2028 NAEP Science Assessment and Item Specifications

Contract # 91995922C0001

NATIONAL ASSESSMENT GOVERNING BOARD

Honorable Beverly Perdue
Chair

Lesley Muldoon
Executive Director

For further information, contact:
National Assessment Governing Board
800 N. Capitol St., NW
Suite 825
Washington, DC 20002-4233
www.nagb.gov

February 2024

The 2028 NAEP Science Assessment and Item Specifications 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 Core Ideas*, and excerpts from *Next Generation Science Standards: For States, by States* and associated appendices, with permission granted by the National Academies Press, Washington, DC.

TABLE OF CONTENTS

TABLE OF EXHIBITS

CHAPTER 1: Overview

The Assessment and Item Specifications document is a companion to the 2028 NAEP Science Assessment Framework. The purpose of the Specifications is to operationalize the Framework by describing the design and development of the NAEP Science Assessment. The primary audience for the Specifications is the National Center for Education Statistics (NCES) and its contractors, who will use this document together with the Framework, to develop the 2028 NAEP Science Assessment.

Throughout the Specifications, the term *Framework* is used to refer to the 2028 NAEP Science Assessment Framework. Other frameworks, such as *A Framework for K–12 Science Education*, are referenced using the full title of the document.

Background on 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 several 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. NAEP scores are always reported at the aggregate level.

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.

Overview of Assessment Design and Item Specifications

The Assessment and Item Specifications that guided the design and development of the NAEP Science Assessments administered since 2009 were established prior to significant shifts in science education. Updates have been made to reflect these changes, both to the Framework, which describes the construct to be measured by the 2028 NAEP Science Assessment, and to these Specifications, which describe the design and development of a fair and accurate measurement of the updated construct.

The organization of the Specifications follows the structure of the Framework—each document is divided into four chapters, followed by a set of appendices. The four-chapter organization of the Specifications is listed below:

Chapter 1 provides an overview of the document.

Chapter 2 defines a three-dimensional construct focused on sensemaking in the contexts of phenomena and problems.

Chapter 3 describes the design of the assessment and development of items, including illustrations that demonstrate key characteristics of the assessment components.

Chapter 4 explains how results of the 2028 NAEP Science Assessment will be reported.

A Note About Sample Items: The sample items included in these Specifications have been created by framework panelists or staff or modified from items originally designed for a different purpose. The sample items were developed to illustrate important aspects and features of these Specifications. Items that are included on operational NAEP assessments go through rigorous item development procedures in a multiyear process, which includes reviews by multiple stakeholders, pretesting, piloting, and iterative revisions. The sample items included in these Specifications have not gone through the item development process for NAEP that will be used for actual items created for the operational NAEP science assessments.

Portions of the following sample items appear in the Framework in addition to the Specifications:

- Human Migration to Appalachia
- Locusts
- Limu Kohu
- Making Soap
- Park Flooding
- Plant Growth
- Permafrost Melting

A New Construct

The construct described in both the Framework and Specifications includes three dimensions: Disciplinary Concepts (DCs), Science and Engineering Practices (SEPs), and Crosscutting Concepts (CCCs). Items on the assessment will be two- or three-dimensional. A two-dimensional item will include a DC and an SEP. A three-dimensional item will include all three dimensions: a DC, an SEP, and a CCC. **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. Items may also draw upon multiple statements within a dimension, including interdisciplinary items. Each item will receive one score representing the integration of the dimensions measured by the item.

Aligning With the Framework and the Assessment and Item Specifications

The assessment should be developed to be aligned to science achievement expectations expressed as performances, defined by the intersection of the three dimensions, and used to solve problems and make sense of phenomena:

- The content of the assessment should be aligned with the content of the Framework and Specifications. The assessment as a whole should reflect the breadth of knowledge covered by the statements within each dimension of the Framework and Specifications.

The content of the assessment should reflect the emphasis described by the clarifications and should not go beyond the assessment boundaries described in this document. The assessment should represent the balance of DCs, SEPs, and CCCs at each grade as described in the Framework and Specifications.

- The complexity on the assessment should be aligned to the complexity framework described in the Framework and Specifications.
- While it is not possible to cover every possible statement within the three dimensions in the Framework and Specifications on one assessment, appropriate alignment between the assessment and the Framework and Specifications at each grade should be maintained in the item pools. The assessment should be built so that the constructs represented by the statements in each dimension area are adequately represented. The breadth and relative emphasis in each dimension, as presented in the Framework and Specifications, should be represented in the assessment as a whole. The developer should avoid under- or over-emphasizing particular discipline groups, practice categories, or crosscutting concepts to the extent feasible, with a goal of ensuring broad, but not necessarily equal, coverage in any given year's item pool and coverage of all statements over time.
- The assessment should represent the balance of types of items specified in the Framework and Specifications and give appropriate emphasis to the ways in which students are expected to use the three dimensions of science through solving problems and making sense of phenomena, situated in real-world contexts.
- Contexts used to engage student interest should be as compelling as possible and should include phenomena in the natural and designed world, as well as problems that lend themselves to engineering design.
- The assessment should report and interpret scores based on the Framework, Specifications, and NAEP achievement level descriptions (ALDs). That is, the assessment should be developed so that scores will reflect both the three dimensions in the Framework and Specifications and the range of performances illustrated in the NAEP science ALDs.
- The assessment should give all students tested a reasonable opportunity to demonstrate the range of scientific reasoning in using the disciplinary concepts, science and engineering practices, and crosscutting concepts covered by the Framework and Specifications.

CHAPTER 2: Dimensions of Science Achievement

The NAEP 2028 Science Assessment measures **science achievement**, the construct described in the Framework as "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." This construct includes the three dimensions of science:

- **NAEP 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.
- **NAEP Science and Engineering Practices** (SEPs) are ways of working to develop scientific explanations of phenomena or design engineering solutions to problems.
- **NAEP Crosscutting Concepts** (CCCs) 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 sections of Chapter 2 describe each dimension at the level of detail needed to develop multidimensional items that engage students in making sense of phenomena and solving problems. Chapter 3 describes the characteristics and types of these items. Many perspectives informed an iterative process used to develop the sets of statements for each dimension and the refinement of individual statements, including internal reviews by project committees and staff, and external feedback gathered from outreach activities with a variety of stakeholders.

NAEP Science Assessment Terminology

In addition to the three dimensions, terms that are broadly used in science education are defined for their specific meanings in the context of the NAEP Science Assessment. These definitions should be used throughout the assessment development process. Terminology used in both the Framework and Specifications is included in a glossary in Appendix D.

- **Clarification:** a statement that enhances NAEP DCs by explaining the emphasis, giving examples, or providing a specific point of detail.
- **Boundary:** a statement that tells the item writer what the item should not cover in relation to the NAEP DC.
- **Include:** within a clarification, the introduction to a list of examples that is not intended to be exhaustive; developers may use these examples for development or may use examples that are similar and do not alter the intent of the list.
- **Illustrative item:** an item used to show one or more ways to interpret the Specifications.

NAEP Science Discipline Groups and Subgroups

The three discipline groups—Physical Science, Life Science, and Earth and Space Sciences—and subgroups within each, are used to organize content and facilitate comparisons of ideas within and across topics, and across grades. The following table summarizes the discipline groups and subgroups described in the 2028 NAEP Science Framework.

Exhibit 2.1. NAEP 2028 Science Discipline Groups and Subgroups

<table>
<tr><th>Physical Science</th><th>Life Science</th><th>Earth and Space Sciences</th></tr>
<tr><td>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</td><td>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</td><td>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</td></tr>
</table>

Statements describe what students should know at each grade for each disciplinary concept. Most items will align to a statement within a single subgroup or part of a statement. However, items may be aligned to statements from multiple subgroups, including statements from different disciplines. As described in the section on multidimensional alignment, items that address statements from more than one discipline should have a primary alignment for the purpose of assigning a reporting category.

Item Distribution

The framework calls for an even distribution of items across the three discipline groups at each of grades 4, 8, and 12. The distribution of items among the three discipline groups reflects the equal importance given to each discipline of science in the assessment. For the 2028 NAEP Science Assessment, the equal distribution is a shift from prior assessments, which placed a greater emphasis on Earth and Space Sciences in grade 8 and less emphasis in grade 12.

Exhibit 2.2. Approximate Distribution of Items by Discipline Group 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%

With respect to crosscutting concepts and science and engineering practices, the emphasis should be on meaningful representation rather than a strictly equal distribution, as described in the sections below.

NAEP Science Disciplinary Concepts

Statements that describe what students should know at each grade for each disciplinary concept were guided primarily by *A Framework for K–12 Science Education*, which was informed by state science standards—the vast majority of which were also guided by *A Framework for K–12 Science Education* and modified by each state through input from internal and external reviews.

Each of the three discipline groups, listed in the definition of the NAEP Science construct, were first broadly described, followed by an iterative process of selecting disciplinary concepts that best represent the central principles for developing successively deeper understanding at grades 4, 8, and 12. Selection of each set of statements and refinement of the individual concepts followed an iterative process intended to limit the breadth of the dimension to a range that could feasibly be assessed by the NAEP Science Assessment at sufficient depth.

Similar Disciplinary Concepts Across Multiple Grade Levels

To ensure grade-appropriate alignment, developers should be aware of similar disciplinary concepts across multiple grade levels. For example, P4.2, P8.3, and P12.3 all describe states of matter. However, using a particle model to explain phenomena related to the properties of matter in a particular state is introduced in grade 8; the energy at the particle level is introduced at grade 12. Developers should compare these DCs, along with the clarification and boundaries at the lower grades, to ensure that phenomena selected for assessment can be modeled and explained using grade-appropriate understandings of the DCs.

Specifications Added to Disciplinary Concept Exhibits

Exhibits for the NAEP Science Disciplinary Concepts from the Framework have been augmented in this document (Exhibits 2.3A, etc.). The Specifications include bracketed text that presents clarifications and boundaries to inform item development. As noted in the section on

NAEP Science Assessment Terminology, clarifications that use the term *include* provide an open-ended list of examples that can be used in items.

Disciplinary Concepts Statements

The following exhibits provide the statements with which items should be aligned and integrated with the other dimensions.

NAEP Physical Science Disciplinary Concepts

Exhibit 2.3A. Topic: Matter and Its Properties

Overarching Question: *How can the great variety of substances and processes of change in matter be explained?*

Subtopics	Grade 4	Grade 8	Grade 12
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.]	
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.	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

Subtopics	Grade 4	Grade 8	Grade 12
Structure of Matter (continued)			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.]
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	

Subtopics	Grade 4	Grade 8	Grade 12
Phases of Matter and Atomic Substructure (continued)		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:** The emphasis 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.]	
Chemical Processes		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. [**Clarification:** The emphasis 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	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. P12.4: A stable molecule has less energy 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 for the initial molecule or molecules to

Subtopics	Grade 4	Grade 8	Grade 12
Chemical Processes (continued)		distinguished at this level, and students are not expected to remember atomic masses.]	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 to 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.]

Subtopics	Grade 4	Grade 8	Grade 12
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.]

Exhibit 2.3B. Topic: Motion and Forces

Overarching Questions: *How can motion be described? What makes the motion of an object change?*

Subtopics	Grade 4	Grade 8	Grade 12
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. [**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.] [**Boundary:** At this grade level, gravity is not discussed.]	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 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.]	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 aligned forces is included but non-collinear forces are treated only qualitatively.]
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

Subtopics	Grade 4	Grade 8	Grade 12
Forces between Objects (continued)			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., $p=mv$]
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.

Exhibit 2.3C. Topic: Energy

Overarching Questions: *Why do we care about keeping track of energy? Why are so many different phenomena associated with energy?*

Subtopics	Grade 4	Grade 8	Grade 12
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:** Emphasis 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.
Kinetic and Potential Energy	P4.7: Objects in motion have energy. The faster a given object is moving, the more energy it has.	P8.12: The energy of motion of particles 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 $\frac{1}{2}mv^2$ for given values of mass m and speed v, but does not include rearranging the	P12.13: Energy is a quantitative property 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.

Subtopics	Grade 4	Grade 8	Grade 12
Kinetic and Potential Energy (continued)		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:** Emphasis is on macroscopic objects and their mass, charge, and magnetic properties.]	[**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. When energy is transferred into or out of a system it can change the kinetic energy, the potential energy, or both within that system. Increased thermal motion with a corresponding rise in temperature is an example of the former. Phase change, when a material changes state (solid, liquid, or gas) as energy is added to or removed from a material but temperature does not change, is an example of the latter.] [**Boundary:** The quantum model of radiation as a flux of particles is not introduced.]
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

Subtopics	Grade 4	Grade 8	Grade 12
Thermal and Radiant Energy (continued)		[**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.]	atmosphere. [**Clarification:** Various gasses, known as greenhouse gasses, present in the atmosphere absorb and reradiate 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.]

Subtopics	Grade 4	Grade 8	Grade 12
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.

Exhibit 2.3D. Topic: Waves and Their Roles as Carriers of Information

Overarching Questions: *How can information be encoded, sent over long distances, and decoded? What physical phenomena do we use to do this?*

Subtopics	Grade 4	Grade 8	Grade 12
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 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, Wi-Fi, 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.]	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 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).
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.]	

Subtopics	Grade 4	Grade 8	Grade 12
Sound Waves (continued)		[**Boundary:** Assessment should include details of sound wave forms.]	
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.

NAEP Life Science Disciplinary Concepts

Exhibit 2.4A. Topic: From Molecules to Organisms: Structures and Processes

Overarching Question: *How do organisms live, grow, respond to their environment, and reproduce?*

Subtopics	Grade 4	Grade 8	Grade 12
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.] 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.	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:** Emphasis is on the idea that all life functions, including growth and reproduction, are a result of chemical reactions.] [**Boundary:** Assessment is limited to the following types of molecules: water, proteins, carbohydrates, lipids, and nucleic acids. 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 synthesis. Assessment does not include the names, steps, or specific processes involved in chemical reactions.] L12.2: Multicellular organisms have a hierarchical organization, in which its systems support functions necessary for

Subtopics	Grade 4	Grade 8	Grade 12
Structure and Function of Living Things (continued)		[**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.]	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.]

Subtopics	Grade 4	Grade 8	Grade 12
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 cell division 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.]
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:** Emphasis is on what plants and animals need. Examples could include that animals need to take in food, but plants do not, the different kinds of food needed by different types	L8.4: Photosynthesizers (i.e., 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.] [**Boundary:** Assessment does not include the biochemical mechanisms of	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

Subtopics	Grade 4	Grade 8	Grade 12
Matter and Energy in Organisms (continued)	of animals, the requirement of plants to have light, and that all living things need water.]	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.]	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

Subtopics	Grade 4	Grade 8	Grade 12
Matter and Energy in Organisms (continued)			include specific matter or energy transfer steps.]

Exhibit 2.4B. Topic: Ecosystems: Interactions, Energy, and Dynamics

Overarching Question: *How and why do organisms interact with their environment, and what are the effects of these interactions?*

Subtopics	Grade 4	Grade 8	Grade 12
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 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.	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. [**Clarification:** These limits result from such factors as the availability of living and nonliving resources and from such challenges as predation, competition, and disease.]
Cycles of Matter and Energy Transfer	L4.5: Much of the matter (materials) organisms need to grow and survive comes from other organisms and that same matter is used again later by other organisms. [**Clarification:** Emphasis is on the reuse	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	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

Subtopics	Grade 4	Grade 8	Grade 12
Cycles of Matter and Energy Transfer (continued)	of matter from one organism to another.]	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.] [**Boundary:** Assessment should not include identification of microscopic organisms.]	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.]
Ecosystem Dynamic, 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 and natural or human-induced disturbances. Examples of ways ecosystem health could be measured include ecosystem services (cleaning	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,

Subtopics	Grade 4	Grade 8	Grade 12
Ecosystem Dynamic, Functioning, and Resilience (continued)		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 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.]	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.] 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.]

Exhibit 2.4C. Topic: Heredity: Inheritance and Variation of Traits

Overarching Questions: *How are the characteristics of one generation passed to the next? How can individuals of the same species and even siblings have different characteristics?*

Subtopics	Grade 4	Grade 8	Grade 12
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 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.]	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. [**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.]	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 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.]
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.	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

Subtopics	Grade 4	Grade 8	Grade 12
Variation (continued)		Variations of inherited traits between parent and offspring arise from the inherited subset of chromosomes (and therefore genes). [**Clarification:** Emphasis is on conceptual understanding that changes in genetic material may result in making different proteins.] [**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 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.	mutations, which are a source of genetic variation. Environmental factors can also cause mutations in genes, and mutations can be inherited. [**Boundary:** Assessment does not include the steps of meiosis.] L12.16: Environmental factors affect 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.]

Exhibit 2.4D. Topic: Biological Evolution: Unity and Diversity

Overarching Questions: *How can there be so many similarities among organisms yet so many different kinds of plants, animals, and microorganisms? How does biodiversity affect humans?*

Subtopics	Grade 4	Grade 8	Grade 12
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 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.]	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 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.]	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 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.]
Mechanisms of Change	L4.9: Species change over time. Sometimes the differences in characteristics between individuals of the same species provide advantages in	L8.15: Adaptation by natural selection acting over generations is one important process by which species change over time in response to	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

Subtopics	Grade 4	Grade 8	Grade 12
Mechanisms of Change (continued)	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 offspring. Examples of environmental changes could include changes in land characteristics, water distribution, temperature, food, and other organisms.]	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 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.]	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 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.]

Subtopics	Grade 4	Grade 8	Grade 12
Mechanisms of Change (continued)			[**Boundary:** Assessment does not include other mechanisms of evolution, such as genetic drift, gene flow through migration, and coevolution.]

NAEP Earth and Space Sciences Disciplinary Concepts

Exhibit 2.5A. Topic: Universe, Solar System, and Earth

Overarching Question: *How do we explain Earth's relationship to objects in space?*

Subtopics	Grade 4	Grade 8	Grade 12
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; assessment does not include the names or shape of specific constellations.]	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 seafloor cores.] [**Boundary:** Items should not assess knowledge of any specific episode of time or change in paleoclimate.]
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	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	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

Subtopics	Grade 4	Grade 8	Grade 12
Solar System (continued)	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.]	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.]	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.]
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 redshift and blueshift, 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

Subtopics	Grade 4	Grade 8	Grade 12
Formation of the Universe (continued)			immediately following the Big Bang.]

Exhibit 2.5B. Topic: Earth's Systems

Overarching Question: *What are the Earth's systems, and how do they change?*

Subtopics	Grade 4	Grade 8	Grade 12
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 ranges, earthquakes, and volcanoes.] [**Boundary:** Assessment should focus on observable features and phenomena on the surface of the Earth.]	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. [**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.]	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. 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.]
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	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 occurred in terms of relative time. [**Clarification:** Emphasis is on how evidence from rock strata is used to	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

Subtopics	Grade 4	Grade 8	Grade 12
Earth's History (continued)	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 knowledge of specific fossil organisms. The expression of time is limited to sequences that represent relative time.]	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.]	rocks, and minerals.] [**Boundary:** Calculations of age dates should be limited to geometric progressions (doubling/halving, such as in half-lives).]
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.]

Subtopics	Grade 4	Grade 8	Grade 12
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).] E8.9: 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.]

Exhibit 2.5C. Topic: Earth and Human Activity

Overarching Question: *How do Earth's system processes and human activities affect each other?*

Subtopics	Grade 4	Grade 8	Grade 12
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 system 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 extraction 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.]
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	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.]

Subtopics	Grade 4	Grade 8	Grade 12
Natural Hazards (continued)		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.]	
Human Impacts on Earth's 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

Subtopics	Grade 4	Grade 8	Grade 12
Human Impacts on Earth's Systems (continued)			two global regions (continental scale).]
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 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.]	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 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.]

NAEP Science and Engineering Practices

Students use the practices of science and engineering to apply disciplinary concepts and crosscutting concepts as they demonstrate sensemaking of phenomena and solve problems. The list of practices described in *A Framework for K–12 Science Education* is now used in science classrooms across the country, and thus the NAEP Science and Engineering Practices reflect this list. Exhibits 2.7–2.14 define the expectations of students at grades 4, 8, and 12 in using these practices.

The eight 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

While prior assessments have included practices, the 2028 NAEP Science Assessment represents a shift away from identifying and using concepts, and toward a broader range of practices that describe how scientists and engineers, and therefore students, engage in doing science. The sub-statements in the Framework 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. When the same wording in a sub-statement is used for an SEP at different grade levels, the sophistication of student performances is expected to change based on the disciplinary concept. When the disciplinary concept is more complex, the use of the SEP becomes more complex.

Operationalizing the NAEP Science and Engineering Practices

The science and engineering practices describe the thought processes or scientific reasoning in which students engage as they apply concepts to make sense of a phenomenon or solve a problem. Multidimensional items require students to use a practice rather than just describe the practice or process itself and to integrate the practice together with the other dimensions. This section, together with the corresponding section of the Framework, describes the practices in several levels of detail to communicate how these practices describe the work of scientists, engineers, and others in STEM and related professions, the intent of shifts in science education to more closely reflect students' progress toward mastery of these practices, and the ways in which the assessment is designed to elicit evidence of this progress.

Balance of Science and Engineering Practices

In order to limit the total number of assessment targets, development for the 2028 NAEP assessment will pair the eight practices into the four categories shown in Exhibit 2.6. The practices in each pair are often used with common purpose, described by the practice categories of Investigating, Analyzing, Explaining, and Evaluating. Items should assess the Science and Engineering Practice within each category and not the category itself. Categories are only used to ensure balance across the assessment.

Because the selection of a practice for an item is based on appropriate use of the practice for applying a disciplinary concept and a crosscutting concept, the balance of science and engineering practices is not precisely prescribed. However, to appropriately represent the construct of science achievement, each SEP category will be used in at least 10 percent of items. Although this threshold applies to grades 4, 8, and 12, the practice categories may not be assessed equally across all three grades. For example, emphasis on analyzing may increase in higher grades as students develop the broader mathematical understanding used to engage in the practices within this category.

Exhibit 2.6. Balance of NAEP Science and Engineering Practice Categories

Practice category	Percentage of items	Science and Engineering Practices
Investigating	At least 10%	Asking Questions and Defining Problems
		Planning and Carrying Out Investigations
Analyzing	At least 10%	Analyzing and Interpreting Data
		Using Mathematics and Computational Thinking
Explaining	At least 10%	Developing and Using Models
		Constructing Explanations and Designing Solutions
Evaluating	At least 10%	Engaging in Argument From Evidence
		Obtaining, Evaluating, and Communicating Information

More guidance around pairing crosscutting concepts with disciplinary concepts and science and engineering practices can be found in Exhibit 2.19.

Science and Engineering Practice Statements

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 in Exhibit 2.7 are those that can reasonably be the practice element of a science or engineering assessment item.

Exhibit 2.7. Asking Questions and Defining Problems

Aspects of the NAEP SEP	Grade 4	Grade 8	Grade 12
Asking questions to inform an investigation or develop an explanation or model of phenomena	S4.1: 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).] S4.2: Ask "what if" questions about a system or phenomenon being observed that could be investigated empirically.	S8.1: Ask questions to clarify and/or refine an observation, model, or explanation of phenomena; or to clarify and/or refine an engineering problem. S8.2: Ask questions that can be answered with empirical evidence to investigate relationships between variables in a system model or in phenomena.	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. S12.2: 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.
Asking questions as part of understanding, evaluating, and/or challenging the work of others	S4.3: Ask questions to clarify an argument or interpretation of a data set.	S8.3: Ask questions to clarify or respectfully challenge the evidence and/or the premise(s) of an argument or interpretation of a data set.	S12.3: 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.
Defining a design problem that addresses a need	S4.4: Define a design problem to provide a solution for a situation	S8.4: Define a design problem that considers relevant scientific	S12.4: Define a design problem that involves the development of a

Aspects of the NAEP SEP	Grade 4	Grade 8	Grade 12
Defining a design problem that addresses a need (continued)	people want to change that can be solved through the development of a new or improved object or tool.	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.	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 for collecting data. To plan for 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. Planning for field observations 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 planning for engineering investigations to test design solutions, students select tools, materials, and processes relative to constraints and criteria.

NAEP Science Assessment items should provide students with tools or instrument-specific information that is needed for successful item completion. Students will not be required to carry out experiments with physical equipment, but simulations or virtual laboratories could be made available for some items.

Exhibit 2.8. Planning and Carrying Out Investigations

Aspects of the NAEP SEP	Grade 4	Grade 8	Grade 12
Developing or revising an investigation plan	S4.5: Plan an investigation to explore a scientific question or design problem taking into consideration appropriate variables and tests.	S8.5: 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.	S12.5: 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	S4.6: Select methods and/or tools for collecting data.	S8.6: Select and evaluate tools to collect, record, and analyze data.	S12.6: Select and evaluate appropriate tools to collect, record, analyze, synthesize, and evaluate data.
Predicting expected outcomes	S4.7: Make predictions about what would happen if a variable changes. S4.8: Predict the outcome of an experiment or a design solution based on a model, a phenomenon, or a design plan.	S8.7: Predict the change in a dependent variable when a change in an independent variable occurs in an investigation or test of a design plan.	S12.7: 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. S12.8: 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

Aspects of the NAEP SEP	Grade 4	Grade 8	Grade 12
Predicting expected outcomes (continued)			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 2.9. Analyzing and Interpreting Data

Aspects of the NAEP SEP	Grade 4	Grade 8	Grade 12
Displaying data to observe patterns and relationships	S4.9: 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.	S8.8: 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).	S12.9: 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	S4.10: Analyze data to determine whether it supports or refutes a claim about a phenomenon or design solution.	S8.9: Analyze data to provide evidence to support or reject a model or explanation or use to improve a design solution.	S12.10: Analyze data to provide evidence to support or reject a model or explanation or use to optimize a design solution relative

Aspects of the NAEP SEP	Grade 4	Grade 8	Grade 12
Analyzing data to support or reject claims about phenomena or improve design solutions (continued)	S4.11: Analyze data from tests of two solutions to the same problem to compare the strengths and weaknesses of how each performs.		to criteria for success.
Evaluating the quality and adequacy of data		S8.10: Evaluate the limitations of the data for the intended use, considering factors such as quantity and quality of the data, the tools used to obtain it, and its presentation.	S12.11: Evaluate whether the data are sufficient in quantity, accuracy and reliability for the purpose intended and 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 will be at or below what is expected on the NAEP Mathematics Assessment. 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.

Exhibit 2.10. Using Mathematics and Computational Thinking

Aspects of the NAEP SEP	Grade 4	Grade 8	Grade 12
Using mathematics	S4.12: Apply simple mathematical concepts and/or processes (such as simple computation, measurement) to a scientific question or a design problem. [**Boundary**:	S8.11: Apply mathematical concepts and/or processes (such as ratio, rate, percent, basic operations, and simple computations) to scientific questions and/or design	S12.12: Apply mathematical techniques (such as functions, statistical reasoning, and computational algorithms) to represent and solve

Aspects of the NAEP SEP	Grade 4	Grade 8	Grade 12
Using mathematics (continued)	Computations involve addition and subtraction of whole numbers only and do not include converting units of measure.]	problems. S8.12: Interpret and use quantities involving ratios based on two different types of units of measure (e.g., speed, density, and population density).	scientific questions and/or design problems. S12.13: Interpret and apply ratios, rates, percentages, and unit conversions in the context of complicated measurement problems involving quantities with derived or compound units (such as mg/mL, kg/m^3, etc.).
Computational thinking	S4.13: Break a process into a series of steps.	S8.13: Use algorithms (a series of ordered steps) to solve a design problem. S8.14: Apply digital tools and/or mathematical concepts and arguments to test and compare proposed solutions to design problems.	S12.14: Apply or revise algorithms when analyzing data or designing, programming, testing, and revising scientific models, explanations, and design solutions. S12.15: 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 a 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 2.11. Developing and Using Models

Aspects of the NAEP SEP	Grade 4	Grade 8	Grade 12
Developing and using models to explain phenomena or design a solution	S4.14: Develop, use, and/or revise a model to describe and explain a phenomenon or describe a design proposal. S4.15: Identify and describe how the parts of a model and the relationships between them represent a phenomenon.	S8.15: 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. S8.16: Use a model to test ideas about phenomena in natural systems or proposed design solutions.	S12.16: 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	S4.16: Identify limitations of a model for a phenomenon in terms of what the model can or cannot yet explain.	S8.17: Evaluate limitations of a model for a phenomenon and propose revisions to address what the model cannot yet explain.	S12.17: 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

Students are expected to apply scientific knowledge to explain phenomena or to develop designs that offer a solution to a problem. Explanations 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. Designed solutions must be supported by tests of the design through prototypes or simulations.

Exhibit 2.12. Constructing Explanations and Designing Solutions

Aspects of the NAEP SEP	Grade 4	Grade 8	Grade 12
Data-based explanations	S4.17: Develop an evidence-based description or explanation supported by evidence and reasoning of a phenomenon or the action of a designed solution.	S8.18: 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.	S12.18: 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	S4.18: Relate an explanation of a phenomenon to a model.	S8.19: Evaluate whether a model provides sufficient explanation of the phenomenon and how the model could be revised to better explain the observations.	S12.19: Evaluate a model-based explanation or a design proposal using empirical evidence and the application of disciplinary concepts.
Designing and comparing solutions	S4.19: Compare multiple possible solutions to a design problem based on how well each is likely to meet the criteria and constraints of the problem.	S8.20: Apply scientific ideas or principles to propose tests or trade-offs needed to optimize a design.	S12.20: Evaluate and/or refine a solution for a 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 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. 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 2.13. Engaging in Argument From Evidence

Aspects of the NAEP SEP	Grade 4	Grade 8	Grade 12
Constructing an argument to support or refute a model, explanation, or design solution	S4.20: Construct and/or support an argument with evidence to support or reject a claim about a phenomenon or a design solution. S4.21: 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.	S8.21: Construct an argument with evidence and scientific reasoning to support or reject a proposed model, explanation, or design solution for a problem. S8.22: Identify evidence that could be used to refute a claim about a phenomenon.	S12.21: Construct an argument with evidence and scientific reasoning to support or reject a proposed model, explanation, or design solution for a problem.
Evaluating and/or improving an argument for an explanation, model, or design solution	S4.22: Evaluate an argument based on the evidence or reasoning it includes.	S8.23: Revise an argument that supports or rejects a model, explanation, or design solution for a problem to address new evidence. S8.24: Compare and critique two arguments on the same question to analyze their fit with the evidence and/or whether they emphasize similar or	S12.22: 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. S12.23: Compare and evaluate the arguments for two competing design solutions, based on design criteria, empirical evidence,

Aspects of the NAEP SEP	Grade 4	Grade 8	Grade 12
Evaluating and/or improving an argument for an explanation, model, or design solution (continued)		different evidence and/or interpretations.	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.

NAEP Science items 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. In addition, reading comprehension is not intended to be explicitly measured by NAEP Science items.

Exhibit 2.14. Obtaining, Evaluating, and Communicating Information

Grade 4	Grade 8	Grade 12
S4.23: Evaluate whether the information presented is	S8.25: Assess the credibility, accuracy, and possible bias of an article on a science topic (e.g.,	S12.24: Assess the credibility, accuracy, and possible bias of an article on a science topic

Grade 4	Grade 8	Grade 12
evidence, an opinion, or a fictional story. S4.24: Evaluate whether the information presented in a text summarizing a graph or table of data accurately reflects the claim that could be made from the data.	based on where it is found, the qualifications of the source, and/or the evidence given to make the claim). S8.26: Evaluate information from two different sources to determine whether there are conflicts between them. S8.27: Identify and critique standard flaws in science-related arguments (e.g., poor assumptions, cause vs. correlation, faulty explanations, or overgeneralizations from limited data).	(e.g., based on where it is found, the qualifications of the source, and/or the evidence given to make the claim). S12.25: Evaluate scientific and/or technical information from multiple sources, assessing the evidence used by and the information on qualifications and expertise of each source. S12.26: Identify and critique standard flaws in science-related arguments (e.g., poor assumptions, cause vs. correlation, faulty explanations, or overgeneralizations from limited data).

Multidimensional Alignment and Illustrations

All items will be at least two dimensional; each two-dimensional item will align to at least one disciplinary concept and one science and engineering practice. The following grade 4 item set, Moving Basketball, assesses NAEP SEPs that come from each of the four practice categories.

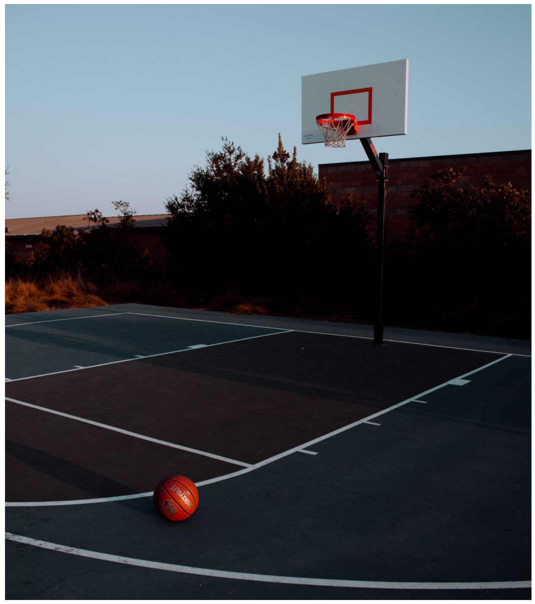
Exhibit 2.15. Moving Basketball SEPs

Practice category	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 Communications Information

Exhibit 2.16. Moving Basketball

Item Set ID: Moving Basketball

Aliya and Rishi go to an empty basketball court, as shown.[1]



They notice that a ball on the ground is moving, even though no one is pushing it. They make the following observations as they try to figure out why the ball is moving.

Aliya and Rishi's Observations
• It is very windy today—the wind is pushing us! • The ball moves more when we feel a lot of wind on our faces. • There are no other people nearby who might have been using the ball.

Item 1

Aliya claims that the wind is applying a force to the ball. Which observation would support Aliya's claim?

A. The ball moves after the wind stops blowing.
B. The ball moves against the wind as the wind blows toward the ball.
C. When the wind blows gently, the ball stays in place.
D. When the wind blows hard, the ball moves in the direction the wind blows.

Key: D

[1] Photo by https://unsplash.com/@matiasvizuals

Item 2

Rishi and Aliya decide to design an investigation to figure out if the wind can cause a ball to move. Which investigation plan will be the **most** helpful for figuring out if the wind can cause a ball to move?

A.

Plan 1
1. Place two balls on the basketball court at the same starting point. 2. Push one ball, and do not push the second ball. 3. Record: a. Speed of each ball b. Distance each ball moves 4. Repeat the steps pushing only the second ball and repeat again pushing both balls.

B.

Plan 2
1. Place one ball on the floor of a closed room. 2. Use a fan on low speed to blow air toward the ball. 3. Record: a. Speed of the ball b. Distance the ball moves 4. Repeat the steps with the fan on medium speed and repeat again with the fan on high speed.

C.

Plan 3
1. Place one ball on the basketball court. 2. Use a fan on low speed to blow air toward the ball. 3. Record: a. Speed of the ball b. Distance the ball moves

D.

Plan 4
1. Place two balls on the floor of a closed room at the same starting point. 2. Push one ball, and do not push the second ball. 3. Record: a. Speed of each ball b. Distance each ball moves

Key: B

Item 3

Aliya and Rishi gathered information while they watched the effects of the wind on the ball. Which of the following data **best supports the claim that wind causes unequal forces acting on the ball**?

A.

Trial [or observation]	Strength of wind	Ball movement
1	Very strong	Very fast
2	Strong	Fast
3	Weak	Slow

B.

Trial [or observation]	Strength of wind	Ball movement
1	Weak	Fast
2	Very weak	Fast
3	No wind	Fast

C.

Trial [or observation]	Strength of wind	Ball movement
1	No wind	Slow
2	Very weak	Slow
3	No wind	None

D.

Trial [or observation]	Strength of wind	Ball movement
1	Very strong	Very fast
2	Strong	Very fast
3	Strong	Slow

Key: A

Item 4

After observing the ball for some time, Aliya and Rishi notice that the ball moves for a while and then stops and later moves again. Rishi tells Aliya that sometimes he feels the wind blowing against his face, but the ball does not move.

Use your understanding of balanced and unbalanced forces to explain what is happening to the ball when the wind blows, but the ball does not move.

Exhibit 2.17. Moving Basketball Item 4 Constructed Response Scoring Notes

A complete answer includes reasoning that there are always multiple forces acting on the ball. When the wind blows and the ball's motion changes, the wind is exerting more force than the other forces acting on the ball. When the wind blows but the ball's motion does not change, the force from the wind on the ball is equal (or balanced) to the other forces acting on the ball. Emphasis for scoring is on reasoning with the idea of balanced and unbalanced forces, not on the use of exact vocabulary (i.e., balanced/unbalanced, equal/unequal, and similar language is acceptable).

NAEP Science Crosscutting Concepts

Broad concepts common to nearly all fields of science and engineering—unifying concepts or themes—have long been recognized as useful to science education for organizing and connecting knowledge (American Association for the Advancement of Science, 1990; National Research Council, 1996); their explicit inclusion as targets for the alignment of items has not. Recent shifts in science education called for in *A Framework for K–12 Science Education* have renewed emphasis on these crosscutting concepts, and most states' academic standards for science have included statements describing how items assess these concepts. The seven crosscutting concepts for the NAEP Science Assessment are the following:

- 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 2.20–2.26 define the expectations of students at grades 4, 8, and 12 for using these crosscutting concepts. One or more statements within each crosscutting concept and grade level describe in detail ways students may be asked to apply each crosscutting concept in items. Individual items aligned to a crosscutting concept address at least one of the statements; the assessment overall includes items aligned to at least one statement listed for each of the crosscutting concepts but need not include items aligned to every statement.

