that uses a chain of cause-and-effect associations between the changes	In **life science**, students should be able to integrate disciplinary concepts, science practices, and crosscutting concepts to engage in the *high degree of sensemaking* of phenomena when provided *limited prompting* to perform tasks such as: • evaluating the information provided by two different sources that compare the structures and processes used by organisms for both sexual and asexual reproduction to determine whether they provide conflicting information, • evaluating data to provide evidence to support or reject a model that demonstrates that the atoms that make up the organisms in an ecosystem are cycled repeatedly between the living and nonliving parts of the ecosystem, • constructing an argument using evidence to explain that genetic mutations may result in changes in the structure and function of the proteins encoded by genes, • revising an explanation that uses a chain of cause-and-

	a species undergoes over time in response to changes in environmental conditions.	effect associations between the changes a species undergoes over time in response to changes in environmental conditions and that heritable traits that support successful survival and reproduction become more common.
In **Earth and space sciences**, students should be able to integrate disciplinary concepts, science practices, and crosscutting concepts to engage in *simple sensemaking* of phenomena when provided *a well-defined set of actions* to perform tasks such as: • using a model to describe observable patterns in the motion of objects in the sky relative to Earth, • identifying evidence that can be used to refute a claim about the relative times of major events in Earth's history based on fossil records, • constructing graphical displays of data to identify relationships between the interactions involving sunlight, the ocean, the atmosphere, or landforms and the weather patterns in a given location, • identifying evidence to support an argument for how observable	In **Earth and space sciences**, students should be able to integrate disciplinary concepts, science practices, and crosscutting concepts to engage in the *moderate degree of sensemaking* of phenomena when provided *substantial cues* to perform tasks such as: • developing a model to test ideas about observable patterns in the motion of objects in the sky relative to Earth, • making a claim about the relative time of major events in Earth's history based on fossil records, • interpreting graphical displays of data to identify relationships between the interactions involving sunlight, the ocean, the atmosphere, ice, or landforms and the weather patterns in a given location, • constructing an argument using evidence for how observable	In **Earth and space sciences**, students should be able to integrate disciplinary concepts, science practices, and crosscutting concepts to engage in the *high degree of sensemaking* of phenomena when provided *limited prompting* to perform tasks such as: • revising a model based on observable patterns in the motion of objects in the sky relative to Earth to make predictions about the future motion or positions of objects in the sky, • evaluating evidence that can be used to refute a claim about the relative time of major events in Earth's history based on fossil records and the sequence of rock strata, • evaluating the limitations of data presented in graphical displays to identify relationships between the interactions involving sunlight, the

phenomena that precedes the occurrence of some natural hazards can help forecast future events.	phenomena that precedes the occurrence of some natural hazards can help forecast future events in order to minimize risks.	ocean, the atmosphere, ice, or landforms and the weather patterns in a given location, • comparing and critiquing two arguments that explain how observable phenomena that precedes the occurrence of some natural hazards can help forecast future events and minimize risks to analyze their fit with the evidence or whether they emphasize similar or different evidence.

NAEP Grade 12 Science Achievement Level Descriptions		
NAEP Basic Level	**NAEP Proficient Level**	**NAEP Advanced Level**
Students should be able to demonstrate partial mastery and competency of the knowledge needed to understand science disciplinary concepts and the use of science practices and crosscutting concepts to evaluate the merits of the reasoning and interpretation of real-world situations. Students performing at this level should be able to perform tasks that focus on disciplinary concepts such as: • all matter is made of atoms, • the motion of an object changes only if	Students should be able to demonstrate solid academic performance and competency of the knowledge needed to understand relationships among closely related science disciplinary concepts and the use of science practices and crosscutting concepts to analyze and evaluate the merits of the reasoning and interpretation of real-world situations. Students performing at this level should be able to perform tasks that focus on disciplinary concepts such as:	Students should be able to demonstrate superior performance and competency of the knowledge needed to understand relationships among closely related science disciplinary concepts and the use of science practices and crosscutting concepts to analyze and evaluate the merits of explanations or predictions as applied to real world situations. Students performing at this level should be able to perform tasks that focus on disciplinary concepts such as:

the sum of the forces acting on the object is non-zero, • photosynthesis converts light energy to stored chemical energy, • DNA sequences vary among species but there are many overlaps, • the decay of radioactive isotopes in rocks provides a way to date rock formations, • water's unique properties can help explain weathering, • humans can mitigate negative impacts on Earth's resources through the responsible monitoring and management of natural resources.	• all matter is made of atoms that contain protons, neutrons, and electrons, • momentum is always conserved, • cellular respiration is a chemical process in which the chemical bonds of food molecules and oxygen molecules are broken forming new compounds that can transport energy, • genetic information can be derived from similarities and differences in amino acid sequences, • the decay of radioactive isotopes in rocks from Earth, moon rocks, and meteorites provides a way to date rock formations that can be used as evidence for Earth's formation, • water's unique properties can help explain the erosion of landforms and deposition of sediments, • humans can mitigate negative impacts on Earth's resources and global environment through the responsible monitoring and management of natural resources.	• the electrostatic forces between subatomic particles explain both the structure of isolated atoms, and why atoms combine to form molecules, compounds, and extended materials, • momentum is always conserved because the forces between any two interacting objects are equal and opposite and thus result in equal and opposite changes in momentum, • matter and energy are conserved in photosynthesis and cellular respiration, • genetic information can be derived from similarities / differences in amino acid sequences and from anatomical and embryological evidence, • the decay of radioactive isotopes in rocks from Earth, moon rocks, and meteorites provides a way to date rock formations that can be used as evidence for Earth's formation and early history, • water's unique properties can help explain erosion, deposition, frost wedging and their effects on landforms, • humans can mitigate negative impacts on

		Earth's resources and global environment through the responsible monitoring and management of natural resources but when the sources of such problems are not well understood, some actions could magnify the problems.
Students require a well-defined set of actions to be able to apply science practices and crosscutting concepts such as: • asking questions that arise from examining an explanation of a phenomenon, • using a simple model of a system that includes mathematical relationships to describe phenomena, • planning an investigation that will produce data that can support the scientific explanation of a phenomenon, • using evidence to support an argument about a proposed explanation for structure-function relationships in a system, • identifying mathematical algorithms when analyzing data that show quantitative relationships between variables, • analyzing representations of	Students require substantial cues to be able to apply science practices and crosscutting concepts such as: • asking questions that arise from examining patterns in data to identify additional evidence needed to support the pattern, • developing a simple model of a system that includes scale, proportion, and mathematical relationships to describe phenomena, • planning an investigation that considers the appropriate variables to control to produce data that can be used as evidence for cause-and-effect relationships in a phenomenon, • constructing an argument with evidence and scientific reasoning to support a proposed explanation for structure-function	Students require limited cueing to be able to apply science practices and crosscutting concepts such as: • asking investigable questions to frame a hypothesis about cause and effect relationships in a phenomenon, • revising a model of a system that includes scale, proportion, and mathematical relationships to explain phenomena, • evaluating the design of an investigation to produce data that can be used as evidence for cause-and-effect relationships in a phenomenon considering possible confounding variables, • revising an argument to support or reject a proposed explanation for structure-function relationships in a system to address new evidence, • interpreting and applying mathematical

data sets from an investigation using tools or technologies and including probability, • identifying evidence-based relationships between variables to support an explanation of a phenomenon, • using disciplinary concepts to identify standard flaws in science-related arguments due to faulty explanations.	relationships in a system, • applying simple statistical reasoning to represent and solve scientific questions or reveal patterns in data, • analyzing patterns in data to provide evidence to support or reject a model, • constructing an explanation of a phenomenon that uses a chain of evidence-based associations between variables in a phenomenon, • using multiple scientific sources to obtain evidence to support an argument describing cause-and-effect relationships in a phenomenon.	concepts or processes in the context of complicated measurement problems to represent and solve scientific questions or reveal patterns in data, • analyzing data to provide evidence of cause-and effect relationships to support or reject a model, • revising an explanation of a phenomenon that uses a chain of cause-and-effect associations between factors to account for relationships between variables in a phenomenon, • evaluating a science-related argument to identify standard flaws related to cause vs. correlation.
In **physical science**, students should be able to integrate disciplinary concepts, science practices, and crosscutting concepts to engage in *simple sensemaking* of phenomena when provided *a well-defined set of actions* to perform tasks such as: • identifying data to support an explanation for the relationship between temperature and the motion and rate of	In **physical science**, students should be able to integrate disciplinary concepts, science practices, and crosscutting concepts to engage in the *moderate degree of sensemaking* of phenomena when provided *substantial cues* to perform tasks such as: • using patterns in data to support an explanation of the relationship between temperature, the rate of collisions between	In **physical science**, students should be able to integrate disciplinary concepts, science practices, and crosscutting concepts to engage in the *high degree of sensemaking* of phenomena when provided *limited prompting* to perform tasks such as: • analyzing data to reveal patterns to support or reject an explanation of the relationship between temperature, and the

collisions between atoms and/or molecules in liquids or gases, • selecting the appropriate tools to collect data that can serve as quantitative evidence for the relationship between the relative magnitudes of two aligned forces acting on an object and the change in motion of the object, • constructing representations of data sets to identify the relationship between the energy available within a system and the motion and interactions of matter and radiation within that system, • identifying scientific questions that arise from examining models to explain the relationship between frequency, wavelength, and the speed at which waves travel through different media.	atoms and/or molecules in liquids or gases and the effect on chemical reactions, • identify control variables and confounding variables in the design of an investigation about the relationship between the relative magnitudes of two aligned forces acting on an object and the change in motion of the object, • analyzing representations of data to reveal patterns that can serve as evidence to explain how the energy available within a system depends on the motion and interactions of matter and radiation within that system, • asking questions that arise from examining models about the relationship between frequency, wavelength, and the speed at which waves travel through different media.	motion and rate of collisions between atoms and/or molecules in liquids or gases and the effect on chemical reactions, • planning an investigation to produce data that can serve as evidence for an explanation of the relationship between the relative magnitudes of two aligned forces acting on an object and the change in motion of the object, • analyzing data to provide evidence to support of reject a model that illustrates how the energy available within a system depends on the motion and interactions of matter and radiation within that system, • evaluating questions about the relationship between frequency, wavelength, and the speed at which waves travel through different media to determine whether they are investigable.
In **life science**, students should be able to integrate disciplinary concepts, science practices, and crosscutting concepts to engage in *simple sensemaking* of phenomena when provided *a well-defined set of actions* to perform tasks such as:	In **life science**, students should be able to integrate disciplinary concepts, science practices, and crosscutting concepts to engage in the *moderate degree of sensemaking* of phenomena when provided *substantial cues* to perform tasks such	In **life science**, students should be able to integrate disciplinary concepts, science practices, and crosscutting concepts to engage in the *high degree of sensemaking* of phenomena when provided *limited prompting* to perform tasks such as:

• assessing the credibility of a source that explains the role of meiosis in genetic variation based on the evidence it provides, • identifying data that can provide evidence to support a model that illustrates that only a small fraction of the matter consumed at lower trophic levels is transferred to upper trophic levels in a food web, • identifying evidence to support an argument about how DNA codes for the formation of proteins that determine traits, • identifying data to provide evidence to support an explanation about how natural selection can result from competition for resources.	as: • evaluating the scientific information provided in multiple sources that explain the role of meiosis in genetic variation based on the evidence they provide, • analyzing data to provide evidence to support or reject a model that illustrates that only a small fraction of the matter consumed at lower trophic levels is transferred to upper trophic levels in a food web, • using patterns in DNA sequences to construct an argument about how DNA codes for the formation of proteins that determine traits, • analyzing patterns in data to provide evidence to support an explanation about how natural selection can result from competition for resources.	• Identifying and critiquing standard flaws in science-related arguments about how environmental factors affect heritable traits and the probability of occurrence of those traits in a population due to faulty evidence, • evaluating whether the data available is enough to support or reject a model that illustrates that only a small fraction of the matter consumed at lower trophic levels is transferred to upper trophic levels in a food web, • using patterns in DNA sequences to revise an argument to support or reject an explanation about how DNA codes for the formation of proteins that determine traits, • evaluating whether the quality of the data used as evidence is sufficient to support an explanation about how natural selection can result from competition for resources.
In **Earth and space sciences**, students should be able to integrate disciplinary concepts, science practices, and crosscutting concepts to engage in *simple sensemaking* of phenomena	In **Earth and space sciences**, students should be able to integrate disciplinary concepts, science practices, and crosscutting concepts to engage in the *moderate degree of sensemaking* of	In **Earth and space sciences**, students should be able to integrate disciplinary concepts, science practices, and crosscutting concepts to engage in the *high degree of sensemaking* of phenomena

when provided *a well-defined set of actions* to perform tasks such as: • using data to describe how the changes in the orientation of Earth's axis of rotation have altered the intensity and distribution of sunlight falling on the planet, • identifying evidence that can be used to support a claim about how dating rock formations provide evidence of Earth's early history, • using models to describe how ocean and atmospheric circulations influence weather, • identifying evidence to support an argument for how the size and location of human populations have been impacted by natural hazards.	phenomena when provided *substantial cues* to perform tasks such as: • using patterns in global temperature data to explain how the changes in the shape of Earth's orbit and the orientation of its axis of rotation have altered the intensity and distribution of sunlight falling on the planet, • making a claim supported with evidence about how measurements of the decay of radioactive elements in minerals and rocks provide evidence of Earth's early history, • developing models to explain how ocean and atmospheric circulations influence climate on a global scale, • constructing an argument supported by evidence for how the size and location of human populations have been impacted by natural hazards.	when provided *limited prompting* to perform tasks such as: • analyzing global temperature data to reveal patterns that can support a claim about how the changes in the shape of Earth's orbit and the orientation of its axis of rotation have altered the intensity and distribution of sunlight falling on the planet, • revising a claim supported with evidence about how measurements of the decay of radioactive elements in minerals and rocks provide evidence for Earth's formation, • using models to explain how the absorption, reflection, storage, and redistribution of energy from the Sun that reaches the Earth's surface lead to temporal patterns in Earth's climate, • revising an argument to support or reject an explanation for how the size and location of human populations have been impacted by natural hazards to address a counterclaim.