Operationalizing the NAEP Science Crosscutting Concepts

Items aligned to a statement in the crosscutting concepts dimension are written in ways that elicit explicit evidence of their use. For example, an item stem may present a dataset for which many patterns can be identified. However, if a correct response to the item is possible without using any of these patterns, the item is not aligned to the crosscutting concept Patterns. If a correct response to the item is only possible through recognizing or applying patterns to the data, the item is aligned, whether or not the item mentions the term *pattern*. Exhibit 2.18 includes additional approaches to eliciting evidence of crosscutting concepts.

Exhibit 2.18. Applying Crosscutting Concepts to Disciplinary Concepts

Crosscutting Concept	Application
Mechanisms and Explanation: Cause and Effect	Students use cause and effect reasoning about the disciplinary concept. For example, in an item aligned to P8.3, S8.15, and C8.3, students could be asked to develop a model to predict that a temperature change will cause a change in the motion of the molecules in a substance and a change in the physical state of the substance.
Scale, Proportion, and Quantity	Students use different scales to reason about the disciplinary concept. For example, in an item aligned to E12.6, S12.18, and C12.4, students could be asked to use data as evidence to construct an explanation of changes to a landform using interactions between water and various materials that make up the landform, including how interactions occurring over shorter time periods led to changes in the landform over long periods of time.
Scale, Proportion, and Quantity	Students compare the functions of a system at different scales to reason about the disciplinary concept. For example, in an item aligned to L8.2, S8.24, and C8.5, students could be asked to compare arguments about functions within a complex multicellular organism using evidence at the level of the organism as well as evidence at the cellular, tissue, and organ system level.

Approaches to operationalizing the NAEP CCCs may, for some items, need to limit the complexity of a three-dimensional item. One such approach would be to provide a description of some aspects of the concepts within a crosscutting concept. To be aligned, the item should then require students to use the description together with other aspects of the crosscutting concept to respond. Examples of ways to operationalize the NAEP CCCs while limiting complexity include the following:

- describing a pattern in the data before asking students to use the pattern to predict the timing of a natural hazard
- defining a system before asking students to quantify forces within said system or model energy inputs and outputs of the system
- providing some functions of the structures in a system before asking students to use these relationships to modify the functions or update the designs of the structures

Crosscutting concepts overlap with each other and with the other dimensions. Items that can be answered correctly with or without using the intended crosscutting concept could undermine the claim that students are able to apply that crosscutting concept. Assessments that are designed to elicit evidence that students can use crosscutting concepts include items that explicitly ask students to provide such evidence.

When applying a disciplinary concept to make sense of a phenomenon or solve a problem, students may find natural combinations of the other dimensions, as shown in Exhibit 2.19. Assessment developers may find these combinations helpful when seeking to integrate a crosscutting concept into an item. These combinations are intended solely as guidance and in no way restrict the integration of dimensions in items on the assessment.

Exhibit 2.19. Natural Combinations of Science and Engineering Practices with Crosscutting Concepts

Practice Category	Science and Engineering Practice	Crosscutting Concept
Investigating	Asking Questions and Defining Problems	All CCCs readily combine
	Planning and Carrying Out Investigations	Cause and Effect: Mechanism and Explanation Scale, Proportion, and Quantity Energy and Matter: Flows, Cycles, and Conservation
Analyzing	Analyzing and Interpreting Data	Patterns Cause and Effect: Mechanism and Explanation Scale, Proportion, and Quantity
	Using Mathematics and Computational Thinking	Patterns Scale, Proportion, and Quantity
Explaining	Developing and Using Models	Systems and System Models Cause and Effect: Mechanism and Explanation Structure and Function Stability and Change
	Constructing Explanations and Designing Solutions	Systems and System Models Cause and Effect: Mechanism and Explanation Energy and Matter: Flows, Cycles, and Conservation
Evaluating	Engaging in Argument From Evidence	Systems and System Models Cause and Effect: Mechanism and Explanation Stability and Change Scale, Proportion, and Quantity
	Obtaining, Evaluating, and Communicating Information	All CCCs readily combine

Balance of Crosscutting Concepts

There is no target range for the distributions of items aligned to the crosscutting concepts; instead, assessment development should emphasize the appropriate use of the crosscutting concept in three-dimensional items that are also aligned to the other two dimensions. Every crosscutting concept should be included in the assessment. As noted in the section on multidimensional alignment above, CCCs are included in three-dimensional items.

Crosscutting Concept Statements

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 its 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.

Exhibit 2.20. 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
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. [**Boundary:** Statistical displays are limited to bar graphs and pictographs for categorical data and line plots for measurement data (whole number measurements only).]	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.	C12.1: 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.

Mechanisms and Explanation: Cause and Effect

Cause and effect involves the search for the underlying cause of a phenomenon. 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. 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 use ideas related to 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). The NAEP CCC of Mechanisms and Explanation: Cause and Effect is also closely associated with the NAEP SEP 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 2.21. 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
C4.2: Cause-and-effect relationships are routinely identified, tested, and used to explain changes. C4.3: Events that occur together with regularity might have a cause-and-effect relationship or might have some other shared explanation.	C8.2: Relationships can be classified as causal or correlational, and correlation does not necessarily imply causation. C8.3: Cause-and-effect relationships may be used to predict phenomena in natural or designed systems. C8.4: Phenomena may have more than one cause, and some cause-and-effect relationships in systems can only be described using probability.	C12.2: Empirical evidence is required to differentiate between cause and correlation and make claims about specific causes and effects. 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.

Scale, Proportion, and Quantity

Scale, proportion, and quantity 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.

An understanding of scale involves understanding not only 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 a 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. Metric units of measure are used for grades 4, 8, and 12.

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 or 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 NAEP CCC of Scale, Proportion, and Quantity figures prominently in the NAEP SEPs of Using Mathematics and Computational Thinking and of 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 NAEP CCC also figures prominently in the NAEP SEP 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 2.22. 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
C4.4: Natural objects and/or observable phenomena exist from the very small to the immensely large or from very short to very long time periods.	C8.5: 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. C8.6: Time, space, and energy phenomena can be observed at various scales using models to study systems. Proportional 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.	C12.4: 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). C12.5: 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 thinking 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 so, 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. 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 can be 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.

Exhibit 2.23. 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
C4.5: 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.	C8.7: 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 to analyze and	C12.6: 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

Grade 4	Grade 8	Grade 12
	explain a phenomenon, and, after testing, to make predictions about the phenomenon. C8.8: Systems may interact with other systems; they may have subsystems and be a part of larger more complex systems. C8.9: Engineers design systems to achieve particular functions or do specific items. An engineering design plan includes a system model. Engineers also use system models to troubleshoot system failures.	models may include equations that describe relationships between relevant quantities in the system. C12.7: 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 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. It is informative to track the transfers of matter and energy within, into, or out of a system.

In many systems, there also are cycles of various type—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 2.24. 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
C4.6: To understand the function of a system, it is often	C8.10: Matter is conserved because atoms are conserved in	C12.8: Flows of matter and transfers of energy into, out of,

Grade 4	Grade 8	Grade 12
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 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.]	physical and chemical processes. C8.11: Energy manifests itself to our observation in multiple different ways, including in mechanical, thermal, electrical, and chemical processes. Energy can transfer between these different observed effects and between objects or systems. C8.12: 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.	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 or transfer of energy into or out of the system. C12.9: 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 depends on the shapes and relationships of key parts as well as on the properties of the materials from which they are made. 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.

Understanding how a bicycle works involves examining the structures and their functions at the scale of the frame, wheels, pedals, and so on. However, building a 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 change the heaviness of the bicycle by using 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 can be designed and fabricated.

Exhibit 2.25. 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
C4.7: Different materials have different substructures, which can influence how they behave (function). C4.8: Within any system, natural or designed, the structures of objects, their composition, influences the overall function of the system and its subsystems.	C8.13: 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. C8.14: Structures can be designed to serve particular functions by taking into account properties of different materials and how materials can be shaped and used.	C12.10: 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. C12.11: 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. 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. 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 NAEP DCs in Chapter 2 of the Framework, the sample items in Chapter 3 of the Framework, and here in the NAEP Science Assessment and Item Specifications.

Exhibit 2.26. 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
C4.9: Change in conditions can be described or predicted 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).]	C8.15: Stability or change over time in a system depends on external conditions as well as on relationships and conditions within the system. C8.16: Systems can appear stable on one time scale but viewed on a longer time scale are seen to be changing.	C12.12: Rates of change are quantifiable and are important quantities to consider in modeling any system. C12.13: Feedback mechanisms within a system are important elements for explaining or designing for either the stability or instability of the system. C12.14: 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.

Multidimensional Alignment and Illustrations

Three-dimensional items align to at least one statement from each of the dimensions: disciplinary concepts, science and engineering practices, and crosscutting concepts. Before writing an item, developers will find it helpful to first draft a science achievement expectation that combines the statements from the targeted dimensions. The following multipart, discrete item requires grade

12 students to use a NAEP DC and SEP together with the NAEP CCC Patterns in their response to the questions. The science achievement expectation drafted for this item is "Students can analyze and interpret patterns in data about a phenomenon to determine the likelihood that certain factors determine carrying capacities within ecosystems." Exhibit 3.34 in Chapter 3 provides additional examples of science achievement expectations from physical, life, and earth and space sciences, along with the targeted statements from each dimension and a rationale for the combination.

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 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 NAEP CCC element, in conjunction with the other two dimensions. The NAEP DC, SEP, and CCC are all very straightforward and clearly cued to students.

The same context can support items that align with different SEPs and CCCs. For example, the short and extended constructed response item in Exhibit A.18 illustrates how this context can be used to elicit evidence of students' understanding of parts of different SEPs and a different CCC.

Exhibit 2.27. Wild Dogs

Item ID: Wild Dogs (adapted from OpenSciEd)

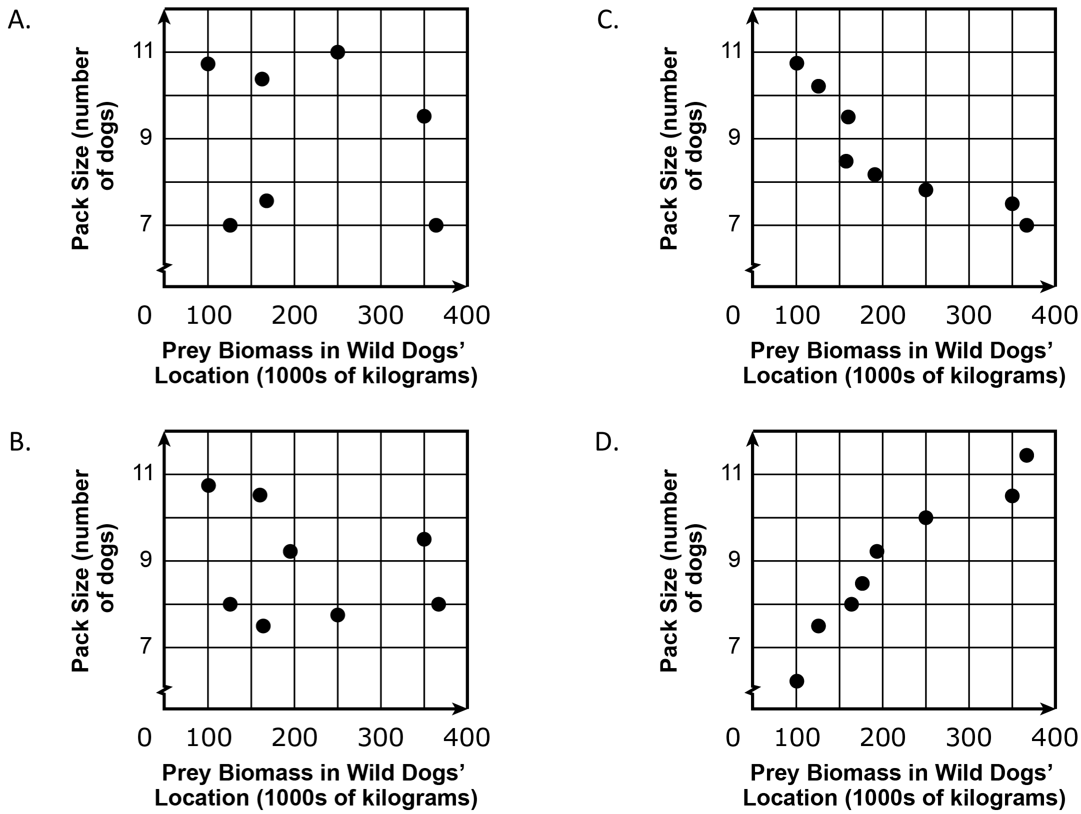
There has been a fast decline in African wild dog populations across Africa. Conservation experts are trying to establish a new population in a national park in Malawi and are trying to figure out how many wild dogs the area can support. Pictures of an African Wild Dog and its primary prey, the antelope, are shown.[2]

African Wild Dog

Antelope

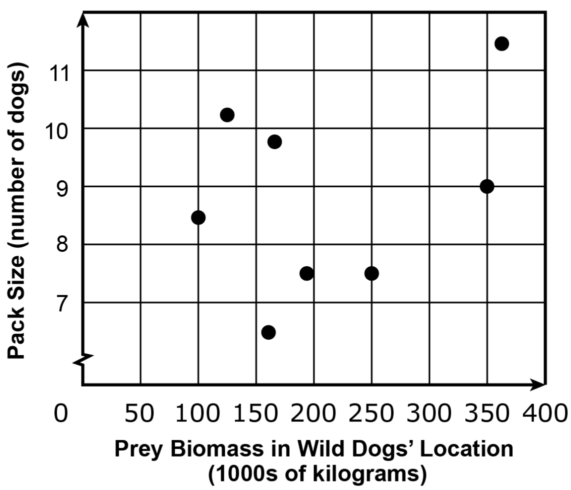
[2] African Wild Dog by Derek Keats, CC BY 2.0, via Wikimedia Commons; Hein waschefort, CC BY-SA 3.0, via Wikimedia Commons

Part A. The amount of food available, or prey biomass, in the national park is a limiting factor in African wild dog population size. Which graph represents the relationship you would expect to see between prey biomass and population size for African wild dogs in the national park if prey biomass is a limiting factor?



Key: D

Part B. The graph shows the actual relationship between African wild dog pack size and prey biomass that scientists observed in a similar territory. Based on the graph, complete the sentence to describe whether prey biomass plays an important role in determining the carrying capacity of African wild dogs in this location.



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.

Use the graph to support your answer:

Exhibit 2.28. Wild Dogs Part B Constructed Response Scoring Notes

Student reasoning should indicate that they correctly interpret the graph as having no clear relationship between prey biomass available and pack size, indicating that prey biomass does not play a very important role in determining carrying capacity in this area.

- Some students may note that there is a slight trend toward a positive relationship between prey biomass and pack size. While the relationship is limited at best, some students may see it as enough to warrant noting. If students do note this relationship, they should be awarded credit as long as it is clear that they are reading and interpreting the graph correctly and using an accurate understanding of carrying capacity to make the connection between food availability and pack size.

CHAPTER 3: Overview of the Assessment Design

This chapter provides an overview of the major components of the science assessment design, which includes the types of assessment items and how they can be used to expand the ways in which students are asked to demonstrate what they know and are able to do in science. In addition, this chapter describes how the assessment is distributed across the disciplines and practices described in Chapter 2.

Previous NAEP Science assessments included content knowledge and four practices. In order for students to demonstrate what they know and can do with respect to the range of disciplinary concepts, crosscutting concepts, and practices in the Framework, the 2028 NAEP Science Assessment should include two and three-dimensional items that may be discrete or may be components of an item set or scenario-based task. The sensemaking required by items is always based on a connection between the item and a phenomenon, which may be included within the item or, in the case of item sets and scenario-based tasks, presented in one or more stimuli.

Items must be at least two dimensional, requiring students to engage in at least one science and engineering practice in order to apply a disciplinary concept to the phenomenon. Items that are three dimensional also require students to apply a crosscutting concept. Two-dimensional items are designed to gather evidence that students can use the practices to apply their understanding of a disciplinary concept. Items that ask students to recall component knowledge that is learned through the use of practices would tend to assess the practice separately from the other dimensions rather than requiring the student to use the practice in an integrated way. Three-dimensional items also gather evidence that students can use the practices to apply their understanding of a crosscutting concept.

The 2028 NAEP Science Assessment should continue to be developed to allow for the widest possible range of students to demonstrate sensemaking in science by following the principles of universal design for assessment.

Types of Items

The two primary types of items on the 2028 NAEP Science Assessment are selected response and constructed response. Selected response items include a range of formats that all require a student to select one or more response options from a given, limited set of choices. Constructed response items require students to create a response that could be numerical or text-based; graphical responses could be included within parts of multipart items that also include constructed response parts. However, students would not be expected to include free drawing in any response.

Balance by Response Type

The assessment will consist of about 65 percent selected response items and 35 percent constructed response items. Since items requiring a constructed response 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 3.1. Approximate Distribution of Items by Response Type

Type of Response	Distribution
Selected response	65% of items
Constructed response	35% of items

Selected Response

Within the range of selected-response formats are more traditional single-selection and multiple-selection multiple choice items, as well as a variety of innovative formats made possible by digital test administration, often labeled "technology-enhanced" in state assessment programs. The following list includes the formats that may be developed as discrete items, used as components of multipart items, grouped around a common stimulus in an item set, or interwoven in a storyline in a scenario-based task:

- **Single-selection multiple choice:** Students respond by selecting a single choice from a set of given choices.
- **Multiple-selection multiple choice:** Students respond by selecting two or more choices that meet the condition stated in the stem of the item.
- **Matching:** Students respond by inserting (i.e., dragging and dropping) one or more source elements (e.g., an image) into target fields (e.g., a table).
- **Zones:** Students respond by selecting one or more regions on a graphic stimulus.
- **Grid:** Students evaluate statements, such as claims or explanations, or classify components of a system based on their properties or interactions. The answer is entered by selecting cells in a table.
- **Inline choice:** Students respond by selecting one option from one or more drop-down menus that may appear in various sections of an item.

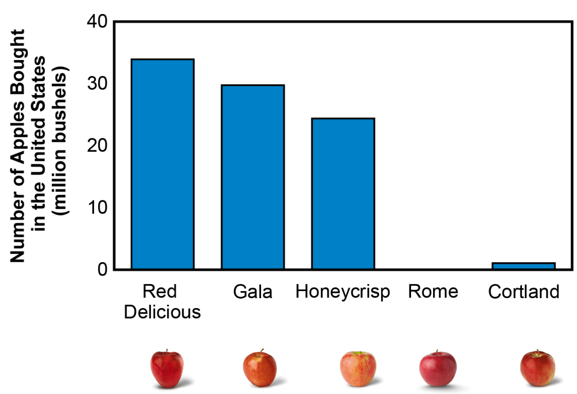
Exhibit 3.2 provides an example of a discrete, single-selection multiple choice item that assesses all three dimensions at grade 4. In this item, students analyze patterns in data to determine which parent apples are most likely to produce offspring with desirable characteristics.

Exhibit 3.2. Selecting Apples

Item ID: Selecting Apples

Millions of people buy apples to eat every day. Apples are getting harder to grow because of extreme changes in temperature that are happening in many apple-growing areas. The graph shows the amounts of the five types of apples that people in the United States (U.S.) bought most often in one year. The table shows information about each of the types of apples shown in the graph.[3]

Apples People in U.S. Bought Most Often



Apple Facts

Type of apple	Characteristics	What temperature will the apples grow in?
Red Delicious	Sweet, crisp	Mild
Gala	Sweet, crisp	Mild
Honeycrisp	Sweet/tart, crisp	Hot, mild, cold
Rome	Tangy, dense/soft	Hot
Cortland	Tart, a little crisp	Hot, mild, cold

A farmer wants to use two of the apples in the graph and table as parents to produce a new apple that has these characteristics:

- It is an apple that people buy most often at stores in the U.S.
- The apple will grow in the widest range of temperatures.

Based on the information in the graph and the table, which two types of apples could be used as parents to produce a new apple that has these characteristics?

A. Red Delicious and Gala
B. Red Delicious and Rome
C. Rome and Cortland
D. Gala and Honeycrisp

Key: D

Constructed Response

Constructed response items for the NAEP Science assessment include two primary formats: short and extended constructed response. Constructed response formats may be discrete, parts of multipart items, and may be included in item sets and scenario-based tasks. Within the short constructed response format are a variety of ways that students might respond:

[3] Apple images from New York Apple Association, applesfromny.com

- **Short constructed response:** Students respond by giving a short response, from a single word or number to a few sentences. For example, students may label a model or system, classify data, or describe a pattern in a system or dataset.
- **Extended constructed response:** Students respond by giving a description or explanation that requires more than a few words. For example, students may explain a system model with supporting evidence, synthesize information from multiple sources, or describe a process with multiple components or interactions.

Exhibit 3.3 provides an example of a discrete, constructed response item that assesses all three dimensions at grade 12. In this item, students construct an argument for why the temperature of a system (phenomenon) does not change during a phase change.

Exhibit 3.3. Melting Ice

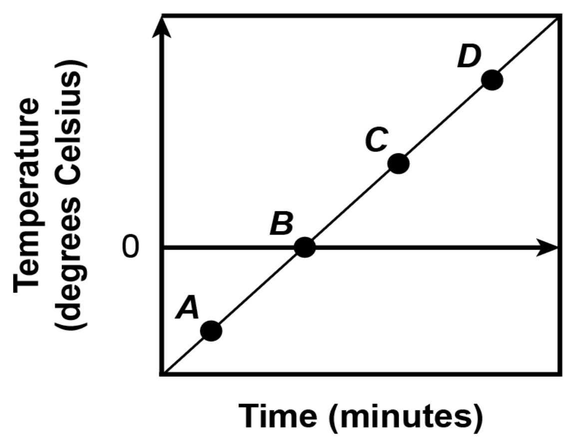
Item ID: Melting Ice (adapted from CREATE for STEM, Michigan State University)

When ice cubes are left out in the sun in hot weather, they melt. The table shows the observations a student made while watching ice cubes melt. The student created the graph to show how he expected the temperature of the ice cubes to change while they melted.

Student's Observations

Time Segment	Observation
A–B	Solid ice
B–C	Mixture of ice and water
C–D	Liquid water

Student's Predicted Temperature Changes



Use your understanding of what happens at the molecular level during a phase change to make an argument that supports or refutes the student's temperature change prediction. In your argument, be sure to state whether you agree or disagree with the student's prediction and describe what happens during each time segment (A–B, B–C, C–D), including the following:

- energy flows
- phases of matter
- molecular motion
- forces between molecules
- changes in temperature

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 whether 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. Exhibit 3.4 provides an example of part of a preliminary scoring guide developed for the item in Exhibit 3.3.

Exhibit 3.4. Melting Ice Constructed Response Scoring Notes

Student responses should indicate that the linear prediction of temperature increase over time does not consider the energy required for phase changes (melting) at the molecular level. Student responses should show that they understand that temperature is a measurement of energy—as temperature of a system rises, this leads to more molecular motion. During these phase changes, energy is absorbed to break intermolecular forces and transition between phases, resulting in a plateau in the temperature-time graph—the temperature does not increase because the additional energy is being used to overcome the intermolecular forces.

Item Characteristics

The specific components of an item are determined by item type. Two components are constant across all item types: (1) the item stem and (2) the response. The item stem is the introduction to the item and the question asked of, or directive given to, students. In discrete items, the item stem should include a stimulus that presents a compelling phenomenon or problem and provides all of the necessary information for students to respond, clearly laying out for the students what is being asked and the expected response method. Alternatively, items may split up into parts or may be arranged in item sets and scenario-based tasks; the stimulus is separate from the item stems in these types of arrangements. For items with multiple parts, items that are parts of item sets, and items that are parts of scenario-based tasks, the stimulus provides some of the necessary information for students to respond to the individual parts or items.

A **discrete item (DI)** is a single, standalone item. Students need to be able to read the stimulus/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. Discrete items may be any of the types listed above in the section on types of items.

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 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.

The parts of a multipart item may be preceded by a stimulus that presents the phenomenon and contextual information. Alternatively, the phenomenon may be presented in the first item part. Subsequent parts draw upon this same phenomenon and each item part may present different aspects of the phenomenon. The parts may be of different types, including any of the types listed in the section on item types.

A multipart item provides an opportunity to assess the dimensions separately or together; each part of the multipart item may be one, two, or three dimensional. The majority of multipart items would be expected to make use of their multipart construction to assess all three dimensions.

Although a one-dimensional item is not allowed, an item part could be one dimensional if it were only part of a multipart item that was, overall, at least two dimensional. A one-dimensional item part would not automatically be easier than a two-dimensional item part. However, item parts that assess a single dimension may afford developers with additional flexibility, together with other factors that reduce difficulty, to provide unique opportunities for students performing at the *Basic* and *Approaching Basic* levels to demonstrate understanding not available in items that have only one part that is required to be at least two dimensional.

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 during pilot testing because it is found to not be functioning as intended, the other items in the set may be preserved. 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. Item sets should include at least one item that is three dimensional.

The item set in Exhibit 3.5 illustrates how a robust phenomenon or problem to solve can give rise to items across multiple disciplines (e.g., Life Science and Earth and Space Sciences). As 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: Locust swarms can cause a lot of damage. Students are asked to consider what is making swarming worse and what solutions there might be to this problem.

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 middle and high school levels in both Life Science and Earth and Space Sciences. This version of the item set was selected to show how a wider range of NAEP 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.

Scoring an item set such as this one involves scoring each item independently. That the item is part of a set does not affect how the score on the item is included in the total score on the assessment. Within a multipart item, each part may have a maximum score of 1 or more points. The total score for the item is calculated by adding the scores on each part; this total is then aggregated with the scores on all the other items on the assessment.

Exhibit 3.5. Locusts

Item ID: Locusts

Common stimulus (available with all items in the task)

A desert locust is an insect that undergoes changes to its body in certain environmental conditions. Figure 1 describes some differences between two modes of a desert locust.

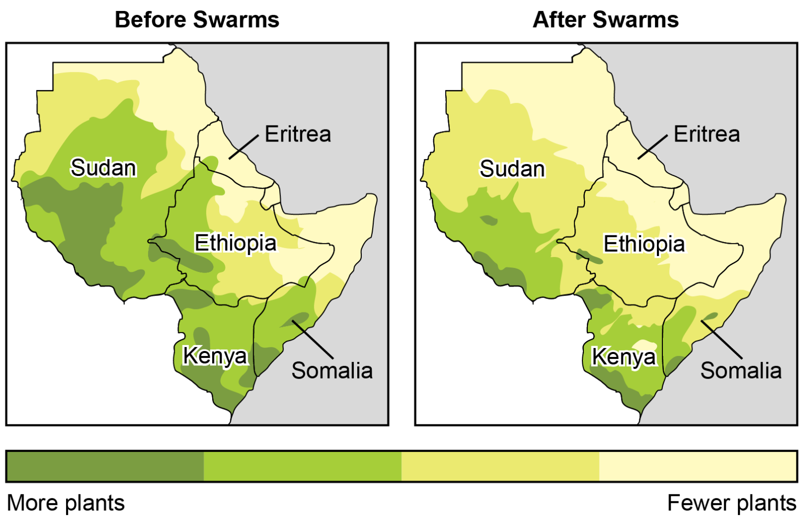
Figure 1. Desert Locust[4]

Mode 1: Grasshopper (Dry, warm or cool weather)	Mode 2: Locust (Wet/rainy, warm or hot weather)
• Behave independently • Stay away from other desert locusts • Mostly walk slowly and jump • Limited 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.	• Behave as a united group (swarms) • Gather together with 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 six months.

[4] Bernard DuPont Bird Locust Attribution-ShareAlike (CC BY-SA 2.0); Magnus Ullman, CC BY-SA 3.0, via Wikimedia Commons

When these insects are in Mode 2, they are able to swarm. A single swarm of locusts can cover an area of up to 100 square miles, with 40 to 80 million locusts in each square mile. Swarms can travel up to 100 miles a day. Figure 2 shows the effect of three months of locust swarms on available vegetation in an area of Africa.

Figure 2. Available Vegetation Before and After Locust Swarms



Item 1

An individual desert locust eats about 2 grams of food each day. Using the data provided, complete the sentences by choosing the correct answers from each dropdown menu.

When desert locusts are in Mode 1, the available vegetation in the areas where they live will likely be **most** similar to the (before/after) swarms map shown in Figure 2.

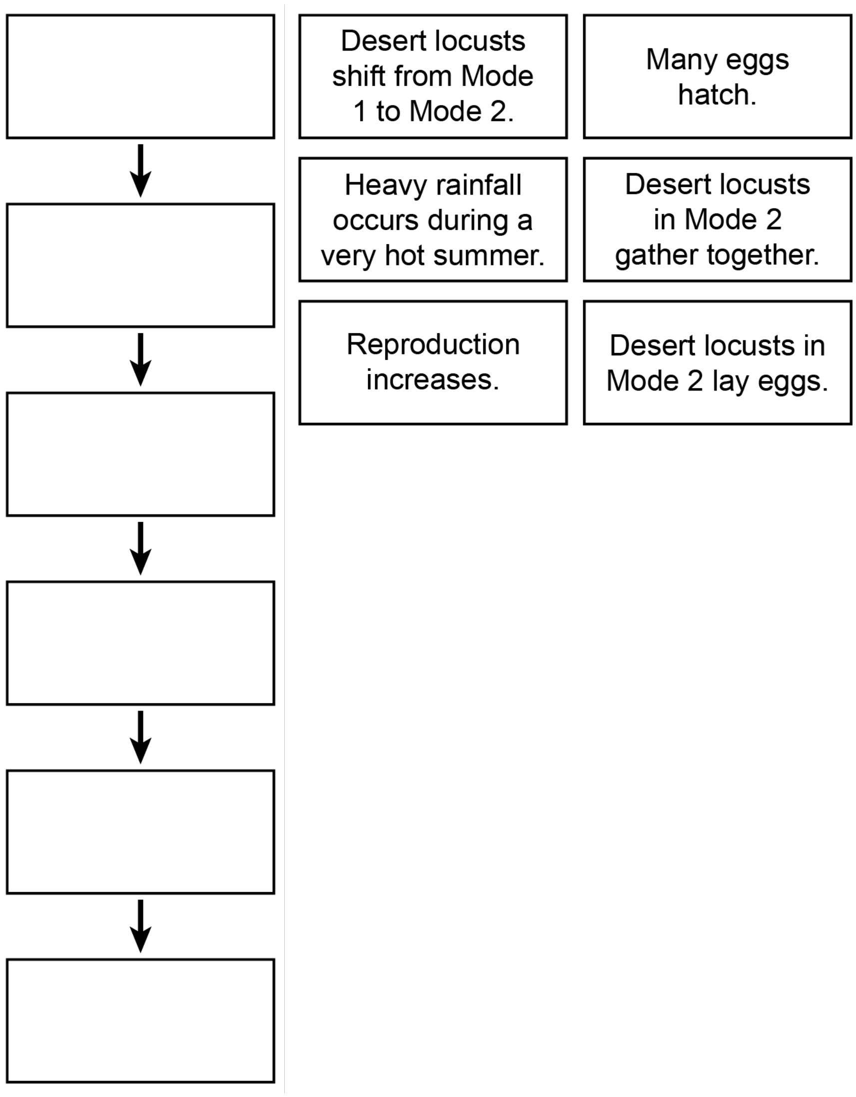
Key: before

This is because, in Mode 1, desert locusts eat a (small/large) amount of plants compared to all the plants available in an area and, therefore, have a (small/large) impact on an area's available plant life.

Key: small; small

Item 2

Use the statements to develop a model to show how locust populations in Mode 2 can become large enough to swarm. Drag the statements into the correct boxes to complete the model. Each statement will be used once.



Key: Heavy rainfall occurs during a very hot summer. Many eggs hatch. Desert locusts shift from Mode 1 to Mode 2. Desert locusts in Mode 2 gather together. Reproduction increases. Desert locusts in Mode 2 lay eggs.

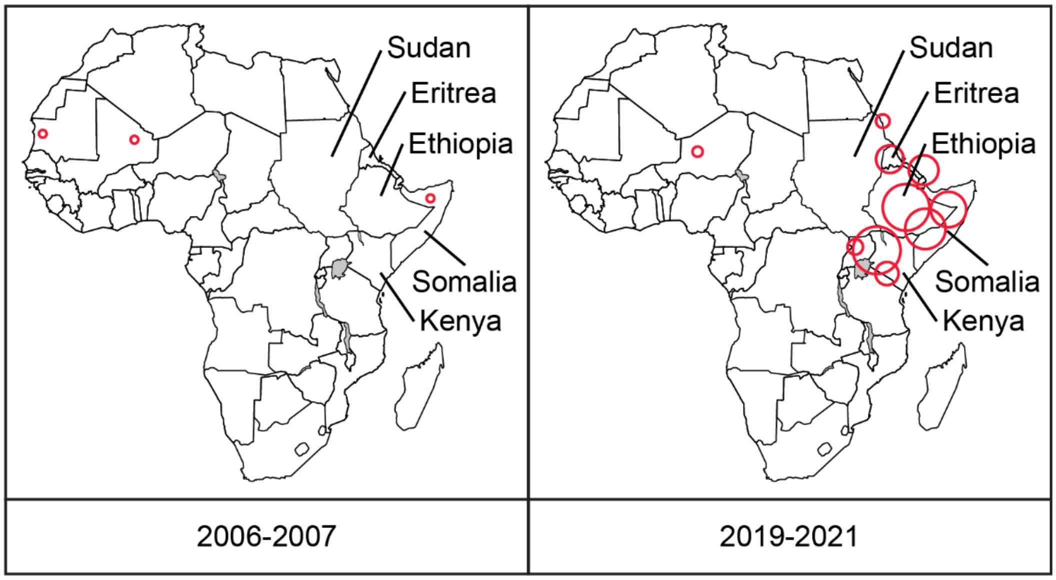
Item 3

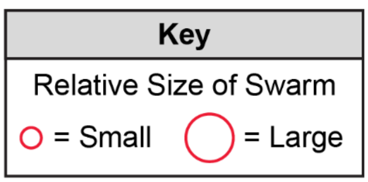
Use the information provided as well as your understanding of scale and quantity to explain why only desert locusts in Mode 2 affect available vegetation as shown in Figure 2.

Item 4

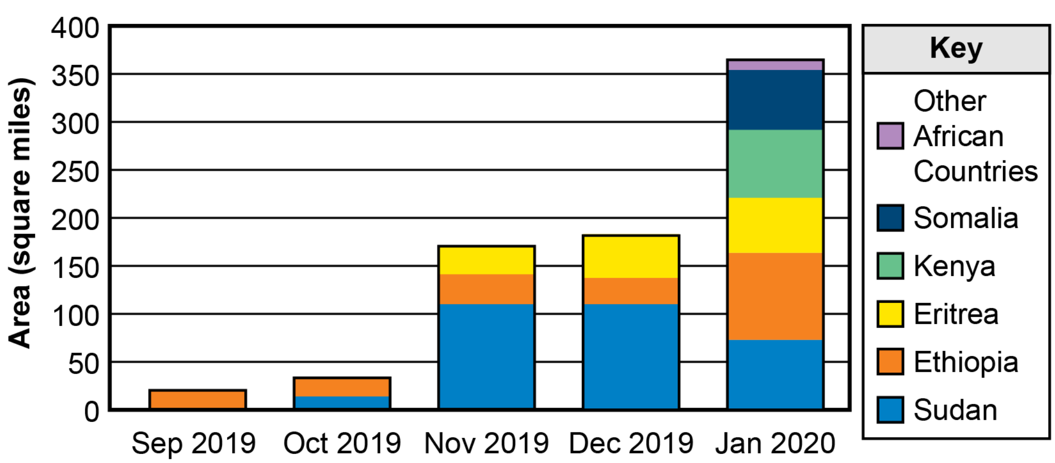
Recently, eastern African countries such as Kenya, Somalia, and Ethiopia have experienced larger and longer-lasting swarms of desert locusts. Over a few months, swarms of desert locusts can eat 50–80% of crops being grown in the area, leaving only a small amount of food for people living in those countries. Red circles on the maps show areas affected by locusts in Mode 2. The bar graph shows the sizes of the areas affected in each African country during one growing season.

Observed Locust Swarms in Africa





Area Affected by Locust Swarms in African Countries



Part A

Some scientists claim that global climate changes are causing desert locusts to switch to Mode 2 more frequently and for longer periods of time. While investigating this claim, the scientists collected the evidence shown in the table.

Select the checkboxes beside each piece of evidence in the table that supports the scientists' claim. Leave blank the checkboxes beside pieces of evidence that do not support the scientists' claim.

Evidence	Yes, this evidence supports the claim.
There were 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.	☐

Key: Boxes 2 and 3 should be checked.

Part B

Explain why each piece of evidence you chose in Part A supports the scientists' claim.

Part C

Use the information provided and your understanding of the impact of human activities on global climate change to make and support a claim about how human activities will **most likely** influence populations of desert locusts in the future.

Item 5

The charts show two proposed solutions for reducing the impacts of desert locust swarms on crops.