APPENDIX B: Sample Items

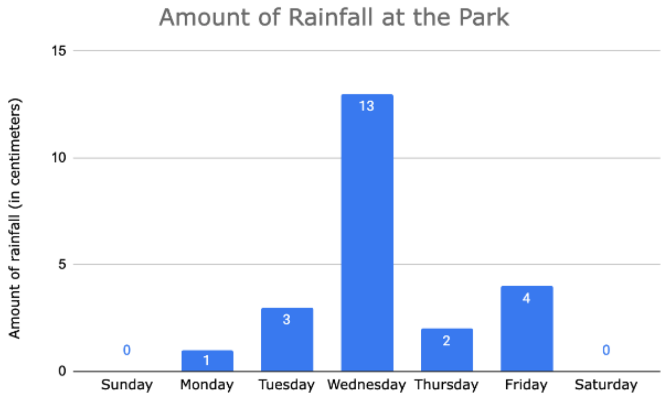
Discrete Items: Single and Multipart

Item ID: Park Flooding (adapted from FABLES)

- **Grade and discipline**: 4th grade, ESS
- **Item type**: Single part, Single Select Multiple Choice
- **Alignment**: this item is a **2D item**, measuring parts of:
    - DC: E4.9: Patterns in when and where weather conditions occur can be used to make predictions about the kind of weather that can be expected in a region
    - SEP: S4.13: Analyze data to determine whether it supports or refutes a claim about a phenomenon or design solution.
- **Item-level claim (derived from targeted dimensions):** Students can analyze and interpret simple data to determine patterns in weather conditions to make a claim about a phenomenon.
- **Phenomenon/uncertainty**: a park flooded one day when it was raining, but not others
- **Complexity**: This item is an example of a Low DC, Low SEP complexity item.
- **Additional detail**: This item requires students to make sense of a very simple phenomenon: a local park that has flooded. The phenomenon is presented through simple text, an image, and a simple graph—this provides students with enough information to demonstrate the targeted DC and SEP in service of sensemaking, without unnecessary reading or cognitive load. This phenomenon is an example of an everyday phenomenon that many students may have directly experienced, or have sufficient experiences to understand.

Exhibit B.1. Park Flooding, Version 1

People visiting a local park one day were surprised to find out that the park was flooded, and will be closed for the day. Use the data below to figure out what caused flooding on that day.



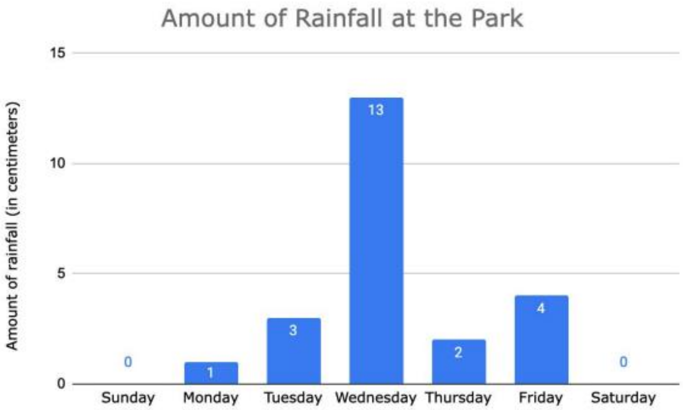
Based on the data and your understanding of weather, which day was most likely the day that the park flooded?

a. Sunday
b. Tuesday
c. Wednesday
d. Thursday

The following version of this item shows a modification that provides additional evidence of student understanding of both the DC and SEP in service of sensemaking. The additional component of the discrete item does not change the alignment or complexity of the item, but does add time to complete and some additional reading load. Over the range of the assessment, there may be some times that the trade-off of more comprehensive evidence is worth the additional time needed to complete the items. Many of the items that follow involve multiple parts to illustrate how to capture more evidence of student understanding through discrete items; however, many of these items could be limited to the first part to serve as a single part discrete item if test developers determine that this would provide sufficient evidence of student understanding.

Exhibit B.2. Park Flooding, Version 2

People visiting a local park one day were surprised to find out that the park was flooded, and will be closed for the day. Use the data below to figure out what caused flooding on that day.



Part A. Based on the graph, which day do you think the park flooded?
- A. Sunday
- B. Monday
- C. Tuesday
- D. Wednesday

Part B: Which piece of evidence best supports your answer in part A?
- A. It rained more on Wednesday than any other day in the week.
- B. It was rainy on Tuesday
- C. It did not rain on Saturday or Sunday
- D. It started raining on Monday

Item ID: Cleopatra's Needle, Version 1 (Adapted from a released NAEP item)
- **Grade and discipline:** 8th grade, ESS
- **Item type:** Selected response, In-line choice
- **Alignment:** this item is a **2D item**, measuring parts of:
    - DC: E8.7: Water continually cycles within and among land, ocean, and atmosphere. Water's movements, both on the land and underground, are driven by gravity and change the land on and below Earth's surface.
    - SEP: S4.13: Analyze data to determine whether it supports or refutes a claim about a phenomenon or design solution.
- **Item-level claim (derived from targeted dimensions):** Students can analyze and interpret simple data about weather conditions to make a claim about a phenomenon, using their understanding of how water changes rocks on Earth's surface.
- **Phenomenon/uncertainty:** a rock structure showed little evidence of deterioration for a very long time under some weather conditions, but began crumbling rapidly under others.

- **Complexity:** This item is an example of a Low DC, Low SEP complexity item.
- **Additional detail:** This item requires students to make sense of a simple phenomenon: that Cleopatra's needle began deteriorating rapidly when it was moved to a new climate, despite having stood without damage for thousands of years prior. The phenomenon is presented through simple text, an image, and a very simple data set—this provides students with enough information to demonstrate the targeted DC and SEP in service of sensemaking, without unnecessary reading or cognitive load. The context presents only simple data that are directly relevant to determining the answer, ensuring relatively low cognitive complexity. This phenomenon leverages a simple observation that includes ties to other countries and historical periods and figures who are important to many specific communities around the world.

Exhibit B.3. Cleopatra's Needle, Version 1

Cleopatra's Needle is a large monument made of rock that stood in Alexandria, a city in the Egyptian desert, for thousands of years. Then it was moved to New York City's Central Park in 1881. After only a few years in New York, its surface began crumbling. Why would the monument begin crumbling so quickly, when it remained intact for so long?

Alexandria, Egypt

New York City, USA

Table 1. Weather data in Alexandria and New York City

	Average amount of rainfall per year (inches)	Average wind speed over a year (miles/hour)
Alexandria, Egypt	7.1	7.1
New York City, USA	44.8	6.8

Use the picture, the data in Table 1 to decide which answer best explains why Cleopatra's needle started crumbling after it was moved to New York City from Alexandria.

(More/Less) rainfall in New York City over a year leads to (more/less) rock break down, causing the changes we see to Cleopatra's needle.

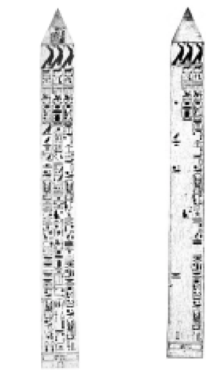
Item ID: Cleopatra's Needle, Version 2 (Adapted from a released NAEP item)

- **Grade and discipline**: 8th grade, ESS
- **Item type**: Multi-part discrete item; Constructed response, short text, extended text
- **Alignment**: this item is a **2D item**, measuring parts of:
    - DC: E8.7: Water continually cycles within and among land, ocean, and atmosphere. Water's movements, both on the land and underground, are driven by

gravity and change the land on and below Earth's surface.
    - SEP: S8.19: Construct or revise an explanation that uses a chain of cause and effect or evidence-based associations between factors to account for the qualitative or quantitative relationships between variables in a phenomenon.
- **Item-level claim (derived from targeted dimensions):** Students can make and support a claim about a phenomenon with evidence from data about weather conditions and reasoning about how water changes Earth's surface.
- **Phenomenon/uncertainty:** a rock structure showed little evidence of deterioration for a very long time under some weather conditions, but began crumbling rapidly under others.
- **Complexity**: This item is an example of a Low DC, medium SEP complexity item.
- **Additional detail**: This item uses the same context and phenomenon as the previous item, but engages students in deeper uses of SEPs by requiring students to make and support an original claim based on data and their understanding of erosion and weathering. The DC used is relatively low level, but the SEP requires students to use multiple components of SEPs in service of sensemaking. The inclusion of Part B acts as both a scaffold and a way to surface understanding from students without extended writing.

Exhibit B.4. Cleopatra's Needle, Version 2

Cleopatra's Needle is a large monument made of rock that stood in the city of Alexandria, in the Egyptian desert for thousands of years. Then it was moved to New York City's Central Park in 1881. After only a few years in New York, its surface began crumbling. Why would the monument begin crumbling so quickly, when it remained intact for so long?



Alexandria, Egypt　New York City, USA

Table 1. Weather data in Alexandria and New York City

	Average amount of rainfall per year (inches)	Average wind speed over a year (miles/hour)
Alexandria, Egypt	7.1	7.1
New York City, USA	44.8	6.8

Part A. Make a claim about why Cleopatra's needle is crumbling in New York City, but did not crumble in Alexandria.

Part B. Select one piece of evidence that supports your claim.

a. There is more precipitation in New York City than in Alexandria.
b. There is less precipitation in New York City than in Alexandria.
c. The photograph of Cleopatra's Needle shows more weathering in Alexandria than in New York City.
d. There are lower wind speeds in New York City than in Alexandria

Part C. Explain how the evidence you selected supports your claim. Use your understanding of how water affects rocks.

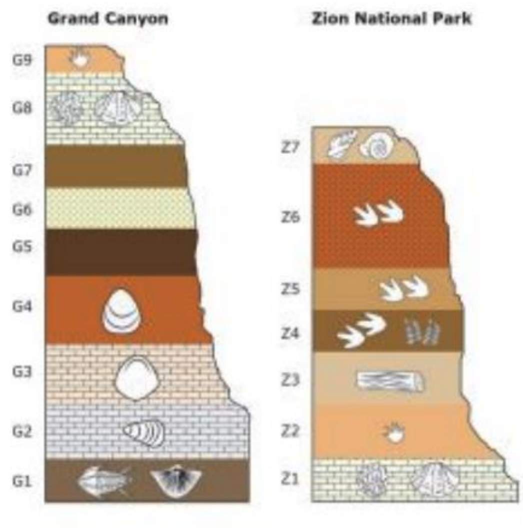
Item ID: Colorado Plateau, Version 1 (Adapted from New Meridian)

- **Grade and discipline**: 8th grade, ESS
- **Item type**: Multi-part, Selected Response, Single Select Multiple Choice and multi-select multiple choice

- **Alignment**: this item is a **2D item**, measuring parts of:
    - DC: E8.5: The geologic time scale interpreted from fossils and the sequence of rock strata provides a way to reconstruct how and when major events in Earth's history in terms of relative time.
    - SEP: S.8.12: Construct, analyze, and/or interpret graphical displays of data and/or large data sets from an investigation (e.g., maps, charts, graphs, and/or tables) to identify relationships between variables (linear vs. nonlinear relationships, causal vs. correlational relationships, and temporal and spatial relationships).
- **Item-level claim (derived from targeted dimensions):** Students can interpret visual representations of fossils and sequence of rock strata to reconstruct relative timelines of phenomena.
- **Phenomenon/uncertainty:** There are multiple rock structures in the Colorado Plateau that have some similar and some different rock layers and fossils—when did they develop relative to one another?
- **Complexity**: This item is an example of a Low DC, low SEP complexity item.
- **Additional detail**: This item requires students to make sense of a simple phenomenon: different rock structures have different layers and fossils within a region. The phenomenon is presented through simple text and an image that serves to both visually represent the phenomenon as well as provide data for interpretation, limiting the amount of time and interpretation students need to engage. The context presents only simple data that are directly relevant to determining the answer, and asks students to only consider two rock formations with a straightforward rock layer/fossil relationship, ensuring relatively low cognitive complexity. Students are directed to choose layers as evidence, and the options within the MC items provide additional direction for how to engage the SEP and DC. This phenomenon leverages a simple real-world observation that is authentically of interest to scientific communities.

Exhibit B.5. Colorado Plateau, Version 1

The Grand Canyon and Zion Canyon are land areas on the Colorado Plateau. Scientists are interested in studying and comparing these areas to better understand how the Earth has changed over very long periods of time.



Part A. Based on the rock layers and fossil data, which national park is older?

A. Grand Canyon
B. Zion National Park

Part B. Which rock layers provide evidence to support your choice?

a. G4 and Z6
b. G7 and Z4
c. G9 and Z2
d. G1 and Z7

Item ID: Colorado Plateau, Version 2 (Adapted from New Meridian)

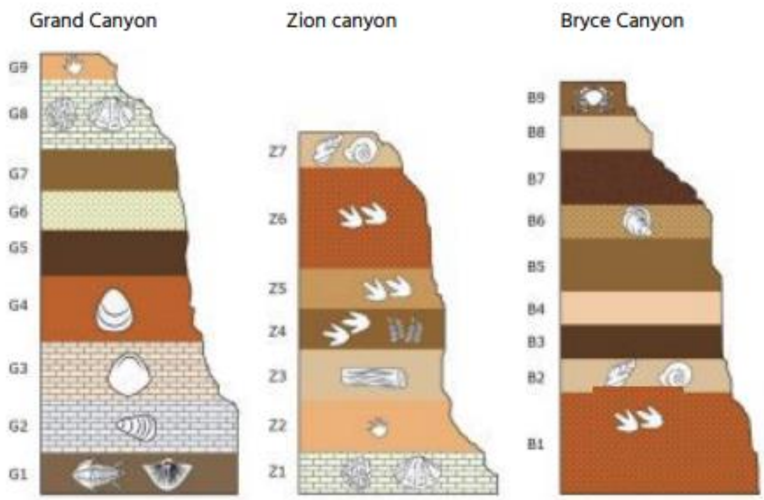
- **Grade and discipline**: 8th grade, ESS
- **Item type**: Multi-part, order items (or drag and drop), constructed response, short text.
- **Could part A be used as a stand-alone item?** Yes.
- **Alignment**: this item is a **2D item**, measuring parts of:
  - DC: E8.5: The geologic time scale interpreted from fossils and the sequence of rock strata provides a way to reconstruct how and when major events in Earth's history in terms of relative time.
  - SEP: S.8.12: Construct, analyze, and/or interpret graphical displays of data and/or large data sets from an investigation (e.g., maps, charts, graphs, and/or tables) to identify relationships between variables (linear vs. nonlinear relationships, causal vs. correlational relationships, and temporal and spatial relationships).
- **Item-level claim (derived from targeted dimensions):** Students can interpret visual

representations of fossils and sequence of rock strata to reconstruct relative timelines of phenomena.

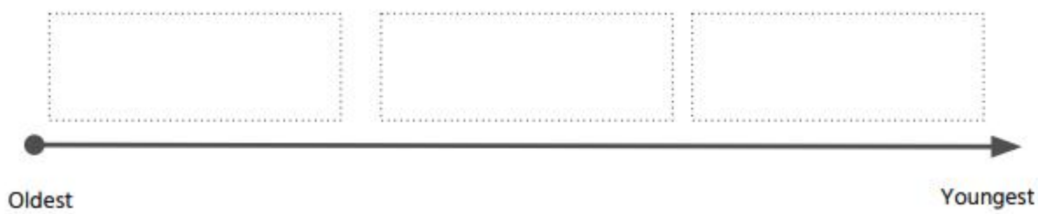
- **Phenomenon/uncertainty:** There are multiple rock structures in the Colorado Plateau that have some similar and some different rock layers and fossils—when did they develop relative to one another?
- **Complexity**: This item is an example of a Low DC, medium SEP complexity item.
- **Additional detail**: This item leverages a similar—but expanded—context to the previous item. By adding a third rock structure, students are engaged in more authentic sensemaking with the SEP, and have to negotiate additional data to make the judgments about relative ages of each structure. By asking students to order 3 structures and provide original artifacts as evidence, the item complexity and degree of sensemaking required increases. The phenomenon continues to be presented through simple text and an image that serves to both visually represent the phenomenon as well as provide data for interpretation, limiting the amount of time and interpretation students need to engage. The context presents only simple data that are directly relevant to determining the answer. Like the previous item, this phenomenon leverages a simple real-world observation that is authentically of interest to scientific communities (as evidenced by numerous articles and federally provided grants on the topic).

Exhibit B.6. Colorado Plateau, Version 2

The Grand Canyon, Zion Canyon, and Bryce Canyon are land areas on the Colorado Plateau. Together they represent almost 2 billion years of Earth's history. Scientists are interested in studying and comparing different areas to better understand how the Earth has changed over very long periods of time.



Part A. **Based on the rock layers and fossil data, which land area is the oldest? Move the national parks to show the correct order of age from oldest to youngest.**



Grand Canyon | Zion Canyon | Bryce Canyon

Part B. Provide 2 pieces of evidence to support your claim about the relative ages of the three locations.

Item ID: Limu Kohu (adapted from University of Hawaii)

- **Grade and discipline**: 8th grade, ESS
- **Item type**: Multi-part, Single Select Multiple choice, short response constructed response.
- **Could part A be used as a stand-alone item?** Yes.
- **Alignment**: this item is a **2D item**, measuring parts of:
  - DC: E8.12: Human activities have significantly altered the biosphere, atmosphere, and geosphere, sometimes damaging or destroying ecosystems and causing the extinction of organisms. Human choices can minimize harm to other organisms and risks to the health of the regional environment.
  - SEP: S8.23: Identify evidence that could be used to refute a claim about a phenomenon.
- **Item-level claim (derived from targeted dimensions):** Students can evaluate evidence about phenomena involving human impacts on the natural world, using their understanding of how human activities can significantly alter the biosphere.
- **Phenomenon/uncertainty:** what human activities are contributing to the decline of limu kohu?
- **Complexity**: This item is an example of a medium DC, low SEP complexity item.
- **Additional detail**: This item includes a few notable features. While the item includes a considerable amount of text for students to process, the text is presented simply. Technology permitting, this item could be modified to include a video of Limu Kohu harvesting, and possibly video interviews with elders to provide the evidence through a different modality. Because of the nature of the DC, items targeting this DC will almost always have to provide considerable contextual information or evidence (as exemplified here) for students to be able to meaningfully engage their understanding of the DC in service of sensemaking. The example provided here is intended to show one way this could be assessed in a discrete item; however, developers may conclude that this DC is best assessed via item sets and scenario based tasks. In addition to highlighting some implications of certain DCs for item design, this item foregrounds certain features of culturally relevant science assessments, including (1) the use of native/home language in the item (i.e., Limu Kohu is the Hawai'ian language term for this seaweed species), (2) use of non-traditional evidence sources that have been useful in university-based science endeavors (i.e., the use of multi-generational/elder accounts as evidence, as used by Stanford botanist Dr. Isabella Aiona Abbott), and (3) explicitly addressing a problem that is meaningful to specific communities (loss of Limu Kohu is very important to Hawaiian communities, and is representative of a broader conversation of the loss of indigenous foodways currently happening).

Exhibit B.7. Limu Kohu

Limu kohu is a type of seaweed that is native to the waters around Honolulu, Hawai'i, and is an important part of food systems as well as cultural and religious practices. Although Limu Kohu was easy to find for hundreds of years, Limu kohu populations around Honolulu have now been rapidly declining over the past 60 years.

Limu Kohu, found off of the coast of Honolulu.

Observations from Generations of Hawaiian Elders about Limu Kohu Growth and Harvesting
• Needs warm water and high salinity to grow • Grow and reproduces well on the edges of coral reefs • When limu kohu is trimmed, it regrows • When the base of the limu kohu is harvested, it cannot regrow

Part A. The following human activities are occurring in the area. Which is mostly likely to cause the **least** harm on limu kohu populations?

A. Industrial ways of harvesting limu picks all parts of the plant, including the base.
B. Many restaurants are using traditional harvesting practices that only harvest the top of the limu kohu.
C. Industry in the area is creating run-off that is changing the temperature and salinity of coastal regions.
D. Ships coming into the area have introduced invasive seaweed species that need the same resources as limu kohu.

Part B. Support your answer, using the information provided and your understanding of human impacts on the environment.

Item ID: Permafrost (adapted from OSE)

- **Grade and discipline**: 12th grade, PS
- **Item type**: Selected Response, Drag and Drop (or Ordering)
- **Alignment**: this item is a **3D item**, measuring parts of:
  - DC: P12.14: When sunlight is absorbed at Earth's surface it is eventually re-radiated as infrared radiation that transfers heat into the atmosphere. The average temperature of the atmosphere is determined by how long the energy stays in the system until it is reradiated into space from the top of the atmosphere.
  - SEP: S8.5: Develop, use, and/or revise a model to describe, explain, and/or

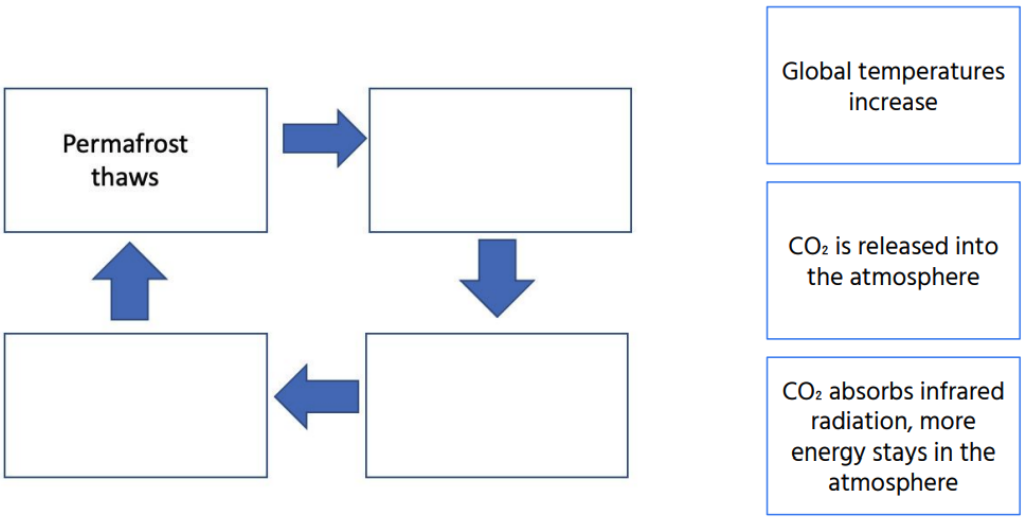
predict phenomena by identifying relationships among parts and or quantities in a system, including both visible and invisible quantities.
    - CCC: C12.14: Feedback mechanisms within a system are important elements for explaining or designing for either the stability or instability of the system.
- **Item-level claim (derived from targeted dimensions):** Students can complete a model explaining the causal chain of events that lead to a phenomenon using their understanding of how the release of greenhouse gases leads to increased temperatures on Earth.
- **Phenomenon/uncertainty:** how can permafrost melting lead to increased global temperatures?
- **Complexity**: This item is an example of a low DC, low SEP complexity, low CCC complexity item.
- **Additional detail**: This item presents students with a simple, real-world phenomenon that is deeply relevant to many current scientific endeavors. While students are sensemaking—they need to use their understanding of PS and developing models to figure out what the appropriate causal chain of events is that would lead to global temperature rises—this is highly directed, leading to a low complexity item across all three dimensions. The item uses very simple language, but still conveys an uncertainty and a problem that matters to many specific communities, both within the scientific space as well as those who are most affected by global temperature changes.

Exhibit B.8. Permafrost, Version 1

Permafrost is a layer of soil and ice that is just below the surface in the Arctic. Historically, permafrost stayed frozen for many years at a time. However, in some areas permafrost is now melting, which can cause many changes to Earth's surface and living things.

One major concern is that permafrost contains carbon dioxide. The carbon dioxide is trapped when permafrost remains frozen, but is released as it melts.

Use the statements and arrows below to develop a model that shows the relationships between thawing permafrost and rising global temperatures.



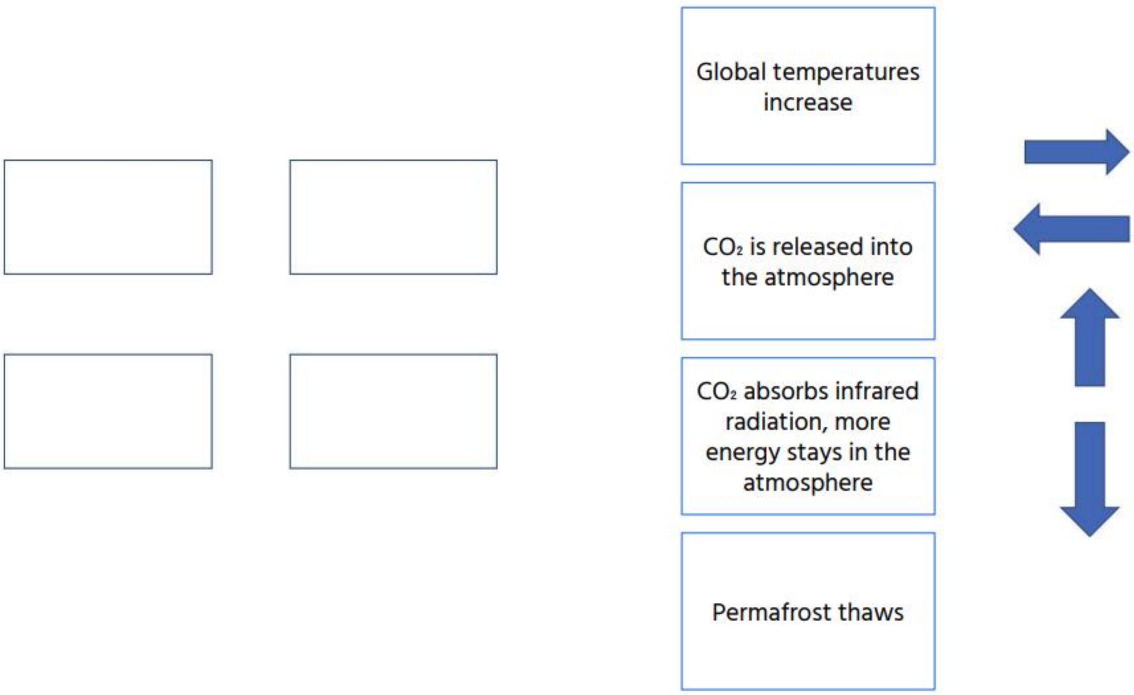
In the following two versions of this item, components of student responses are modified such that the complexity of the item (and the degree of sensemaking) is increased, without altering the context of the items. The first modification (Exhibit X) asks students to more independently develop a model. This both requires more sensemaking with all three dimensions as students figure out, with more limited cueing, what the causal relationships might be. It also allows students to provide more nuanced models—while some students will provide the simplest version of the model, others might indicate multiple sub-feedback loops that contribute to rising temperatures. This item is an example of a modification that leads to a medium DC, medium SEP, medium CCC complexity item.