Option 1: Spraying Chemical Pesticides Airplanes are used to spray chemical pesticides over affected areas.	**Option 2: Spraying Biological Pesticides** A variety of methods are used to spray biological pesticides over affected areas.
Considerations for Option 1	**Considerations for Option 2**
• Works 99% of the time. • Locusts are killed within 24 hours. • Chemical pesticides can be toxic to other plants and animals in the sprayed area. • Animals (including humans) who eat the dead locusts or sprayed plants can become ill. Some do not survive. • Using this method is somewhat expensive.	• Effectiveness of this method is unknown and is still being tested. • Takes 1–3 weeks to kill locusts. • There is no impact on mammals or crops. • May kill other insects that are closely related to locusts. • Using this method is very expensive.

Scientists predict that Kenya will experience locust swarms again this year. Use the information provided about Option 1 and Option 2 as well as your understanding of the needs of ecosystems to choose the better solution for Kenya. Use scientific reasoning and at least 2 pieces of evidence to support the solution you choose.

Exhibit 3.6. Locusts Constructed Response Scoring Notes

Item	Scoring Notes
3	Student responses should include reasoning that connects the shift to Mode 2 and associated behavioral impacts (they are closer together, reproduce more) as a positive feedback loop that leads to relatively huge population growth. Note that students DO NOT need to use terms like positive feedback loop as long as the concept is present in their response. Students may also colloquially use language like “exponential growth” to indicate the scale at which the population is growing.
4B	Student responses connect the evidence to the provided claim, using reasoning that shows an understanding of the targeted DCs. Note that the emphasis here is on reasoned connection between the evidence and claim, using an accurate understanding of the DC.

Item	Scoring Notes
4C	Students make a claim that human activities will most likely lead to increased swarming. In their argument, they include reasoning that connects provided evidence to the claim, grounded in an understanding of the relationship between human activities, changes to climate/temperature, and impact on locust swarming. While students may include multiple reasons/pieces of evidence, one example connecting evidence to the claim or a generalized response that summarizes across the evidence provided is sufficient here.
5	Students may choose either Option 1 or Option 2 and support their choice with two pieces of evidence provided, along with reasoning grounded in the targeted DC. Note that students may also bring reasoning that reflects their own priorities/values (e.g., monetary cost, impact on animals or environment, impact on peoples' livelihoods, etc.). These are acceptable as elements of student reasoning, as long as they are connected to evidence provided and consistent with scientific reasoning.

A **scenario-based task (SBT)** includes 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. Scenario-based tasks should include at least one item that is three dimensional.

A note about SBTs: While scenario-based tasks can be rich opportunities for student sensemaking, they are often more resource-intensive to develop. SBTs should be used judiciously, with a particular focus on those performances that are difficult to assess in other contexts.

These may include the following:

- **Scenarios that require considerable contextual information to fairly surface the intended targets.** Assessing some NAEP SEPs (e.g., Using Mathematics and Computational Thinking; Obtaining, Evaluating, and Communicating Information), DCs, and CCCs requires a large amount of contextual information for students to engage in the task. This is often because the assessment target itself does not specify specific contexts, methods, models, or experiences students should focus on in instruction; as a result, this information needs to be provided in the item, with enough context so that any student can understand the information. While it is feasible to assess these NAEP SEPs with discrete items, the amount of information students need makes it difficult to justify for a single item. In some cases, this can be addressed through an item set; in others, SBTs will be particularly helpful.
- **Phenomena/problems and assessment targets that require iteration.** Particularly in grades 8 and 12, students are expected to iteratively engage with information, updating

their sensemaking with new/multiple sources of information. SBTs can be particularly useful in these contexts.

- **High-complexity and increasingly complete performances.** Many high-complexity performances involve cascades of NAEP SEPs, CCCs, and DCs. These often draw on NAEP DCs from multiple domains, SEPs that connect with each other (e.g., Constructing Explanations and Designing Solutions based on multiple sources of information of varying credibility), and complex CCCs. Student performances on items early in the SBT may include fewer components of these DCs, SEPs, and CCCs than items later in the SBT, and over the course of the entire SBT, student performances may provide evidence that addresses these dimensions in a more complete way. These performances are often particularly necessary for addressing authentic phenomena and problems using the specified dimensions in grades 8 and 12.

SBTs can include different item types in combinations that 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.

The stimuli in scenario-based tasks are more complex than those of item sets and may provide students with opportunities to engage in a broader range of science practices, for example, by allowing students to use a variety of datasets that describe science systems under study in an investigation or several related investigations. Multimedia components can provide opportunities for students to make observations of science systems in action, including systems and components of systems that may otherwise be difficult to observe because they are too small (subatomic particles and organelles), too large (ocean currents and clusters of galaxies), too fast (lightning), or too slow (mountains forming).

Developing a scenario-based task is complex due to the interrelationships of the parts and the need to maintain a cohesive storyline through the process of developing and iteratively revising stimuli and items. This process can be aided by steps prior to the development of the items, such as outlining the stimuli and drafting ideas for items that describe how students will use DCs, SEPs, and CCCs to make sense of the phenomenon, solve a problem, or address other aspects of the context such as using a model that is presented in a diagram or analyzing a dataset in a table or graph. The examples of science achievement expectations in Exhibit 3.34 provide examples of multidimensional combinations. However, when drafting ideas for items during the outlining process for an SBT, developers should take additional steps after writing a science achievement expectation: (1) decide how the multidimensional performance described by the science achievement expectation (i.e., the item) will require students to interact with the stimulus materials and (2) determine the characteristics of a successful interaction (i.e., a description of the expected response). Outlining stimuli and items is particularly important for SBTs to organize the component stimuli and item ideas into a coherent story that builds toward more cognitively complex sensemaking.

Additionally, the following steps can be used to structure the process of developing an SBT:

- Develop authentic, relevant, and compelling presentations of the phenomenon and context.
- Select and iteratively refine these presentations to ensure interest and appropriateness for students at the grade level.
- Select and iteratively revise the phenomenon or problem and context to ensure they support sensemaking through the use of the relevant disciplinary concepts, crosscutting concepts, and practices.
- Ensure that the required sensemaking is consistent with the statements for the three dimensions at the appropriate grade level and does not exceed any boundaries for assessment.

Making Sense of Phenomena and Solving Problems

Sensemaking in science is at the heart of the construct measured by the NAEP Science Assessment. Items that contribute to the measurement of this construct require students to make sense of phenomena that are real-world events or processes that provide a setting for an item or set of items. These phenomena describe observations that students are likely to make and are motivated to understand or may present problems that students are motivated to solve. Sensemaking occurs when students apply their previously developed ideas, abilities, and experiences to address unexplained phenomena or unsolved problems. Developers select phenomena and problems that are simple or complex, depending on how many items they will be used with and the intended difficulty and complexity of these items.

A simple example of water vapor condensing on a cool surface may be sufficient for a single item; a more complex phenomenon of a cloud forming just above the top of a mountain may be better suited to a set of related items. In the simple example, the observation presented to students motivates the need to know and demonstrate that knowledge by explaining how energy transfer affects the kinetic energy of particles, which in turn causes a change in the state of matter. There is more complexity available to discuss in this simple example; other practices could be used prior to or in support of the explanation. However, the items would likely be somewhat repetitive variations assessing the same concept, potentially leading to overrepresentation on the assessment. In the more complex example, there is greater opportunity to explore a wider range of scientific ideas, from the behavior of individual particles to the interactions of Earth's systems and implications for the uneven distribution of water resources around the globe. An important feature of these two examples is that both phenomena could be used with items that are easy or difficult or items that are simple or complex, depending on how each item requires a student to use a science idea to make sense of the phenomenon.

Features of Phenomena and Problems Used in Item Contexts

As described by the 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 using NAEP DCs, SEPs, and CCCs. To do so, all two-dimensional and three-dimensional items are designed to focus on 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.

Compelling phenomena and problem-based contexts present authentic uncertainty of a situation in ways that give students something to make sense of. Phenomena and problems that are considered compelling often have some of the following features:

- a specific instance
- authentic uncertainty
- relevance to particular communities

Note that compelling phenomena and problems in assessment contexts focus on explicit relevance, not on individual student interest or a specified degree of impact. A phenomenon may be compelling without each student being deeply invested in the outcome or without the phenomenon having a huge global impact.

Phenomena provide a setting for an item or set of items. They should be chosen to engage student attention and sensemaking that requires the targeted NAEP DCs, SEPs, and CCCs for a satisfying explanation or effective solution. Problems are meaningful challenges that present a situation requiring new or improved technologies or processes. Where appropriate, phenomenon and problem descriptions should include the impact, such as effects on people, animals, or the environment. Phenomena may result from both human-designed and natural processes and systems. To serve as the context of an item, phenomena and problems must involve a NAEP DC 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. Compelling phenomena and problems promote student engagement in items and demonstration of their knowledge and skills. They do so by providing contexts that make the authentic uncertainty of a situation clear to students and by giving students something puzzling to solve.

Criteria for Selecting High-Quality Science Phenomena 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 and that assessments are accessible, and providing 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 with different and diverse 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- or 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 NAEP DCs and SEPs 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.
- Use an authentic question or other prompt that leads the student to use the targeted NAEP DC and SEP (and CCC when appropriate) to explain the phenomenon or figure out a solution to the problem.
- Provide 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 by 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.

The selection of phenomena provides the test developer with an extraordinary means of increasing assessment quality. A low-quality assessment includes few, mostly superficial phenomena that are not relevant to students and that lack context or are presented in contexts devoid of meaning. These superficial phenomena afford little opportunity for sensemaking. Items associated with these phenomena seldom ask students to do more than describe related scientific content and often promote the assessment of rote content or procedural knowledge for which the phenomena may not even be needed to respond.

In contrast, a high-quality assessment presents phenomena that prompt students to wonder, to think of questions, and to consider tentative explanations for their observations as their prior knowledge is activated. This step may seem superfluous in a process focused on gathering the most evidence of student knowledge in the shortest time; it is, in fact, essential for an assessment designed to gather evidence of deep understanding of complex content.

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.

For example, a problem could begin with the phenomenon of water droplets forming on the inside of a window. Students could be required to both make sense of the phenomenon and address a problem. The latter may include defining the problem, such as thermal energy transfer, oxidation of metal parts, or mold growth, and designing a solution, such as a better-insulating window, finishing metal parts with a protective coating, or reducing the humidity of the indoor air.

Items may provide students with opportunities to apply their knowledge to solve real-world problems. These problems may be presented as the focus of one or more items, or they may be introduced within the context of a focal phenomenon. Problems motivate students to design solutions that require application of scientific principles described in the disciplinary concepts.

The NAEP Science Assessment should be developed using phenomena and engineering problems that are relevant and meaningful to students. Phenomena or problems can be made accessible by describing them in simple terms, using specialized terminology only when necessary and only when that terminology is consistent with the expectations described in the Framework. Phenomena can be made accessible by presenting them in contexts where a student is likely to observe them.

Phenomena may provide "windows" into science systems that can be modeled and explained through an understanding of unseen causal mechanisms. They may prompt students to define problems and design solutions. They may include relationships and interactions that are described in data and understood through the analysis and interpretation of that data, including the application of mathematical and computational models to the data. These phenomena afford opportunities to ask or direct students to apply their prior knowledge as they engage in the practices of science in the performance required by an item.

Phenomena and problems also provide opportunities for students to use crosscutting concepts. For example, students may recognize patterns in multiple observations of a phenomenon, including how those patterns change with scale, proportion, or the quantities of components within a system. Using crosscutting concepts, students can also organize their knowledge to interpret phenomena and solve problems.

Scientific Accuracy

Because the 2028 NAEP Science Assessment will present detailed information about real-world phenomena and problems, the development of stimuli must be supported by reliable sources to establish their scientific accuracy, such as published reference resources, peer-reviewed research, and academic databases. These sources are used to confirm the accuracy of information in stimuli, not as sources of excerpts. Verbatim text, though not expressly prohibited, would likely not be needed or be appropriate for use in the NAEP Science Assessment due to the complex language and jargon common to such sources.

Science is subject to change; new research may uncover important findings that can affect the accuracy of assessment content. To ensure fidelity to scientific principles over many administrations, phenomena and problems should be based on time-tested and well-established findings in scientific literature. Current events in science may provide exciting topics for science instruction but can be problematic in large-scale assessment should attempts to replicate new findings fail and content based on outdated research no longer be usable.

Situating Phenomena and Problems in Contexts

Phenomena and problems require context. The selection and development of these contexts afford another mechanism for increasing assessment quality. A high-quality assessment uses the context in one or more stimuli and items to prompt students to move from a narrow "window" provided by a phenomenon statement to a broader experience that makes visible the parts of the natural world in which the phenomena is likely to occur. Like phenomena, contexts vary from simple to complex, depending on their intended use. Contexts can provide background information that students can synthesize with their prior knowledge as they make sense of a phenomenon. Rich contexts can make accessible phenomena that would otherwise be too challenging for students to understand alone. Contexts can help support the relevance of phenomena and problems and lend them meaning. Exhibit 3.7 shows an example where the context makes the phenomenon meaningful.

Exhibit 3.7. Phenomenon and Context Example

Phenomenon	The same type of plants in different parts of a garden are different sizes.
Context	A community garden is planted in an area that has a concrete wall on one side. A gardener notices that people are growing the same plants in different parts of the garden. A simple data display shows that tomatoes are larger farther away from the wall, but the lettuce heads are larger close to the wall.
Item Description	Students ask a question about how much sunlight each part of the garden receives.

As shown in the exhibit, the context supports sensemaking in ways that provide students with opportunities to demonstrate their understanding about how sunlight provides energy that plants need and their ability to ask a question that could be investigated.

Phenomena prompt students to bring in their disciplinary understanding that is needed to respond to items; the items must be designed to require only the knowledge that is described in the Framework. Some of the specific knowledge about a phenomenon may only be available to students who have prior experience with that phenomenon. Items that require this specific knowledge threaten the validity of the assessment argument. A high-quality assessment can bridge the gap between the prior knowledge that is expected and the specific information that students may not have and is instead provided by the context. Care must be exercised in designing these contexts to ensure that the means of interpreting contextual information is available to students through the use of the practices, crosscutting concepts, or both. Further, contextual information must be limited to only what is needed to respond to any associated items. Gratuitous contexts may be distracting and may add unnecessary cognitive demand and time to the assessment. Exhibit 3.8 shows an example where the context adds information that is unnecessary.

Exhibit 3.8. Phenomenon and Context Non-Example

Phenomenon	The same plants in different parts of a garden are different sizes.
Context	A gardener notices that there are two types of tomatoes growing in the garden and one type is consistently taller than the other. A model shows how energy moves through the parts of the tomato plant.

The context in this exhibit provides some information that would not be useful. If students were asked to use the model to describe the function of the parts of the tomato plant, they would not need to know anything about the types of tomato plants. If students were asked to explain how variations of traits in tomato plants affect their growth, the model of a single tomato plant would not support students in making sense of the phenomenon.

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. 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 an 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.

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. Some examples of the considerations follow:

- Use only as many words as needed to convey a compelling and necessary context.
- Use narrative, expository, and scientific types of language and vary it to make it appropriate for the context.
- Use everyday 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.
- Research about how students learn should be taken into consideration in scoring guidance.

Developers will need to ensure that text complexity does not become the primary determining factor in student performance (i.e., the science assessment should not become a reading assessment).

Text complexity and length of stimuli should be appropriate to the grade assessed. Qualitative approaches and quantitative tools to measure grade-level appropriateness should be used to ensure that the reading level of any text is lower than the grade assessed to allow students to focus on applying their science knowledge. The structure of each sentence should be leveraged to present a single, coherent idea; limit the number of clauses; and use the highest frequency vocabulary that will do the job.

Appropriate use of data displays and other visual media can present content in parallel to text, making the content accessible and limiting the need for long blocks of text. In addition, logical organization, bulleted information, and interactive elements can support students' ability to access parts of longer stimuli that are typical in item sets and scenario-based tasks.

It is important to note that NAEP Science is an assessment of science in English. The considerations listed above not only benefit all students but have the potential to reduce the cognitive load for students who may be simultaneously learning English and science in schools (e.g., designated English Learners, emergent bilingual learners, etc.).

It should also be noted that reading and writing (and speaking and listening) in science are central elements of many of the NAEP SEPs and therefore are part of expectations for science learning and achievement. NAEP SEPs such as Obtaining, Evaluating, and Communicating Information; Engaging in Argument From Evidence; and Constructing Explanations and Designing Solutions require students to develop and demonstrate disciplinary literacy in science. While items should take care to limit any gratuitous language use or complexity, it is equally important that NAEP not overly limit language engagement so that the construct cannot be meaningfully assessed. As described in the section on accessibility, a combination of universal design approaches to development and accommodations are needed to provide opportunities for all students to demonstrate what they know and can do in science.

Deciding What to Include and Exclude in the Context

The Framework is designed to enable students to demonstrate their conceptual understanding of NAEP DCs and CCCs and to use SEPs. 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 that bodies have systems that work together, 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 the Specifications 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. Details provided to students in item contexts should be carefully constructed in developmentally appropriate ways, accounting for reading and cognitive load, as well as assumptions about prior knowledge and schema development of the targeted students for the test. The following three illustrative contexts demonstrate ways that context can support students in applying their conceptual understanding. The amount and complexity of the context varies across these three examples to demonstrate what is appropriate for different grade levels, different targeted complexities of items, and different numbers of items (discrete or item set).

Illustrative Contexts

- **Park Flooding** includes a simple phenomenon that can be used for a discrete or multipart item.

- **Human Migration to Appalachia** includes an example of a richer phenomenon with authentic, compelling phenomena that can support an item set.
- **Locusts** includes an example of a robust phenomenon that can be used not only for multiple items (item set) but across multiple disciplines.

The first of these illustrative contexts is included in Exhibit 3.9, an item that requires students to make sense of a simple phenomenon: a local park has flooded. This phenomenon is an example of an everyday phenomenon that many students may have directly experienced or have sufficient experiences to understand. This item elicits simple sensemaking because students must understand that flooding is most likely to occur on the day with the most rain, rather than just any day with rain. This requires interpreting the context as well as the data, albeit in a very straightforward way. In grade 4, students are still developing the concept of flooding and its relationship to rain; while this item elicits sensemaking based on developmentally appropriate and expected schema for 4th graders, this same performance would not be considered sensemaking in grades 8 and 12.

Exhibit 3.9. Park Flooding, Version 1

Item ID: Park Flooding (adapted from Formative Assessment Bundling Literacy and Elementary Science)
Phenomenon: A park flooded when it was raining one day but not other days.

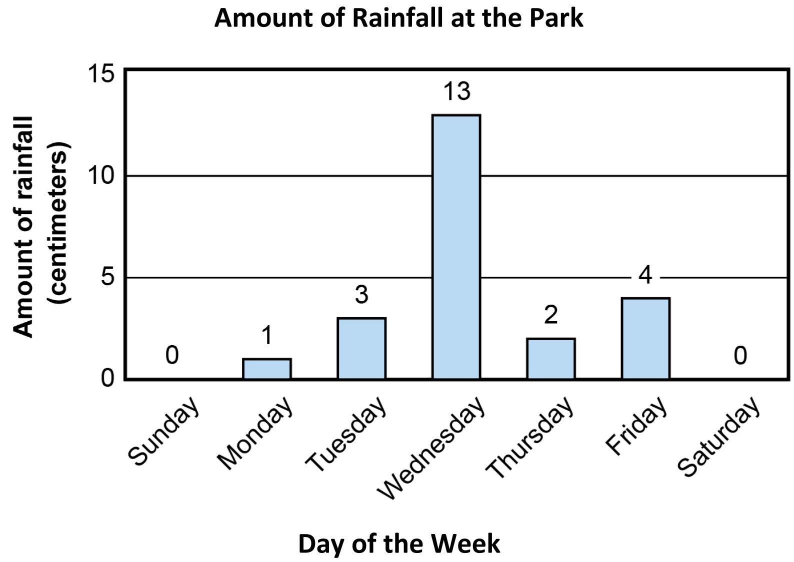
People visiting a local park noticed that the park was flooded and was closed for the day. The picture shows the flooded park.

Flooded Park[5]

The park was closed only on the day the flooding happened. The bar graph shows the rainfall for each day of that week.

---

[5] Betty Longbottom / Flooded Playground! - Cliffe Avenue / CC BY-SA 2.0



Based on the data, on which day was the park **most likely** flooded?

A. Sunday
B. Monday
C. Tuesday
D. Wednesday

Key: D

The second illustrative context, Exhibit 3.10, presents a stimulus with a meaningful phenomenon and problem context that poses considerable challenges and presents authentic and compelling problems. Sample items that use this stimulus are provided in Exhibit 3.16.

Exhibit 3.10. Human Migration to Appalachia

Phenomenon/Problem: Human migration to Appalachia has been greater than predicted by computer models. The model used to make the prediction needs to be revised to better reflect the factors that influence migration into different regions of Appalachia.

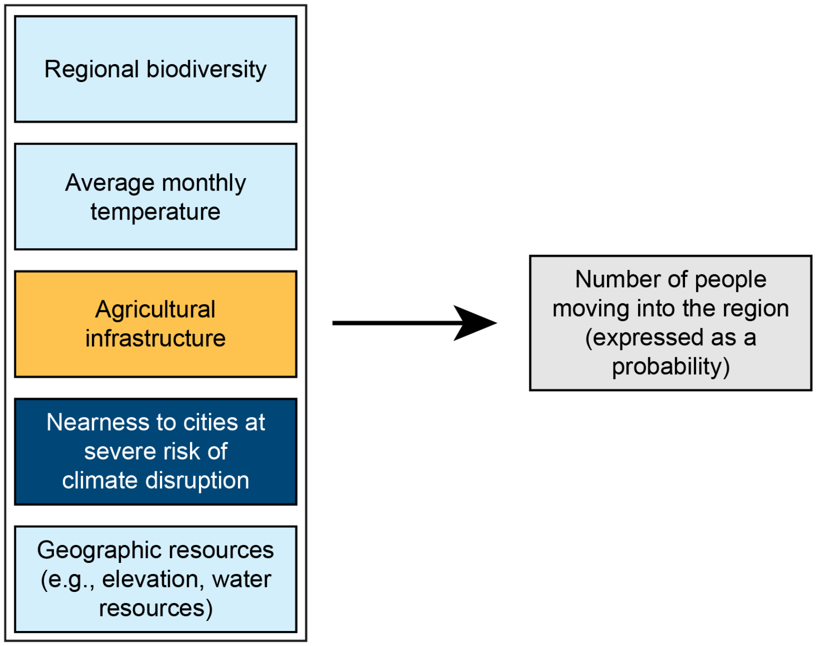
Appalachia is considered “climate resilient.” This means that the area can successfully handle the impacts of changes to climate and can prevent those impacts from growing worse. The green areas in Figure 1 show where Appalachia is located in the United States.

Figure 1. Map of Appalachia

Computational models predict that many people will move into the Appalachian region over the next 20 years as they seek to find places to live that are safer and more stable.

Figure 2 shows one model local leaders are using to predict how many people will move into Appalachia. Blue indicates factors that are expected to increase migration, and orange indicates factors expected to decrease migration. Darker colors indicate more weight on that factor in the model. Agricultural infrastructure includes farms, markets and businesses that support farms, and the transportation and communication systems in the area.

Figure 2. Computational Model for Predicting Human Migration



When this model was tested against recent population growth due to migration into two locations in Appalachia, leaders noticed some differences between what the model predicted and what the data showed. The table shows these differences for the two locations, along with information about how high or low each location is rated on several factors.

Predicted and Actual Population Growth in Pittsburgh and the Shenandoah Valley

Location	Pittsburgh	Shenandoah Valley
Predicted population growth	high	low
Actual population growth	low	high
Relative biodiversity	low	high
Average monthly temperature range	29–73° F	32–74° F
Relative agricultural infrastructure	low	medium
Nearness to climate-impacted urban centers	high	high
Access to usable water	medium	high

The third illustrative example, Exhibit 3.11, is a stimulus that presents a meaningful phenomenon and problem context that deeply matters to many people around the world and is posing considerable challenges right now. It was selected to show how a wider range of NAEP SEPs and CCCs can be engaged in items across an item set than the examples above, including items across multiple disciplines (e.g., Life Science and Earth and Space Sciences) and some that are often difficult to assess. An item set based on this stimulus could also be used to address the underlying biology (e.g., genetics, specialized subsystems) connected to the physiological changes locusts undergo; research on potential solutions; impacts on biodiversity in regions with swarming; consideration of patterns of locust swarming going back thousands of years (stability and change); consideration of whether current upticks are significant or not (more sophisticated data analysis); and so on. This context could support items for either grades 8 or 12 in both Life Science and Earth and Space Sciences. Possible NAEP DCs, SEPs, and CCCs that could be included with the stimulus are listed in Appendix A.

Exhibit 3.11. Locusts

Phenomenon and engineering design problem: Locust swarms can cause a lot of damage.

A desert locust is an insect that undergoes changes to its body in certain environmental conditions. Figure 1 describes some differences between two modes of a desert locust.

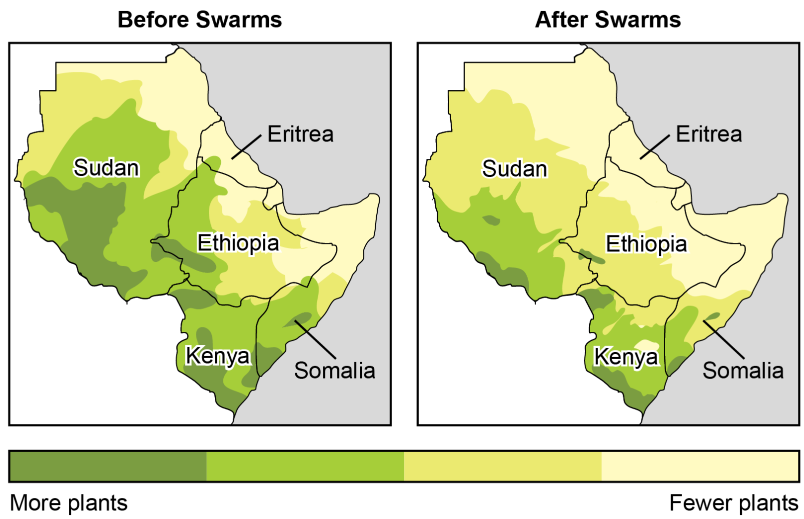
Figure 1. Desert Locust[6]

Mode 1: Grasshopper (Dry, warm or cool weather)	Mode 2: Locust (Wet/rainy, warm or hot weather)
• Behave independently • Stay away from other desert locusts • Mostly walk slowly and jump • Limited 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.	• Behave as a united group (swarms) • Gather together with 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 six months.
	

[6] Bernard DuPont Bird Locust Attribution-ShareAlike (CC BY-SA 2.0); Magnus Ullman, CC BY-SA 3.0, via Wikimedia Commons

When these insects are in Mode 2, they are able to swarm. A single swarm of locusts can cover an area of up to 100 square miles, with 40 to 80 million locusts in each square mile. Swarms can travel up to 100 miles a day. Figure 2 shows the effect of three months of locust swarms on available vegetation in an area of Africa.

Figure 2. Available Vegetation Before and After Locust Swarms



Item Development

Chapter 2 describes the NAEP Science Disciplinary Concepts in each of three discipline groups at each grade level: Physical Science, Life Science, and Earth and Space Sciences. Chapter 2 also describes the seven NAEP CCCs and eight SEPs that are the targets for assessing science across grade levels. Chapter 2, along with the guidelines in this chapter, focuses on realizing the intent of the Framework in developing items used on the assessment.

The guidelines offered here highlight only some of the critical considerations in item development, concentrating on topics specific to the NAEP Science Assessment. Item writers should refer to directions for developing items provided by the Governing Board and its designees in addition to the information in this document.

General Principles of Item Writing

NAEP items will be developed in accordance with recommended practice and the Governing Board Item Development and Review Policy (National Assessment Governing Board, 2002).

Features of Multidimensional Items

As described previously, measuring the construct described in these Specifications requires that each item should require students to bring together NAEP DCs, SEPs, and when possible, CCCs to successfully address a question or accomplish a task. 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 NAEP DC?
- Does the item require students to demonstrate their understanding of the NAEP DC through application of an SEP and, if included, 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 multipart item should be at least two dimensional and three dimensional if appropriate. Item sets and scenario-based tasks should provide evidence of students' ability to use the three dimensions together to explain a phenomenon or address a problem by including at least one item that is three dimensional. Each item (discrete or individual items within item sets and scenario-based tasks) will receive one score representing the integration of the dimensions measured by the item. Consider an item that requires students to make sense of the following phenomenon:

- Teosinte is a type of grass related to corn that typically grows in Mexico and Central America, an environment where the sun is directly overhead at midday in summer (Sánchez González et al., 2018).

A cognitively complex and challenging item would first present the phenomenon, support sensemaking of this phenomenon with context (such as a map or other data display) and then present and ask students to evaluate tests of a solution to the effects of climate on plant growth by comparing teosinte growing in its native habitat to teosinte growing in multiple locations outside of this environment.

A strategy that may be useful for limiting the complexity of multidimensional items would be assessing dimensions separately within a multidimensional item. For example, an item that asks students to compare the outcomes of different amounts of sunlight on the growth of teosinte and then asks the student to describe the variables in this experiment is first asking the student to apply the disciplinary concept and then to apply the science and engineering practice.

The following four illustrative examples demonstrate how items require students to use two or three dimensions together in different types of discrete items and how an item set can require students to use different combinations of dimensions for different items. The first two items involve multiple parts to illustrate how to capture more evidence of student understanding through discrete items; however, 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.

Illustrative Examples

- **Plant Growth** is a two-dimensional discrete, multipart item.
- **Park Flooding** is a two-dimensional discrete, multipart item.
- **Making Soap** is a three-dimensional discrete item.

- **Human Migration to Appalachia** is an item set with items across multiple disciplines, using the three dimensions together.

In Exhibit 3.12, an example of a two-dimensional discrete item, students have to apply their understanding of what plants need to grow to make a prediction. Students do not need to understand a specific NAEP CCC element to respond to the item—the NAEP SEP and DC are sufficient to respond to this question. While the item is an implicit example of the NAEP CCC Mechanisms and Explanation: Cause and Effect, students do not need to explicitly bring an understanding of cause-and-effect relationships to respond and, therefore, the item is not considered three dimensional. This item provides an example of lower-level sensemaking with the NAEP DC and SEP—while students do (a) need to understand that plants need water and air to grow and (b) need to be able to use this information to evaluate a phenomenon across multiple (experimental) conditions, they are very closely applying simple grade-appropriate NAEP DCs and SEPs. The NAEP SEP is engaged in service of surfacing NAEP DC understanding, rather than expanding the nature of how students explain the phenomenon. This level of sensemaking would be appropriate to surface understanding from students who have had the opportunity to begin developing an understanding of the grade 4 NAEP DC and SEP.

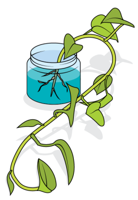
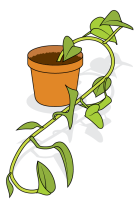
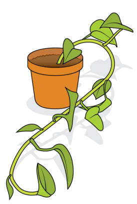
Two additional versions of the item appear in Appendix A, Exhibits A.13 and A.15. These versions demonstrate how the item can be modified to increase the complexity of the item with respect to specific dimensions and overall.

Exhibit 3.12. Plant Growth

Item ID: Plant growth (adapted from Next Generation Science Assessments)

The plants shown were placed in a classroom on the same day. They are all the same kind of plant. The plants were placed on the same side of the room near a window, so they receive the same amount of light each day. Students in the class want to find out what the plants need the most in order to grow. They grow the plants using the conditions shown in the table.

Conditions for Growing Plants

Plant	Planted in Soil	Water
 Plant A	No	Water added regularly for one month
 Plant B	Yes	Water added regularly for one month
 Plant C	Yes	No water added

Part A

Which plant will likely grow the **least** over the next month?

A. Plant A
B. Plant B
C. Plant C

Key: C

Part B

Provide one reason the plant you chose in Part A will grow the least over the next month.

Exhibit 3.13. Plant Growth Part B Constructed Response Scoring Notes

- Reasons students provide should leverage understanding of what plants need to grow (water, air, minerals from soil)
- Note that while a complete answer might include comparisons among plants (e.g., Plant A and B have X, but Plant C does not), this is not a requirement.
- Possible reasons include:
  - Plant C does not get water
  - Plant C does not get minerals
  - Plant C does not get water or minerals
- Students should receive credit as long as their reason supports their choice, with an accurate understanding of plant needs for growth.

Exhibit 3.14 presents version 2 of the item in Exhibit 3.9. This version shows a modification that provides additional evidence of student understanding of both the NAEP DC and SEP in service of sensemaking. The additional component of the second discrete item does not change the alignment or complexity of the item but does add time to complete and some additional reading load. This trade-off between more comprehensive evidence and the additional time needed to complete the items may be valuable at some times over the range of the assessment.

Exhibit 3.14. Park Flooding, Version 2

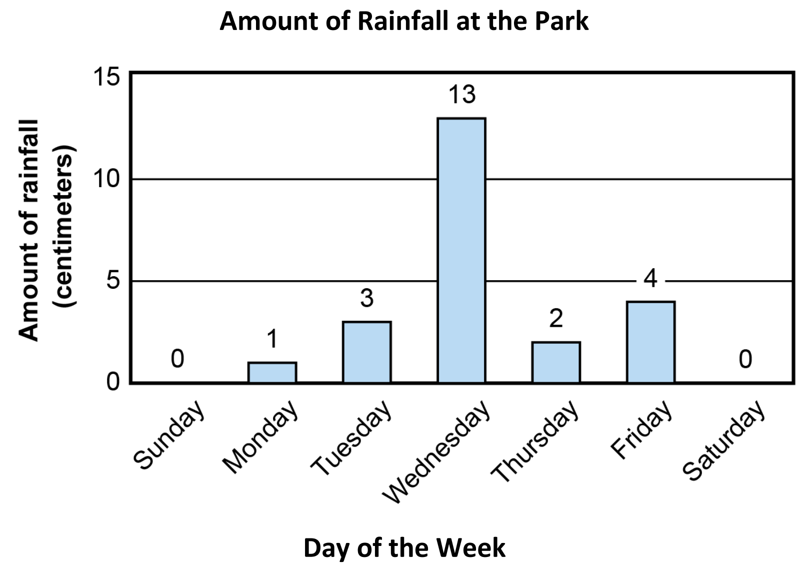
Item ID: Park Flooding (adapted from Formative Assessment Bundling Literacy and Elementary Science)

People visiting a local park noticed that the park was flooded and was closed for the day. The picture shows the flooded park.

Flooded Park[7]

The park was closed only on the day the flooding happened. The bar graph shows the rainfall for each day of that week.

[7] Betty Longbottom / Flooded Playground! - Cliffe Avenue / CC BY-SA 2.0



Part A

Based on the data, on which day was the park **<u>most likely</u>** flooded?

A. Sunday
B. Monday
C. Tuesday
D. Wednesday

Key: D

Part B

Based on your understanding of weather, which piece of evidence **<u>best</u>** supports your answer in Part A?

A. This day was rainy.
B. The rain started falling on this day.
C. This day had more rainfall than any other day in the week did.
D. The amount of rainfall on this day was lower than on any other day.

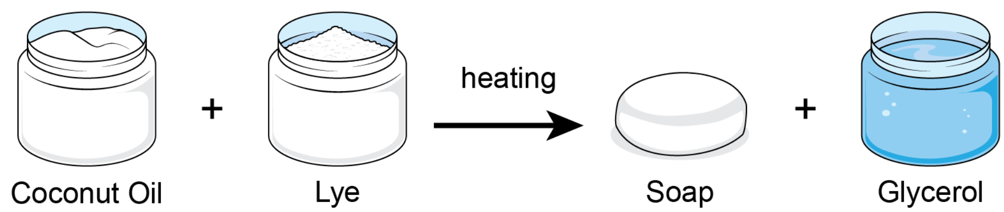
Key: C

In Exhibit 3.15, students have to apply their understanding of chemical reactions to analyze data while looking for patterns among the specific characteristic properties that will indicate that a chemical reaction has occurred. This performance requires that students use their understanding of the NAEP DC, use the NAEP SEP to analyze data, and use the NAEP CCC to look for patterns to figure out whether a chemical reaction has occurred. This is a three-dimensional performance that requires relatively simple sensemaking with multiple dimensions.

Exhibit 3.15. Making Soap

Item ID: Making Soap (adapted from the Next Generation Science Assessment Project)

One way to make soap is to heat a combination of coconut oil and lye. The diagram shows a simplified model of the soapmaking process.



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

Properties of Soapmaking Substances

Substance	Mass (g)	Odor	Density (g/cm^3)	Melting point (°C)
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 from the density of glycerol.
C. The total mass of soap and glycerol is the same as the total mass of coconut oil and lye.
D. The melting points of soap and glycerol are different from the melting points of coconut oil and lye.

Key: D

Exhibit 3.16 is an item set for grade 12 that demonstrates how NAEP DCs, SEPs, and CCCs can be used in the service of sensemaking. The items leverage simple uses of NAEP SEPs and CCCs, allowing a wider range of students to access and engage with the rich context.

It should be noted that this item set could be expanded to more deeply and comprehensively assess (a) related Earth and Space Sciences DCs (e.g., how the geography of the Appalachian region has contributed to climate resilience over time), (b) the Life Science DCs (e.g., how ecosystem dynamics contribute to resilient biodiversity), or (c) their integrated use (e.g., how geographic features that have evolved over time have led to adaptations and speciation, contributing to the rich biological systems in the area and impacts on human civilizations). These could be explored in further independent items within an item set or in related items within a scenario-based task.

Exhibit 3.16. Human Migration to Appalachia

Item ID: Human Migration to Appalachia

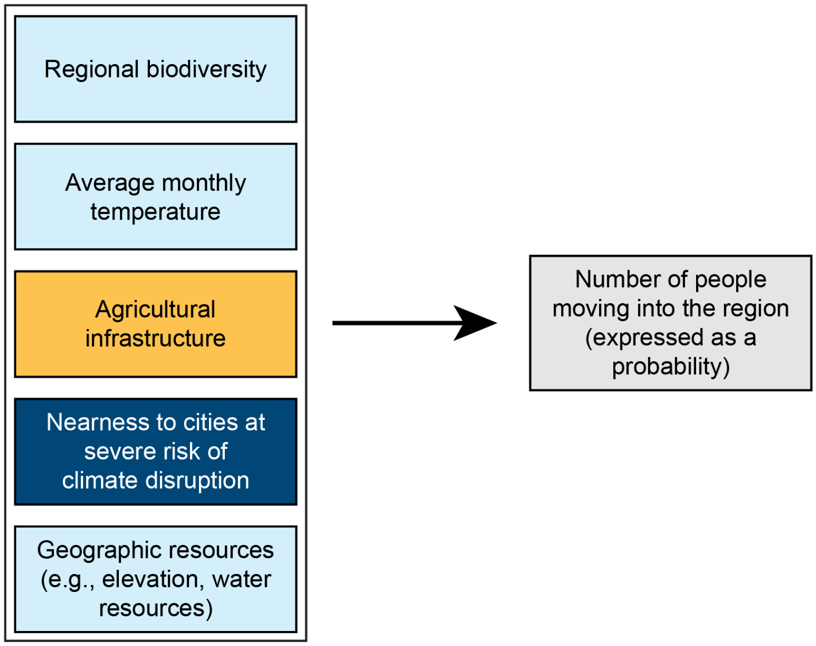
Appalachia is considered “climate resilient.” This means that the area can successfully handle the impacts of changes to climate and can prevent those impacts from growing worse. The green areas in Figure 1 show where Appalachia is located in the United States.

Figure 1. Map of Appalachia

Computational models predict that many people will move into the Appalachian region over the next 20 years as they seek to find places to live that are safer and more stable.

Figure 2 shows one model local leaders are using to predict how many people will move into Appalachia. Blue indicates factors that are expected to increase migration, and orange indicates factors expected to decrease migration. Darker colors indicate more weight on that factor in the model. Agricultural infrastructure includes farms, markets and businesses that support farms, and the transportation and communication systems in the area.

Figure 2. Computational Model for Predicting Human Migration



When this model was tested against recent population growth due to migration into two locations in Appalachia, leaders noticed some differences between what the model predicted and what the data showed. The table shows these differences for the two locations, along with information about how high or low each location is rated on several factors.

Predicted and Actual Population Growth in Pittsburgh and the Shenandoah Valley

Location	Pittsburgh	Shenandoah Valley
Predicted population growth	high	low
Actual population growth	low	high
Relative biodiversity	low	high
Average monthly temperature range	29–73° F	32–74° F
Relative agricultural infrastructure	low	medium
Nearness to climate-impacted urban centers	high	high
Access to usable water	medium	high

Item 1
Based on your understanding of what human societies need to be successful, which idea **best** explains why the model prediction was different from the observed data?

A. Pittsburgh has highly limited access to water. The model did not account for how this shortage would limit the growth of local businesses.
B. Human societies require reasonable annual temperature ranges. The model incorrectly assumed that average monthly temperature was a less important factor.
C. Human societies require access to food and water. The model did not account for how important access to food and water would be for human migration to Appalachia.
D. The Shenandoah Valley has very little access to fresh fruits and vegetables. The model incorrectly assumed that agricultural infrastructure would not promote migration.

Key: C

Item 2
Human migration to areas like Appalachia can result in rural gentrification. People who are currently living in low-cost, natural resource–rich areas are forced to leave, resulting in the loss of access to the resources and communities they have actively contributed to developing and maintaining. As a result, they can no longer enjoy the benefits of these regions.

Describe **one** way you could revise or build on the computational model to better understand how migration into Appalachia could impact current residents' access to natural resources. Be sure to explain how the change you describe will provide a better understanding of how migration into Appalachia could impact current residents' access to natural resources.

Exhibit 3.17. Human Migration to Appalachia Item 2 Constructed Response Scoring Notes

Emphasis here is on an understanding of how to revise the model parameters to better understand more nuanced population-environment dynamics. Student ideas can be wide reaching but should be justifiable as at least one of the following: (a) updating the model to better understand who moves into and out of the region, and/or (b) determining the feedback impacts of population growth on biodiversity, and natural resources such as water. This item specifically focuses on the computational reasoning aspect of this SEP and can include either quantitative or qualitative reasoning from students. Appropriate lines of reasoning here can include, but are not limited to, the following:

- updating the model to account for demographic subgroups
- considering housing costs/other metrics for socioeconomic status as part of the inputs and/or outputs, as a mechanism to better understand the relative wealth/characteristics of who lives in the area
- calculations of migration out of Appalachian regions
- feedback mechanisms that influence biodiversity and natural resource availability (Note: This could be specific to particular resources, or general at the level of the categories included in the model.)
- relative factors for scaling variables (quantitative or qualitative)

Assessing the Full Range of Student Performance

It is important that the NAEP Science Assessment 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 (National Academies of Sciences, Engineering, and Medicine, 2017; National Research Council, 2005; 2007). While traditional approaches to assessment often assume that rote understanding or simple procedural skills (e.g., definitions, facts, lab skills) are less 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. For the purpose of the Specifications, increasing sophistication refers to a student expression of understanding that is more thorough, more precise, more accurate, and more coherent throughout. Likewise, assessments intended to surface what students who have not yet mastered grade-level expectations know and are able to do may do so more effectively by varying the complexity 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 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 nature of the phenomenon or problem context;
- the sophistication of language, graphics, or mathematical elements, and their presentation together as appropriate;
- 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 et al. (2015); and WestEd et al. (2019). The purpose of the complexity framework is to inform item development so 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 degree and nature of sensemaking 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 also how each dimension contributes to sensemaking, items can be designed more intentionally. For example, some items provide substantial scaffolding for engaging in the NAEP SEP, with limited cueing for the NAEP DC, while other items engage in a lower level of sensemaking with NAEP DCs while providing students the opportunity to engage with the NAEP SEP and CCC more deeply. In some items, the NAEP 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 nonroutine lens on a phenomenon or problem (e.g., asking students to examine a seemingly causal relationship that is correlational). These are important considerations for developing a balanced assessment that can intentionally surface a range of student thinking.

Exhibit 3.18. Complexity of Multidimensional 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	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	Students make decisions about which CCC to use to organize their approach to / reasoning within an item or task. Students explicitly use the CCCs to	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,

High (cont.)	disciplines). Students use DCs to address a significant uncertainty, with many possible alternative accounts.	sophisticated thinking that expands the nature of sensemaking. Students use SEPs to navigate complex interactions among multiple components of phenomena and problems.	expand sensemaking. With limited prompting, students use CCCs to navigate phenomena and problems with significant uncertainty and many possible alternative accounts	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 sophisticated use of a single DC component. Students use DCs to address a moderate uncertainty, with limited alternative accounts.	Students are cued to use specific SEPs and components of SEPs to address the item. Students use a single SEP component in support of authentic sensemaking, Students use SEPs to navigate simple interactions among components of phenomena and problems.	Students are cued to use a specific CCC component CCCs serve to focus student thinking within the item With guidance, students use the CCCs to navigate simple interactions among components of phenomena and problems with moderate uncertainty	Students are provided substantial cues for addressing the phenomenon or problem. They are prompted with specific DCs, SEPs, and CCCs, and provided 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	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 lower-level 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 applications of DCs, SEPS, and CCCs, often involving one or two scaffolded steps. Students address a low degree of

Low (cont.)	uncertainty.			uncertainty with a single possible account.

The following illustrative examples demonstrate ways that items may vary in complexity, including two versions of a low-complexity item, a series of different versions of an item with different complexities, and an item set with items that vary in complexity.

Please see Appendix A for more on the complexity of each sample item and additional detail that provides guidance on how these questions are addressed during item development and what sensemaking the items are designed to elicit. Each discrete item and each multipart item should be at least two dimensional and three dimensional if appropriate. 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.

Illustrative Examples for Varying Levels of Complexity

- **Park Flooding** presents two versions of a low-complexity item.
- **Cleopatra's Needle** and **Permafrost** present ways to vary complexity across different versions of an item.
- **Human Migration to Appalachia** presents ways to vary complexity across an item set that is medium complexity overall.

For example, Exhibit 3.19 illustrates a low-complexity grade 4 item assessing Earth and Space Sciences with low DC complexity and low SEP complexity. The phenomenon is presented through simple text, an image, and a simple graph—this provides students with enough information to demonstrate the targeted NAEP DC and SEP in service of sensemaking, without unnecessary reading or cognitive load.

Exhibit 3.19. Park Flooding, Version 1

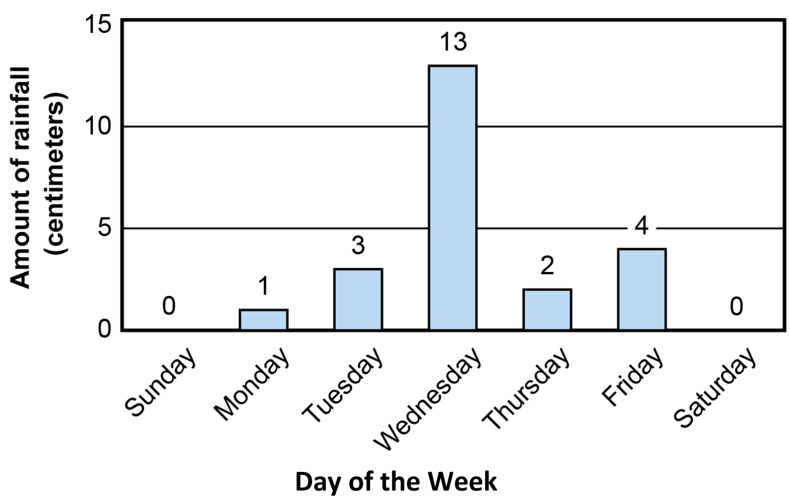
Item ID: Park Flooding (adapted from Formative Assessment Bundling Literacy and Elementary Science)

People visiting a local park noticed that the park was flooded and was closed for the day. The picture shows the flooded park.

Flooded Park[8]

The park was closed only on the day the flooding happened. The bar graph shows the rainfall for each day of that week.

Amount of Rainfall at the Park



Based on the data, on which day was the park **most likely** flooded?

A. Sunday
B. Monday
C. Tuesday
D. Wednesday

Key: D

The following version of this item shows a modification that provides additional evidence of student understanding of both the NAEP DC and SEP in service of sensemaking. The additional

[8] Betty Longbottom / Flooded Playground! - Cliffe Avenue / CC BY-SA 2.0

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 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.

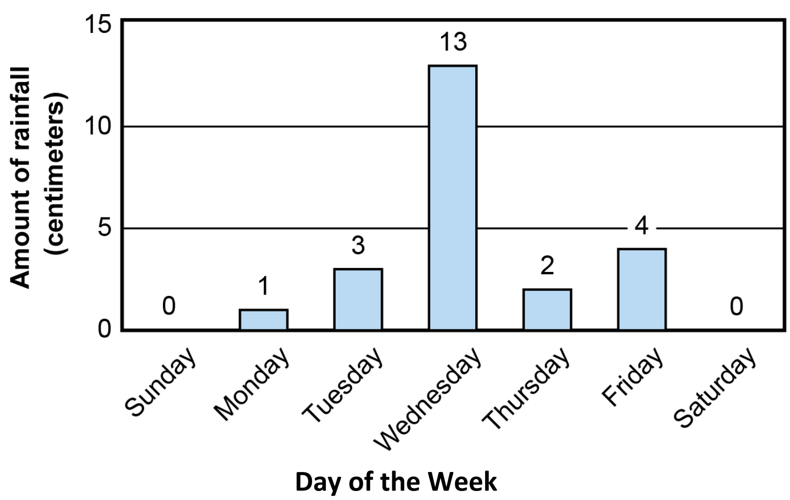
Exhibit 3.20. Park Flooding, Version 2

People visiting a local park noticed that the park was flooded and was closed for the day. The picture shows the flooded park.

Flooded Park[9]

The park was closed only on the day the flooding happened. The bar graph shows the rainfall for each day of that week.

Amount of Rainfall at the Park



9 Betty Longbottom / Flooded Playground! - Cliffe Avenue / CC BY-SA 2.0

Part A

Based on the data, on which day was the park **most likely** flooded?

A. Sunday
B. Monday
C. Tuesday
D. Wednesday

Key: D

Part B

Based on your understanding of weather, which piece of evidence **best** supports your answer in Part A?

A. This day was rainy.
B. The rain started falling on this day.
C. This day had more rainfall than any other day in the week did.
D. The amount of rainfall on this day was lower than on any other day.

Key: C

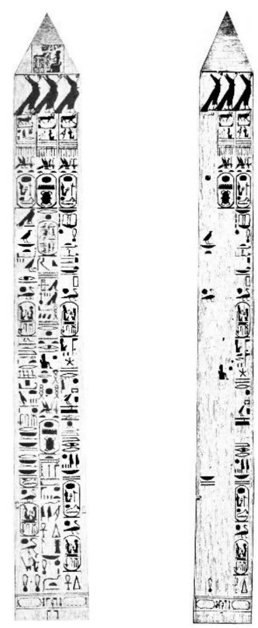
Exhibit 3.21 presents an example of a low DC, low SEP complexity item. This item requires students to make sense of a simple phenomenon: that Cleopatra's needle began deteriorating rapidly when it was moved to a new climate, despite having stood without damage for thousands of years prior. The phenomenon is presented through simple text, an image, and a very simple data set—this provides students with enough information to demonstrate the targeted NAEP 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 complexity. This phenomenon also leverages an observation that includes ties to other countries and historical periods and figures who are important to many specific communities around the world. To better represent cultural responsiveness in this context, there could be several possible extensions of the item, such as exploring the tradeoffs of moving a monument from one location to another and asking students to suggest actions that could be taken in the future to protect pieces of history.

Exhibit 3.21. Cleopatra’s Needle, Version 1

Item ID: Cleopatra’s Needle, Version 1 (adapted from a released NAEP item)

Cleopatra’s Needle is a large monument made of rock. For thousands of years, Cleopatra’s Needle stood in Alexandria, a city in the Egyptian desert. In 1881, Cleopatra’s Needle was moved to Central Park in New York City. After only a few years in New York City, the surface of the monument began crumbling. The diagram shows how Cleopatra’s Needle looked when it stood in Alexandria and how it looked after a few years in New York City. The table shows average yearly weather data for the two locations where Cleopatra’s Needle stood.

Cleopatra’s Needle



Alexandria — New York City

Weather Data in Alexandria and New York City

Location	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 pictures and the weather data to determine why Cleopatra’s Needle began crumbling after it was moved to New York City. Then, use the drop-down menus to correctly complete the sentences.

The (amount of rainfall/wind speed) in New York City over a year is (greater than/less than) that in Alexandria, Egypt.

This leads to (more/less) weathering of rock and causes the crumbling of Cleopatra’s Needle.

Key: amount of rainfall; greater than; more

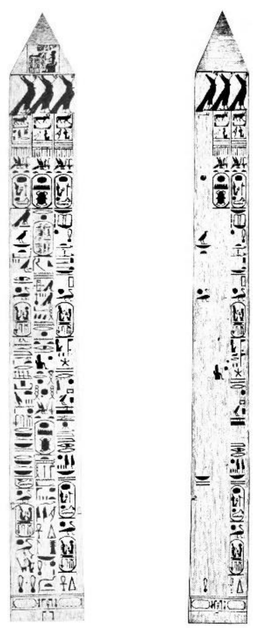
Exhibit 3.22 uses the same context and phenomenon as the previous item but engages students in deeper uses of NAEP SEPs by requiring students to make and support an original claim based on data and their understanding of erosion and weathering. The NAEP DC used is relatively low level, but the NAEP SEP requires students to use multiple components of NAEP 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 3.22. Cleopatra’s Needle, Version 2

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

Cleopatra’s Needle is a large monument made of rock. For thousands of years, Cleopatra’s Needle stood in Alexandria, a city in the Egyptian desert. In 1881, Cleopatra’s Needle was moved to Central Park in New York City. After only a few years in New York City, the surface of the monument began crumbling. The diagram shows how Cleopatra’s Needle looked when it stood in Alexandria and how it looked after a few years in New York City. The table shows average yearly weather data for the two locations where Cleopatra’s Needle stood.

Cleopatra’s Needle



Alexandria New York City

Weather Data in Alexandria and New York City

Location	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

Use the picture and the data in the table to make a claim about why Cleopatra’s Needle began crumbling in New York City when it did not crumble in Alexandria.

Part B

Use the picture and the data in the table to:

- Describe one piece of evidence from the picture or the table that supports the claim you made in Part A.
- Explain how the evidence supports the claim.

Exhibit 3.23. Cleopatra's Needle Version 2 Constructed Response Scoring Notes

Part A:

- Claims should reflect an analysis of the relative damage to the stone surface and the different weather conditions between New York City and Alexandria.
- Claims should focus on the impact of water changing the surface of rock (stone/monument) rather than wind (note that wind is included here so that students have to analyze the data and specifically make the claim that water is more likely to be responsible here).
- Student claims should be interpreted relative to Part B—for example, students may make a claim such as "It began to crumble because the weather conditions were different." Credit should be awarded if their response in part B indicates that they understand that water has changed the surface of the rock.

Part B:

- Evidence should derive from the data provided, for example, observations comparing the state of the monument in Alexandria versus New York City, the differential rainfall, or the differential rainfall and the similar windspeed (a complete argument often includes refuting possible counterclaims; while this question does not require this, students should not be penalized if they include evidence that refutes a counterclaim, such as the wind is driving degradation of the rock, as long as the reasoning makes clear that students interpreted the data appropriately).
- Rationales should include both
  - how the evidence supports their specific claim and
  - reasoning that connects the differential rainfall to increased damage in New York, using the idea that water can change a rock's surface as part of the reasoning.
- Students may include a counter argument for wind, saying that wind also could have this effect, but because wind speeds are similar in the two areas, there is no evidence that wind is the cause.
- Some students may provide a more general claim (e.g., weather conditions are different) and may provide reasoning that draws on DC or CCC understanding that is beyond the data that are provided here (e.g., some students could posit that wind may have led to damage if historically there was substantially less robust winds in Alexandria, and the similar wind speeds are only a recent observation [connecting to the CCC Stability and Change] or could cite a different environmental cause that is plausible). Students could be given partial credit for appropriate reasoning with the DC/CCC but should not be given full credit without connecting specifically to the data provided.

Exhibit 3.24 illustrates a low-complexity grade 12 item with low DC complexity, low SEP complexity, and low CCC complexity. By comparison to Exhibit 3.21 this shows how adding a dimension does not necessarily increase the complexity. While students cannot respond to this item without bringing some understanding of the DC, SEP, and CCC, it is heavily supported, thus limiting the amount of sensemaking students engage in with any dimensions. Students are given the relationship they need to map, only relevant statements to move, and a very structured and nonquantitative model illustrating feedback loops. It should be noted that this item leans into the intentional progressions built into the dimensions: the SEP and CCC are appropriate for grade 12 at a low level, and they are related to expectations and performances at lower grade levels. This allows a way to reduce the complexity of the item, remain consistent with the grade-level targets, and account for the fact that lower-performing students may have a less sophisticated, less well-developed understanding of the targeted dimensions that more closely approximates sophisticated performances at lower grade bands.

Exhibit 3.24. Permafrost, Version 1

Item ID: Permafrost Melting (adapted from OpenSciEd)

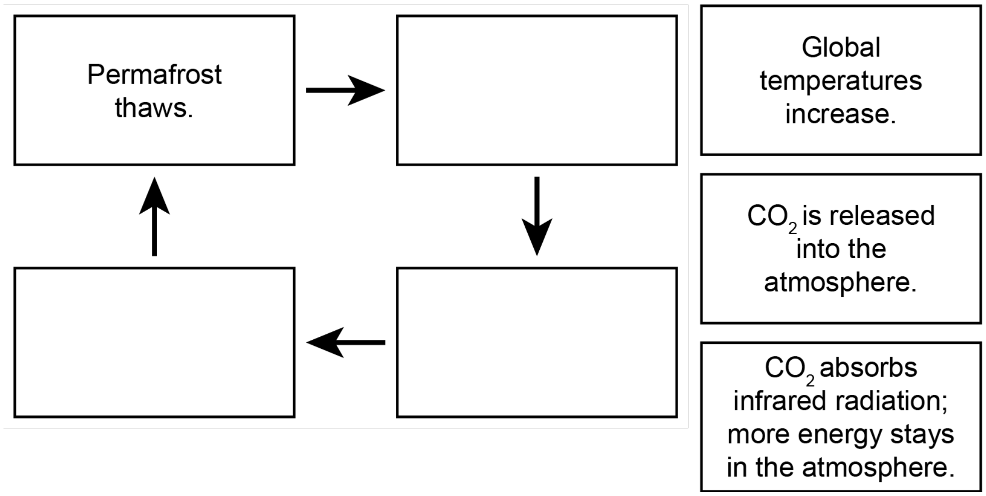
Permafrost is a layer of soil and ice that is just below the surface in the Arctic. An example of permafrost is shown in the picture.

Permafrost[10]

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.

Use the statements to develop a model that shows the relationships between thawing permafrost and rising global temperatures. Drag the statements into the correct boxes to complete the model. Each statement will be used once.

[10] Permafrost from OpenSciEd, CC BY-SA 4.0



Key: Permafrost thaws. CO_2 is released into the atmosphere. CO_2 absorbs infrared radiation; more energy stays in the atmosphere. Global temperatures increase.

The following two versions of the item are modified to be higher complexity by (a) requiring students to develop the model with significantly less support (Exhibit 3.25), and/or (b) asking students to consider implication and limitations of the model (Exhibit 3.26). By making these modifications, students are more independently responsible for figuring out why melting permafrost contributes to rising global temperatures by using their understanding of modeling, feedback loops, and energy in systems. In these variations, the complexity has been increased, but there is still substantial cueing in the form of the provided statements. A more complex item might ask students to develop an original model, without any prompting statements.

Exhibit 3.25. Permafrost, Version 2

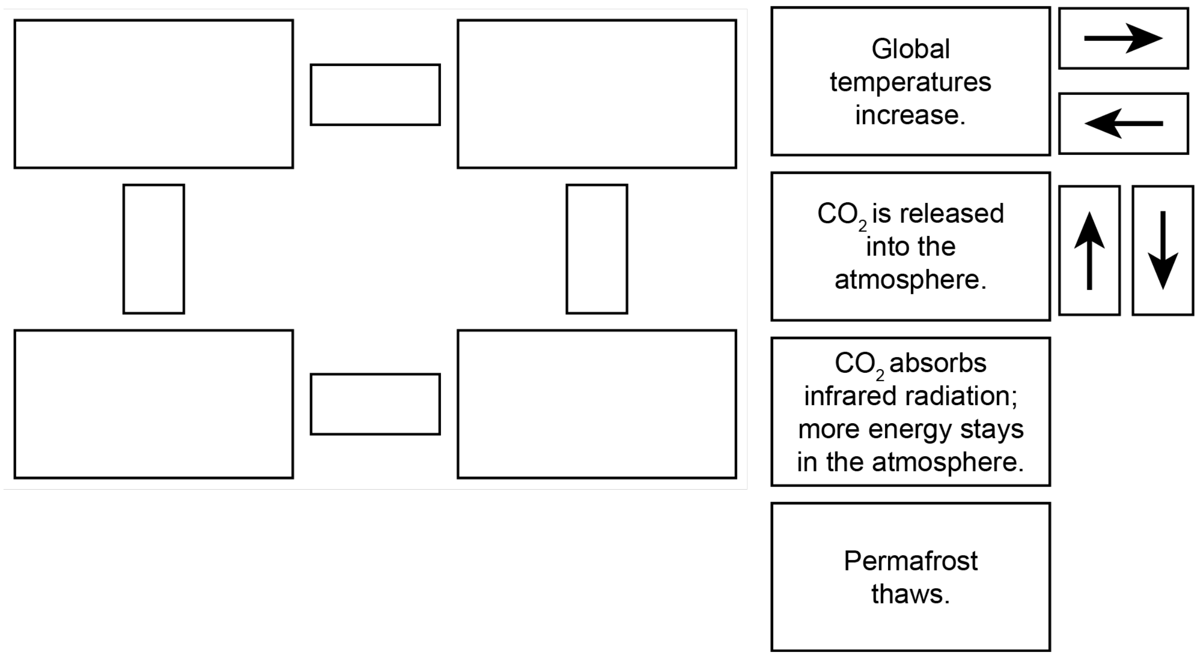
Permafrost is a layer of soil and ice that is just below the surface in the Arctic. An example of permafrost is shown in the picture.

Permafrost[11]

[11] Permafrost from OpenSciEd, [CC BY-SA 4.0]

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.

Use the statements and arrows to develop a model that shows the relationships between thawing permafrost and rising global temperatures. Drag the statements and arrows into the correct boxes to develop the model. Each statement and each arrow will be used once.



Key: Global temperatures increase. → Permafrost thaws. → CO_2 is released into the atmosphere. → CO_2 absorbs infrared radiation; more energy stays in the atmosphere. → (Responses can start with any of the draggers.)

Exhibit 3.26. Permafrost, Version 3

Permafrost is a layer of soil and ice that is just below the surface in the Arctic. An example of permafrost is shown in the picture.

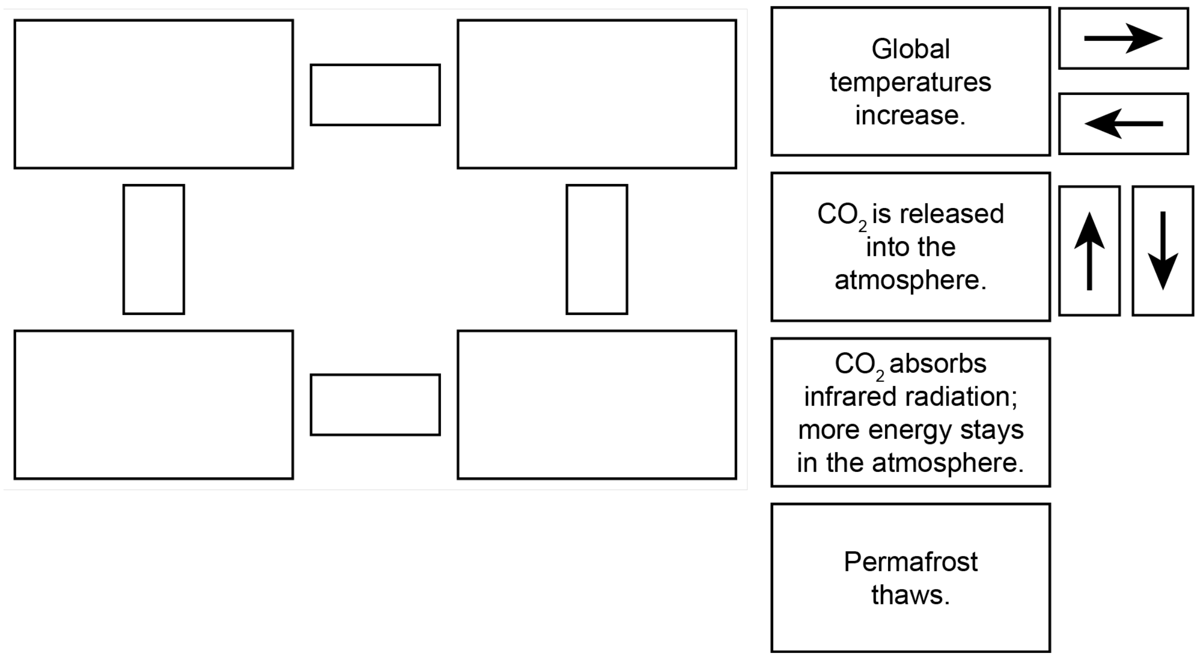
Permafrost[12]

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.

[12] Permafrost from OpenSciEd, CC BY-SA 4.0

Part A

Use the statements to develop a model that shows the relationships between thawing permafrost and rising global temperatures. Drag the statements and arrows into the correct boxes to complete the model. Each statement and arrow will be used once.



Key: Global temperatures increase. → Permafrost thaws. → CO_2 is released into the atmosphere. → CO_2 absorbs infrared radiation; more energy stays in the atmosphere. → (Responses can start with any of the draggers.)

Part B

Complete the sentence by choosing the correct answer from the drop-down menu.

The model predicts that the rate of melting of the permafrost will likely (increase / stay the same / decrease) over the next 50 years.

Key: increase

Part C

Describe a limitation of the model you developed in Part A.

Exhibit 3.27. Permafrost Part C Constructed Response Scoring Notes

- Students provide one limitation of the model, for example:
    - The model is missing factors such as interactions with other systems and feedback loops.
    - Thawing may occur at different rates in different locations around the world.
    - There may be a time lag between temperature increases and thawing.

Exhibit 3.28 presents ways to vary complexity across an item set that is medium complexity overall. It should be noted that this item set could be expanded to more deeply and comprehensively assess (a) related Earth and Space Sciences DCs (e.g., how the geography of the Appalachian region has contributed to climate resilience over time), (b) the Life Science DCs (e.g., how ecosystem dynamics contribute to resilient biodiversity), or (c) their integrated use (e.g., how geographic features that have evolved over time have led to adaptations and speciation, contributing to the rich biological systems in the area and impacts on human civilizations). These could be explored in further independent items within an item set or in related items within a scenario-based task.

Exhibit 3.28. Human Migration to Appalachia

Item ID: Human Migration to Appalachia

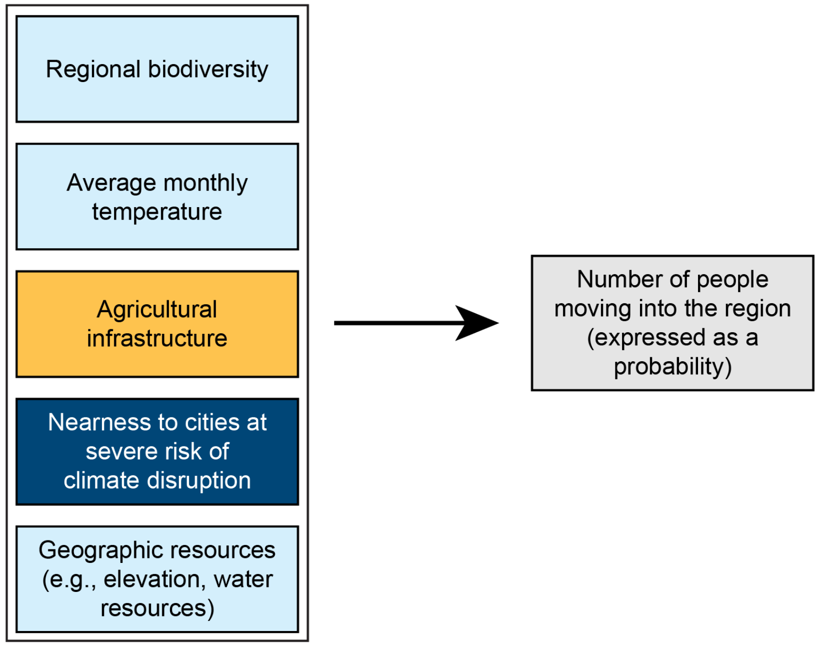
Appalachia is considered "climate resilient." This means that the area can successfully handle the impacts of changes to climate and can prevent those impacts from growing worse. The green areas in Figure 1 show where Appalachia is located in the United States.

Figure 1. Map of Appalachia

Computational models predict that many people will move into the Appalachian region over the next 20 years as they seek to find places to live that are safer and more stable.

Figure 2 shows one model local leaders are using to predict how many people will move into Appalachia. Blue indicates factors that are expected to increase migration, and orange indicates factors expected to decrease migration. Darker colors indicate more weight on that factor in the model. Agricultural infrastructure includes farms, markets and businesses that support farms, and the transportation and communication systems in the area.

Figure 2. Computational Model for Predicting Human Migration



When this model was tested against recent population growth due to migration into two locations in Appalachia, leaders noticed some differences between what the model predicted and what the data showed. The table shows these differences for the two locations, along with information about how high or low each location is rated on several factors.

Predicted and Actual Population Growth in Pittsburgh and the Shenandoah Valley

Location	Pittsburgh	Shenandoah Valley
Predicted population growth	high	low
Actual population growth	low	high
Relative biodiversity	low	high
Average monthly temperature range	29–73° F	32–74° F
Relative agricultural infrastructure	low	medium
Nearness to climate-impacted urban centers	high	high
Access to usable water	medium	high

Item 1

Based on your understanding of what human societies need to be successful, which idea **best** explains why the model prediction was different from the observed data?

A. Pittsburgh has highly limited access to water. The model did not account for how this shortage would limit the growth of local businesses.
B. Human societies require reasonable annual temperature ranges. The model incorrectly assumed that average monthly temperature was a less important factor.
C. Human societies require access to food and water. The model did not account for how important access to food and water would be for human migration to Appalachia.
D. The Shenandoah Valley has very little access to fresh fruits and vegetables. The model incorrectly assumed that agricultural infrastructure would not promote migration.

Key: C

Item 2

Human migration to areas such as Appalachia can result in rural gentrification. People who are currently living in low-cost, natural resource–rich areas are forced to leave, resulting in the loss of access to the resources and communities they have actively contributed to developing and maintaining. As a result, they can no longer enjoy the benefits of these regions.

Describe **one** way you could revise or build on the computational model to better understand how migration into Appalachia could impact current residents' access to natural resources. Be sure to explain how the change you describe will provide a better understanding of how migration into Appalachia could impact current residents' access to natural resources.

Exhibit 3.29. Human Migration to Appalachia Item 2 Constructed Response Scoring Notes

Emphasis here is on an understanding of how to revise the model parameters to better understand more nuanced population-environment dynamics. Student ideas can be wide-reaching but should be justifiable as at least one of the following: (a) updating the model to better understand who moves into and out of the region, and/or (b) determining the feedback impacts of population growth on biodiversity, and natural resources such as water. This item specifically focuses on the computational reasoning aspect of this SEP and can include either quantitative or qualitative reasoning from students. Appropriate lines of reasoning here can include, but are not limited to, the following: • updating the model to account for demographic subgroups • considering housing costs / other metrics for socioeconomic status as part of the inputs and/or outputs, as a mechanism to better understand the relative wealth/characteristics of who lives in the area • calculations of migration out of Appalachian regions • feedback mechanisms that influence biodiversity and natural resource availability (Note: This could be specific to particular resources, or general at the level of the categories included in the model.) • relative factors for scaling variables (quantitative or qualitative)

Appendix A includes additional examples of items at a range of complexity levels, including how a given item can be modified.

Complexity, Opportunity to Learn, and Surfacing the Full Range of Student Performance

One major goal of the complexity framework is to provide a way to vary the nature of items so that all students who have had the opportunity to learn some science can make their understanding visible on the NAEP Science Assessment. The complexity framework assumes that every item still needs to require grade-level NAEP DCs and SEPs. What varies are the degree and sophistication of the sensemaking required to respond to the item, which we expect will scale with increasing proficiency (e.g., students who have a robust grasp of science will be able to successfully engage with increasingly complex items, across a range of NAEP DCs, SEPs, and CCCs). Some students may not demonstrate robust science understanding simply because they did not have sufficient opportunity to learn the science being assessed by NAEP. The complexity framework attends to this to some degree by leveraging the intentional NAEP DC, SEP, and CCC progressions across grades 4, 8, and 12. The grade 8 NAEP SEP expectation may reflect a lower level of sophistication of a similar target at grade 12, so that an item that represents performance at the *NAEP Advanced* level for grade 8 may be similar to an item that represents performance at the *NAEP Basic* level for grade 12. If a student has not had the opportunity to develop their understanding beyond the grade 8 expectation, some information about their science understanding can still be captured. However, if students are performing at the lower end of the scale because they have not had opportunity to learn the targeted science, even low-complexity items may not surface their understanding.

Reflecting a Wide Range of Students

It is essential for the 2028 NAEP Science Assessment 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. Funds of knowledge are considered to be the historical accumulation of abilities, bodies of knowledge, assets, and cultural ways of interacting that a student might possess. Below are definitions and general principles for culturally relevant contexts for NAEP Science, followed by a list of 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, languages, and gender.
- 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 so that the assessment is 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.
- When contexts focus on legitimate interests of communities, it is more likely that all students will be engaged with the items. A culturally relevant lens asks whether 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

Phenomena and context provide opportunities to engage students with a wide range of experiences and identities. The NAEP Science Assessment should include a broad range of contexts to increase the likelihood that students will encounter descriptions that are relevant and meaningful, including those that reflect their experiences and identities, and those that provide windows into the experiences and identities of others:

- 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 people of color 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, and so on.

Stimuli should be analyzed to avoid bias and sensitivity issues that could negatively affect a students' testing experience. Topics to avoid include

- contexts that prompt a negative emotional response, such as dangerous situations;
- unsafe practices, particularly in descriptions of school or laboratory settings; and
- stereotypes or other forms of bias against specific groups, for example, as defined by gender, age, income, education, ethnicity, culture, or religion.