Exhibit B.9. Permafrost, Version 2

Permafrost is a layer of soil and ice that is just below the surface in the Arctic. Historically, permafrost stayed frozen for many years at a time. However, in some areas permafrost is now melting, which can cause many changes to Earth's surface and living things.

One major concern is that permafrost contains carbon dioxide. The carbon dioxide is trapped when permafrost remains frozen, but is released as it melts.

Use the statements and arrows below to develop a model that shows the relationships between thawing permafrost and rising global temperatures. Drag the statements and arrows to show causal relationships. You may use arrows more than once.



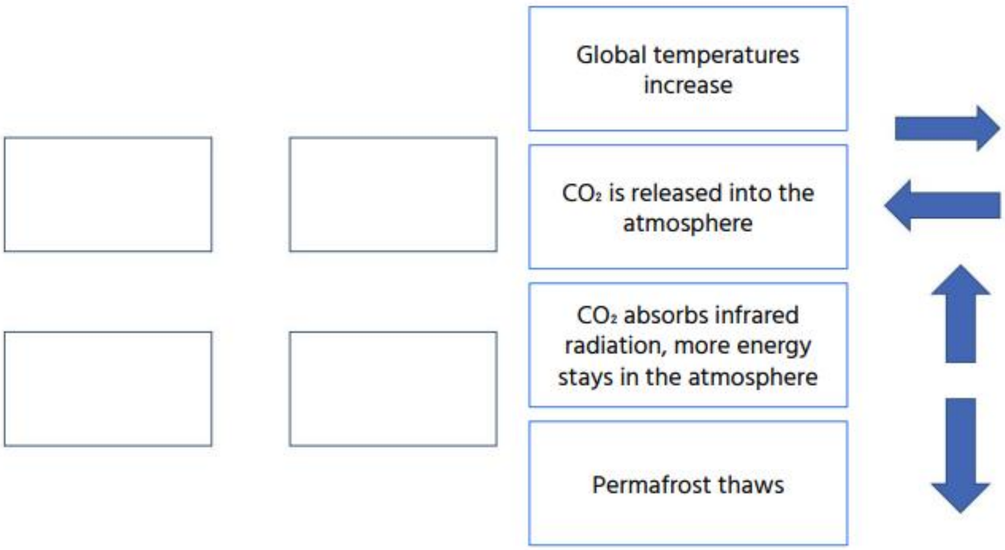
In the following example, this same item is modified once more to illustrate how more of the SEP and an additional CCC can be assessed within this phenomenon and context frame. By adding two parts to this item—a single select multiple choice and a short constructed response—students have to apply their DC understanding through the lens of stability and change (part B) and show a deeper understanding of modeling (Part C). Together, this final version is a more robust (albeit more time intensive) assessment of a three-dimensional target. Please note that parts B and C could be used with either version of the initial model development shown previously.

Exhibit B.10. Permafrost, Version 3

Permafrost is a layer of soil and ice that is just below the surface in the Arctic. Historically, permafrost stayed frozen for many years at a time. However, in some areas permafrost is now melting, which can cause many changes to Earth's surface and living things.

One major concern is that permafrost contains carbon dioxide. The carbon dioxide is trapped when permafrost remains frozen, but is released as it melts.

Use the statements and arrows below to develop a model that shows the relationships between thawing permafrost and rising global temperatures. Drag the statements and arrows to show causal relationships. You may use arrows more than once.



Part B. What does your model predict about how the rate of melting of the permafrost might change over the next 50 years?

The rate of melting will:

A. Increase
B. Stay the same
C. Decrease

Part C. What is one limitation of your model?

Item ID: Plant Growth (adapted from NGSA)

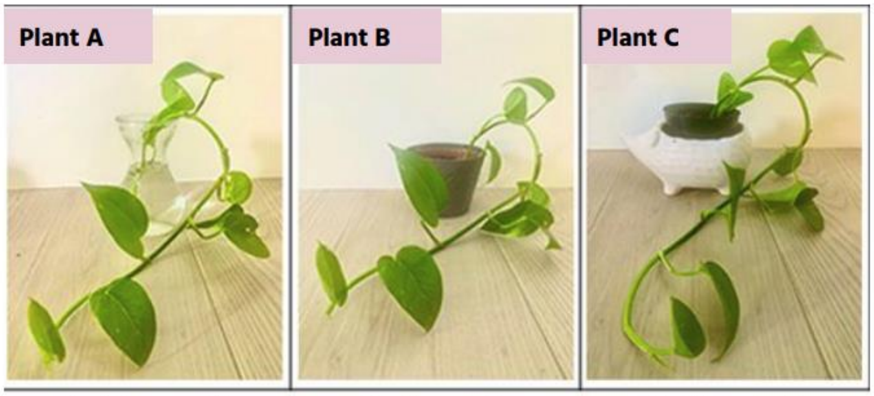
- **Grade and discipline**: 4th grade, LS
- **Item type**: Multipart item, Single Select multiple choice and short constructed response
- **Could part A be used as a stand-alone item?** Yes (with the possible addition of one

other answer choice, such as all the plants would grow the same amount)

- **Alignment**: this item is a **2D item**, measuring parts of:
  - DC: L4.3: All animals need food, water, and air in order to live and grow. They obtain their food from their surroundings – from plants or from other animals. Plants need air, water, minerals (in the soil), and light to live and grow.
  - SEP: S8:10: Predict the change in a dependent variable when a change in an independent variable occurs.
- **Item-level claim (derived from targeted dimensions):** Students can make a prediction about a phenomenon, using their understanding of what plants need to survive and grow.
- **Phenomenon/uncertainty:** how do soil and water conditions affect plant growth?
- **Complexity**: This item is an example of a low DC, low SEP complexity.
- **Additional detail**: This item presents students with a simple phenomenon that students will likely have some prior knowledge of (e.g., having seen plants in their classrooms, homes, parks, etc). This phenomenon is not presented as sophisticated scientific investigation, but rather as something that early elementary students may directly experience—different plants growing in their classroom. The item itself still requires students to bring an understanding of what plants need to grow to make a reasoned prediction, but both the SEP and DC are highly guided.

Exhibit B.11. Plant Growth, Version 1

The plants below were recently placed in a classroom. They are all the same kind of plant, are on the same side of the room, near a window, and receive the same amount of light. Students decide to vary the water and soil conditions to figure out what plants need to grow.



Plant in vase with water and no soil	Plant in pot with soil and water	Plant in pot with soil and no water

Part A. Which plant do you expect will grow **<u>the least</u>** over one month?

a. Plant A
b. Plant B
c. Plant C

Part B. Provide one reason why you think that plant will grow less than the other plants in the investigation.

The previous item focuses on a central component of the DC—that plants need water to survive—but does not necessarily provide evidence of the full DC, which includes a requirement for minerals from soil. The following example shows a slight modification that more completely elicits understanding of the DC. It should be noted that this item is largely at the same complexity levels as the previous item, but because it involves an additional component of disciplinary understanding, may be more challenging for some students.

Exhibit B.12. Plant Growth, Version 2

The plants below were recently placed in a classroom. Students decide to vary the water and soil conditions to figure out what plants need to grow. They are all the same kind of plant, are on the same side of the room, near a window, and receive the same amount of light.

Plant in vase with water and no soil | Plant in pot with soil and water | Plant in pot with soil and no water

Part A. Which plant do you think will grow the **<u>most</u>** over the month?

A. Plant A
B. Plant B
C. Plant C
D. Plant A and B will grow the same amount

Part B. Provide one reason to support your answer in Part A.

Because this phenomenon is so simple and intrinsically connected to the DC, it can be challenging to modify this item to require significantly more complex sensemaking with the DC. However, the item could be modified to engage the SEP at a higher degree of complexity, as well as the CCC. The following single select multiple choice example illustrates this, leveraging the same context and phenomenon, with additional practice complexity. This item is a **3D** item with low DC complexity, medium SEP complexity, and low CCC complexity. In this example, the SEP complexity is increased by adding simple data analysis to the prediction expected of students. Doing so also elicits some evidence of:

- an additional SEP element: S4.13: Analyze data to determine whether it supports or refutes a claim about a phenomenon or design solution; and
- CCC element: C4.1: Similarities and differences in patterns can be used to sort, classify,

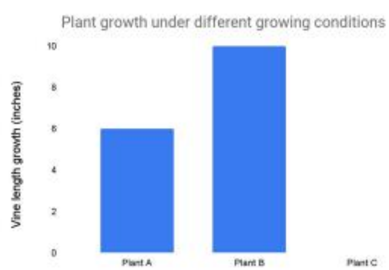
communicate, predict, and explain, with various representations (such as physical graphs or diagrams) to describe and analyze features of simple natural phenomena and designed products.

Exhibit B.13. Plant Growth, Version 3

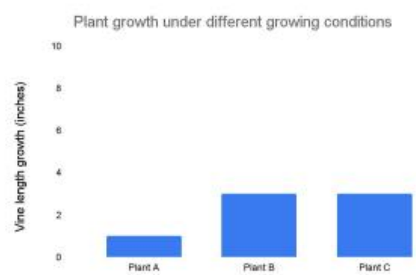
The plants below were recently placed in a classroom. Students decide to vary the water and soil conditions to figure out what plants need to grow. They are all the same kind of plant, are on the same side of the room, near a window, and receive the same amount of light.

Plant in vase with water and no soil | Plant in pot with soil and water | Plant in pot with soil and no water

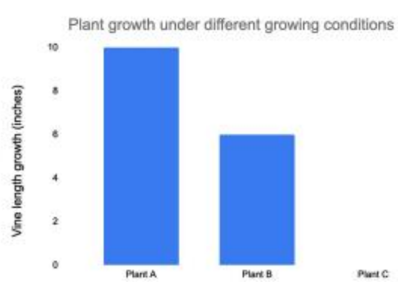
Which graph shows what you expect to see in terms of plant growth over one month?



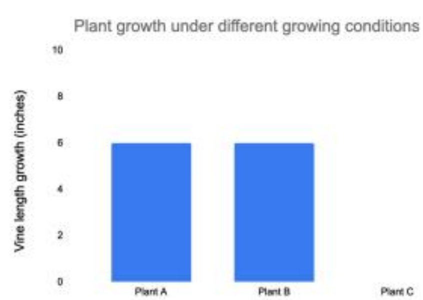
A



B



C



D

Item ID: Wild Dogs (adapted from OpenSciEd)

- **Grade and discipline**: 12th grade, LS

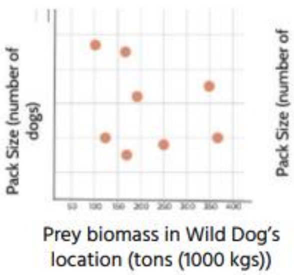
- **Item type**: Multipart item, Single Select multiple choice and short constructed response
- **Could part A be used as a stand-alone item?** Yes (with the possible addition of one other answer choice, such as a flat relationships between pack size and prey biomass—note that Part A alone is a largely 2D item, with patterns engaged much more implicitly)
- **Alignment**: this item is a **3D item**, measuring parts of:
    - DC: L12.8: Ecosystems have carrying capacities, which are limits to the numbers of organisms and populations they can support. Organisms would have the capacity to produce populations of great size were it not for the fact that environments and resources are finite. This fundamental tension affects the abundance (number of individuals) of species in any given ecosystem.
    - SEP: S8.12: Construct, analyze, and/or interpret graphical displays of data and/or large data sets from an investigation (e.g., maps, charts, graphs, and/or tables) to identify relationships between variables (linear vs. nonlinear relationships, causal vs. correlational relationships, and temporal and spatial relationships).
    - CCC: C12.1: Mathematical and statistical/probabilistic representations are needed to identify or describe patterns and to build models and explain mechanisms.
- **Item-level claim (derived from targeted dimensions):** Students can analyze and interpret patterns in data about a phenomenon to determine the likelihood that certain factors determine carrying capacities within ecosystems.
- **Phenomenon/uncertainty:** is prey availability the primary determinant of wild dog carrying capacity in this region?
- **Complexity**: This item is an example of a low DC, low SEP, low CCC complexity item.
- **Additional detail**: This item presents students with an authentic phenomenon that matters to specific communities. By asking students to evaluate data to determine whether prey availability is an important determinant of carrying capacity in this region, the item requires that students use their understanding of patterns (low level) together with their understanding of carrying capacity and data analysis to appropriately interpret the provided data. Because part B asks students to more explicitly consider the likelihood and degree of contribution of prey availability to carrying capacity, students must draw more explicitly on the specific high-school level CCC element, in conjunction with the other two dimensions. The DC, SEP, and CCC are all very straightforward and clearly cued to students.

Exhibit B.14. Wild Dogs, Version 1

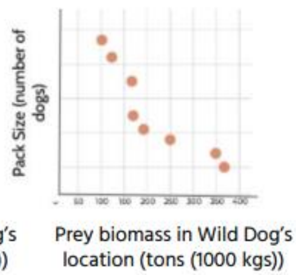
There has been a fast decline in African wild dog populations across Africa. Conservation experts are trying to establish a new population in a national park in Malawi, and are trying to figure out how many wild dogs the territory can support.

African wild dogs primarily eat antelope. **Prey biomass** is an indicator of how much antelope is available within a given location.

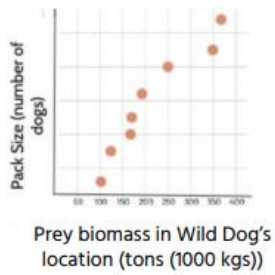
Part A. Scientists believe that the amount of food available could be a limiting factor in African wild dog pack size. Which graph shows the relationship you would expect between prey biomass and the population of African wild dogs?



A

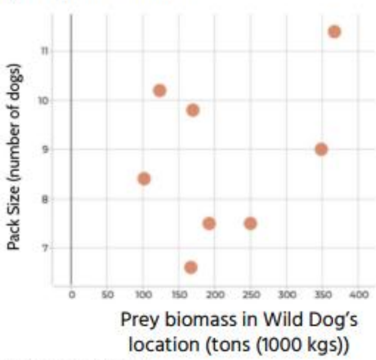


B



C

Part B. The graph below shows the actual relationship between African wild dog pack size and prey biomass that scientists observed in a similar territory. Based on the graph, does prey biomass play an important role in determining the carrying capacity of African wild dogs in this location? Use data to support your answer.



Data Source: Mills & Gorman (1997)

Prey biomass (does/**does not**) play an important role in determining the carrying capacity of African wild dogs in this location.

Support your answer here.

The same context can support items that align with different SEPs and CCCs. For example, the following short and extended constructed item illustrates how this context can be used to elicit evidence of students' understanding of parts of the following dimensions, in addition to the DC:

- SEP element: S12.1: Ask questions that arise from examining a model, an explanation, or a design plan to clarify and/or identify additional needed information or tests, and
- CCC element: C12.3: Cause-and-effect relationships can explain and predict complex

natural and human-designed systems. Such explanations may require examining and modeling small scale mechanisms within the system.

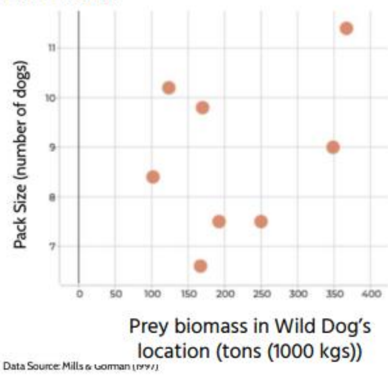
In this example, Part A could be used as a stand-alone question.

Exhibit B.15. Wild Dogs, Version 2

There has been a fast decline in African wild dog populations across Africa. Conservation experts are trying to establish a new population in a national park in Malawi, and are trying to figure out how many wild dogs the territory can support.

African wild dogs primarily eat antelope. **Prey biomass** is an indicator of how much antelope is available within a given location.

The graph below shows the relationship between African wild dog pack size and prey biomass that scientists observed in a similar territory. Based on the data provided and your understanding of ecosystem dynamics, write two questions that could help you better understand what determines carrying capacity in this location.



Question 1.

Question 2.

Part B. Describe how these questions will help you better understand how many wild dogs this location could support.

Item ID: Drinking Water

- **Grade and discipline**: 4th grade, PS
- **Item type**: Multiple-Select multiple choice
- **Alignment**: this item is a **2D item**, measuring parts of:
  - DC: P4.3: Unequal forces acting on an object can change its motion or forces can balance against other forces to hold the object in place.
  - SEP: S4.20: Construct and/or support an argument with evidence to support or

reject a claim about a phenomenon or a design solution.

- **Item-level claim (derived from targeted dimensions):** Students can identify evidence that supports a claim that unequal forces lead to the observation of a phenomenon.
- **Phenomenon/uncertainty:** does sucking on a straw apply a force to the liquid in a cup?
- **Complexity**: This item is an example of a low DC, low SEP complexity item.
- **Additional detail**: This item presents students with a simple, authentic phenomenon that is likely to be familiar to many students. The context is presented using very simple language and with an image that illustrates the phenomenon. Student sensemaking with the DC and SEP is highly cued through the provided answer choices.

Exhibit B.16. Drinking Water

A student is drinking water from a glass with a straw. She notices that the water stays still until she sucks on the straw. When she sucks on the straw, the water moves up through the straw to her mouth.

Select two pieces of evidence that the student is applying a force on the water when she sucks on the straw.

- ❑ **The water moves when she sucks on the straw.**
- ❑ **The water stays still before the student uses the straw.**
- ❑ She can see the water in the glass.
- ❑ The water stays cold in the glass.

Item ID: Making Soap (modified from NGSA)

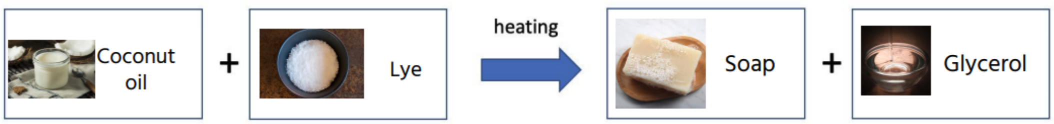
- **Grade and discipline**: 8^{th} grade, PS
- **Item type**: Single-Select multiple choice
- **Alignment**: this item is a **3D item**, measuring parts of:
  - DC: P8.4: In a chemical reaction, the atoms of the reacting substances are regrouped in characteristic ways into new substances with different properties. Atoms only rearrange. As such the amount of matter does not change.
  - SEP: S8.13: Analyze data to provide evidence to support or reject a model or explanation or use to improve a design solution.
  - CCC: C8.1: Patterns in data can be identified and represented using graphs, charts, and tables. Analyzing patterns can help identify cause and effect relationships and estimate probabilities of events.
- **Item-level claim (derived from targeted dimensions):** Students can analyze data to find patterns that indicate whether a phenomenon involves a chemical reaction based on the characteristic properties of substances.
- **Phenomenon/uncertainty:** what evidence indicates that a chemical reaction produces

soap from lye and coconut oil?

- **Complexity**: This item is an example of a low DC, low SEP, low CCC complexity item.
- **Additional detail**: This item presents students with a simple, authentic phenomenon centered around a product (soap) nearly all students will be familiar with, with images to make the process more accessible to a wide range of learners. The context is presented using very simple language, and only presents the data that are directly relevant to the phenomenon. All answer choices are correct, reducing cognitive load of needing to check them for accuracy. While the SEP and DC are very prominent in the sensemaking required in this item, the CCC is also required because students need to be able to identify and distinguish the patterns most relevant to actually identifying that a chemical reaction occurred.

Exhibit B.17. Making Soap

One way to make soap is to heat a combination of coconut oil and lye.



The data table below shows properties of each substance in the process.

Sample	Mass (g)	Odor	Density (g/cm3)	Melting point (℃)
Coconut oil	100	Coconut	0.93	27
Lye	20	Odorless	2.13	318
Soap	115	Coconut	0.95	48
Glycerol	5	Odorless	1.26	17.8

Which data provide evidence that making soap involves a chemical reaction?

A. Coconut oil and soap both smell like coconut
B. The density of soap is different than the density of glycerol
C. The melting points of soap and glycerol are very different than the melting points of coconut oil and lye
D. The total mass of soap + glycerol is the same as the total mass of coconut oil + lye.

Item ID: Rusting Nails (modified from MI and AL)

- **Grade and discipline**: 8th grade, PS
- **Item type**: Multi-part, Single-Select multiple choice
- **Alignment**: this item is a **2D item**, measuring parts of:
  - DC: P8.4: In a chemical reaction, the atoms of the reacting substances are regrouped in characteristic ways into new substances with different properties.

Atoms only rearrange. As such the amount of matter does not change.

  - SEP: S8.5: Develop, use, and/or revise a model to describe, explain, and/or predict phenomena by identifying relationships among parts and or quantities in a system, including both visible and invisible quantities.

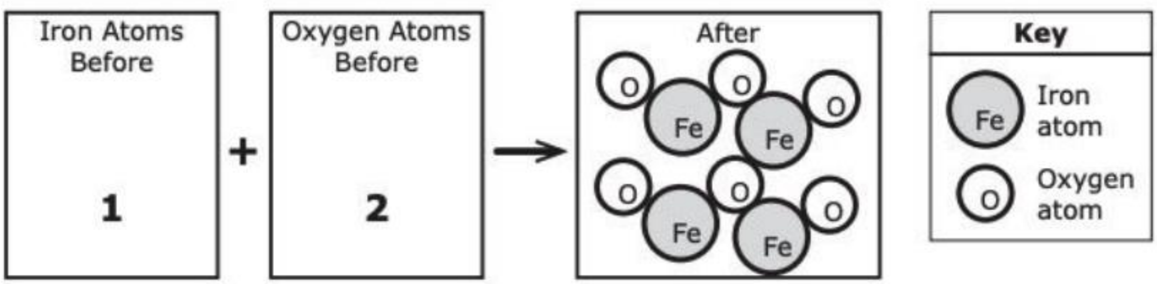
- **Item-level claim (derived from targeted dimensions):** Students can complete a model showing how atoms rearrange in a chemical reaction to account for a phenomenon.
- **Phenomenon/uncertainty:** is rusting evidence of a chemical reaction?
- **Complexity**: This item is an example of a low DC, low SEP complexity item.
- **Additional detail**: This item presents students with a simple, authentic phenomenon, with images to make the process more accessible to a wide range of learners. The SEP and DC are both highly directed, resulting in a very low complexity item. This item can be easily modified to engage modeling SEP more authentically by changing the item type to a drag and drop, where students drag and drop the correct number of Fe and O atoms into the appropriate boxes.

Exhibit B.18. Rusting Nails

When a bucket of iron nails are left outside, the iron in the nails reacts with oxygen in the air to produce rust.

Before **After**

Answer the questions below to complete the model below to describe how iron nails become rusty.



Part A. How many iron atoms should go in box 1?

a. 1
b. 2
c. 4
d. 6

Part B. How many oxygen atoms should go in box 2?

a. 1
b. 2
c. 4
d. 6

Item ID: Melting Ice

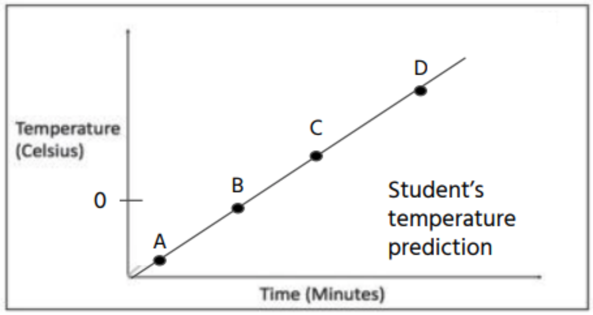
- **Grade and discipline**: 12th grade, PS
- **Item type**: Constructed response
- **Alignment**: this item is a **3D item**, measuring parts of:
  - DC: P12.3: In gasses or liquids, the motion of atoms or molecules leads to collisions between them. Such collisions are necessary for chemical processes to occur. Higher rates of collisions occur at higher temperatures because atoms are typically moving faster, and at higher pressure in a gas because the atoms are closer together.

    - SEP: S12.21: Construct an argument with evidence and scientific reasoning to support or reject a proposed model, explanation, or design solution for a problem.
    - CCC: Cause-and-effect relationships can explain and predict complex natural and human-designed systems. Such explanations may require examining and modeling small scale mechanisms within the system.
- **Item-level claim (derived from targeted dimensions):** Students can construct an argument for why the temperature of a system (phenomenon) does not change during a phase change.
- **Phenomenon/uncertainty:** why doesn't the temperature of the system continue changing during a phase change?
- **Complexity**: This item is an example of a low DC, low SEP complexity item, low CCC.
- **Additional detail**: This item largely requires students to make their understanding of the molecular processes underlying phase changes clear, using the SEP and CCC to do so. As a result, the overall complexity of the item is quite low.

Exhibit B.19. Melting Ice

When ice cubes are left out in the sun in hot weather, they melt. One student makes the following observations and draws what he predicts the temperature of water looks like over time.

Time point	Observation
A-B	Solid ice
B-C	Mixture of ice and water
C-D	Liquid water



Do you agree with the student's prediction? Use your understanding what happens at the molecular level during a phase change to make an argument that supports or refutes the student's prediction. In your argument, include what happens at each segment (A-B, B-C, C-D), including:

- Energy flows
- Phases of matter
- What is happening to molecular motion
- The role of forces between molecules
- Expected changes in temperature

You may use diagrams to support your argument.
I (agree/**disagree**) with the student's prediction.