Some disciplinary concepts include content, such as hurricanes and earthquakes, that has the potential to introduce sensitivity issues. This content should be included in ways that avoid specific, recent events and predictions of risk for specific regions to minimize the impact on students during the assessment. Elements to consider include the following:

- Contexts that 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, and so on
- Solutions and explanations presented to and generated by students through the assessment items that improve people's lives and livelihoods
- Item contexts that 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 that include diverse representations of who is considered a scientist and engineer
- Contexts that position non-White people as (a) more than a stereotyped experience and (b) powerful doers and contributors to science and the broader world
- Contexts that do not include gratuitous or superficial representation of diverse races, ethnicities, genders, and so on

The following two illustrative examples provide culturally relevant contexts and items. Each example draws upon different cultures and provides access to cultural information in different ways. In both cases, the culturally relevant information is necessary information for students to respond to the items.

Illustrative Examples

- **Limu Kohu**
- **Human Migration to Appalachia**

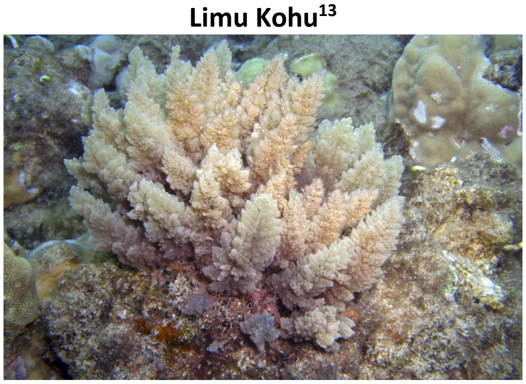
The first of these two examples, Exhibit 3.30, is an example of a two-dimensional grade 8 item. Several features of culturally responsive items are included, such as (a) the use of native/home language in the item (i.e., *limu kohu* is the Hawaiian language term for this seaweed species), (b) use of nontraditional evidence sources that have been useful in university-based science endeavors (i.e., the use of multigenerational/elder accounts as evidence, as used by Stanford botanist Dr. Isabella Aiona Abbott), and (c) 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 about the loss of indigenous foodways currently happening).

Exhibit 3.30. Limu Kohu

Item ID: Limu Kohu (adapted from the State Performance Assessment Learning Community)

Limu kohu is a type of seaweed that is native to the waters around Honolulu, Hawaiʻi. It 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 been rapidly declining over the past 60 years. An example of limu kohu seaweed is shown in the picture. The table describes observations of limu kohu.

Limu Kohu[13]



Observations from Generations of Hawaiian Elders about Limu Kohu Growth and Harvesting
• Limu kohu needs warm water and high salinity to grow. • Limu kohu grows 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.

[13] MDC Seamarc Maldives, CC BY-SA 4.0, via Wikimedia Commons

Part A

Which human activity is **least** likely to cause harm to limu kohu populations?

A. Companies using industrial methods of harvesting limu kohu remove the whole limu kohu plant.
B. Restaurants using traditional methods of harvesting limu kohu remove the top of the limu kohu.
C. Industrial runoff changes the temperature and salinity of the water in coastal regions where limu kohu live.
D. Ships visiting Hawaiʻi introduce invasive seaweed species that use the same resources as limu kohu into coastal regions where limu kohu live.

Key: B

Part B

Use the information provided and your understanding of the impacts of human activities on the environment to support your answer to Part A.

Exhibit 3.31. Limu Kohu Part B Constructed Response Scoring Notes

Students provide one statement that is based on the information provided, shows understanding of the impacts of human activities on the environment, and that supports the answer to Part A. For example:

- Traditional harvesting practices that focus on harvesting only the top portions of limu kohu are less destructive compared to other methods that involve uprooting the entire plant. When only the top is harvested, the base of the plant remains intact, allowing the limu kohu to potentially regrow.
- The information provided indicates that limu kohu has the ability to regrow when it is trimmed. This suggests that harvesting the upper parts of the seaweed allows it to regenerate, ensuring the sustainability of the population over time.

The second example, Exhibit 3.32, attends to cultural relevance by highlighting a specific community (in this case, the many largely rural communities that make up Appalachia) through an asset-based lens. This is particularly important because these specific rural communities are often portrayed and understood through deficit-oriented lenses (e.g., focusing on poverty, lack of educational resources and college degrees, economically less advantageous career options, etc.).

Exhibit 3.32. Human Migration to Appalachia

Item ID: Human Migration to Appalachia

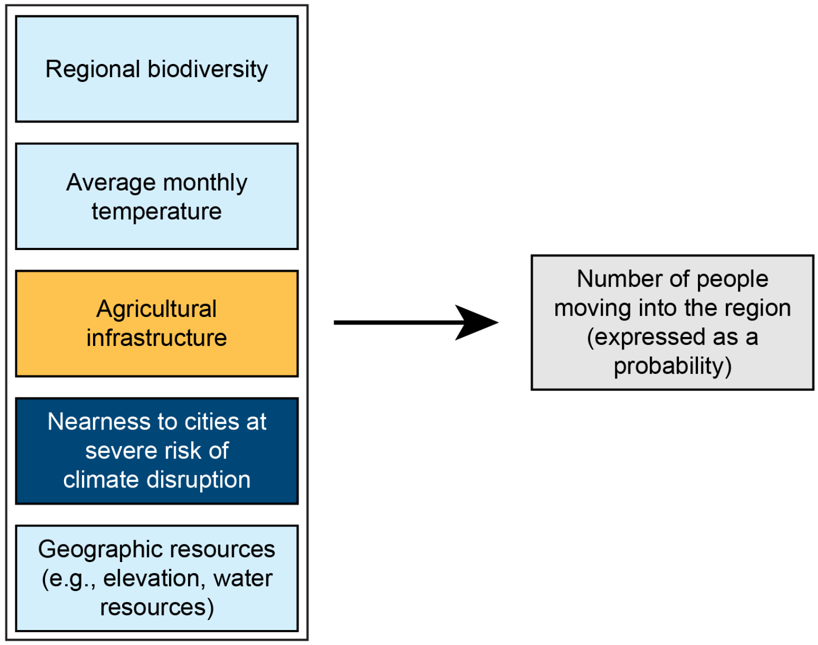
Appalachia is considered "climate resilient." This means that the area can successfully handle the impacts of changes to climate and can prevent those impacts from growing worse. The green areas in Figure 1 show where Appalachia is located in the United States.

Figure 1. Map of Appalachia

Computational models predict that many people will move into the Appalachian region over the next 20 years as they seek to find places to live that are safer and more stable.

Figure 2 shows one model local leaders are using to predict how many people will move into Appalachia. Blue indicates factors that are expected to increase migration, and orange indicates factors expected to decrease migration. Darker colors indicate more weight on that factor in the model. Agricultural infrastructure includes farms, markets and businesses that support farms, and the transportation and communication systems in the area.

Figure 2. Computational Model for Predicting Human Migration



When this model was tested against recent population growth due to migration into two locations in Appalachia, leaders noticed some differences between what the model predicted and what the data showed. The table shows these differences for the two locations, along with information about how high or low each location is rated on several factors.

When this model was tested against recent population growth due to migration into two locations in Appalachia, leaders noticed some differences between what the model predicted and what the data showed. The table shows these differences for the two locations, along with information about how high or low each location is rated on several factors.

Predicted and Actual Population Growth in Pittsburgh and the Shenandoah Valley

Location	Pittsburgh	Shenandoah Valley
Predicted population growth	high	low
Actual population growth	low	high
Relative biodiversity	low	high
Average monthly temperature range	29–73° F	32–74° F
Relative agricultural infrastructure	low	medium
Nearness to climate-impacted urban centers	high	high
Access to usable water	medium	high

Item 1

Based on your understanding of what human societies need to be successful, which idea **best** explains why the model prediction was different from the observed data?

A. Pittsburgh has highly limited access to water. The model did not account for how this shortage would limit the growth of local businesses.
B. Human societies require reasonable annual temperature ranges. The model incorrectly assumed that average monthly temperature was a less important factor.
C. Human societies require access to food and water. The model did not account for how important access to food and water would be for human migration to Appalachia.
D. The Shenandoah Valley has very little access to fresh fruits and vegetables. The model incorrectly assumed that agricultural infrastructure would not promote migration.

Key: C

Item 2

Human migration to areas such as Appalachia can result in rural gentrification. People who are currently living in low-cost, natural resource–rich areas are forced to leave, resulting in the loss of access to the resources and communities they have actively contributed to developing and maintaining. As a result, they can no longer enjoy the benefits of these regions.

Describe **one** way you could revise or build on the computational model to better understand how migration into Appalachia could impact current residents' access to natural resources. Be sure to explain how the change you describe will provide a better understanding of how migration into Appalachia could impact current residents' access to natural resources.

Exhibit 3.33. Human Migration to Appalachia Item 2 Constructed Response Scoring Notes

Emphasis here is on an understanding of how to revise the model parameters to better understand more nuanced population-environment dynamics. Student ideas can be wide-reaching but should be justifiable as at least one of the following: (a) updating the model to better understand who moves into and out of the region, and/or (b) determining the feedback impacts of population growth on biodiversity, and natural resources such as water. This item specifically focuses on the computational reasoning aspect of this SEP and can include either quantitative or qualitative reasoning from students. Appropriate lines of reasoning here can include, but are not limited to, the following: • updating the model to account for demographic subgroups • considering housing costs / other metrics for socioeconomic status as part of the inputs and/or outputs, as a mechanism to better understand the relative wealth/characteristics of who lives in the area • calculations of migration out of Appalachian regions • feedback mechanisms that influence biodiversity and natural resource availability (Note: This could be specific to particular resources, or general at the level of the categories included in the model.) • relative factors for scaling variables (quantitative or qualitative)

For additional examples, please see Appendix A.

Science Achievement 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 write a science achievement expectation expressed as a *performance*—something that the student can be expected to do to indicate they understand the targeted NAEP DC and can apply it via the cued associated NAEP SEP (and CCC, when possible). This process begins after the selection of one or more DCs and a grade-appropriate phenomenon or problem that students can make sense of or solve using the DCs. Once the DCs have been identified, it is then possible to find a realistic context for use in the assessment—a phenomenon that requires explanation, such as the changing length of day over the seasons, as in the grade 4 example in Exhibit 3.34 or a problem calling for an engineering design, such as the grade 12 example. The grade 8 expectation could be implemented through either a phenomenon or a problem.

The next step toward writing a science achievement expectation follows by considering the CCCs and SEPs that students can use to apply the knowledge described by the DC. Developers should consider the following questions when writing a science achievement expectation:

- Does the performance described in the science achievement expectation fit the purposes of sensemaking or problem-solving.

- Does the integration of the three dimensions describe a performance that is grade appropriate?
- What complexity and difficulty are needed in the science achievement expectation to address particular complexity and achievement levels?

The following are examples that can be used to build items for grades 4, 8 and 12.

Exhibit 3.34. Examples of Science Achievement Expectations

Science Achievement 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/temperature cues for plant life cycles, etc.).
Grade 8 Physical Science Ask questions about the interactions between objects to determine how changes in their motions are	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 moving	C8.8: Systems may interact with other systems; they may have subsystems and be a part of larger more complex systems.	S8.2: Ask questions that can be answered with empirical evidence to investigate relationships between variables in a	Students can begin to ask questions to develop a qualitative understanding of forces at entry points to making sense of

Science Achievement Expectation	DC	CCC	SEP	Rationale
determined by the sum of the forces acting on each object.	in a straight line with the same speed and direction as before.		system model or in phenomena.	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 grade 12 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.

As the item developer gathers information and images about particular contexts, thinks about the level of complexity and difficulty that are needed in the performance, and chooses an appropriate item type, the item starts to take shape. Each exhibit in the Specifications is accompanied by a description of the multidimensional understanding it assesses and an explanation. Developers can use these paragraphs to follow the development process described in this section.

Digital Tools

The NAEP Science Assessment based on the Specifications will be administered via computer. A digital environment provides opportunities to include a number of digital tools—and, at times, science-specific tools—students can use to respond to the items. The 2028 NAEP Science Assessment will include digital tools to support NAEP DCs, SEPs, and CCCs.

The testing environment will need to provide a computational tool equivalent to a four-function calculator. 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 tutorials for each grade level can be found on the National Center for Education Statistics website.

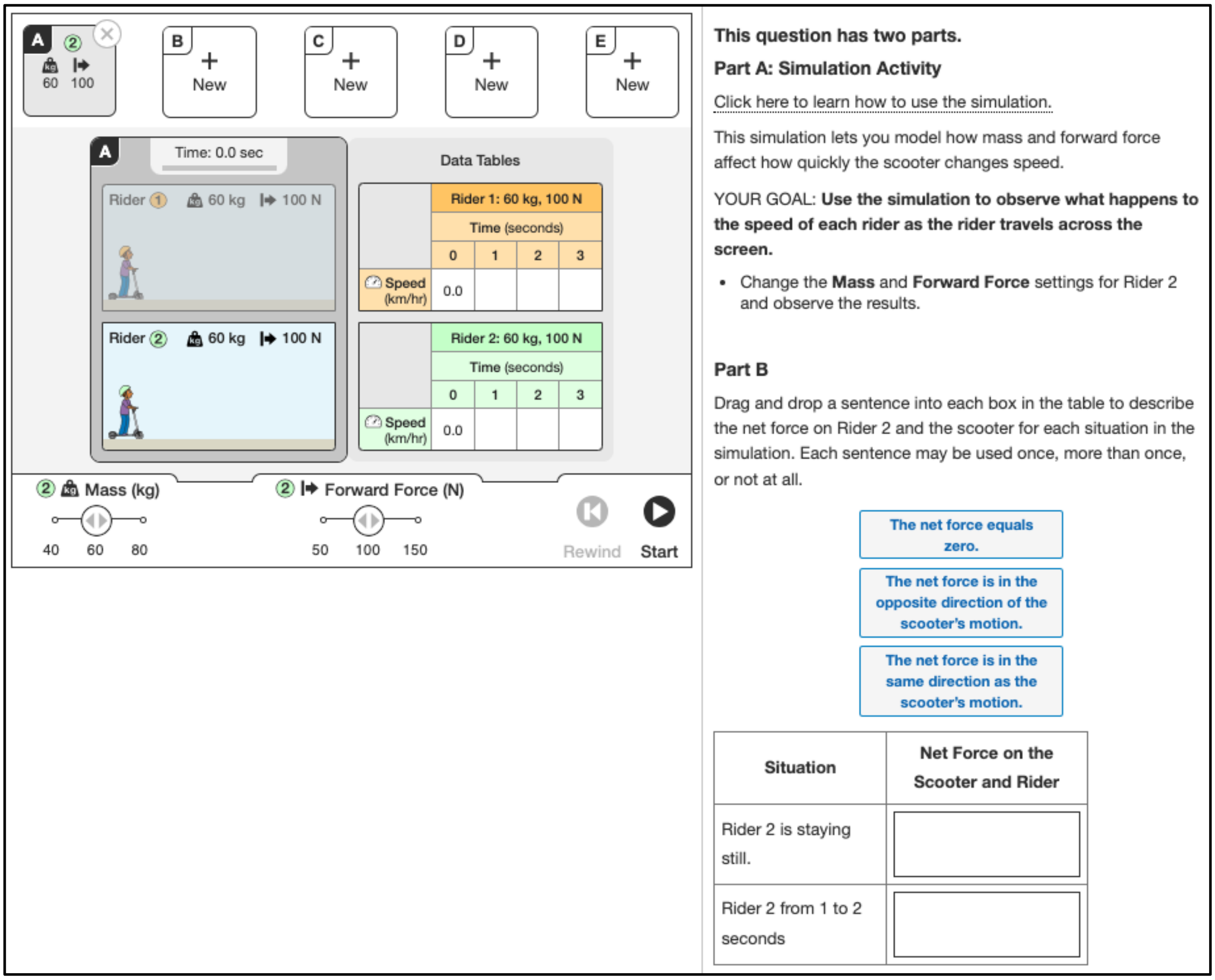
All digital NAEP assessments include system tools, which are always available and common across all NAEP assessments. There are also science and mathematics tools, which are specific to and available only for certain items on NAEP science assessments. The materials and accompanying items are carefully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

The illustrations in the Specifications are static screenshots to illustrate examples of these digital tools; however, the screenshots 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)

The following example (Exhibit 3.35) 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. Please note that this simulation is intended to illustrate the use of a simulation within a set of items. These examples were taken from a larger item set or scenario-based task and, as standalone simulations, do not feature the phenomenon described earlier in the task. This item, therefore, is not intended to be an example of a possible NAEP item.

Exhibit 3.35. Sample Simulation from a Multidimensional Item Set[14]



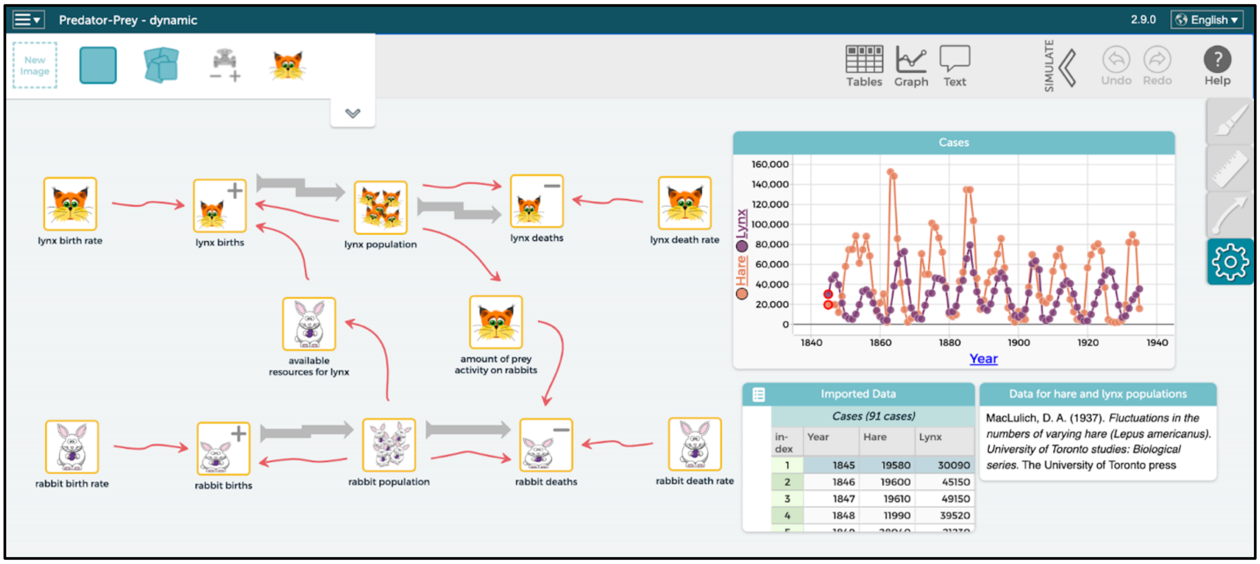
Similarly, the example below (Exhibit 3.36) shows how digital tools might be used to allow students to construct dynamic models that they independently develop. This example leverages SageModeler, a free, open-source, web-based systems dynamics modeling tool commonly used in science education. This tool allows students to define variables, relationships, and degree of influence, and to run models and collect data.

The exhibit shows a model of the interactions between two populations and the resulting changes to the populations over time. This example could be used with items that assess ecosystems NAEP DCs; analyzing, explaining, and evaluating SEPs; and relevant CCCs, including but not limited to Patterns, Cause and Effect: Mechanism and Explanation, Systems and System Models, Energy and Matter: Flows, Cycles, and Conservation, and Stability and change. The complexity of the model could be tailored to the targeted grade and complexity level.

[14] Image used with permission from Massachusetts Department of Elementary and Secondary Education

A modeling tool such as this could be used to model any system in which understanding how the components interact and how those interactions lead to changes in the overall system is appropriate for assessing the NAEP DCs, SEPs, and CCCs. Please note that this item is not intended to be an example of a possible NAEP item.

Exhibit 3.36. Sample Modeling Tool (SageModeler)[15]



Process Data for Future NAEP Science Assessments

A key challenge is the need to capture enough information about science content knowledge, crosscutting concepts, and practices for a reliable and valid assessment. When this happens within the context of scenario-based tasks, which require more time for engagement and completion, data may be available from fewer items per student. However, other sources of evidence may be available from *process data*, interactions between the student and the assessment other than their responses to items.

While there is potential to collect and use process data to measure student performance, carefully constructing and validating these measures may be challenging to achieve in practice and require more iterations than the typical assessment development cycle provides. A special study in the area of science assessment is needed to explore and fully realize the potential of process data within scenario-based tasks, described in Appendix E.

Accessibility

Science assessments should be developed to allow for the participation of the widest possible range of students. For NAEP, the interpretation of scores should lead to valid inferences about levels of performance of the nation's students, as well as to valid comparisons across states and urban districts. All students should have the opportunity to demonstrate their ability to use

[15] Image generated using SageModeler (https://sagemodeler.concord.org/), which was developed at the Concord Consortium and Michigan State University.

practices to apply disciplinary concepts and crosscutting concepts that the NAEP Science Assessment is intended to measure. To this end, item writing should follow the principles of universal design and sound testing practices as recommended by the National Center on Educational Outcomes (Thompson, et al., 2005). These include attention to the population being assessed, precise definition of the constructs being assessed, review for fairness and accessibility of item content, clarity of the language and graphics used throughout the assessments, and the provision of accommodations without changing the constructs being assessed.

Although application of universal design principles to the item development process considers the ways in which the population being assessed can demonstrate sensemaking in science, the use of such principles does not remove the need for accommodations altogether. With this in mind, items should be written to allow for necessary accommodations, including the use of online tools available to students during test administration, without changing the constructs being assessed.

The NAEP Science Assessment is designed to measure student achievement across the nation. Consequently, NAEP incorporates inclusive policies and practices into every aspect of the assessment, including selection of students, participation in the assessment administration, and valid and effective accommodations. NAEP is administered to a sample of students who represent the student population of the nation, regardless of race/ethnicity, socioeconomic status, disability, status as an English Learner, or any other demographic or geographic factors. Therefore, the NAEP Science Assessment provides an opportunity for participating students to demonstrate sensemaking in science, including students who have learned science in a variety of ways, followed different curricula, and used different instructional materials; students who have mastered disciplinary concepts, crosscutting concepts, and practices to varying degrees; students with a variety of disabilities; and students who are English Learners. The related design issue is the development of a large-scale assessment that measures the science achievement of students who come to the assessment with different experiences, strengths, and challenges; who approach science from different perspectives; and who have different ways of displaying their sensemaking in science.

NAEP uses two methods to design an accessible assessment program that provides accommodations for students with special needs. The first is addressed by careful item and delivery design with the full consideration of the range of participating students. For many students with disabilities and students whose native language is not English, the standard administration of the NAEP assessment will be most appropriate. For other students with disabilities (SD students) and some English Learners (EL students), NAEP allows for a variety of accommodations, which can be used alone or in combination.

Some accommodations are built-in features, called Universal Design Elements, of the NAEP system tools that are available to all students. Other accommodations, such as additional assessment time, are offered for specific eligible students. Available accommodations fall into four categories (see National Assessment Governing Board, n.d., for specific information about accommodations):

- standard NAEP Practice, available in almost all NAEP assessments for SD and EL students
- other accommodations for SD students that require special presentation, such as Braille or sign language
- other accommodations for EL students
- Universal Design Elements that are built-in features of the computer-based assessments available to all students

For more detailed information about accommodations, see the Accommodations page on the NCES website and the Governing Board's Policy Statement *NAEP Testing and Reporting on Students with Disabilities and English Language Learners* (National Assessment Governing Board, 2014).

The Role of Mathematics in Items That Assess Science

Because mathematics is a part of science, it is necessary to characterize the mathematical skills that assessment items will draw upon at each grade level of the NAEP Science assessment. Mathematics required for understanding and responding to assessment items should not be so complex that an item is assessing mathematical understanding and facility rather than science understanding. Calculations should be simple so that the computation time required is small.

In general, science items at each grade level should be designed to assess the construct of science achievement as described by the Framework and clarified in the Specifications. Much of this description is within the SEP Mathematics and Computational Thinking, including statements, clarifications, and boundaries. The following are other examples where the use of mathematics is described:

- The SEP Developing and Using Models includes aspects of mathematical modeling.
- The SEP Analyzing and Interpreting Data includes aspects of statistics and probability.
- The CCC Scale, Proportion, and Quantity includes aspects of measurement.

Within the Framework, any aspect of mathematics that is appropriate for a given grade level is usually appropriate for higher grade levels.

The following are appropriate at grades 8 and 12:

- fraction or decimal quantities
- signed quantities
- unit conversions
- the coordinate plane
- rate, ratio, proportion, or percent
- volume
- probability

The following are appropriate only at grade 12:

- exponents
- scientific notation
- modeling nonlinear relationships
- modeling relationships among more than two variables

No mathematics knowledge is required beyond the statements in the Framework. The following principles provide item developers with additional guidance about the appropriate role of mathematics in items that assess science:

- Students at any grade may use a calculator to perform any calculations that may be helpful or required for successfully completing test items on the NAEP Science Assessment. More information about calculators is included in the section on Digital Tools.
- Calculations should be included as needed for measuring the construct of science achievement. Any necessary mathematical formulas or facts should be provided to the student.
- Items should be designed so that all calculations could be performed using multidigit whole numbers and/or decimals, depending on the grade level.
    - At grade 4, operations only ever involve positive whole numbers, including quotients.
    - At grade 8, operations only ever involve positive whole numbers and decimals.
    - At grade 12, operations could involve signed whole numbers and decimals.
- Items that represent quantities or relationships as fractions should be designed to convert to decimals so that fraction calculations are not required.

These principles apply to quantities that students are required to use in calculations. For example, signed quantities may be needed in an item at grade 8, but the quantities would only be represented or interpreted, not operated on.

Matrix Sampling

The design of NAEP uses matrix sampling to enable a broad and deep assessment of students' knowledge and skills that also minimizes the time burden on schools and students. Matrix sampling is a sampling plan in which different samples of students take different samples of items. Students taking part in the assessment do not all receive the same items. Matrix sampling greatly increases the capacity to obtain information across a much broader range of the objectives than would otherwise be possible.

Balance of the Assessment

A high-quality science assessment is based on relevant and meaningful phenomena and problems that provide opportunities for students to show that they are able to use the practices to apply disciplinary concepts and crosscutting concepts in items. The integration of the dimensions of science requires particular attention to the balance of the assessment, as each item is classified in two or three categories for the purposes of maintaining specific distributions.

This balance of the disciplinary concepts, science and engineering practices, and crosscutting concepts is introduced in Chapter 2; a summary of all the balance dimensions, including item type, follows.

Exhibit 3.37. Summary of Balance Dimensions

Balance Dimension	Distribution
Discipline Group	33.3% each
Practice Category	At least 10% each
Crosscutting Concept	At least one item each
Item Type	65% SR, 35% CR

The assessment should be built so that the constructs represented by the statements in each dimension area are adequately represented. The breadth and relative emphasis in each dimension, as presented in the Framework and Specifications, should be represented on the assessment as a whole. The developer should avoid under- or over-emphasizing particular discipline groups, practice categories, or crosscutting concepts to the extent feasible, with a goal of ensuring broad, but not necessarily equal, coverage in any given year’s item pool and coverage of all statements over time.

CHAPTER 4: Reporting NAEP 2028 Science Results

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 typically 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 TUDA program 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, and Los Angeles Unified School Districts and New York City Public Schools). 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 school district 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.

Reporting Scale Scores and Achievement Levels

NAEP typically 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 reasoning and problem-solving on the assessment. These reporting categories harken back to the claims described in Section 1D of the Framework (pages 4–5) with their focus on sensemaking and multidimensionality. 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 NAEP DCs in **Physical Science**, in combination with NAEP SEPs and CCCs.
- **Sensemaking in Life Science**: The student reasons scientifically using NAEP DCs in **Life Science**, in combination with NAEP SEPs and CCCs.
- **Sensemaking in Earth and Space Sciences**: The student reasons scientifically using NAEP DCs in **Earth and Space Sciences**, in combination with NAEP SEPs and CCCs.

Despite the focus on multiple dimensions, NAEP will not report on any of the three dimensions separately, as they work together. That is, there will be no separate scores for students' knowledge of NAEP DCs, SEPs, or CCCs. 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 NAEP DCs 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—NAEP DCs, SEPs, and CCCs—and does not prioritize knowledge of the NAEP DCs.

Because alignment to a statement within one of the three discipline groups is the basis for assigning an item to one of the reporting categories described above, all items will align to at least one disciplinary concept. Items may align to multiple disciplinary concepts and, for interdisciplinary items that span discipline groups, one disciplinary concept will be emphasized within the item to ensure the response to the item clearly provides evidence of one disciplinary group for the purpose of reporting.

NAEP also reports percentages of students who attain each of the three achievement levels: *NAEP Basic*, *NAEP Proficient*, and *NAEP Advanced*. Cut scores define the minimum score needed for placement within a NAEP achievement level; scores and associated achievement levels are reported at the aggregate level.

NAEP Achievement Level Descriptions

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). ALDs specific to the 2028 NAEP Science Framework are included in Appendix C. 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).

Exhibit 4.1. Generic Achievement Level Policy Definitions for NAEP

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

Reporting 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.

ALDs specific to the 2028 NAEP Science Framework describe in greater detail the performance expected at each achievement level. These should be used to guide item development. Exhibit 4.2 provides sample excerpts of the descriptions for grade 4 included in the Framework. These excerpts demonstrate the overall differences between levels.

Exhibit 4.2. Excerpts of Achievement Level Descriptions for NAEP Science at Grade 4

NAEP Basic	*NAEP Proficient*	*NAEP Advanced*
Students at this level should be able to demonstrate partial mastery and competency in making sense of common phenomena or designing solutions using science and engineering practices and/or crosscutting concepts together with disciplinary concepts such as: A. different types of matter (materials) have different properties, B. a force acting on an object at rest can move the object, C. water and light are needed for a plant's survival . . .	Students at this level should be able to demonstrate solid academic performance and competency in making sense of phenomena or designing solutions using science and engineering practices and/or crosscutting concepts together with disciplinary concepts such as: A. matter (materials) can be classified based on its properties, B. a change in motion requires unequal forces acting on an object, C. varying amounts of water and light may affect a plant's growth . . .	Students at this level should be able to demonstrate superior performance and competency in making sense of complex phenomena or designing solutions using science and engineering practices and/or crosscutting concepts together with disciplinary concepts such as: A. different types of matter (materials) have multiple different properties, and it is therefore necessary to consider pros and cons when selecting a material for a specific purpose, B. two objects can exert forces on each other and these forces contribute to motion, C. some animals obtain the matter they need for growth and survival from plants or from other animals . . .

Following the first administration of the science assessment based on the updated framework, **new cut scores** will need to be defined. The ALDs in Appendix A will provide the basis for initial stages of the standard setting process for the 2028 NAEP Science Assessment.

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 (National Assessment Governing Board, 2018). The Governing Board policy does not extend to creating ALDs for performance below the *NAEP Basic* level.

New Reporting ALDs will be created after the assessment has been administered to specify how students have demonstrated competency at each achievement level by examining the characteristics of the items that map to each achievement level.

Items Showcasing Achievement Level Descriptions

The following three examples showcase ways that items might be expected to perform at a particular achievement level and how particular variations of each item would perform at the other achievement levels. These items were selected to include different grade levels, disciplines of science, and achievement levels.

The first of these examples, Exhibit 4.3, is a grade4 Physical Science item expected to perform at the *NAEP Proficient* level. The item reflects the following physical sciences example in the proficient ALD: “describing patterns in data to support the claim that the motion of an object is related to its energy.” Note that the options are designed to present choices that target different combinations of ability with respect to the NAEP DC, SEP, and CCC. Students must both analyze the data *and* use a pattern to explain their analysis. The dimensions are being explicitly measured by including options that might seem correct to a student who is only using one of these dimensions and not the other. For example, a student who understands that the evidence supports the idea that the faster an object is moving, the more energy it has, but does not understand how to apply the NAEP CCC to that evidence, might select option C—this response does not use a pattern to explain the analysis. Students who use the DC and SEP only, and not the CCC, may select this option. The emphasis word *best* in the item stem allows the developer to elicit evidence of partial understandings while only having one correct answer. Although the item overall assesses a proficient level of understanding, the CCC is not assessed at a proficient level. An item that is designed to prompt students to use a pattern without explicitly telling the student to do so would more likely reach the proficient level of the CCC.

The context was selected to be familiar to a wide range of students because softball is a common activity that takes place at school. The phenomenon “A softball player notices that older players throw the ball faster” is an observation that a student is likely to make. Data are included to support this phenomenon and the use of a NAEP SEP and CCC. To ensure the plausibility of the data, the speeds used for each age in the table were gathered from a training program for softball pitchers (Pourciau, 2023). The names selected for the item are purposefully representative of names used across a range of cultural and ethnic groups and are likely to be recognized by users of other languages. The names were also selected to avoid similarity with other words in English (Solano-Flores et al., 2014)

An item at the *Basic* level with the same alignment might involve students in organizing data to reveal patterns related to the motion of an object and its energy. The item could present all of the data needed to describe the pattern, for example, by providing data on both speed and energy. The item might use a context that is less cognitively complex, for example, by using objects that provide the energy of motion themselves (such as a battery-powered toy) or by using a variable in the data that is more commonly associated with energy (such as rolling marbles from different heights on a ramp).

An item at the advanced level with the same alignment might involve students in analyzing patterns in data gathered for two different objects to explain the relationship between the motion of an object and its energy. The item could use a context that is less familiar and more complex—for example, spacecraft designed to reach different parts of the solar system.

Exhibit 4.3. Softball

Item ID: Softball

Shakira plays softball at her school. She notices that some players can throw the ball faster than others. The table shows how fast Shakira and three other players at her school throw a softball.

Name	Age (years)	Speed of throw (meters per second)
Shakira	13	20
Rosa	12	18
Alissa	11	17
Yuki	10	16

Which statement **best** explains a pattern in the data?

A. The younger players threw a softball that had more mass than the softball the older players threw.
B. The softball had more energy when the older players threw it than when the younger players did.
C. Some players used more energy to throw the softball.
D. Softballs can have different masses.

Key: B

The item in Exhibit 4.4 is expected to perform at the *NAEP Advanced* level because it uses an unfamiliar context and includes multiple, dependent parts that require different combinations of the dimensions for a complete response. Further, the selected response item parts reflect advanced performance by requiring students to select the correct answers among plausible, attractive distractors. This grade 8 item reflects the following Earth and Space Sciences example in the advanced ALD "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."

The context draws upon Pueblo Indian culture from the Southwest United States, where people planned and built structures oriented to the patterns of the apparent motions of the sun and moon. The context provides an example that supports an expanded understanding of how science is done and who contributes to science (U.S. Department of the Interior, 2015). The phenomenon is "On the first day of spring, the sun is visible through a structure called 'Casa Rinconada,' located in a pueblo in the Southwest United States."

An item with a similar alignment but reflecting the *NAEP Proficient* level might involve students developing a model to test ideas about observable patterns in the motion of objects in the sky relative to Earth. The item could use the familiar example of the area around a school building while adding complexity relative to the basic example by requiring students to develop a model based on the context instead of presenting the model in the item.

An item with a similar alignment but reflecting the *NAEP Basic* level might involve students using a model to describe observable patterns in the motion of objects in the sky relative to Earth. The item could use a more familiar and less complex context, such as a model of the area around a school building. Using the same scenario, a basic item could be an entry point into an item set. For example, a basic item could ask students to use the model with the wire and foam ball to describe the sun's motion across the sky.

An item with a similar alignment but reflecting performance below the *NAEP Basic* level might involve students selecting from a set of images of models. This item might need less information in the stimulus. To ensure alignment, while providing access to students performing at this level, the item could use additional parts that target different dimensions; some of the parts might be accessible to students scoring around the level of the first cut score, while other parts might be slightly more difficult.

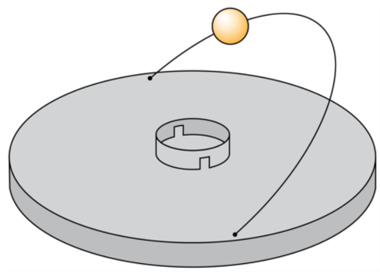
Exhibit 4.4. Spring Sunlight

Item ID: Spring Sunlight

The picture shows the sun on the first day of spring. The sun is visible through a structure called "Casa Rinconada," located in a pueblo in the Southwest United States.[16]

[16] Source: National Parks Service

The picture shows a model two students plan to use to predict when the sun will be visible through the structure.



The model has the following characteristics:

- A small copy of the structure is in the center of the model.
- A curved wire is placed into holes near the edge of the model.
- The wire is attached to the model at an angle.
- A foam ball moves along the wire.

Part A

Which statement **best** describes how the students could modify the model to improve their predictions?

A. Add another wire below the model so that the foam ball can be moved in a complete circle.
B. Add holes so that the foam ball can be moved between different locations on the edge of the model.
C. Change the angle of the wire so the foam ball can be placed above the structure in the middle of the model.
D. Change the length of the wire so the foam ball can be moved to different distances from the structure.

Key: B

Part B

Which statement describes a pattern in the motion of the sun that the model would **most likely** show with the modification identified in Part A?

A. The sun is visible only during the day.
B. The sun always moves in the same direction.
C. The sun is directly above the structure once each day.
D. The sun rises in different locations in different seasons.

Key: D

Exhibit 4.5 presents a grade 12 Life Science item that is expected to perform at the *NAEP Basic* level because it uses a familiar context, albinism, and a relatively low complexity relationship between DNA and a lack of melanin. The item reflects the following parts of examples in the basic ALDs: “using evidence to support an argument about a proposed explanation …” and “… genes code for the formation of proteins that determine traits.”

An item with a similar alignment but reflecting the *NAEP Proficient* level might involve students using patterns in DNA sequences to construct an argument about how DNA codes for the formation of proteins that determine traits. A proficient item could require a more difficult and complex analysis of the evidence, constructing an argument, and an understanding of a more complex relationship between DNA sequences and the formation of proteins.

An item with a similar alignment but reflecting the *NAEP Advanced* level might involve students in 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. An advanced item could require an unfamiliar function of DNA, revising an argument, and interpretation of more complex patterns of changes to genes and proteins.

Exhibit 4.5. Albino Mice

Item ID: Albino Mice

> Melanin is a type of protein that determines the color of the skin and hair of mice and other animals. Albino mice have no melanin and therefore, their skin and hair have no color.
>
> Which statement **best** describes evidence to support the argument that albinism is caused by a change to a gene in mice?
>
> A. The offspring of a tan mouse and an albino mouse are tan.
> B. The offspring of two albino mice are missing cells that make melanin.
> C. The offspring of a tan mouse and an albino mouse all have some melanin.
> D. The offspring of two albino mice have DNA that is unable to code for melanin.
>
> Key: D

Contextual Variables

NAEP legislation[17] requires reporting according to various student populations (303[b][2][G]), including:

- gender,
- race/ethnicity,
- eligibility for free/reduced-price lunch,
- students with disabilities, and
- English 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 also 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 and inadequate educational opportunities.

The *Standards for Educational and Psychological Testing* (American Educational Research Association et al., 2014) recommend that reports of group differences in assessment performance

[17] National Assessment of Educational Progress Authorization Act, Pub. L. 107-279, 3 U.S.C. §§ 301–305 (2002)

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.

Exhibit 4.6 summarizes the important components of NAEP reporting.

Exhibit 4.6. Components of NAEP Reporting

Component	Key Characteristics

How information is reported	• Results are published mainly online with an interactive report card
	• Interactive data tools include:
	    ○ performance associated with all collected contextual variables,
	    ○ profiles of states and TUDA districts,
	    ○ item maps, and
	    ○ sample questions.

What is reported	• Percentage of students by achievement level and scale score • Sample responses to illustrate achievement level definitions • Performance of various subgroups at the national level • Contextual information from NAEP questionnaires

Science-Specific Contextual Variables

Research 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.

Two NRC reports have summarized much of what is known in this domain, namely *Science and Engineering for Grades 6–12* (National Academies of Sciences, Engineering, and Medicine, 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. These and other studies, including the *Framework for K-12 Science Education* (National Research Council, 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. NAEP science assessment, along with contextual information about the experiences of the tested students is intended to measure how well that need is being met. Contextual information is critical to interpreting its results.

Research has informed an expanded view of the factors that shape opportunities to learn (Anderson et al., 2018; National Research Council, 2015; Penuel et al., 2015) 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—human, material, and social resources that shape student access to science (Brown, 2019; National Research Council, 2015).

Studies have demonstrated that what students learn is shaped by the availability of various science programs, curricula, extracurricular activities geared toward science, proximity to a science museum or a science and technology center, 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 (Anderson et al., 2018; National Research Council, 2015; Penuel et al., 2015). Teachers' and administrators' beliefs about what science is, how one learns science, and who can learn science also affect student learning (Anderson et al., 2018; Brown, 2019). 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 (Brown, 2019). Steele et al. (2002) 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. Student questionnaires have a strict time limit of 15 minutes. There are also limitations on the content of the questions that can be included on the student questionnaire. Questions about some factors may not be appropriate in the NAEP context. The NAEP legislation prohibits the collection of information on personal or family beliefs and attitudes and specifies that only information directly related to the appraisal of academic achievement may be collected. Given the constraints, not all topics can be addressed.

To support prioritization and ensure that NAEP results have appropriate context for interpretation, the 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 SEPs and NAEP CCCs. Interpreting students' achievement requires a basic understanding of what NAEP DCs, SEPs, and CCCs students have engaged with. Given variation across states in standards and frameworks, this information is crucial.
- **Teacher factors.** Research demonstrates that teacher quality is the most important in-school factor in predicting student achievement. This framework prioritizes the collection of data on teacher preparation and professional development.
- **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 school climate, instructional leadership, additional instructional personnel, time, technology, curriculum, and materials. The 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 emphasize 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.

Conclusion

NAEP Science Assessment results provide periodic snapshots of science achievement across large groups of students. This performance is specific to a large-scale assessment developed for the entire nation and is moderated by the contexts in which students learn and demonstrate ability in science. NAEP Science Assessment results are most appropriately used as one element within a range of data sources at multiple levels of the nation's educational system.

APPENDIX A: Sample Items and Scoring Notes

Appendix A presents sample items, including those included in the body of the Specifications and additional items, along with complete metadata that describe the grade, discipline, item type, alignment, item-level claim, phenomenon and engineering design problem, and complexity. For some items, multiple versions are presented to illustrate how modifications can alter alignment and complexity.

Discrete Items: Single and Multipart

The following items are discrete items, meaning that they produce a single score that would be aggregated with the scores of all the items on the assessment for a total test score. Many of the discrete items involve multiple parts to illustrate how to capture more evidence of student understanding, which would be scored together to produce a single score for the overall item. Many of these items could be limited to the first part to illustrate a single part discrete item, which test developers may determine would provide sufficient evidence of student understanding.

Physical Science

Exhibit A.1 presents 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 NAEP DC and SEP is highly cued through the provided answer choices. Note that selecting "was higher" would make each sentence true but would not support the claim.

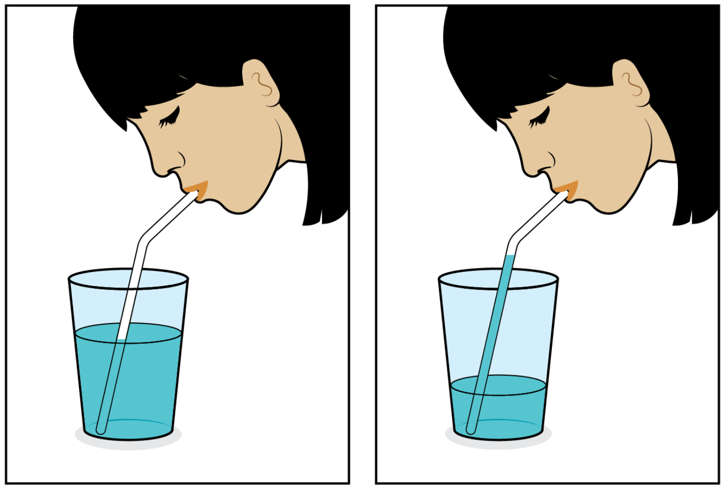
Exhibit A.1. Drinking Water

Item ID: Drinking Water
Grade and discipline: Grade 4, Physical Science
Item type: Inline 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: Water moves when a student sucks on a straw and stays still when the student does not suck on the straw.
Complexity: This item is an example of a low DC, low SEP complexity item.

A student is using a straw to drink water from a cup, as shown in the pictures.



The student notices that the water in the cup remains still until she starts drinking. When she is drinking, the water moves up through the straw to her mouth and the level of the water in the cup goes down.

The student makes a claim that she applies a force to the water only while she is drinking the water. Complete the sentence by using the dropdown menus to select the evidence that supports the student's claim.

Before the student started drinking, the water in the cup (was higher/remained still). When the student was drinking, the water in the straw (was higher/started moving).

Key: remained still; started moving

Exhibit A.2 presents students with a simple, authentic phenomenon centered on 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 true statements, reducing cognitive load of needing to check them for accuracy.

The item requires a three-dimensional performance: students have to apply their understanding of chemical reactions to analyze data, while looking for patterns among the specific characteristic properties that will indicate that a chemical reaction has occurred. This performance requires that students use their understanding of the NAEP DC, use the NAEP SEP to analyze data, and use the NAEP CCC to look for patterns to figure out whether a chemical reaction has occurred—relatively simple sensemaking with multiple dimensions.

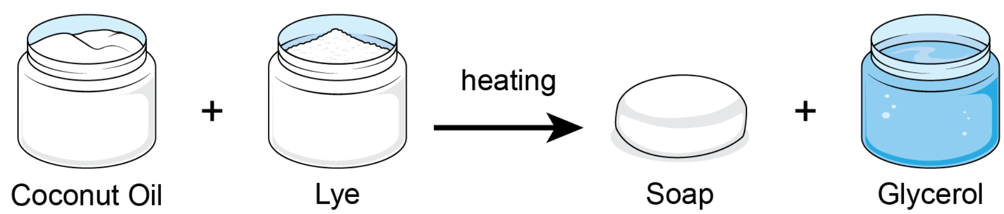
Exhibit A.2. Making Soap

Item ID: Making Soap (adapted from Next Generation Science Assessments)
Grade and discipline: Grade 8, Physical Science
Item type: Single-selection 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: S8.9: 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: Heating a combination of coconut oil and lye produces soap.
Complexity: This item is an example of a low DC, low SEP, low CCC complexity item.

One way to make soap is to heat a combination of coconut oil and lye. The diagram shows a simplified model of the soapmaking process.



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

Properties of Soapmaking Substances

Substance	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 from the density of glycerol.
C. The total mass of soap and glycerol is the same as the total mass of coconut oil and lye.
D. The melting points of soap and glycerol are different from the melting points of coconut oil and lye.

Key: D

Exhibit A.3 presents students with a simple, authentic phenomenon, with images to make the process more accessible to a wide range of learners. Students' use of the NAEP SEP and DC is highly directed, resulting in a very low complexity item. This item can be easily modified to engage the NAEP SEP of Developing and Using Models 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 A.3. Rusting Nails

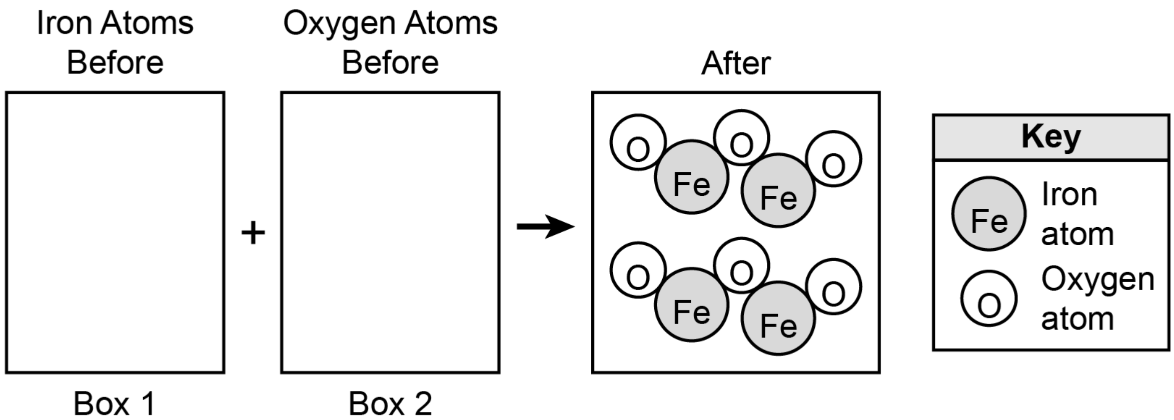
Item ID: Rusting Nails (adapted from M-STEP Grade 8 Science Sample Items and Alabama Math, Science, and Technology Initiative)
Grade and discipline: Grade 8, Physical Science
Item type: Multipart, single-selection multiple choice
Alignment: This item is a **2D item**, measuring parts of:

- 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: S8.15: 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: When iron nails are left outside, they rust. Students are asked to determine how the nails became rusty.
Complexity: This item is an example of a low DC, low SEP complexity item.

When iron nails are left outside, the iron (Fe) in the nails reacts with oxygen (O_2) in the air to produce rust (Fe_2O_3). The pictures show the nails before and after the reaction.

A partially complete model of the chemical reaction between iron and oxygen is shown.



Part A

Based on the information shown in the model of the chemical reaction, which number of iron atoms should be placed in Box 1?

A. 1
B. 2
C. 4
D. 6

Key: C

Part B

Based on the information shown in the model of the chemical reaction, which number of oxygen atoms should be placed in Box 2?

A. 1
B. 2
C. 4
D. 6

Key: D

Exhibit A.4 provides an example of a discrete, extended constructed response item that assesses all three dimensions at grade 12. In this item, students construct an argument for why the temperature of a system (phenomenon) does not change during a phase change.

Exhibit A.4. Melting Ice

Item ID: Melting Ice (adapted from CREATE for STEM, Michigan State University)
Grade and discipline: Grade 12, Physical Science
Item type: Extended constructed response
Alignment: This item is a **3D item**, measuring parts of:

- DC: P12.13: Energy is a quantitative property 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.
- 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: 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.

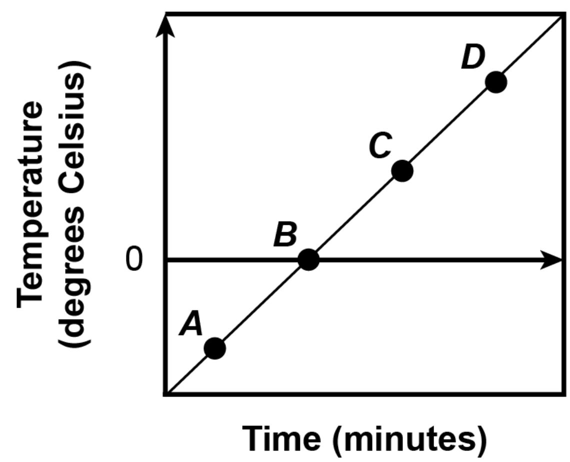
Item-level claim (derived from targeted dimensions): Students can construct an argument for why the temperature of a system (phenomenon) does not change during a phase change.
Phenomenon: The temperature of the system doesn't continue changing during a phase change. Students are asked to predict this phenomenon and support their prediction with an argument.
Complexity: This item is an example of a medium DC, medium SEP, low CCC complexity item, as the item cues students to demonstrate their understanding of the molecular processes underlying phase changes in a nonroutine example of an incorrect prediction, using a NAEP SEP and CCC to do so.

When ice cubes are left out in the sun in hot weather, they melt. The table shows the observations a student made while watching ice cubes melt. The student created the graph to show how he expected the temperature of the ice cubes to change while they melted.

Student's Observations

Time Segment	Observation
A–B	Solid ice
B–C	Mixture of ice and water
C–D	Liquid water

Student's Predicted Temperature Changes



Use your understanding of what happens at the molecular level during a phase change to make an argument that supports or refutes the student's temperature change prediction. In your argument, be sure to state whether you agree or disagree with the student's prediction and describe what happens during each time segment (A–B, B–C, C–D), including the following:

- energy flows
- phases of matter
- molecular motion
- forces between molecules
- changes in temperature

Exhibit A.5. Melting Ice Constructed Response Scoring Notes

Student responses should indicate that the linear prediction of temperature increase over time does not consider the energy required for phase changes (melting) at the molecular level. Student responses should show that they understand that temperature is a measurement of energy—as temperature of a system rises, this leads to more molecular motion. During these phase changes, energy is absorbed to break intermolecular forces and transition between phases, resulting in a plateau in the temperature-time graph—the temperature does not increase because the additional energy is being used to overcome the intermolecular forces.

Exhibit A.6 illustrates a low complexity grade 12 item with low DC complexity, low SEP complexity, and low CCC complexity. While students cannot respond to this item without bringing some understanding of the DC, SEP, and CCC, it is heavily supported, thus limiting the amount of sensemaking students engage in with any dimensions. Students are given the relationship they need to map, only relevant statements to move, and a very structured and nonquantitative model illustrating feedback loops. Note that this item leans into the intentional progressions built into the dimensions: the SEP and CCC are appropriate for grade 12 at a low level, and they are related to expectations and performances at lower grade levels. This allows a way to reduce the complexity of the item, remain consistent with the grade-level targets, and account for the fact that lower-performing students may have a less sophisticated, less well-developed understanding of the targeted dimensions that more closely approximates sophisticated performances at lower grade bands.

Exhibit A.6. Permafrost, Version 1

Item ID: Permafrost Melting (adapted from OpenSciEd)
Grade and discipline: Grade 12, Physical Science
Item type: Matching
Alignment: This item is a **3D item**, measuring parts of:

- DC: P12.14: When sunlight is absorbed at Earth's surface, it is eventually reradiated 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: S12.16: 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.
- CCC: C12.13: Feedback mechanisms within a system are important elements for explaining or designing for either the stability or instability of the system.

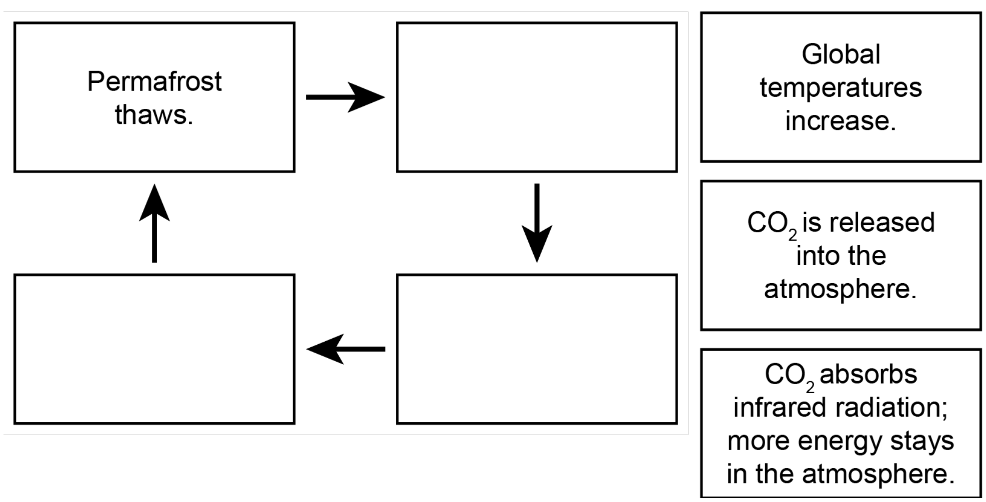
Item-level claim (derived from targeted dimensions): Students can complete a model explaining the causal chain of events that lead to a phenomenon using their understanding of how the release of greenhouse gasses leads to increased temperatures on Earth.
Phenomenon: Permafrost is melting, and that melting seems to be associated with increased global temperatures.
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. An example of permafrost is shown in the picture.

Permafrost[18]

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.

Use the statements to develop a model that shows the relationships between thawing permafrost and rising global temperatures. Drag the statements into the correct boxes to complete the model. Each statement will be used once.



Key: Permafrost thaws. CO_2 is released into the atmosphere. CO_2 absorbs infrared radiation; more energy stays in the atmosphere. Global temperatures increase.

The following two versions of the item are modified to be higher complexity by (a) requiring students to develop the model with significantly less support (Exhibit A.7), and/or (b) asking students to consider implication and limitations of the model (Exhibit A.8). By making these modifications, students are more independently responsible for figuring out why permafrost

[18] Permafrost from OpenSciEd, CC BY-SA 4.0

melting contributes to rising global temperatures by using their understanding of modeling, feedback loops, and energy in systems. In these variations, the complexity has been increased, but there is still substantial cueing in the form of the provided statements. A more complex item might ask students to develop an original model, without any prompting statements.

Exhibit A.7. Permafrost, Version 2

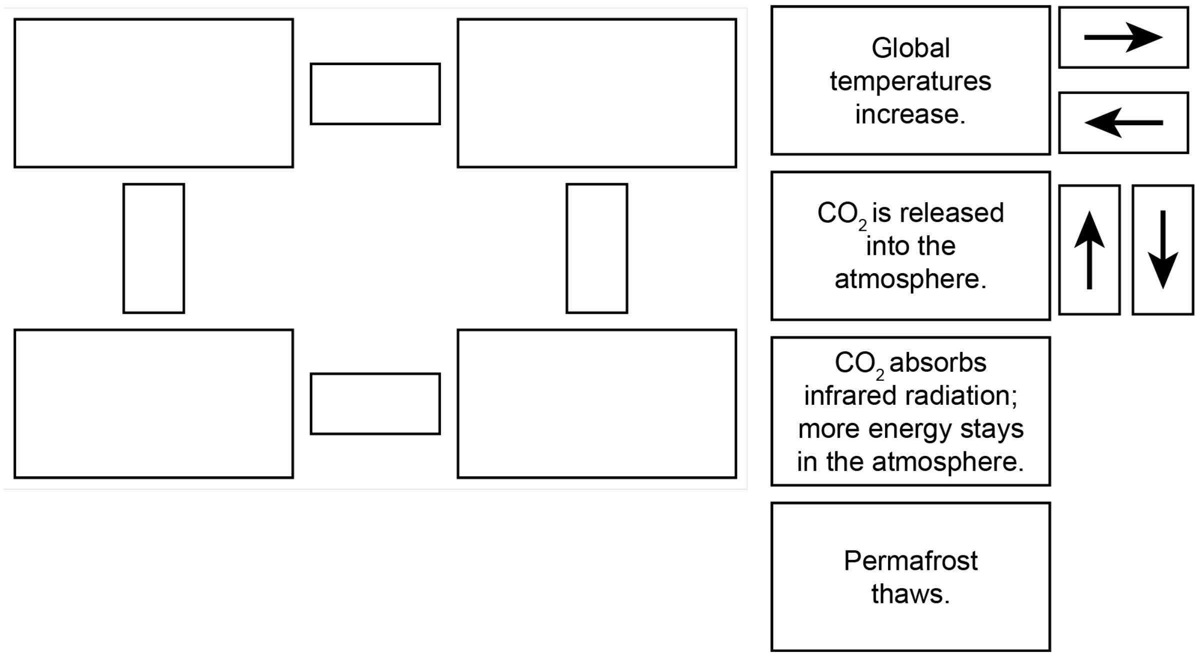
Permafrost is a layer of soil and ice that is just below the surface in the Arctic. An example of permafrost is shown in the picture.

Permafrost[19]

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.

Use the statements and arrows to develop a model that shows the relationships between thawing permafrost and rising global temperatures. Drag the statements and arrows into the correct boxes to develop the model. Each statement and each arrow will be used once.

[19] Permafrost from OpenSciEd, CC BY-SA 4.0



Key: Global temperatures increase. → Permafrost thaws. → CO_2 is released into the atmosphere. → CO_2 absorbs infrared radiation; more energy stays in the atmosphere. → (Responses can start with any of the draggers.)

Exhibit A.8. Permafrost, Version 3

Permafrost is a layer of soil and ice that is just below the surface in the Arctic. An example of permafrost is shown in the picture.

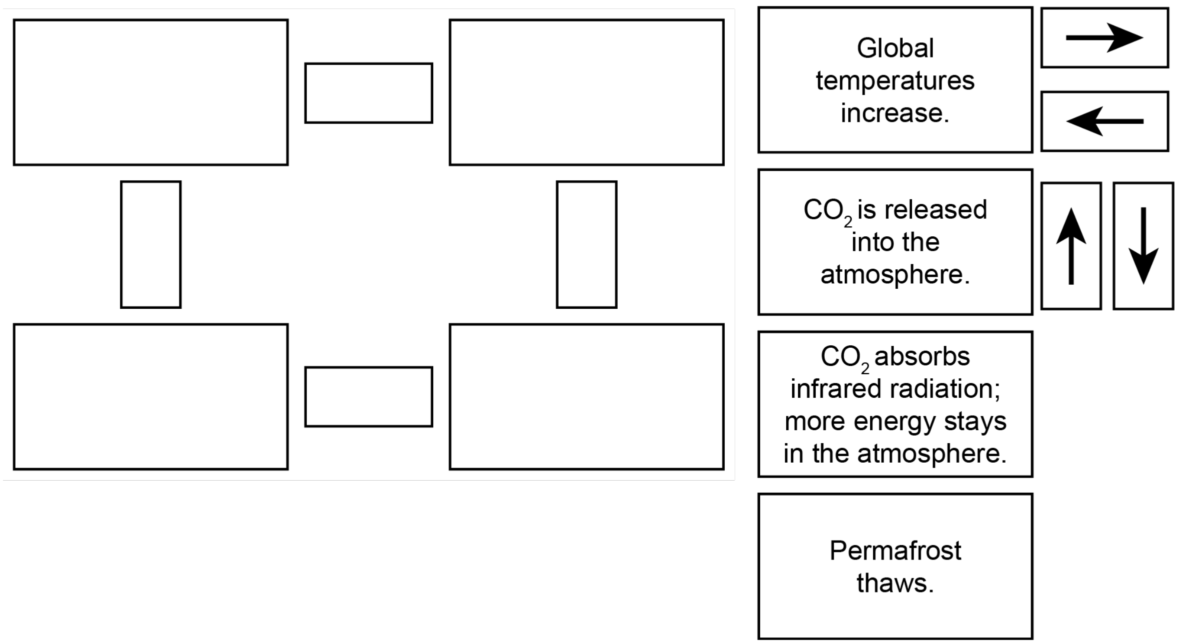
Permafrost[20]

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.

[20] Permafrost from OpenSciEd, CC BY-SA 4.0

Part A

Use the statements to develop a model that shows the relationships between thawing permafrost and rising global temperatures. Drag the statements and arrows into the correct boxes to complete the model. Each statement and arrow will be used once.



Key: Global temperatures increase. → Permafrost thaws. → CO_2 is released into the atmosphere. → CO_2 absorbs infrared radiation; more energy stays in the atmosphere. → (Responses can start with any of the draggers.)

Part B

Complete the sentence by choosing the correct answer from the drop-down menu.

The model predicts that the rate of melting of the permafrost will likely (increase / stay the same / decrease) over the next 50 years.

Key: increase

Part C

Describe a limitation of the model you developed in Part A.

Exhibit A.9. Permafrost Part C Constructed Response Scoring Notes

- Students provide one limitation of the model, for example:
    - The model is missing factors such as interactions with other systems and feedback loops.
    - Thawing may occur at different rates in different locations around the world.
    - There may be a time lag between temperature increases and thawing.

Exhibit A.10 presents a grade 4 Physical Science item expected to perform at the *NAEP Proficient* level. The item reflects the following physical science example in the proficient ALD "describing patterns in data to support the claim that the motion of an object is related to its energy." Note that the options are designed to present choices that target different combinations of ability with respect to the NAEP DC, SEP, and CCC. Students must both analyze the data *and* use a pattern to explain their analysis. The dimensions are being explicitly measured by including options that might seem correct to a student who is only using one of these dimensions and not the other. For example, a student who understands that the evidence supports the idea that the faster an object is moving, the more energy it has, but does not understand how to apply the NAEP CCC to that evidence, might select option C—this response does not use a pattern to explain the analysis. Students who use the DC and SEP only, and not the CCC, may select this option. The emphasis word *best* in the item stem allows the developer to elicit evidence of partial understandings while only having one correct answer. Although the item overall assesses a proficient level of understanding, the CCC is not assessed at a proficient level. An item that is designed to prompt students to use a pattern without explicitly telling the student to do so would more likely reach the proficient level of the CCC.

The context was selected to be familiar to a wide range of students because softball is a common activity that takes place at school. The phenomenon, "A softball player notices that older players throw the ball faster" is an observation that a student is likely to make. Data are included to support this phenomenon and the use of a NAEP SEP and CCC. To ensure the plausibility of the data, the speeds used for each age in the table were gathered from a training program for softball pitchers (Pourciau, 2023). The names selected for the item are purposefully representative of names used across a range of cultural and ethnic groups and likely to be recognized by users of other languages. The names were also selected to avoid similarity with other words in English (Solano-Flores et al., 2014)

An item at the *Basic* level with the same alignment might involve students in organizing data to reveal patterns related to the motion of an object and its energy. The item could present all of the data needed to describe the pattern, for example, by providing data on both speed and energy. The item might use a context that is less cognitively complex, for example, by using objects that provide the energy of motion themselves (such as a battery-powered toy) or by using a variable in the data that is more commonly associated with energy (such as rolling marbles from different heights on a ramp).

An item at the advanced level with the same alignment might involve students in analyzing patterns in data gathered for two different objects to explain the relationship between the motion of an object and its energy. The item could use a context that is less familiar and more complex—for example, spacecraft designed to reach different parts of the solar system.

Exhibit A.10. Softball

Item ID: Softball
Grade and discipline: Grade 4 Physical Science
Item type: Single-selection multiple choice
Alignment: This item is a **3D item**, measuring parts of:

- DC: P4.7: Objects in motion have energy. The faster a given object is moving, the more energy it has.
- SEP: S4.10: Analyze data to determine whether it supports or refutes a claim about a phenomenon or design solution.
- CCC: 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.

Phenomenon: A softball player notices that older players throw the ball faster.

Shakira plays softball at her school. She notices that some players can throw the ball faster than others. The table shows how fast Shakira and three other players at her school throw a softball.

Name	Age (years)	Speed of throw (meters per second)
Shakira	13	20
Rosa	12	18
Alissa	11	17
Yuki	10	16

Which statement **best** explains a pattern in the data?

A. The younger players threw a softball that had more mass than the softball the older players threw.
B. The softball had more energy when the older players threw it than when the younger players did.
C. Some players used more energy to throw the softball.
D. Softballs can have different masses.

Key: B

Life Science

In Exhibit A.11, an example of a two-dimensional discrete item, students have to apply their understanding of what plants need to grow to make a prediction. Students do not need to understand a specific NAEP CCC element to respond to the item—the NAEP SEP and DC are sufficient to respond to this question. While the item is an implicit example of the NAEP CCC Mechanisms and Explanation: Cause and Effect, students do not need to explicitly bring an understanding of cause-and-effect relationships to respond, and therefore, the item is not considered three-dimensional. This item provides an example of lower-level sensemaking with the NAEP DC and SEP—while students do (a) need to understand that plants need water and air to grow and (b) need to be able to use this information to evaluate a phenomenon across multiple (experimental) conditions, they are very closely applying simple grade-appropriate NAEP DCs and SEPs. The NAEP SEP is engaged in service of surfacing NAEP DC understanding rather than expanding the nature of how students explain the phenomenon. This level of sensemaking would be appropriate to surface understanding from students who have had the opportunity to begin developing an understanding of the grade 4 NAEP DC and SEP.

The 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 a 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 understand what plants need to grow to make a reasoned prediction, but both the NAEP SEP and DC are highly guided.

Exhibit A.11. Plant Growth, Version 1

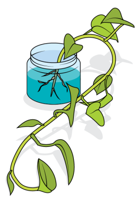
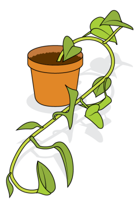
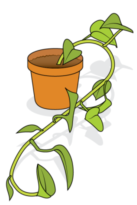
Item ID: Plant Growth (adapted from the Next Generation Science Assessment Project)
Grade and discipline: Grade 4, Life Science
Item type: Multipart item, single-selection multiple choice and short constructed response
Could part A be used as a stand-alone item? Yes (with the possible addition of one 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: S4.8: Predict the outcome of an experiment or a design solution based on a model, a phenomenon, or a design plan.

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: Soil and water conditions affect plant growth.
Complexity: This item is an example of a low DC, low SEP complexity.

The plants shown were placed in a classroom on the same day. They are all the same kind of plant. The plants were placed on the same side of the room near a window, so they receive the same amount of light each day. Students in the class want to find out what the plants need the most in order to grow. They grow the plants using the conditions shown in the table.

Conditions for Growing Plants

Plant	Planted in Soil	Water
 Plant A	No	Water added regularly for one month
 Plant B	Yes	Water added regularly for one month
 Plant C	Yes	No water added

Part A

Which plant will likely grow the **<u>least</u>** over the next month?

A. Plant A
B. Plant B
C. Plant C

Key: C

Part B

Provide one reason the plant you chose in Part A will grow the least over the next month.

Exhibit A.12. Plant Growth Part B Constructed Response Scoring Notes

- Reasons students provide should leverage understanding of what plants need to grow (water, air, minerals from soil)
- Note that while a complete answer might include comparisons among plants (e.g., Plant A and B have X, but Plant C does not), this is not a requirement.
- Possible reasons include:
  - Plant C does not get water
  - Plant C does not get minerals
  - Plant C does not get water or minerals
- Students should receive credit as long as their reason supports their choice, with an accurate understanding of plant needs for growth.

The following two additional versions of the item demonstrate how the item can be modified to increase the complexity of the item with respect to specific dimensions and overall. The version of the item in Exhibit A.13 focuses on a central component of the NAEP DC—that plants need water to survive. It does not provide evidence of the full NAEP DC, which includes a requirement for minerals from soil. The following example shows a slight modification that more completely elicits understanding of the NAEP DC. It should be noted that this item is largely at the same complexity level as the previous item but because it involves an additional component of disciplinary understanding, may be more challenging for some students.

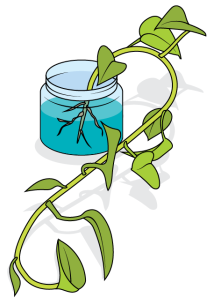
Exhibit A.13. Plant Growth, Version 2

Part A

The plants shown were placed in a classroom on the same day. They are all the same kind of plant. The plants were placed on the same side of the room near a window, so they receive the same amount of light each day. Students in the class decide to give different amounts of soil and water to each plant to find out what plants need the most in order to grow. The soil has minerals in it.

Label two of the plants to show which plant will grow the most and which plant will grow the least. Drag the words into the correct boxes. Each word will be used once.

most | least

		
Plant A: Plant in jar with water and no soil	Plant B: Plant in pot with soil and water	Plant C: Plant in pot with soil and no water

Key: Plant B most; Plant C least

Part B

Provide one reason that explains why the labels you chose for the two plants in Part A are correct.

Exhibit A.14. Plant Growth Part B Constructed Response Scoring Notes

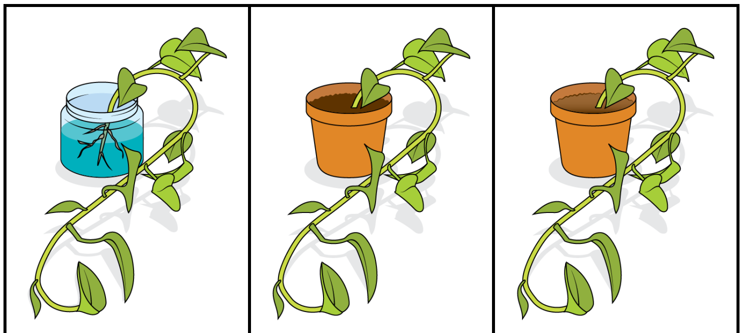
- Student reasoning should connect the needs of plants (water, air, and minerals from soil) to their relative growth. This may include:
  - Plants need water and minerals from soil, so the plant with access to both will grow the best.
  - Without water, plants die, so the plant with no water will grow the least.
  - Even though plants need minerals and water to grow, without water, plants die, so the plant without water will grow the least.

Because this phenomenon is so simple and intrinsically connected to the NAEP DC, it can be challenging to modify this item to require significantly more complex sensemaking with the NAEP DC. However, the item could be modified to engage the NAEP SEP at a higher degree of complexity, as well as the NAEP CCC. The following single-selection multiple choice example illustrates this, leveraging the same context and phenomenon, with additional SEP 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 NAEP SEP element: S4.10: Analyze data to determine whether it supports or refutes a claim about a phenomenon or design solution, and
- NAEP 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 A.15. Plant Growth, Version 3

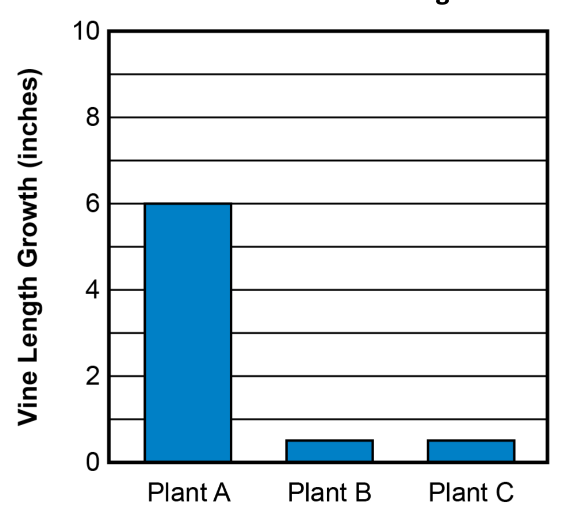
The plants shown were placed in a classroom on the same day. They are all the same kind of plant. The plants were placed on the same side of the room near a window, so they receive the same amount of light each day. Students in the class decide to give different amounts of soil and water to each plant to find out what plants need the most in order to grow.



Plant A: Plant in jar with water and no soil	Plant B: Plant in pot with soil and water	Plant C: Plant in pot with soil and no water

The bar graph shows how much Plant A will likely grow over the next month. Complete the graph by dragging the bars to heights that show how much Plant B and Plant C will likely grow over the next month.

Plant Growth Under Different Growing Conditions



Key: Plant B > 6; Plant C < 6

Exhibit A.16 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 NAEP CCC element, in conjunction with the other two dimensions. The NAEP DC, SEP, and CCC are all very straightforward and clearly cued to students. In this example, Part A could be used as a stand-alone question.