Item ID: Water vs. Ethanol

- **Grade and discipline**: 12th grade, PS
- **Item type**: Single Select Multiple choice
- **Alignment**: this item is a **2D item**, measuring parts of:
  - DC: P12.3: In gasses or liquids, the motion of atoms or molecules leads to collisions between them. Such collisions are necessary for chemical processes to occur. Higher rates of collisions occur at higher temperatures because atoms are typically moving faster, and at higher pressure in a gas because the atoms are closer together.
  - SEP: S12.12: Analyze data to provide evidence to support or reject a model or explanation or use to optimize a design solution relative to criteria for success.
- **Item-level claim (derived from targeted dimensions):** Students can evaluate and interpret graphs to show how the temperature of a system (phenomenon) is expected to

change given different phases of matter.

- **Phenomenon/uncertainty:** what is happening to the temperature of ethanol vs. water at 0 degrees C?
- **Complexity**: This item is an example of a low DC, low SEP complexity item
- **Additional detail**: This item asks students to use their understanding of temperature during a phase change to make a simple prediction of the temperature/time relationship for two different substances. This requires a simple application of the DC and SEP.

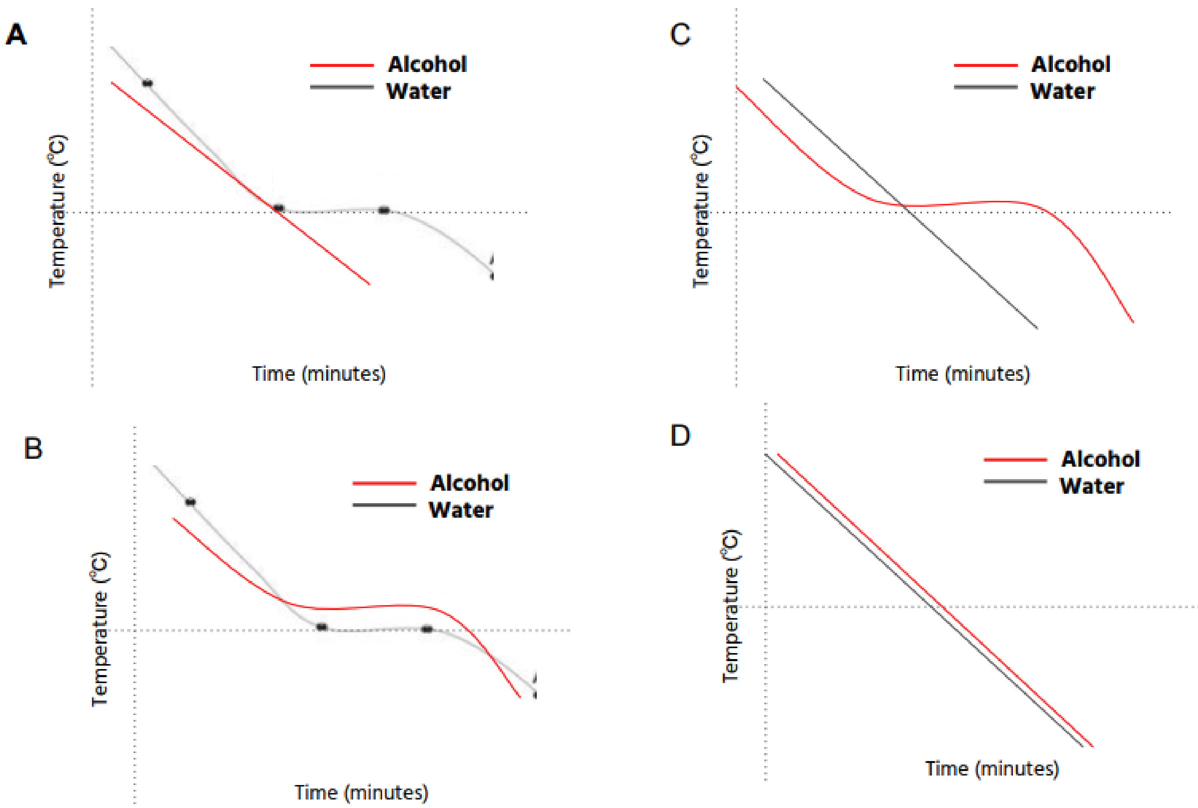
Exhibit B.20. Water vs. Ethanol

The picture shows water and alcohol when they have cooled to 0 degrees Celsius in a freezer.

Water: mixture of solid and liquid

Rubbing Alcohol: liquid

Which of the following temperature/time graphs shows the patterns you expect to observe for water and ethanol placed in a freezer?



Item Sets

For each of the item sets described here, individual items may be able to be used as stand-alone items, but are designed to “hang together” as a coherent set from the student perspective. Item sets often assess related DCs, SEPs, and CCCs, leveraging a common stimulus and/or context.

Here, item sets are described first as a whole set, with item-specific details provided as relevant.

Item Set ID: Seaside City

- **Grade and discipline**: 8th grade, LS
- **Number of Items**: 3
- **Item types**: Single Select Multiple choice, Multiple Select Grid, short constructed response, In-line choice
- **Alignment and Complexity**: together, this item set is a **3D item set at**
- **(overall) medium-high complexity**, measuring parts of:

	DC	SEP	CCC
Item 1	L8.7: In any ecosystem, organisms and populations with similar requirements for food, water, oxygen, or other resources may compete with each other for limited resources, access to which consequently constrains their growth and reproduction. DC complexity: medium	S8.6: Use a model to test ideas about phenomena in natural systems or proposed design solutions. SEP complexity: medium	-
Item 2	8.10: Ecosystems are dynamic in nature; their characteristics can vary over time. Disruptions to any physical or biological component of an ecosystem can lead to shifts in all its populations, therefore helping or hurting the health of the	S8.23: Identify evidence that could be used to refute a claim about a phenomenon. S8.25: Compare and critique two arguments on the same question to analyze their fit with the evidence and/or whether they emphasize similar or	C8.4: Phenomena may have more than one cause, and some cause-and-effect relationships in systems can only be described using probability. CCC complexity: medium

	ecosystem, including its biodiversity DC complexity: high	different evidence and/or interpretations. SEP complexity: medium	
Item 3	L8.7: In any ecosystem, organisms and populations with similar requirements for food, water, oxygen, or other resources may compete with each other for limited resources, access to which consequently constrains their growth and reproduction. DC complexity: medium	S8.5: Develop, use, and/or revise a model to describe, explain, and/or predict phenomena by identifying relationships among parts and or quantities in a system, including both visible and invisible quantities. SEP complexity: medium	C8.3: Cause-and-effect relationships may be used to predict phenomena in natural or designed systems. CCC complexity: medium

- **Phenomenon/uncertainty:** What is causing seaweed observed seaweed decline?
- **Additional detail**: This item set asks students to engage in meaningful sensemaking using the three dimensions. The set can easily be reduced in complexity by reducing the amount of data provided, reducing the number of evidence choices provided, and making distractors more obviously incorrect; an example of how to do so is included following the initial complete item set. It should be noted that in this set, Item 2 is doing considerable heavy lifting—the parts of item 2 are dependent on one another, but item 2 could likely be the foundation of a compelling scenario-based task in its own right, with further building out (e.g., adding additional data to support possible alternative claims, possibly designing an investigation or obtaining and evaluating existing information to propose which claim best fits the data, etc).

Exhibit B.21. Seaside City Item Set

Common stimulus (available with all items in the set):

Seaside City is a popular vacation spot. Each year, more tourists visit and more people come to live there. Many other animals and plants also live in the sea near the city.

Recently, people who live in Seaside City have noticed that the beaches used to have a lot of seaweed, but they rarely see seaweed anymore. Seaweed is very important to the local ecosystem because it is a major food source and safe environment for many animals living in the region.

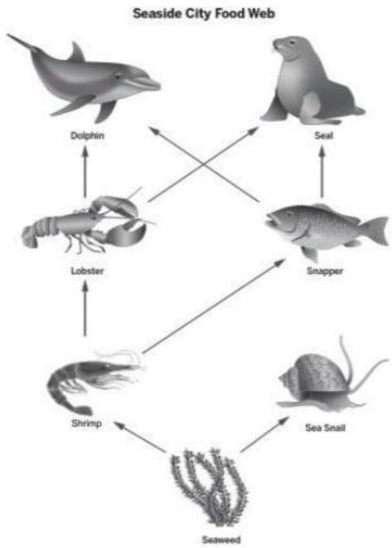
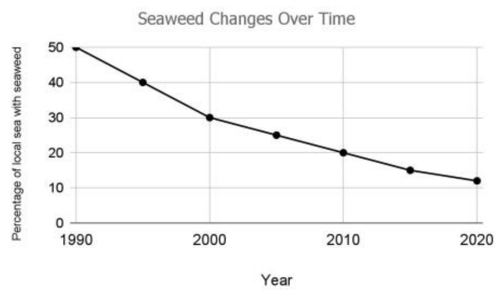


Figure 1: a diagram showing feeding relationships in the sea near Seaside city. Not all living things in this ecosystem are represented.



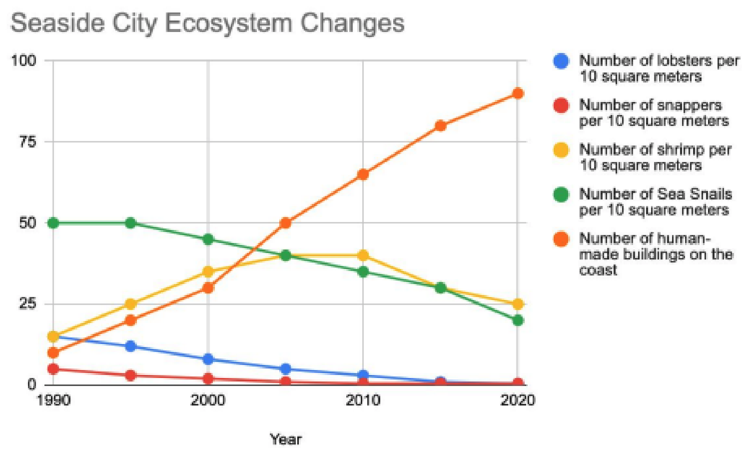
Item 1:

Based on the information provided, which observation would you expect to be true in this ecosystem?

a. the total number of dolphins was higher in 2020 than in 1990
b. individual seals were larger in 2020 than in 1990
c. there was more competition among sea snails and shrimp in 2020 than in 1990
d. Lobsters ate more seals in 2020 than in 1990

Item 2:
Some people claim that **the amount of seaweed near Seaside City is declining because people are eating more lobsters and snappers**.

Part A. Scientists have made the following observations. Use your understanding of how organisms interact in ecosystems and the information provided to select three (3) statements to support the idea and three (3) statements that do not support the claim.



Observations of human impacts
• Increased fishing to provide food for people • Disruptions to the local environment (land and sea) while building homes and businesses for people • Decline of some species as human activities impact the resources they need to survive and reproduce

Evidence	**Supports**	**Does not support**
Both snapper and lobster populations have decreased from 1990-2020.	☑	❑
Lobster and snapper both eat many organisms (including shrimp) that feed on seaweed.	☑	❑
At a similar area nearby, the shrimp population is staying constant while seaweed is declining at a similar rate.	❑	☑
The number of shrimp is higher in 2020 than in 1990	☑	❑
The shrimp population increases initially but then declines.	❑	☑
As the number of human-made buildings on the coast increases, the amount of seaweed decreases.	❑	☑

170

Part B. What is one different idea that could explain the seaweed decline? Select a possible alternative claim, based on the evidence provided, the ecosystem model, and your understanding of interactions in ecosystems.

a. Changes in the sea snail population are leading to declines in the amount of seaweed.
b. Increased human construction along the coast is destroying seaweed in the area.
c. There is a disease killing seaweed.

Part C: What additional information would you need to know to determine whether your new claim provides the best explanation for what is causing seaweed decline near Seaside City? Provide at least 2 additional pieces of information, and describe why they would be needed.

Part D. If your claim is correct, how would you account for the relationship between snappers, lobsters, and seaweed?

Item 3:

In 2021 Seaside City banned new building construction to prevent further damage to the marine ecosystem off its coast. Students at the local middle school have developed a model to predict what the organism populations might look like over time. Based on the model, students make two predictions about the shrimp population in 2030:

Option A: the shrimp population will remain about the same in 2030 compared to 2020.
Option B: the shrimp population will increase by 2030, compared to 2020.

Choose one option and complete the statement below. **Both options could be correct under the right conditions.**

The model predicts that shrimp population will remain about the same if the seaweed present will [remain the same, increase]. This is because how much the shrimp population grows will [be limited, increase, decrease] if competition for resources [stays the same, increases, decreases] compared to 2020.

The following example illustrates how item 2 could be reduced in complexity. This version is largely 2D, low DC and low SEP complexity, without meaningfully assessing the CCC. This version of the item eliminates students' interpretation of evidence from data, asks students to focus on more limited, straightforward evidence to support a single claim, and does not ask students to consider alternatives—all of these elements reduce the complexity and time required for the item, making it more appropriate for students who have not yet developed a sophisticated understanding of the targeted dimensions. The previous version may better represent more advanced students' facility with multiple dimensions in service of sensemaking. Similar approaches can be taken to reduce the complexity of other items as well.

Exhibit B.22. Modified Item 2 from Seaside City

Item 2:
Some people claim that **the amount of seaweed near Seaside City is declining because people are eating more lobsters and snappers**.

Part A. Scientists have made the following observations. Use your understanding of how organisms interact in ecosystems and the information provided to select three (3) statements to support the idea and three (3) statements that do not support the claim.

Evidence from Scientists' Observations	Supports	Does not support
Lobster and snapper both eat many organisms (including shrimp) that feed on seaweed.	☑	❏
The number of shrimp is higher in 2020 than in 1990	☑	❏
As the number of human-made buildings on the coast increases, the amount of seaweed decreases.	❏	☑

Item Set ID: Formation of Hawai'i (adapted from New Meridian)

- **Grade and discipline**: 8th grade, ESS
- **Number of Items**: 4
- **Item types**: Hotspot, Single Select Multiple choice, extended constructed response
- **Alignment and Complexity**: together, this item set is a **3D item set at low complexity**, measuring parts of:

	DC	SEP	CCC
Item 1	E8.4: The Earth consists of layers, including a solid, rigid outer layer divided into plates, which are always moving very slowly. Interactions between Earth's moving plates result in changes of physical	S8.12: Construct, analyze, and/or interpret graphical displays of data and/or large data sets from an investigation (e.g., maps, charts, graphs, and/or tables) to identify relationships between variables (linear vs. nonlinear	-

	features. DC Complexity: low [note: this item serves as helpful guidance in this item set, but likely should not be used in isolation to assess the DC]	relationships, causal vs. correlational relationships, and temporal and spatial relationships). SEP complexity: low	
Item 2	E8.4: The Earth consists of layers, including a solid, rigid outer layer divided into plates, which are always moving very slowly. Interactions between Earth's moving plates result in changes of physical features. DC complexity: low	S8.5: Develop, use, and/or revise a model to describe, explain, and/or predict phenomena by identifying relationships among parts and or quantities in a system, including both visible and invisible quantities.	-
Item 3	E8.4: The Earth consists of layers, including a solid, rigid outer layer divided into plates, which are always moving very slowly. Interactions between Earth's moving plates result in changes of physical features. DC complexity: low	S8.5: Develop, use, and/or revise a model to describe, explain, and/or predict phenomena by identifying relationships among parts and or quantities in a system, including both visible and invisible quantities. SEP complexity: low	-

Item 4	E8.4: The Earth consists of layers, including a solid, rigid outer layer divided into plates, which are always moving very slowly. Interactions between Earth's moving plates result in changes of physical features. DC complexity: medium	S8.19: Construct or revise an explanation that uses a chain of cause and effect or evidence-based associations between factors to account for the qualitative or quantitative relationships between variables in a phenomenon. SEP complexity: medium	C8.3: Cause-and-effect relationships may be used to predict phenomena in natural or designed systems. CCC complexity: low

- **Phenomenon/uncertainty:** how can Hawaiian islands be such different ages? Why is the Big Island so much larger than other Islands?
- **Additional detail**: This item set includes 4 items focused on the formation of the Hawai'ian Islands, a context that is highly relevant to communities living and working in Hawaii. While the items presented here can be independent, they also serve to coherently support student thinking if presented as as set—the complexity of the items will shift depending on whether they are presented as a set (where prior items serve as cueing/light guidance for future items) vs. if items are presented as stand-alone or out of order.

Exhibit B.23. Formation of Hawai'i item set

Common stimulus (available with all items in the task):

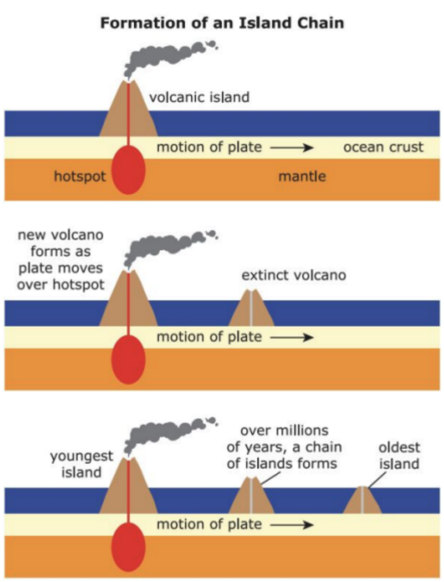
Island	Age (years)
Kaua'i	3.5-5.6 million
O'ahu	2.2-3.3 million
Moloka'i	1.3-1.8 million
Maui	Less than 1 million
Big Island	0.7 million-present day

Hawai'i is made up of several island that have developed over millions of years (table 1). Some of the islands, like Kaua'i, are over 5 million years old, while other islands like the Big Island are still developing today.

174

Item 1: There are more than 100 islands in the island chain. The map below shows some of the largest islands. Based on the data provided, where are oldest islands in the region?

Item 2: The Big Island is the only island in the island chain to have active volcanoes. Use the model to describe why the Big Island has active volcanoes, but the other islands do not.



The island of Hawai'i is:

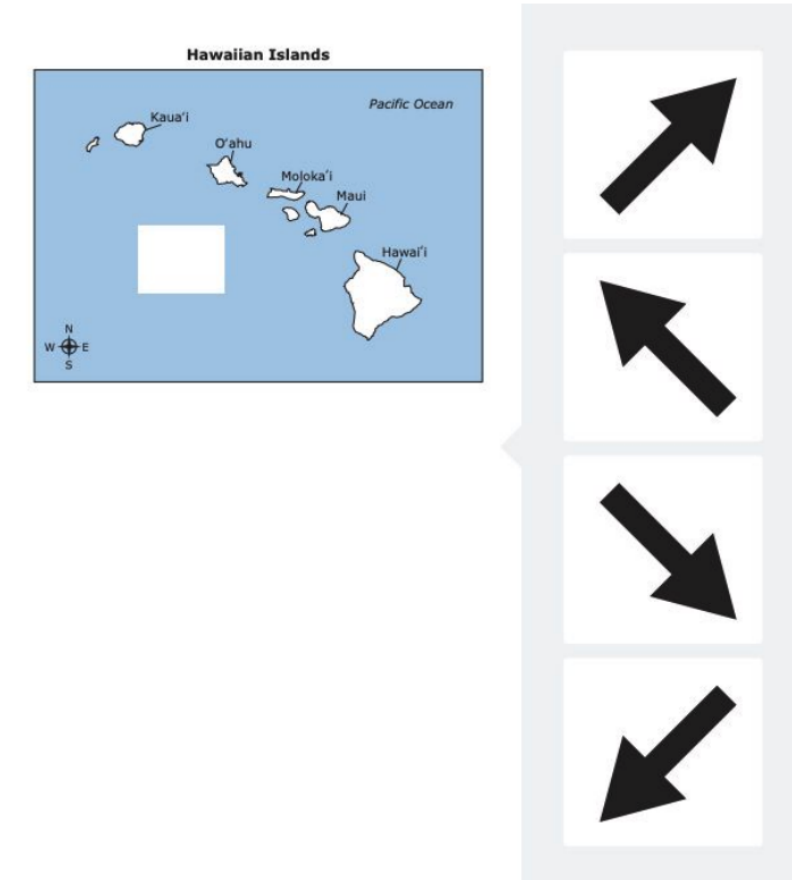
A. closer to the oceanic crust, which allows magma to flow through to the surface.

B. the largest island of the Hawaiian chain, and only large islands can support active volcanoes.

C. the youngest in the island chain, and volcanoes can form only on young islands, not old islands

D. over a large plume of hot magma, which pushes through the oceanic crust and forms the island.

Item 3. Based on your understanding of how Hawai'i has formed, show in which the oceanic plate is moving?



Item 4. Construct and support an explanation for why the Big Island is much larger than older islands like Kaua'i. Use the data provided as well as your understanding of how geologic processes like erosion and plate movements change Earth's surface over time in different ways.

Item Set ID: Locusts

- **Grade and discipline**: 8th grade, ESS and LS
- **Number of Items**: 5
- **Item types**: In line choice, drag and drop, grid, constructed response

- **Alignment and Complexity**: together, this item set is a **3D item set at medium complexity**, measuring parts of:

	DC	SEP	CCC
Item 1	L8.10: Ecosystems are dynamic in nature; their characteristics can vary over time. Disruptions to any physical or biological component of an ecosystem can lead to shifts in all its populations, therefore helping or hurting the health of the ecosystem, including its biodiversity. DC complexity: low	S8.13: Analyze data to provide evidence to support or reject a model or explanation or use to improve a design solution SEP complexity: low	-
Item 2	L8.10: Ecosystems are dynamic in nature; their characteristics can vary over time. Disruptions to any physical or biological component of an ecosystem can lead to shifts in all its populations, therefore helping or hurting the health of the ecosystem, including its biodiversity. DC complexity: low	S8.5: Develop, use, and/or revise a model to describe, explain, and/or predict phenomena by identifying relationships among parts and or quantities in a system, including both visible and invisible quantities. SEP complexity: medium	C8.3: Cause-and-effect relationships may be used to predict phenomena in natural or designed systems. CCC complexity: medium
Item 3	L8.10: Ecosystems are	S8.19: Construct or	C8.5: The observed

	dynamic in nature; their characteristics can vary over time. Disruptions to any physical or biological component of an ecosystem can lead to shifts in all its populations, therefore helping or hurting the health of the ecosystem, including its biodiversity. DC complexity: medium	revise an explanation that uses a chain of cause and effect or evidence-based associations between factors to account for the qualitative or quantitative relationships between variables in a phenomenon. SEP complexity: medium	function of natural and designed systems may change with scale. Phenomena that can be observed at one scale may not be observable at another scale. CCC complexity: medium
Item 4	E8.13: Human activities that release greenhouse gasses, such as production and combustion of fossil fuels, are major factors in the current rise in Earth's temperature. Monitoring the production and reducing the use of fossil fuels can slow the increase in global temperatures as well as the effects of climate change. E8.12: Human activities have significantly altered the biosphere, atmosphere, and geosphere,	S8.23: Identify evidence that could be used to refute a claim about a phenomenon. SEP complexity: low	C8.3: Cause-and-effect relationships may be used to predict phenomena in natural or designed systems. CCC complexity: low

	sometimes damaging or destroying ecosystems and causing the extinction of organisms. Human choices can minimize harm to other organisms and risks to the health of the regional environment. DC complexity: medium		
Item 5	E8.12: Human activities have significantly altered the biosphere, atmosphere, and geosphere, sometimes damaging or destroying ecosystems and causing the extinction of organisms. Human choices can minimize harm to other organisms and risks to the health of the regional environment. DC complexity: medium	S8.25: Compare and critique two arguments on the same question to analyze their fit with the evidence and/or whether they emphasize similar or different evidence and/or interpretations. SEP complexity: high	C8.4: Phenomena may have more than one cause, and some cause-and-effect relationships in systems can only be described using probability. CCC complexity: high

- **Phenomenon/uncertainty:** How can locust swarms cause so much damage? What is making swarming worse? What solutions can we consider?
- **Additional detail**: This item set illustrates how a robust phenomenon can give rise to items across multiple disciplines (e.g., LS and ESS). Like in other examples provided, the complexity of these items can be altered significantly based on item design. This example presents a meaningful phenomenon and problem context that deeply matters to many people around the world, and is posing considerable challenges right now. Other

directions this item set could be taken in include underlying biology (e.g., genetics, specialized subsystems) connected to the physiological changes locusts undergo; research (via a scenario based task) on potential solutions, impacts on biodiversity in regions with swarming, consideration of patterns of locust swarming going back thousands of years (stability and change) and considering whether current upticks are significant or not (more sophisticated data analysis), etc. This context can also easily support items for both MS and HS levels in both LS and ESS. This particular version of the item was selected to show how a wider range of SEPs and CCCs, including some that are often difficult to assess, can be engaged in items across a task. Note that technology permitting, Item 5 would benefit from non-text based sources of information about solutions, such as a video or simulation.

Exhibit B.24. Locusts

Common stimulus (available with all items in the set):

Desert locusts are a form of grasshopper that undergo changes to their bodies in certain environmental conditions (Figure 1).

Desert Locust

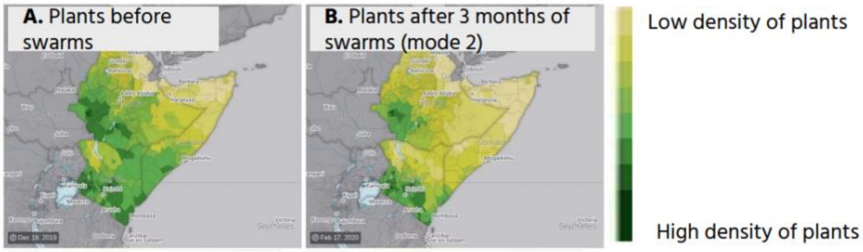
Mode 1: Grasshopper
(Dry, warm or cool weather)

- Behave independently
- Repelled from other desert locusts
- Mostly walk slowly and jump
- Restricted diet
- Small, scattered populations that stay in one place
- Very stable population; females lay eggs but most don't hatch until the environment is wet and hot.

Mode 2: Locust
(Wet/rainy, warm or hot weather)

- Behave as a cohesive unit (swarms)
- Attracted to other desert locusts
- Walk quickly and fly long distances
- Broad diet, including crops
- Tens of billions of locusts in a swarm that can travel up to 100 miles per day
- Population can increase 400x in 6 months.

When these insects are in Mode 2, they are able to swarm. A single swarm can cover up to 100 square miles, with 40 to 80 million locusts in each square mile. Swarms can travel up to 80 miles a day.



Change in vegetation over three months of locust swarms. Dark green shows higher density of plants.

Item 1. Individual locusts can eat about 2g of food each day. Consider the data. In Mode 1, what would you expect the impact on plants in the area to look like? Complete the statements below.

I would expect plants in the area to look more like (**A**/B) when desert locusts are in Mode 1.

In Mode 1, locusts would have a (**small**/large) impact on available plant life in the area. This is because individual locusts eat a (**small**/large) amount of plants relative to all of the plants available.