Exhibit A.16. Wild Dogs

Item ID: Wild Dogs (adapted from OpenSciEd)
Grade and discipline: Grade 12, Life Science
Item type: Multipart item, single-selection 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 relationship 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: S12.9: 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.
- CCC: C12.1: 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.

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 and engineering design problem: Wild dog populations are decreasing across Africa. Students are asked to engage with problem-solving by determining if prey availability is 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.

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 area can support. Pictures of an African Wild Dog and its primary prey, the antelope, are shown.[21]

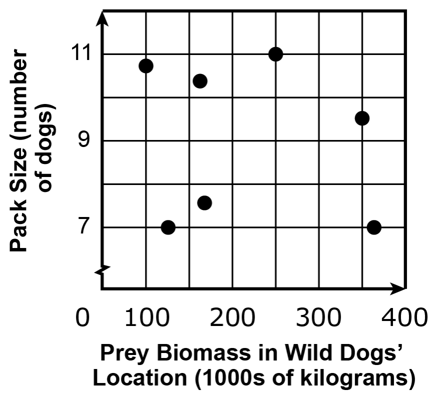
African Wild Dog

Antelope

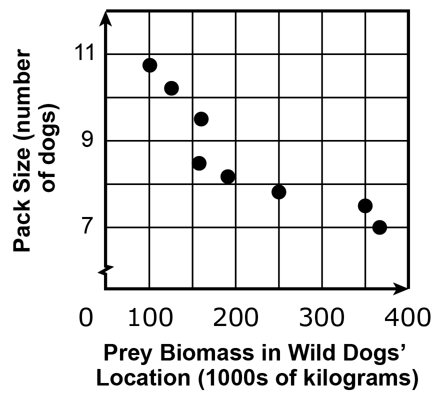
Part A

The amount of food available, or prey biomass, in the national park is a limiting factor in African wild dog population size. Which graph represents the relationship you would expect to see between prey biomass and population size for African wild dogs in the national park if prey biomass is a limiting factor?

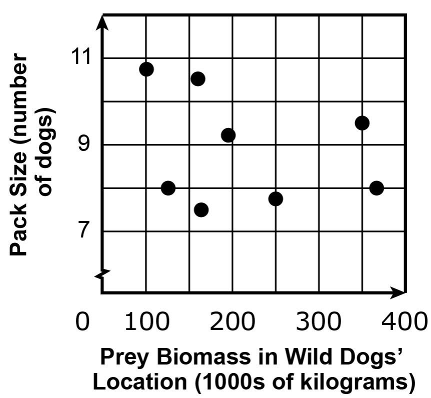
A.



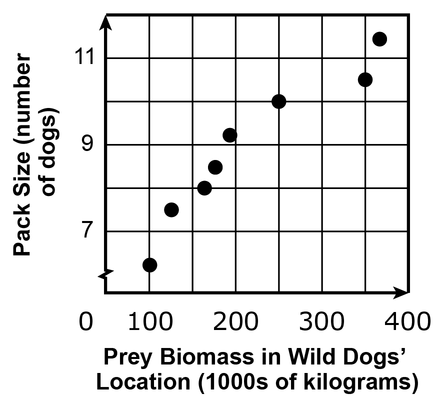
C.



B.



D.

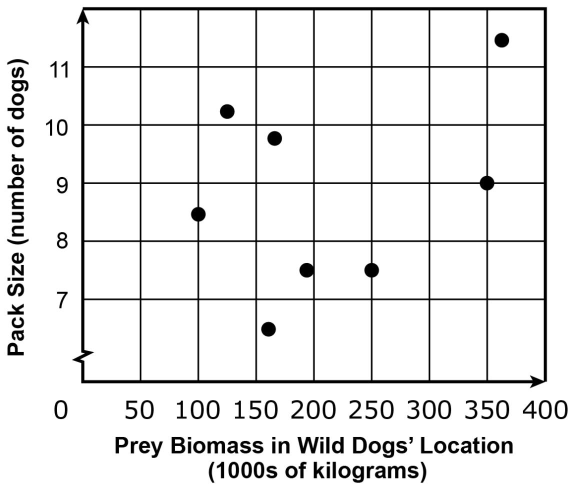


Key: D

[21] African Wild Dog by Derek Keats, CC BY 2.0, via Wikimedia Commons; Hein Waschefort, CC BY-SA 3.0, via Wikimedia Commons

Part B

The graph shows the actual relationship between African wild dog pack size and prey biomass that scientists observed in a similar territory. Based on the graph, complete the sentence to describe whether prey biomass plays an important role in determining the carrying capacity of African wild dogs in this location.



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.

Use the graph to support your answer:

Exhibit A.17. Wild Dogs Part B Constructed Response Scoring Notes

Student reasoning should indicate that they correctly interpret the graph as having no clear relationship between prey biomass available and pack size, indicating that prey biomass does not play a very important role in determining carrying capacity in this area.

- Some students may note that there is a slight trend toward a positive relationship between prey biomass and pack size. While the relationship is limited at best, some students may see it as enough to warrant noting. If students do note this relationship, they should be awarded credit as long as it is clear that they are reading and interpreting the graph correctly and using an accurate understanding of carrying capacity to make the connection between food availability and pack size.

The same context can support items that align with different SEPs and CCCs. For example, the short and extended constructed response items in Exhibit A.18 illustrate how this context can be used to elicit evidence of students' understanding of the DC, in addition to parts of a different SEP and a different CCC:

- NAEP 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.
- NAEP 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 A.18. Wild Dogs, Version 2

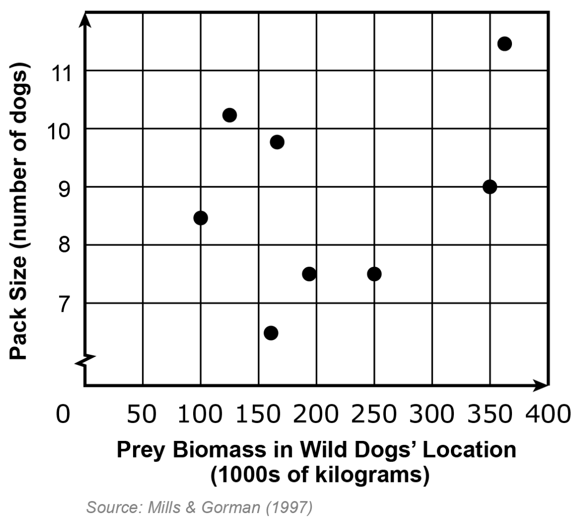
There has been a fast decline in African wild dog populations across Africa. Conservation experts are trying to establish a new population in a national park in Malawi and are trying to figure out how many wild dogs the area can support. Pictures of an African Wild Dog and its primary prey, the antelope, are shown.[22]

African Wild Dog

Antelope

The graph shows the actual relationship between African wild dog pack size and prey biomass that scientists observed in a similar territory. Based on the graph, complete the sentence to describe whether prey biomass plays an important role in determining the carrying capacity of African wild dogs in this location.

Graph 1. Comparison of African Wild Dog Pack Size to Prey Biomass



---

22 African Wild Dog by Derek Keats, CC BY 2.0, via Wikimedia Commons; Hein waschefort, CC BY-SA 3.0, via Wikimedia Commons

Part A

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 the answers to these questions will help you better understand how many wild dogs the national park location could support.

Exhibit A.19. Wild Dogs Constructed Response Scoring Notes

- Student questions and their justification should reflect (1) an understanding that multiple factors (including food availability, predator populations and needs, availability of habitats, mates, etc.) influence the carrying capacity of a region for a given species, and (2) that the data provided suggest that prey biomass is not the overwhelming driver of carrying capacity in this area.
- Student questions may focus on:
  - Better understanding wild dogs' prey/access to food (e.g., asking questions about other possible prey, seasonal availability, wild dogs' feeding processes)
  - Better understanding other factors that could impact carrying capacity (e.g., wild dog predators in the region, habitat availability/destruction/conservation, etc.)
  - Understanding complex ecosystem dynamics that might contribute to carrying capacity
  - Better understanding the data presented, to support more robust interpretation.

Exhibit A.20 requires sensemaking with all three dimensions and asks students to do a great deal of sensemaking (for 4th grade) independently. Apples are a very familiar fruit to nearly all regions and socioeconomic statuses (even internationally) and are an authentically important crop in many different geographies within the United States.

Exhibit A.20. Selecting Apples, Version 1

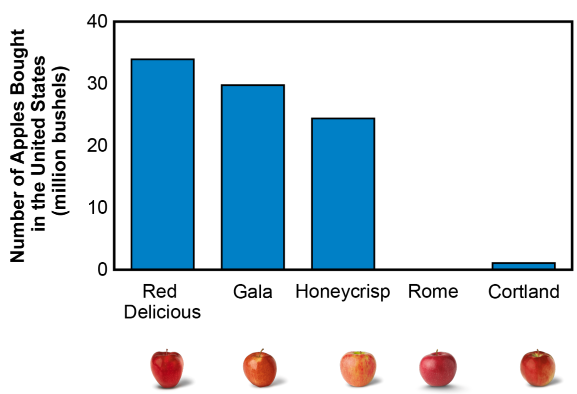
Item ID: Selecting Apples
Grade and discipline: Grade 4, Life Science
Item type: Single-selection multiple choice
Alignment: This item is a **3D item**, measuring parts of:

- DC: 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 both.
- SEP: S4.10: Analyze data to determine whether it supports or refutes a claim about a phenomenon or design solution.
- CCC: 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.

Item-level claim: Students can analyze patterns in simple data to determine which parents are most likely to produce offspring with desirable characteristics.
Complexity: High
Phenomenon and engineering design problem: Apples are a very important food for many people, but they are getting harder to grow. The most popular apple varieties do not grow in many temperatures, which is a problem as temperatures vary more and more in a given area. How can we create a new kind of apple that is both popular and can grow in many different temperatures?

Millions of people buy apples to eat every day. Apples are getting harder to grow because of extreme changes in temperature that are happening in many apple-growing areas. The graph shows the amounts of the five types of apples that people in the United States (U.S.) bought most often in one year. The table shows information about each of the types of apples shown in the graph.[23]

Apples People in U.S. Bought Most Often



Apple Facts

Type of apple	Characteristics	What temperature will the apples grow in?
Red Delicious	Sweet, crisp	Mild
Gala	Sweet, crisp	Mild
Honeycrisp	Sweet/tart, crisp	Hot, mild, cold
Rome	Tangy, dense/soft	Hot
Cortland	Tart, a little crisp	Hot, mild, cold

[23] Apple images from New York Apple Association, applesfromny.com

A farmer wants to use two of the apples in the graph and table as parents to produce a new apple that has these characteristics:

- It is an apple that people buy most often at stores in the U.S.
- The apple will grow in the widest range of temperatures.

Based on the information in the graph and the table, which two types of apples could be used as parents to produce a new apple that has these characteristics?

A. Red Delicious and Gala
B. Red Delicious and Rome
C. Rome and Cortland
D. Gala and Honeycrisp

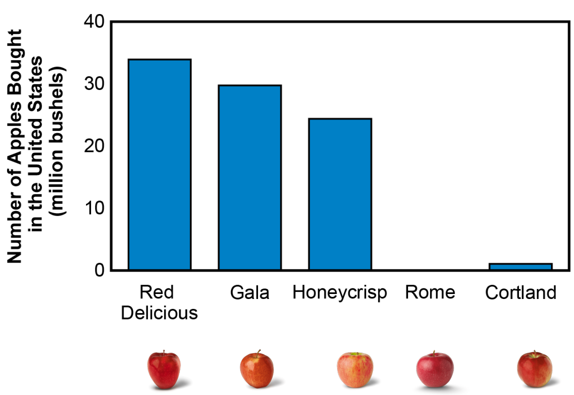
Key: D

The version of the item in Exhibit A.21 shows a more supported version that elicits evidence of the same targets, at a lower complexity.

Exhibit A.21. Selecting Apples, Version 2

Millions of people buy apples to eat every day. Apples are getting harder to grow because of extreme changes in temperature that are happening in many apple-growing areas. The graph shows the amounts of the five types of apples that people in the United States (U.S.) bought most often in one year. The table shows information about each of the types of apples shown in the graph.[24]

Apples People in U.S. Bought Most Often



Apple Facts

Type of apple	Characteristics	What temperature will the apples grow in?
Red Delicious	Sweet, crisp	Mild
Gala	Sweet, crisp	Mild
Honeycrisp	Sweet/tart, crisp	Hot, mild, cold
Rome	Tangy, dense/soft	Hot
Cortland	Tart, a little crisp	Hot, mild, cold

[24] Apple images from New York Apple Association, applesfromny.com

Part A

Based on the information in the graph and the table, which **two** types of apples are the **least likely** to grow in many different temperatures?

A. Rome and Cortland
B. Rome and Gala
C. Red Delicious and Cortland
D. Red Delicious and Gala

Key: D

Part B

A few months ago, apple farmers began selling a new type of apple. The farmers sold more of the new apple than they sold of the Honeycrisp and Cortland apples in the same amount of time. The new apple can be grown in hot and mild temperatures.

Which two types of apples are **most likely** the parents of the new type of apple?

A. Honeycrisp and Rome
B. Honeycrisp and Cortland
C. Red Delicious and Cortland
D. Red Delicious and Honeycrisp

Key: B

Part C

Use the information provided and your understanding of how traits in offspring relate to traits in their parents to support the claim that the two types of apples you chose in Part B are **most likely** to be the parents of the new types of apple.

Exhibit A.22. Selecting Apples Version 2 Constructed Response Scoring Notes

- Student responses should reflect reasoning that the new apple should have characteristics similar to those of its parents.
- Student responses should justify the selection of two apples that, together, match the characteristics of the new apple.
- For example, honeycrisp is popular and grows in many climates, and Cortland grows in many climates, so together they could be parents to a new apple that is both popular and can grow in many climates

Exhibit A.23 presents a grade 12 Life Science item that is expected to perform at the *NAEP Basic* level because it uses a familiar context, albinism, and a relatively low complexity relationship between DNA and a lack of melanin. The item reflects the following parts of examples in the basic ALDs: “using evidence to support an argument about a proposed explanation …” and “… genes code for the formation of proteins that determine traits.”

An item with a similar alignment but reflecting the *NAEP Proficient* level might involve students using patterns in DNA sequences to revise an argument about how DNA codes for the formation of proteins that determine traits. A proficient item could require a more difficult and complex analysis of the evidence and require an understanding of a more complex relationship between DNA sequences and the formation of proteins.

An item with a similar alignment but reflecting the *NAEP Advanced* level might involve students 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. An advanced item could require an unfamiliar function of DNA and require interpretation of more complex patterns of changes to genes and proteins.

Exhibit A.23. Albino Mice

Item ID: Albino Mice
Grade and discipline: Grade 12 Life Science
Item type: Single-selection multiple choice
Alignment: This item is a **2D item**, measuring parts of:

- DC: 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 structural functions, and some have no currently known function.
- 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.

Phenomenon: The skin and hair of albino mice have no color.

Melanin is a type of protein that determines the color of the skin and hair of mice and other animals. Albino mice have no melanin, and therefore, their skin and hair have no color. Which statement **best** describes evidence to support the argument that albinism is caused by a change to a gene in mice? A. The offspring of a tan mouse and an albino mouse are tan. B. The offspring of two albino mice are missing cells that make melanin. C. The offspring of a tan mouse and an albino mouse all have some melanin. D. The offspring of two albino mice have DNA that is unable to code for melanin. Key: D

Earth and Space Sciences

Exhibit A.24 illustrates a low-complexity grade 4 item assessing Earth and Space Sciences with low DC complexity and low SEP complexity. The item requires students to make sense of a simple phenomenon: a local park has flooded. This phenomenon is an example of an everyday phenomenon that many students may have directly experienced or have sufficient experiences to understand. The phenomenon is presented through simple text, an image, and a simple graph—this provides students with enough information to demonstrate the targeted NAEP DC and SEP in service of sensemaking, without unnecessary reading or cognitive load. This item elicits simple sensemaking because students must understand that flooding is most likely to occur on the day with the most rain, rather than just any day with rain. This requires interpreting the context as well as the data, albeit in a very straightforward way. In grade 4, students are still developing the concept of flooding and its relationship to rain; while this item elicits sensemaking based on developmentally appropriate and expected schema for 4th graders, this same performance would not be considered sensemaking in grades 8 and 12.

Exhibit A.24. Park Flooding, Version 1

Item ID: Park Flooding (adapted from Formative Assessment Bundling Literacy and Elementary Science)
Grade and discipline: Grade 4, Earth and Space Sciences
Item type: Single-selection 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: S4.10: Analyze data to determine whether it supports or refutes a claim about a phenomenon or design solution.

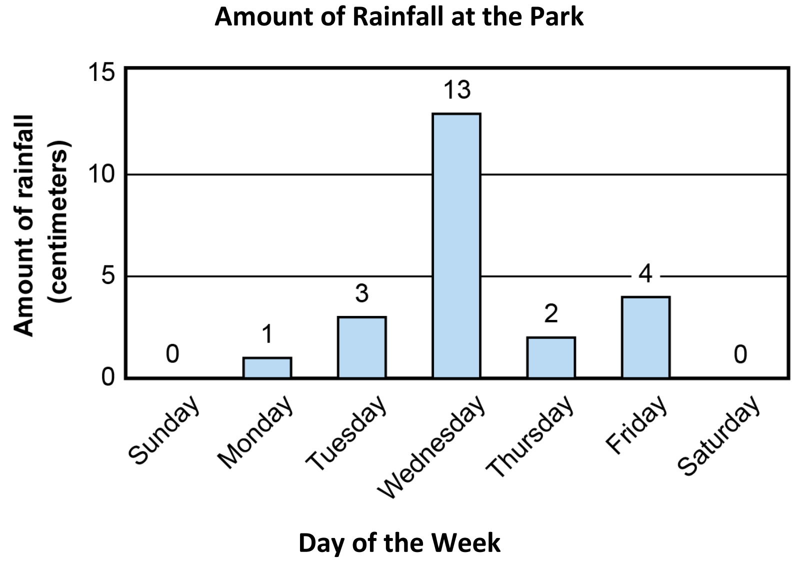
Item-level claim (derived from targeted dimensions): Students can analyze and interpret simple data to determine patterns in weather conditions to make a claim about a phenomenon.
Phenomenon: A park flooded when it was raining one day but not other days.
Complexity: This item is an example of a low DC, low SEP complexity item.

People visiting a local park noticed that the park was flooded and was closed for the day. The picture shows the flooded park.

Flooded Park[25]

The park was closed only on the day the flooding happened. The bar graph shows the rainfall for each day of that week.

[25] Betty Longbottom / Flooded Playground! - Cliffe Avenue / CC BY-SA 2.0



Based on the data, on which day was the park **most likely** flooded?

A. Sunday
B. Monday
C. Tuesday
D. Wednesday

Key: D

The following version of this item shows a modification that provides additional evidence of student understanding of both the NAEP 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 times that the trade-off of more comprehensive evidence is worth the additional time needed to complete the items.

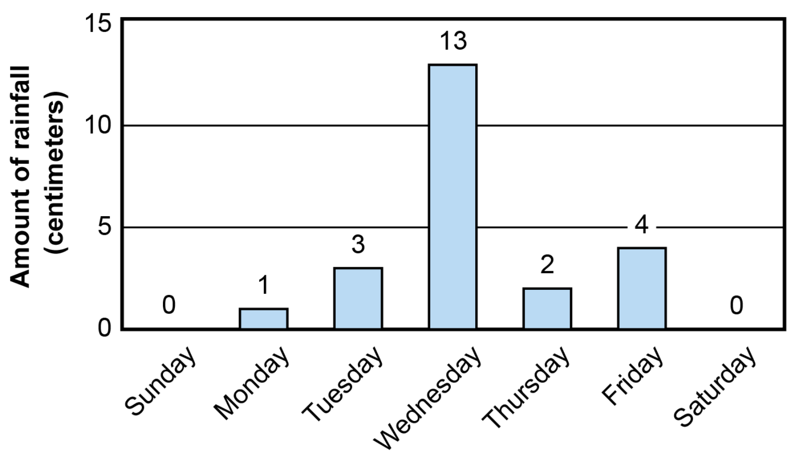
Exhibit A.25. Park Flooding, Version 2

People visiting a local park noticed that the park was flooded and was closed for the day. The picture shows the flooded park.

Flooded Park[26]

The park was closed only on the day the flooding happened. The bar graph shows the rainfall for each day of that week.

Amount of Rainfall at the Park



Day of the Week

Part A

Based on the data, on which day was the park **most likely** flooded?

A. Sunday
B. Monday
C. Tuesday
D. Wednesday

Key: D

---

[26] Betty Longbottom / Flooded Playground! - Cliffe Avenue / CC BY-SA 2.0

Part B

Based on your understanding of weather, which piece of evidence **best** supports your answer in Part A?

A. This day was rainy.
B. The rain started falling on this day.
C. This day had more rainfall than any other day in the week did.
D. The amount of rainfall on this day was lower than on any other day.

Key: C

Exhibit A.26 presents an example of a low DC, low SEP complexity item. 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 NAEP 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 complexity.

Exhibit A.26. Cleopatra's Needle, Version 1

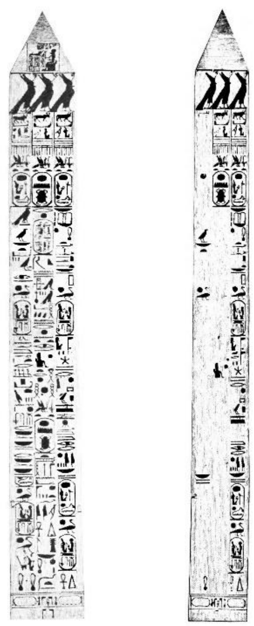
Item ID: Cleopatra's Needle (adapted from a released NAEP item)
Grade and discipline: Grade 8, Earth and Space Sciences
Item type: Inline choice
Alignment: This item is a **2D item**, measuring parts of:

- 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: S8.9: Analyze data to provide evidence to support or reject a model or explanation or use to improve a 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: 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.

Cleopatra’s Needle is a large monument made of rock. For thousands of years, Cleopatra’s Needle stood in Alexandria, a city in the Egyptian desert. In 1881, Cleopatra’s Needle was moved to Central Park in New York City. After only a few years in New York City, the surface of the monument began crumbling. The diagram shows how Cleopatra’s Needle looked when it stood in Alexandria and how it looked after a few years in New York City. The table shows average yearly weather data for the two locations where Cleopatra’s Needle stood.

Cleopatra’s Needle



Alexandria New York City

Weather Data in Alexandria and New York City

Location	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 pictures and the weather data to determine why Cleopatra’s Needle began crumbling after it was moved to New York City. Then, use the drop-down menus to correctly complete the sentences.

The (amount of rainfall/wind speed) in New York City over a year is (greater than/less than) that in Alexandria, Egypt.

This leads to (more/less) weathering of rock and causes the crumbling of Cleopatra’s Needle.

Key: amount of rainfall; greater than; more

Exhibit A.27 uses the same context and phenomenon as the previous item but engages students in deeper uses of a different NAEP SEP by requiring students to make and support an original claim based on data and their understanding of erosion and weathering.

- SEP: S8.18: 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.

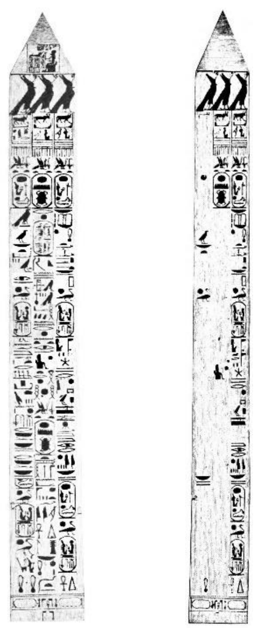
The NAEP DC used is relatively low level, but the NAEP SEP requires students to use multiple components of NAEP SEPs in service of sensemaking at medium complexity. The inclusion of

Part B acts as both a scaffold and a way to surface understanding from students without extended writing.

Exhibit A.27. Cleopatra's Needle, Version 2

Cleopatra's Needle is a large monument made of rock. For thousands of years, Cleopatra's Needle stood in Alexandria, a city in the Egyptian desert. In 1881, Cleopatra's Needle was moved to Central Park in New York City. After only a few years in New York City, the surface of the monument began crumbling. The diagram shows how Cleopatra's Needle looked when it stood in Alexandria and how it looked after a few years in New York City. The table shows average yearly weather data for the two locations where Cleopatra's Needle stood.

Cleopatra's Needle



Alexandria   New York City

Weather Data in Alexandria and New York City

Location	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

Use the picture and the data in the table to make a claim about why Cleopatra's Needle began crumbling in New York City when it did not crumble in Alexandria.

Part B

Use the picture and the data in the table to:

- Describe one piece of evidence from the picture or the table that supports the claim you made in Part A.
- Explain how the evidence supports the claim.

Exhibit A.28. Cleopatra's Needle Version 2 Constructed Response Scoring Notes

Part A

- Claims should reflect an analysis of the relative damage to the stone surface and the different weather conditions between New York City and Alexandria.
- Claims should focus on the impact of water changing the surface of rock (stone/monument), rather than wind (note that wind is included here such that students have to analyze the data and specifically make the claim that water is more likely to be responsible here).
- Student claims should be interpreted relative to Part B—for example, students may make a claim such as "It began to crumble because the weather conditions were different." Credit should be awarded if their response in part B indicates that they understand that water has changed the surface of the rock.

Part B

- Evidence should derive from the data provided, for example, observations comparing the state of the monument in Alexandria versus New York City; the differential rainfall; or the differential rainfall and the similar windspeed (a complete argument often includes refuting possible counterclaims; while this question does not require this, students should not be penalized if they include evidence that refutes a counterclaim, such as the wind is driving degradation of the rock, as long as the reasoning makes clear that students interpreted the data appropriately).
- Rationales should include both:
  - how the evidence supports their specific claim and
  - reasoning that connects the differential rainfall to increased damage in New York, using the idea that water can change a rock's surface as part of the reasoning.
- Students may include a counter argument for wind, saying that wind also could have this effect, but because wind speeds are similar in the two areas, there is no evidence that wind is the cause.
- Some students may provide a more general claim (e.g., weather conditions are different), and may provide reasoning that draws on DC or CCC understanding that is beyond the data that are provided here (e.g., some students could posit that wind may have led to damage if historically there was substantially less robust winds in Alexandria, and the similar wind speeds are only a recent observation [connecting to the CCC Stability and Change] or could cite a different environmental cause that is plausible). Students could be given partial credit for appropriate reasoning with the DC/CCC but should not be given full credit without connecting specifically to the data provided.

Exhibit A.29 is an example of a 2D grade 8 item. Several features of culturally responsive items are included, such as (a) the use of native/home language in the item (i.e., *limu kohu*, which is the name for this seaweed species in the Hawaiian language), (b) use of nontraditional evidence sources that have been useful in university-based science endeavors (i.e., the use of multigenerational/elder accounts as evidence, as used by Stanford botanist Dr. Isabella Aiona Abbott), and (c) 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 about the loss of indigenous foodways currently happening).

Because of the nature of the NAEP 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 NAEP 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 NAEP DC is best assessed via item sets and scenario-based tasks.

Exhibit A.29. Limu Kohu

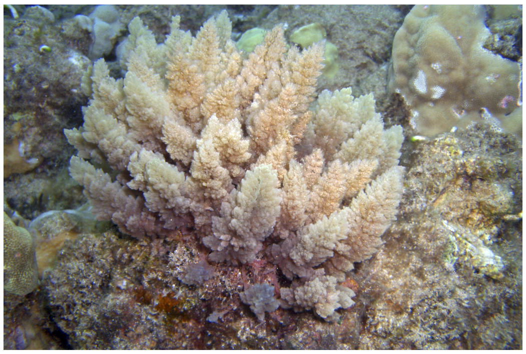
Item ID: Limu Kohu (adapted from the State Performance Assessment Learning Community)
Grade and discipline: Grade 8, Earth and Space Sciences
Item type: Multipart, single-selection multiple choice and short constructed response
Could part A be used as a stand-alone item? Yes.
Alignment: This item is 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: S8.22: 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 and engineering design problem: Limu kohu populations have been declining. Students are asked to make sense of the role of human activities in causing this problem.
Complexity: This item is an example of a medium DC, low SEP complexity item. 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.

Limu kohu is a type of seaweed that is native to the waters around Honolulu, Hawaiʻi. It 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 been rapidly declining over the past 60 years. An example of limu kohu seaweed is shown in the picture. The table describes observations of limu kohu.

Limu Kohu[27]



Observations from Generations of Hawaiian Elders about Limu Kohu Growth and Harvesting
• Limu kohu needs warm water and high salinity to grow. • Limu kohu grows 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

Which human activity is **least** likely to cause harm to limu kohu populations?

A. Companies using industrial methods of harvesting limu kohu remove the whole limu kohu plant.
B. Restaurants using traditional methods of harvesting limu kohu remove the top of the limu kohu.
C. Industrial runoff changes the temperature and salinity of the water in coastal regions where limu kohu live.
D. Ships visiting Hawaiʻi introduce invasive seaweed species that use the same resources as limu kohu into coastal regions where limu kohu live.

Key: B

Part B

Use the information provided and your understanding of the impacts of human activities on the environment to support your answer to Part A.

[27] MDC Seamarc Maldives, CC BY-SA 4.0, via Wikimedia Commons

Exhibit A.30. Limu Kohu Part B Constructed Response Scoring Notes

Students provide one statement that is based on the information provided, shows understanding of the impacts of human activities on the environment, and that supports the answer to Part A. For example:

- Traditional harvesting practices that focus on harvesting only the top portions of limu kohu are less destructive compared to other methods that involve uprooting the entire plant. When only the top is harvested, the base of the plant remains intact, allowing the limu kohu to potentially regrow.
- The information provided indicates that limu kohu has the ability to regrow when it is trimmed. This suggests that harvesting the upper parts of the seaweed allows it to regenerate, ensuring the sustainability of the population over time.

The item in Exhibit A.31 is expected to perform at the *NAEP Advanced* level because it uses an unfamiliar context and includes multiple, dependent parts that require different combinations of the dimensions for a complete response. Further, the selected response item parts reflect advanced performance by requiring students to select the correct answers among plausible, attractive distractors. This grade 8 item reflects the following Earth and Space Sciences example in the advanced ALD: "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."

The context draws upon Pueblo Indian culture from the Southwest United States, where people planned and built structures oriented to the patterns of the apparent motions of the sun and moon. The context provides an example that supports an expanded understanding of how science is done and who contributes to science (U.S. Department of the Interior, 2015). The phenomenon is "On the first day of spring, the sun is visible through a structure called 'Casa Rinconada,' located in a pueblo in the Southwest United States."

An item with a similar alignment but reflecting the *NAEP Proficient* level might involve students developing a model to test ideas about observable patterns in the motion of objects in the sky relative to Earth. The item could use the familiar example of the area around a school building while adding complexity relative to the basic example by requiring students to develop a model based on the context instead of presenting the model in the item.

An item with a similar alignment but reflecting the *NAEP Basic* level might involve students in using a model to describe observable patterns in the motion of objects in the sky relative to Earth. The item could use a more familiar and less complex context, such as a model of the area around a school building. Using the same scenario, a basic item could be an entry point into an item set. For example, a basic item could ask students to use the model with the wire and foam ball to describe the sun's motion across the sky.

An item with a similar alignment but reflecting performance below the *NAEP Basic* level might involve students selecting from a set of images of models. This item might need less information in the stimulus. To ensure alignment, while providing access to students performing at this level, the item could use additional parts that target different dimensions; some of the parts might be accessible to students scoring around the level of the first cut score, while other parts might be slightly more difficult.

Exhibit A.31. Spring Sunlight

Item ID: Spring Sunlight
Grade and discipline: Grade 8 Earth and Space Sciences
Item type: Multipart item, single-selection multiple choice
Alignment: This item is a **3D item**, measuring parts of:

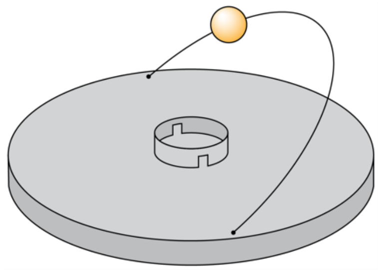
- DC: E8.1: The orbits of Earth around the sun and of the moon around Earth, together with the rotation of Earth on an axis that runs from its north pole to its south pole, cause observable and measurable patterns that can be used to predict apparent motions of the sun and moon and occurrence of tides and seasonal changes through models.
- SEP: S8.15: 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: C8.7: 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, help to make predictions about the phenomenon.

Phenomenon: On the first day of spring, the sun is visible through a structure called "Casa Rinconada," located in a pueblo in the Southwest United States.

The picture shows the sun on the first day of spring. The sun is visible through a structure called "Casa Rinconada," located in a pueblo in the Southwest United States.[28]

[28] Source: National Parks Service

The picture shows a model two students plan to use to predict when the sun will be visible through the structure.



The model has the following characteristics:

- A small copy of the structure is in the center of the model.
- A curved wire is placed into holes near the edge of the model.
- The wire is attached to the model at an angle.
- A foam ball moves along the wire.

Part A

Which statement **best** describes how the students could modify the model to improve their predictions?

A. Add another wire below the model so that the foam ball can be moved in a complete circle.
B. Add holes so that the foam ball can be moved between different locations on the edge of the model.
C. Change the angle of the wire so the foam ball can be placed above the structure in the middle of the model.
D. Change the length of the wire so the foam ball can be moved to different distances from the structure.

Key: B

Part B

Which statement describes a pattern in the motion of the sun that the model would **most likely** show with the modification identified in Part A?

A. The sun is visible only during the day.
B. The sun always moves in the same direction.
C. The sun is directly above the structure once each day.
D. The sun rises in different locations in different seasons.

Key: D

Item Sets

Item sets assess related NAEP DCs, SEPs, and CCCs, leveraging a common stimulus and/or context.

Physical Science

Exhibit A.32 demonstrates how an item set can assess NAEP SEPs that come from each of the four practice categories.

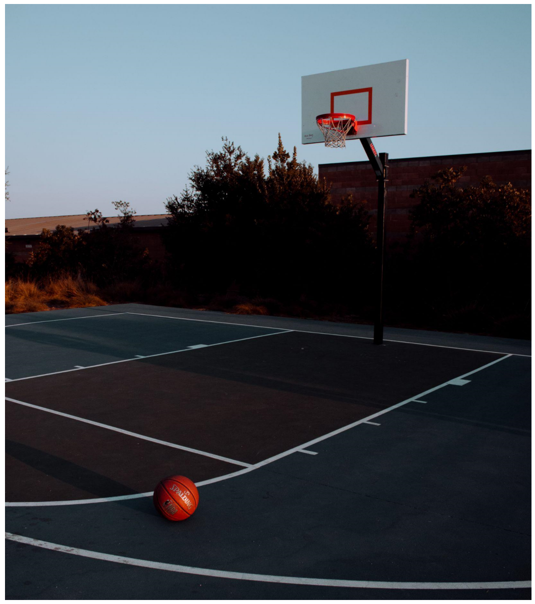
Exhibit A.32. Moving Basketball

Item Set ID: Moving Basketball
Grade and discipline: Grade 4, Physical Science
Number of Items: 4
Item types: Single-selection multiple choice, short constructed response.
Alignment and Complexity: This is a **3D item set** at medium complexity, measuring parts of:

	NAEP DC	NAEP SEP	NAEP CCC
Item 1	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. *Medium complexity*	S4.20: Construct and/or support an argument with evidence to support or reject a claim about a phenomenon or a design solution. *Low complexity*	-
Item 2	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. *High complexity*	S4.5: Plan an investigation to explore a scientific question or design problem taking into consideration appropriate variables and tests. *Medium complexity*	C4.2: Cause-and-effect relationships are routinely identified, tested, and used to explain changes. *Medium complexity*
Item 3	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. *Medium complexity*	S4.10: Analyze data to determine whether it supports or refutes a claim about a phenomenon or design solution. *Medium complexity*	-
Item 4	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. *Medium complexity*	S4.17: Develop an evidence-based description or explanation supported by evidence and reasoning of a phenomenon or the action of a designed solution. *Medium complexity*	C4.2: Cause-and-effect relationships are routinely identified, tested, and used to explain changes. *Low complexity*

Phenomenon: A ball moves when there is no one there to push it.

Aliya and Rishi go to an empty basketball court, as shown.[29]



They notice that a ball on the ground is moving, even though no one is pushing it. They make the following observations as they try to figure out why the ball is moving.

Aliya and Rishi's Observations
• It is very windy today—the wind is pushing us! • The ball moves more when we feel a lot of wind on our faces. • There are no other people nearby who might have been using the ball.

Item 1

Aliya claims that the wind is applying a force to the ball. Which observation would support Aliya's claim?

A. The ball moves after the wind stops blowing.
B. The ball moves against the wind as the wind blows toward the ball.
C. When the wind blows gently, the ball stays in place.
D. When the wind blows hard, the ball moves in the direction the wind blows.

Key: D

[29] Photo by https://unsplash.com/@matiasvizuals

Item 2

Rishi and Aliya decide to design an investigation to figure out if the wind can cause a ball to move. Which investigation plan will be the **most** helpful for figuring out if the wind can cause a ball to move?

A.

Plan 1
1. Place two balls on the basketball court at the same starting point. 2. Push one ball, and do not push the second ball. 3. Record: a. Speed of each ball b. Distance each ball moves 4. Repeat the steps pushing only the second ball and repeat again pushing both balls.

B.