Item 2. Drag the arrows and statements to develop a model to show how locust populations in Mode 2 can become so big. Use the arrows to show causal relationships. You may use arrows more than once, and you may represent multiple causal relationships. [note: this model could be much more nuanced, technology permitting]

Construct your model here.

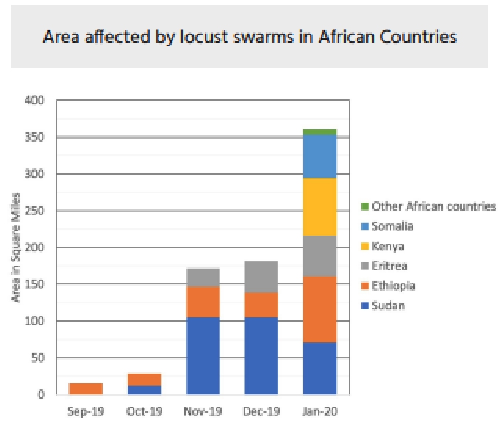
Heavy rainfall in during a very hot summer	Increased mating
Desert locusts shift from mode 1 to mode 2	Desert locusts in Mode 2 lay eggs
Desert locusts in mode 2 attract each other so many locusts are in the same place	Many eggs hatch

Item 3. Describe why locusts in Mode 2, but not Mode 1, have the impact seen in the data. Use your understanding of scale and quantity as well as your understanding of organism needs for survival and reproduction in your answer.

Item 4.

Recently, eastern African countries like Kenya, Somalia, and Ethiopia have experienced larger and longer lasting swarms of desert locusts than before. These swarms pose a significant threat to food security because the locusts can eat 50-80% of all crops grown in the area over just a few months.

Observed locust swarms in Africa	
	
2006-2007	2019-2021



Part A. One hypothesis for why the swarms are getting worse is that **global changes to climate are triggering the switch to the gregarious phase more frequently and for longer periods of time.**

Scientists have collected the following evidence. Which pieces of evidence support this claim?

Evidence	**Yes, this evidence supports the claim.**
There were very heavy rains in 2020 and 2021.	❑
Since 2019, the average temperatures in east African countries have been at their highest in recorded history.	❑
As the average global temperatures have increased, locust swarms have also increased.	❑
Locust swarms have decreased in other areas, like South America, that have also experienced warmer temperatures and more rain.	❑
Less pesticides have been able to be sprayed in eastern African countries due to conflicts in the region.	❑

Part B. Describe why the evidence you chose supports the claim.

Part C. Use [the data and] what you know about how human activities influence the climate to make an argument for how humans could be causing increases in locust swarms. **[note: this item/item part could require a simulation to explore the relationship between temperature/precipitation, atmospheric CO2, extreme weather events, and locust swarms. An alternative, static option is to simply provide the data, and ask students to make sense of it as part of their argument. A third option is to leave the item as is, relying more heavily on DC understanding]**

Item 5.

Consider the following proposed solutions for reducing the impact of locust swarms.

Option 1: Spraying Aerial Chemical Pesticides. Using airplanes, infected countries can spray chemical pesticides over affected areas.

Considerations for Option 1
• Works 99% of the time • Locusts are killed within 24 hours • Chemical pesticides can be toxic to other plants and animals in the area of the spray • Animals (including humans) who eat the dead locusts or affected plants can become sick and may die. • Somewhat expensive

Option 2: Spraying Aerial Biological Pesticides. Using a variety of methods, infected countries can spray biopesticides over affected areas.

Considerations for Option 2
• Effectiveness is unknown, still being tested • Takes 1-3 weeks to kill locusts • No impact on mammals and crops • May kill other related insects. • Very expensive • Can be administered by plane as well as more locally targeted

Scientists predict that Kenya will be experiencing swarms this year. Which solution would you recommend to Kenya? Choose a solution. Support your choice with at least 2 pieces of evidence and reasoning.

APPENDIX C: Glossary

The following terms are used in the NAEP Science Assessment Framework and the NAEP Science Assessment and Item Specifications. Additional terms may be found in the *NAEP Glossary of Terms*.

alignment refers to the coordination of goals, instruction, and assessment in a mutually reinforcing educational system.

constructed response (CR): An item type in which the response is text or mathematical symbols that are entered into a field. Constructed response items are often interactive and are generally scored individually by trained scorers using a detailed rubric that may allow for full or partial credit. These are generally more challenging than selected response items because the correct and alternative answers are not part of the item. Examples of constructed response item types that may be used on the 2028 NAEP Science Assessment are listed below.

- Short Text: The student enters a word or short phrase into a box or completes a sentence.
- Table Text: The student enters text into a table or chart.
- Extended Response: A prompt requires a written response that is several sentences long.
- Numeric Entry: The required response is a number or equation.

context is used to describe all the information presented to a student in framing a task and the prompt that elicits a student response. The same phenomenon or problem can be addressed through many different contexts and thus can frame many tasks. All stimulus information provided to students (e.g., written descriptions, images, videos, simulations, long-form texts, infographics, data tables, graphs, etc.) used to present the phenomenon is considered context, offering background information necessary for students' sensemaking and/or problem-solving.

crosscutting concepts: Crosscutting concepts are ideas that transcend disciplinary boundaries and prove fruitful in explanation, in theory, in observation, and in design. These ideas are conceptual tools that guide effective and reflective practice in all fields of science and engineering.

dimension: As used in the 2028 NAEP Science Framework, dimensions refer to three broad sets of expectations with respect to a student's knowledge and skills: Science and Engineering Practices (SEPs), Disciplinary Concepts (DCs), and Crosscutting Concepts (CCCs).

disciplinary concepts (DCs) are well-tested theories and explanations developed by scientists organized into three major disciplinary groupings: physical science, life science, and Earth and space sciences.

discrete item (DI) is a single, standalone item. Students need to be able to read the stimuli/prompt and answer the question in 2-3 minutes. Compared with other item types, discrete items allow for a large number of items to be included on the assessment, increasing the reliability of the assessment.

engineering is a discipline involved in the definition and solution of problems. Engineering often requires development of a design to solve the problem that meets the criteria for a successful solution within constraints such as time and budget. The term *engineering* includes many areas of application (e.g., medicine, agriculture, infrastructure, environmental management).

evidence is a body of facts or observations that can provide information about whether a belief or proposition is true or valid.

exhibit: This framework refers to tables and figures as "exhibits." Exhibits are numbered consecutively within each chapter. For example, the first three exhibits in Chapter 3 are labeled Exhibit 3.1, Exhibit 3.2 and Exhibit 3.3.

items are the questions students answer, or the tasks they must complete, as part of an educational assessment.

item part: The smallest element requiring a response within an item. For example, a two-part item might consist of a selected-response item part followed by a constructed-response item part that asks the student to explain the answer chosen in the selected-response item part.

item set uses common stimulus material to ask a group of independent questions. Item sets make it possible to take advantage of efficiency by presenting rich and engaging stimulus material, then asking several questions to collect evidence. Since the items do not depend on each other, questions in an item set each receive a separate score.

item subtype: A specific format available within an item type (e.g., multiple choice and multiple select are subtypes of the selected-response item type).

item type is a description of the format of an assessment item. Item types may be categorized by their overall structure and complexity, such as discrete, multipart, item set and scenario-based task. Items may also be categorized by the kind of response required, such as selected response, constructed response, and technology enhanced.

item: An individual assessment element that includes stimulus material, a question/prompt, answer/options or an answer field, scoring criteria, and metadata.

multidimensional refers to Items that integrate two or all three dimensions.

multipart item (MPI) has a few parts that are dependent on each other. For example, a multipart item might ask students to make a choice or decision and follow up with another question to explain their reasoning.

performance expectation (PE): An assessable statement of what students should know and be able to do. Formulating a performance expectation is often the starting point in developing an assessment item.

phenomena: observable events that occur in either natural or human-designed systems. Science assessment items often ask students to explain a phenomenon.

problem: A challenge that arises from a human need or want. In the 2028 NAEP Science Framework, the term problem is used to describe a real-world issue that requires a designed solution; as such, it is an engineering problem.

scenario-based task (SBT) is a sequence of items presented through an unfolding context, often with rich and engaging stimulus material such as images and video. SBTs are often interactive, asking students to respond to several short tasks and questions. However, the task does not have to be interactive to be a scenario-based task. SBTs typically present meaningful and compelling phenomena and problems, including those that require a large amount of background information.

science and engineering practices are ways of working to develop scientific explanations of phenomena or design engineering solutions to problems.

selected response items ask the student to select the best choice (or choices) from a provided group of options. Different types of selected response items that may be used on the 2028 NAEP Science Assessment include the following.

- Single-selection multiple choice: In response to a prompt, students choose a single response from a set of (usually) four or more options.
- Multiple selection multiple choice: Students are prompted to choose two or more responses from a set of (usually) five or more options.
- Matching Table: Students mark their response to a list of statements in a table by marking each option as yes/no, true/false, etc.
- Zone: Students respond to a prompt by marking or dragging a symbol into a different part of the answer space.
- In-line Choice: The student selects a single text option from a drop-down menu within a table or inline text.
- Grid: The student selects points on a grid to complete a task, such as creating lines and shapes, or plotting points.

stem: The item question or prompt to which the student responds.

stimulus: A component of an item set that does not directly require a student response. A stimulus can include text, audio, video, animation/simulation, experimentation, discussion, activity, and/or demonstration.

target: Assessable knowledge and skills. For an item or item part in an item set, the target consists of the evidence statements and associated parts of the dimensions included in the evidence statement for the associated PE.

technology-enhanced item (TEI): A computer-delivered item type in which the response requires specialized computer interaction that is beyond selected-response or constructed-response interactions.

REFERENCES

Achieve (2019a). The Task Annotation Project in Science: Sense-making. Retrieved from https://issuu.com/achieveinc/docs/sense-making_02142019__7_.

Achieve (2019b). A Framework to Evaluate Cognitive Complexity in Science Assessments. Washington, DC: Achieve, Inc. Retrieved from https://www.nextgenscience.org/sites/default/files/Science%20Cognitive%20Complexity%20Framework_Final_093019.pdf.

American Association for the Advancement of Science (1990). Science for all Americans: A project 2061 report on literacy goals in science, mathematics, and technology. Washington, DC: AAAS.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Joint Committee on Standards for Educational and Psychological Testing.

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

Carroll, J. B. (1963). A model of school learning. *Teachers college record*, *64*(8), 1-9.

Carroll, J. B. (1989). The Carroll model: A 25-year retrospective and prospective view. *Educational researcher*, *18*(1), 26-31.

Davis, E. A., & Stephens, A. (2022). Science and Engineering in Preschool through Elementary Grades: The Brilliance of Children and The Strengths of Educators. A Consensus Study Report. *National Academies Press*.

Ferguson, H. B., Bovaird, S., & Mueller, M. P. (2007). The impact of poverty on educational outcomes for children. *Paediatrics & child health*, *12*(8), 701-706.

Joint Task Force on Assessment of the International Reading Association and the National Council of Teachers of English. (2010). *Standards for the assessment of reading and writing* (Rev. ed.). NCTE and Newark, DE: IRA.

Kohlhaas, K., Lin, H. H., & Chu, K. L. (2010). Disaggregated outcomes of gender, ethnicity, and poverty on fifth grade science performance. *RMLE Online*, *33*(7), 1-12.

Lawrence Hall of Science (LHS) and the American Museum of Natural History (AMNH) (2018). Disruptions in Ecosystems: Ecosystem Interactions, Energy, & Dynamics, a model instructional unit and professional development program that aligns with NGSS. https://lawrencehallofscience.org/news/ngss-in-practice-a-model-unit-and-teacher-development/.

National Academies of Sciences, Engineering, and Medicine. (2019). *Science and engineering for grades 6-12: Investigation and design at the center*. National Academies Press.

National Academies of Sciences, Engineering, and Medicine. 2017. *Seeing Students Learn Science: Integrating Assessment and Instruction in the Classroom.* Washington, DC: The National Academies Press. https://doi.org/10.17226/23548.

National Academies of Sciences, Engineering, and Medicine. 2018. *How People Learn II: Learners, Contexts, and Cultures.* Washington, DC: The National Academies Press. https://doi.org/10.17226/24783.

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

National Research Council (2013) Next Generation Science Standards: For States, By States, Volume 1. Washington, DC: National Academies Press.

National Research Council (2013) Next Generation Science Standards: For States, By States, Volume 2. Appendices. Washington, DC: National Academies Press.

National Research Council. 2005. *How Students Learn: Science in the Classroom.* Washington, DC: The National Academies Press. https://doi.org/10.17226/11102.

National Research Council. 2007. *Taking Science to School: Learning and Teaching Science in Grades K-8*. Washington, DC: The National Academies Press. https://doi.org/10.17226/11625.

Steele, C. M., Spencer, S. J., & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. In *Advances in experimental social psychology* (Vol. 34, pp. 379-440). Academic Press.

Tekkumru-Kisa, M., Stein, M. K., & Schunn, C. (2015). A framework for analyzing cognitive demand and content-practices integration: Task analysis guide in science. *Journal of Research in Science Teaching*, *52*(5), 659-685.

WestEd, Center on Standards and Assessment Implementation, & Delaware Department of Education (2019). Cognitive Loading in Three-Dimensional NGSS Assessment: Knowledge, Skills, and Know-How. Retrieved from https://csaa.wested.org/resource/cognitive-loading-in-three-dimensional-ngss-assessment-knowledge-skills-and-know-how/.