Plan 2
1. Place one ball on the floor of a closed room. 2. Use a fan on low speed to blow air toward the ball. 3. Record: a. Speed of the ball b. Distance the ball moves 4. Repeat the steps with the fan on medium speed and repeat again with the fan on high speed.

C.

Plan 3
1. Place one ball on the basketball court. 2. Use a fan on low speed to blow air toward the ball. 3. Record: a. Speed of the ball b. Distance the ball moves

D.

Plan 4
1. Place two balls on the floor of a closed room at the same starting point. 2. Push one ball, and do not push the second ball. 3. Record: a. Speed of each ball b. Distance each ball moves

Key: B

Item 3

Aliya and Rishi gathered information while she watched the effects of the wind on the ball. Which of the following data **best supports the claim that wind causes unequal forces acting on the ball**?

A.

Trial [or observation]	Strength of wind	Ball movement
1	Very strong	Very fast
2	Strong	Fast
3	Weak	Slow

B.

Trial [or observation]	Strength of wind	Ball movement
1	Weak	Fast
2	Very weak	Fast
3	No wind	Fast

C.

Trial [or observation]	Strength of wind	Ball movement
1	No wind	Slow
2	Very weak	Slow
3	No wind	None

D.

Trial [or observation]	Strength of wind	Ball movement
1	Very strong	Very fast
2	Strong	Very fast
3	Strong	Slow

Key: A

Item 4

After observing the ball for some time, Aliya and Rishi notice that the ball moves for a while and then stops and later moves again. Rishi tells Aliya that sometimes he feels the wind blowing against his face, but the ball does not move.

Use your understanding of balanced and unbalanced forces to explain what is happening to the ball when the wind blows, but the ball does not move.

Exhibit A.33. Moving Basketball Item 4 Constructed Response Scoring Notes

A complete answer includes reasoning that there are always multiple forces acting on the ball. When the wind blows and the ball's motion changes, the wind is exerting more force than the other forces acting on the ball. When the wind blows but the ball's motion does not change, the force from the wind on the ball is equal (or balanced) to the other forces acting on the ball. Emphasis for scoring is on reasoning with the idea of balanced and unbalanced forces, not on the use of exact vocabulary (i.e., balanced/unbalanced, equal/unequal, and similar language is acceptable).

Life Science

The item set in Exhibit A.34 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 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 A.34. Seaside City Item Set

Item Set ID: Seaside City (Adapted from the Disruptions in Ecosystems Chapter 5 Assessment Task [pp. 59–63] authored by UC Berkeley's Lawrence Hall of Science.)
Grade and discipline: Grade 8, Life Science
Number of Items: 3
Item types: Single-selection multiple choice, grid, short constructed response, inline choice
Alignment and Complexity: This is a **3D item set** with a medium to high complexity, measuring parts of:

	NAEP DC	NAEP SEP	NAEP CCC
Item 1	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. *Medium complexity*	S8.16: Use a model to test ideas about phenomena in natural systems or proposed design solutions. *Medium complexity*	-
Item 2	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. *High complexity*	S8.22: Identify evidence that could be used to refute a claim about a phenomenon. S8.24: 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. *Medium complexity*	C8.4: Phenomena may have more than one cause, and some cause-and-effect relationships in systems can only be described using probability. *Medium complexity*

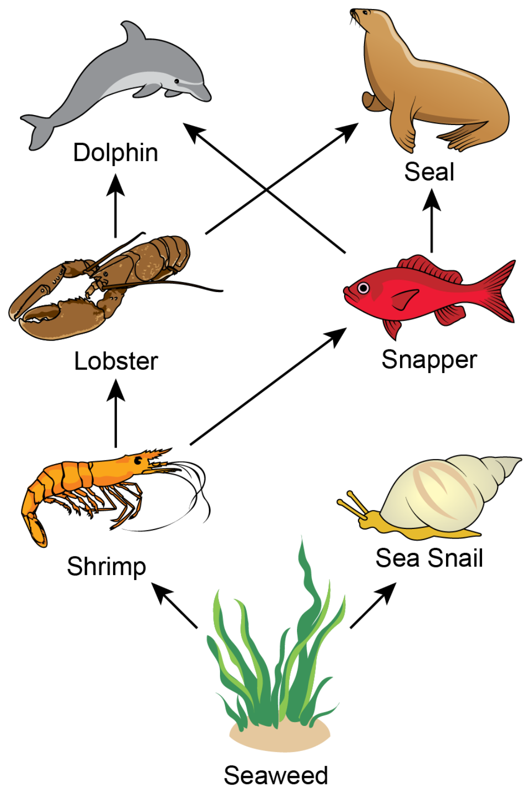
Item 3	L8.6: In any ecosystem, organisms and populations with similar requirements for food, water, oxygen, or other resources may compete with each other for limited resources, access to which consequently constrains their growth and reproduction. *Medium complexity*	S8.15: Develop, use, and/or revise 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. *Medium complexity*	C8.3: Cause-and-effect relationships may be used to predict phenomena in natural or designed systems. *Medium complexity*

Phenomenon and engineering design problem: Beaches used to have a lot of seaweed, but seaweed is rarely seen now. Students are asked to make sense of this problem, as well as evaluate a possible solution.

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

People who have lived in Seaside City for many years remember that the beaches used to have a large amount of seaweed on them all the time. Recently, the people have noticed that it is rare to see any seaweed on the beaches now. Seaweed is very important to the local ecosystem. It is used as a major food source and as a habitat for many animals living in the area. Figure 1 shows a marine food web that operates near Seaside City. Graph 1 shows changes in the seaweed population near Seaside City over time, and Table 1 shows some observations scientists have made around Seaside City over the last 30 years.

Figure 1. Seaside City Food Web



Graph 1. Seaweed Changes Over Time

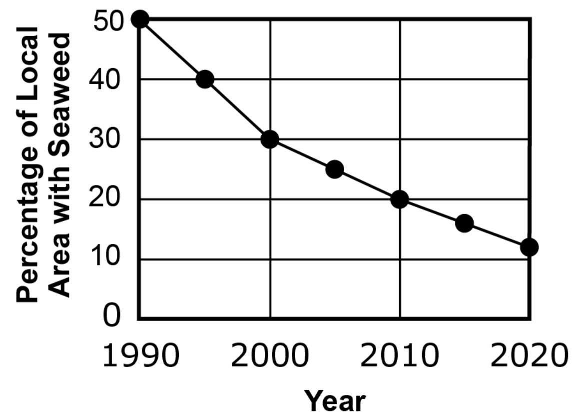


Table 1. Observations of Human Impacts
• Increased fishing to provide food for people • Disruptions to the local environment (land and sea) while building homes and businesses for people • Decline of some species as human activities impact the resources they need to survive and reproduce

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. There was more food available for lobsters in 2020 than in 1990.
C. The average mass of individual seals was larger in 2020 than in 1990.
D. There was more competition among sea snails and shrimp in 2020 than in 1990.

Key: D

Item 2

Some conservationists claim that the amount of seaweed near Seaside City is decreasing because people are eating more lobsters and snappers.

Part A

Scientists studying the ecosystem in and around Seaside City gathered the evidence shown in the table.

Use the information provided and your understanding of how organisms interact in ecosystems to select three (3) statements that support the conservationists' claim and three (3) statements that do not support the claim.

Evidence	Supports	Does not support
Both snapper and lobster populations decreased between 1990 and 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 decreasing at a similar rate.	☐	☐
The number of shrimp was higher in 2020 than in 1990.	☐	☐
The shrimp population increases initially but then decreases.	☐	☐
As the number of human-made buildings on the coast increases, the amount of seaweed decreases.	☐	☐

Key: Rows 1, 2, 4 support; Rows 3, 5, 6 do not support

Part B

Which statement describes another reason for the decrease in seaweed near Seaside City, based on the evidence provided?

A. Natural erosion of sand from coastal beaches is limiting the growth of seaweed.
B. The seaweed in the area has recently (after 2015) been infected with a fatal disease.
C. Changes in the sea snail population are leading to a decrease in the amount of seaweed.
D. Increased human construction along the coast is destroying much of the seaweed in the area.

Key: D

Part C

Describe at least **two** pieces of additional information that you would need to determine whether the best explanation for the decrease in seaweed is the one you selected in Part B. Be sure to explain why the additional information you describe is necessary.

Part D

Assume the reason you selected in Part B is correct. Use the information provided and your understanding of ecosystem dynamics to describe how this will **most likely** affect the relationships between seaweed, snappers, and lobsters near Seaside City.

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 developed a model to predict how the building construction ban will most likely affect organism populations in the future. Based on the model, the students made two predictions about the shrimp population in 2030:

Option A: The shrimp population will remain about the same in 2030 as it is in 2020.

Option B: The shrimp population will be larger in 2030 than it is in 2020.

Use the information provided and your understanding of the relationships between organisms in an ecosystem to complete the sentences. Complete the sentences by choosing the correct answers from the dropdown menus.

The model predicts that the shrimp population will remain about the same if the seaweed population [decreases/increases/remains the same].

This is because the growth of the shrimp population will [be limited, decrease, increase] if competition for resources [stays the same, increases, decreases] compared to what it is in 2020.

Key: remains the same; be limited; stays the same

Exhibit A.35. Seaside City Item 2 Constructed Response Scoring Notes

Part C

- Students identify information that would clarify among multiple possible causal accounts that are consistent with the information provided. A wide range of answers are acceptable, as long as the justification makes it clear (1) that students are have an accurate interpretation of the provided evidence and the relationship to ecosystem dynamics and (2) that the causal links students are investigating with their questions reflect an accurate understanding of how changes in the ecosystem can impact different populations that have interdependent relationships.

Part D
Student response demonstrates the ability to describe how disruptions to a biological component of an ecosystem can lead to shifts in all its population. This may include the following:

- If increased construction is indeed destroying seaweed in the area, it would likely have a disruptive effect on the food web, affecting all organisms within it.
- With less seaweed available, sea snail and shrimp populations might decline due to food scarcity, impacting the snapper population.
- As snapper numbers decrease, lobsters (predators of snappers) might also face food shortages, leading to reduced lobster populations.

Ultimately, a decrease in lobster populations could affect their predators, such as dolphins and seals.

The following example illustrates how item 2 could be reduced in complexity. This version is 2D, low DC and low SEP complexity, without meaningfully assessing the NAEP 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 A.36. Modified Item 2 From Seaside City

Item 2

Some conservationists claim that the amount of seaweed near Seaside City is decreasing because people are eating more lobsters and snappers.

Part A

Scientists studying the ecosystem in and around Seaside City gathered the evidence shown in the table.

Use the information provided and your understanding of how organisms interact in ecosystems to choose whether each statement supports or does not support the conservationists' claim.

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

Key: Rows 1, 2 support; Row 3 does not support

Earth and Space Sciences

The item set in Exhibit A.37 includes four items focused on the formation of the Hawaiian Islands, a context that is highly relevant to communities living and working in Hawaiʻi. While the items presented here can be independent, they also serve to coherently support student thinking if presented as a set—the complexity of the items will shift depending on whether they are presented as a set versus whether items are presented as stand-alone or out of order.

Exhibit A.37. Formation of Hawaiʻi Item Set

Item Set ID: Formation of Hawaiʻi (adapted from the New Meridian Science Exchange)
Grade and discipline: Grade 8, Earth and Space Sciences
Number of Items: 4
Item types: Zones, single-selection multiple choice, extended constructed response
Alignment and Complexity: This is a **3D item set** at low complexity, measuring parts of:

	NAEP DC	NAEP SEP	NAEP 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 features. *Low complexity*	S8.8: 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). *Low complexity*	-
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. *Low complexity*	S8.15: 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. *Low complexity*	-
Item 3	E8.4: The Earth consists of layers, including a solid, rigid outer layer divided into plates, which are always moving very slowly.	S8.15: Develop, use, and/or revise a model to describe, explain, and/or predict phenomena by identifying relationships among parts	-

	Interactions between Earth's moving plates result in changes of physical features. *Low complexity*	and or quantities in a system, including both visible and invisible quantities. *Low complexity*	
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. *Medium complexity*	S8.18: 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. *Medium complexity*	C8.3: Cause-and-effect relationships may be used to predict phenomena in natural or designed systems. *Low complexity*

Phenomenon: The Hawaiian Islands have very different ages, and the Island of Hawaiʻi is much larger than other islands.

Common stimulus (available with all items in the task)

The Hawaiian island chain has several islands that have been developing over millions of years. Some of the islands, such as Kaua'i, began developing over 5 million years ago, while other islands, such as the Island of Hawaiʻi, are still developing today. The table lists the ages of five Hawaiian Islands.

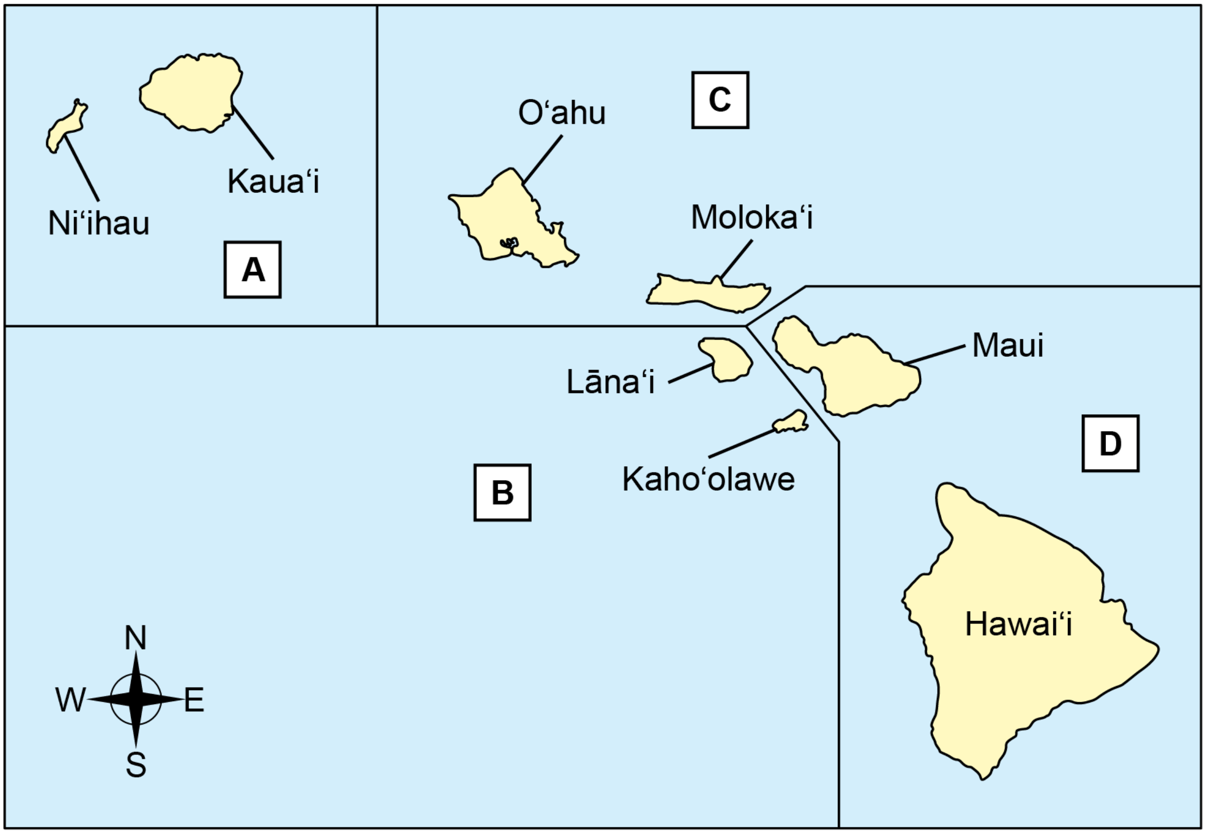
Table 1. Ages of Hawaiian Islands

Kauaʻi	3.5–5.6 million
Oʻahu	2.2–3.3 million
Molokaʻi	1.3–1.8 million
Maui	Less than 1 million
Hawaiʻi	Less than 0.7 million

Item 1

There are more than 100 islands in the Hawaiian island chain. Many of these islands are not shown on most maps of the Hawaiian island chain. The map shows eight of the largest Hawaiian Islands.

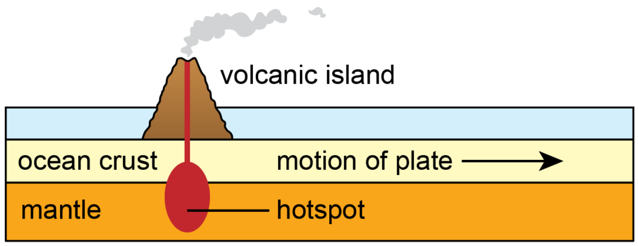
Use the data in Table 1 to select an area on the map where evidence of the oldest Hawaiian Islands is **most likely** to be found.



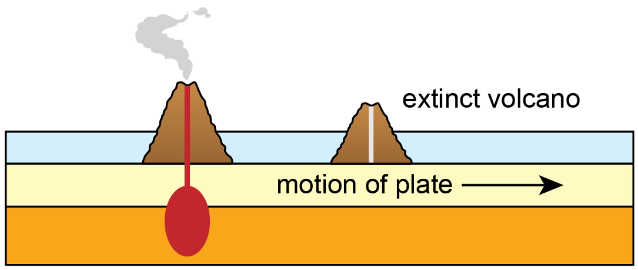
Key: A

Item 2

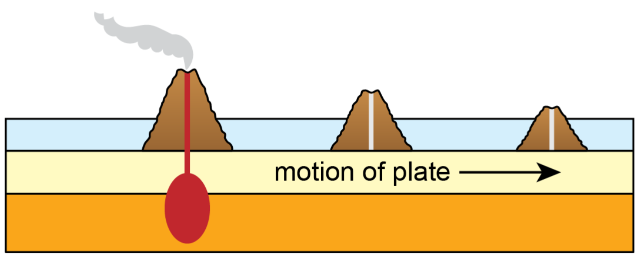
The model shows how an island chain, such as the Hawaiian island chain, is formed.



A tectonic plate moves over a hotspot.



A new volcano forms.



A chain of islands forms over millions of years.

Maui and the Island of Hawaiʻi are the only islands in the Hawaiian island chain that have active volcanoes. Based on the information in Table 1 and in the model, which statement **best** describes why Maui and the Island of Hawaiʻi have active volcanoes, but the other islands do not?

A. Maui and the Island of Hawaiʻi are the largest islands in the chain, and only large islands have active volcanoes.

B. Maui and the Island of Hawaiʻi are the youngest islands in the chain, and volcanoes can only form on young islands.

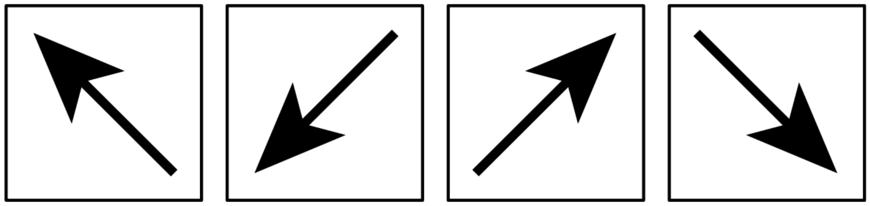
C. Maui and the Island of Hawaiʻi are being built over a plume of hot magma, which pushes through the oceanic crust to form the islands.

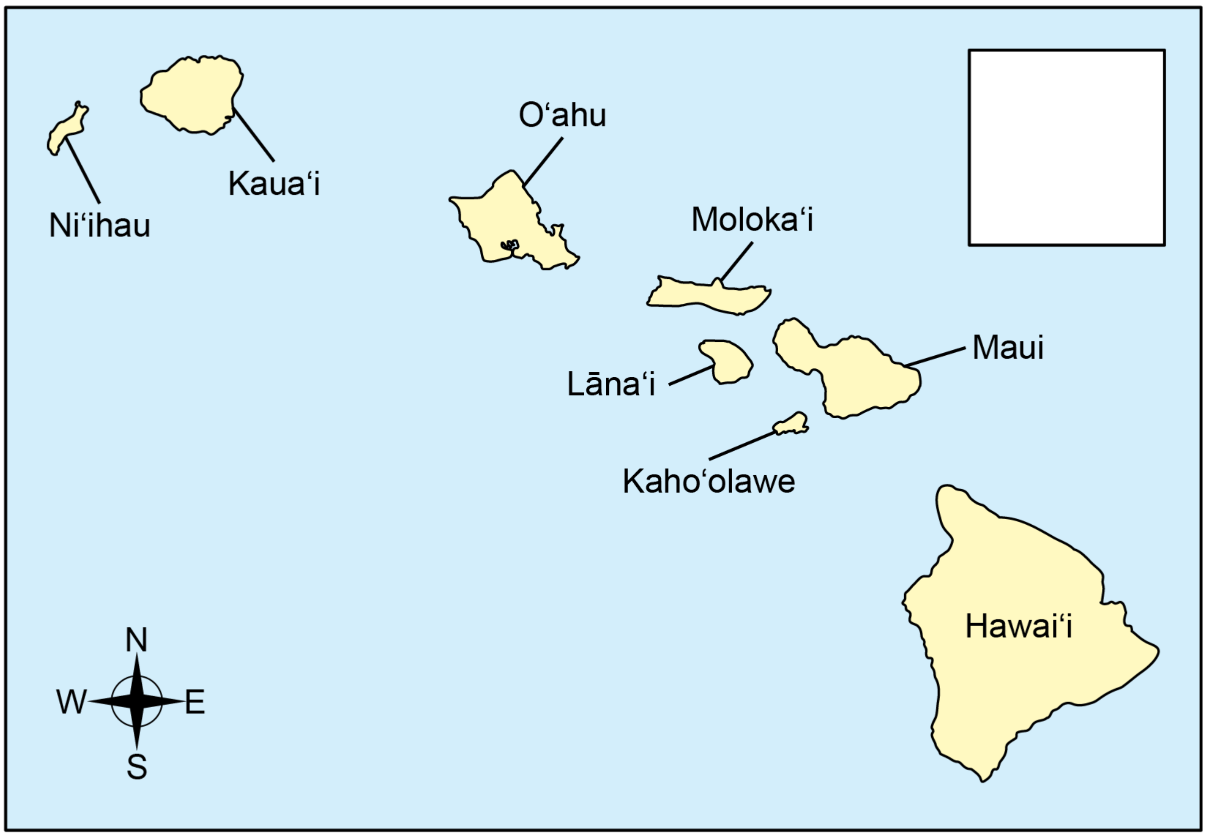
D. Maui and the Island of Hawaiʻi are being built on oceanic crust, which allows plumes of hot magma to flow more easily through it than continental crust.

Key: C

Item 3

Complete the model to show the relationship between the building of island chains, such as the Hawaiian Islands, and the movement of the oceanic plate on which the islands are located. Drag the correct arrow into the box to show the direction of plate motion in the model. Only one arrow will be used.





Key: Arrow 1

Item 4

Use the data provided and your understanding of how geologic processes change Earth's surface over time to explain why the Big Island and Maui are larger than the other islands shown in Table 1.

Exhibit A.38. Formation of Hawai‘i Item 4 Constructed Response Scoring Notes

- Student explanations should cite both data provided in the stimulus and information about how geologic processes (erosion, plate movement, volcanic contributions to island size and growth) contribute to the patterns seen.
- Note that if this item is presented as part of the item set, other items may influence what students include here (e.g., item 2 includes information about active volcanoes); however, this item is not dependent on those previous items, nor should it provide undue advantage to students responding to previous questions in the set.
    - If this item were presented as Part B of item 2, or following item 2, we may expect more discussion of active volcanoes and their contribution to the size of younger islands with active volcanoes.
- Student reasoning should make a connection between size and age of the islands, namely that younger islands tend to be larger. They should support this idea with further DC understanding, focusing on the geologic processes that happen over time. This may include:
    - how erosion over time reduces the size of islands such that older islands have had more erosion impact their size and geography and
    - because newer islands like the Big Island are still forming, geologic processes such as volcanic eruptions driven by plate motion over hotspots lead to massive amounts of deposition, contributing significantly to new/additional landmass.

Multidisciplinary: Life Sciences and Earth and Space Sciences

The item set in Exhibit A.39 illustrates how a robust phenomenon or problem to solve can give rise to items across multiple disciplines (e.g., Life Science and Earth and Space Sciences). As 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: Locust swarms can cause a lot of damage. Students are asked to consider what is making swarming worse and what solutions there might be to this problem.

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 considerations of whether current upticks are significant or not (more sophisticated data analysis), and so on. This context can also easily support items for both middle school and high school levels in both Life Science and Earth and Space Sciences. This version of the item set was selected to show how a wider range of NAEP 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 A.39. Locusts

Item ID: Locusts
Grade and discipline: Grade 8, Earth and Space Sciences and Life Science
Item set with the following item types: Inline choice, matching, grid, constructed response
Alignment and Complexity: This is a **3D item set**, measuring parts of:

	NAEP DC	NAEP SEP	NAEP CCC
Item 1	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. *Low complexity*	S8.9: Analyze data to provide evidence to support or reject a model or explanation or use to improve a design solution. *Low complexity*	
Item 2	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. *Low complexity*	S8.15: 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. *Medium complexity*	C8.3: Cause-and-effect relationships may be used to predict phenomena in natural or designed systems. *Medium complexity*
Item 3	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. *Medium complexity*	S8.18: 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. *Medium complexity*	C8.5: 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. *Medium complexity*

	NAEP DC	NAEP SEP	NAEP CCC
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, 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. *Medium complexity*	S8.22: Identify evidence that could be used to refute a claim about a phenomenon. *Low complexity*	C8.3: Cause-and-effect relationships may be used to predict phenomena in natural or designed systems. *Low complexity*
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. *Medium complexity*	S8.24: 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. *High complexity*	C8.4: Phenomena may have more than one cause, and some cause-and-effect relationships in systems can only be described using probability. *High complexity*

Phenomenon and engineering design problem: Locust swarms can cause a lot of damage. Students are asked to consider what is making swarming worse and possible solutions.

Common stimulus (available with all items in the task)

A desert locust is an insect that undergoes changes to its body in certain environmental conditions. Figure 1 describes some differences between two modes of a desert locust.

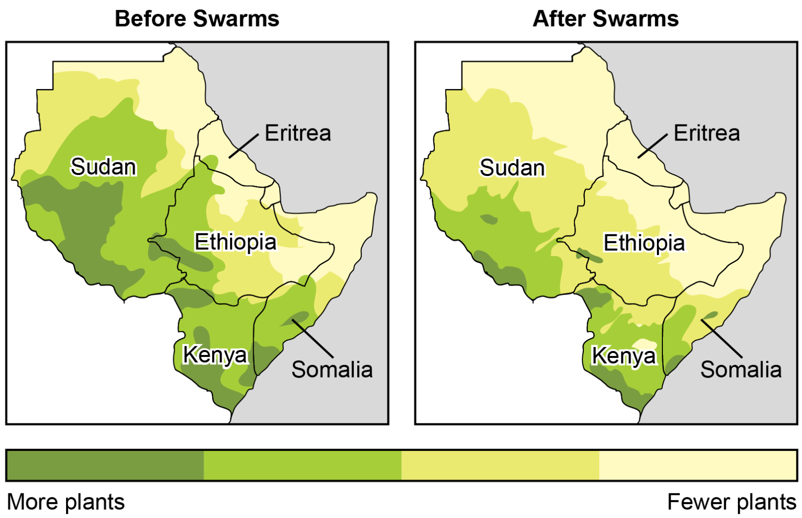
Figure 1. Desert Locust[30]

Mode 1: Grasshopper (Dry, warm or cool weather)	Mode 2: Locust (Wet/rainy, warm or hot weather)
• Behave independently • Stay away from other desert locusts • Mostly walk slowly and jump • Limited 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.	• Behave as a united group (swarms) • Gather together with 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 six months.

[30] Bernard DuPont Bird Locust Attribution-ShareAlike (CC BY-SA 2.0); Magnus Ullman, CC BY-SA 3.0, via Wikimedia Commons

When these insects are in Mode 2, they are able to swarm. A single swarm of locusts can cover an area of up to 100 square miles, with 40 to 80 million locusts in each square mile. Swarms can travel up to 100 miles a day. Figure 2 shows the effect of three months of locust swarms on available vegetation in an area of Africa.

Figure 2. Available Vegetation Before and After Locust Swarms



Item 1

An individual desert locust eats about 2 grams of food each day. Using the data provided, complete the sentences by choosing the correct answers from each dropdown menu.

When desert locusts are in Mode 1, the available vegetation in the areas where they live will likely be **most** similar to the (before/after) swarms map shown in Figure 2.

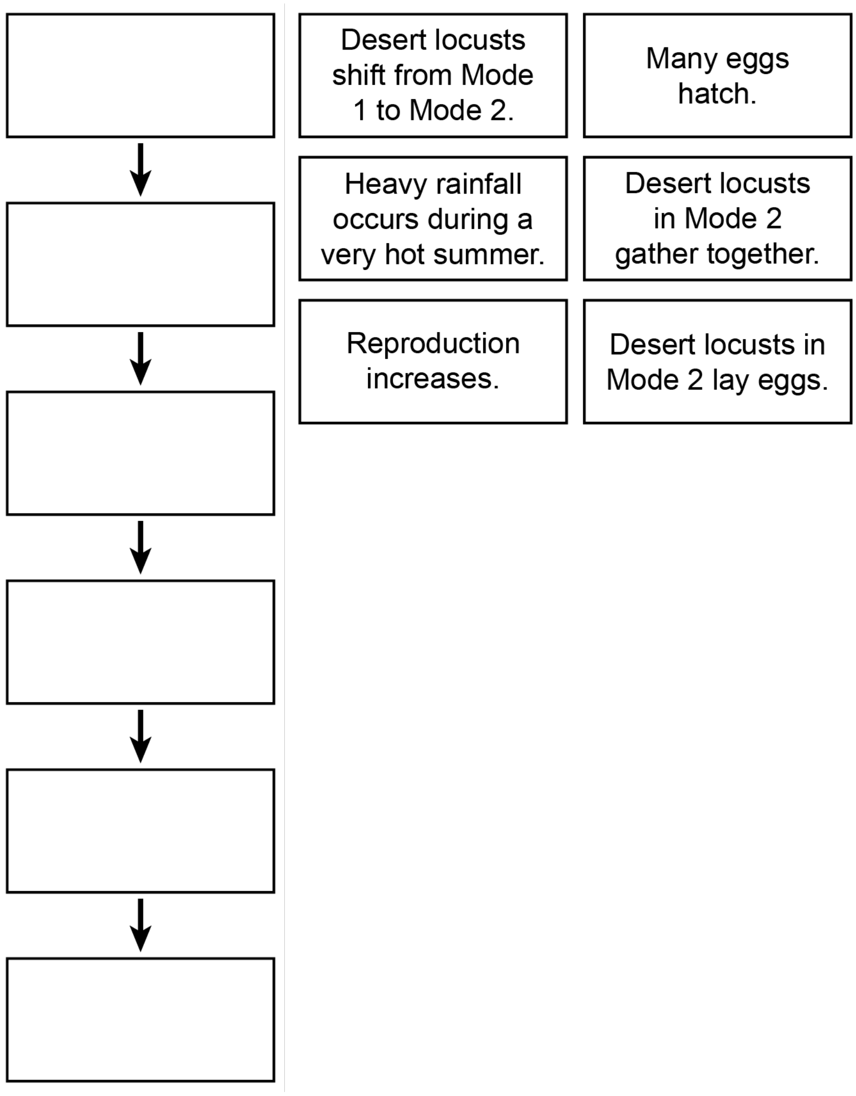
Key: before

This is because, in Mode 1, desert locusts eat a (small/large) amount of plants compared to all the plants available in an area and, therefore, have a (small/large) impact on an area's available plant life.

Key: small; small

Item 2

Use the statements to develop a model to show how locust populations in Mode 2 can become large enough to swarm. Drag the statements into the correct boxes to complete the model. Each statement will be used once.



Key: Heavy rainfall occurs during a very hot summer. Many eggs hatch. Desert locusts shift from Mode 1 to Mode 2. Desert locusts in Mode 2 gather together. Reproduction increases. Desert locusts in Mode 2 lay eggs.

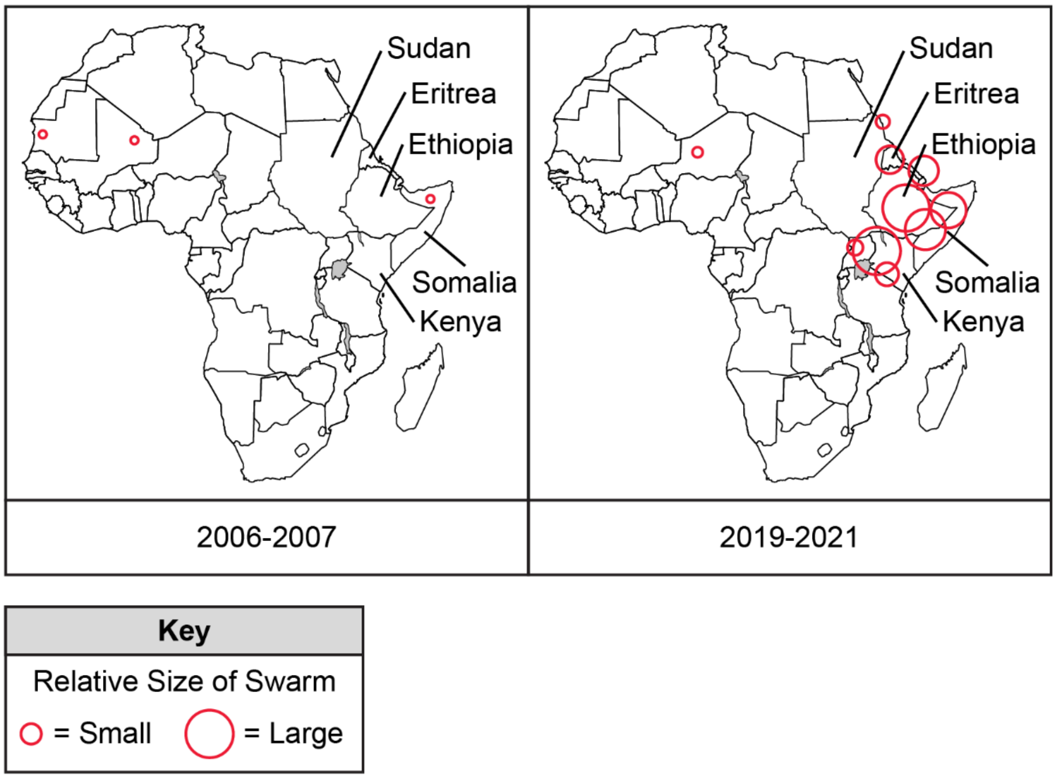
Item 3

Use the information provided as well as your understanding of scale and quantity to explain why only desert locusts in Mode 2 affect available vegetation as shown in Figure 2.

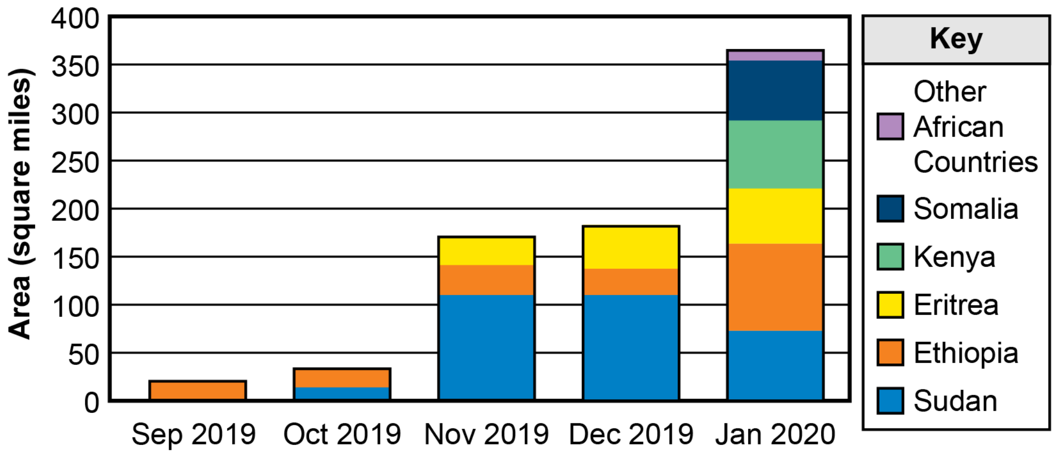
Item 4

Recently, eastern African countries such as Kenya, Somalia, and Ethiopia have experienced larger and longer-lasting swarms of desert locusts. Over a few months, swarms of desert locusts can eat 50–80% of crops being grown in the area, leaving only a small amount of food for people living in those countries. Red circles on the maps show areas affected by locusts in Mode 2. The bar graph shows the sizes of the areas affected in each African country during one growing season.

Observed Locust Swarms in Africa



Area Affected by Locust Swarms in African Countries



Part A

Some scientists claim that global climate changes are causing desert locusts to switch to Mode 2 more frequently and for longer periods of time. While investigating this claim, the scientists collected the evidence shown in the table.

Select the checkboxes beside each piece of evidence in the table that supports the scientists' claim. Leave blank the checkboxes beside pieces of evidence that do not support the scientists' claim.

Evidence	Yes, this evidence supports the claim.
There were 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.	☐

Key: Boxes 2 and 3 should be checked.

Part B

Explain why each piece of evidence you chose in Part A supports the scientists' claim.

Part C

Use the information provided and your understanding of the impact of human activities on global climate change to make and support a claim about how human activities will **most likely** influence populations of desert locusts in the future.

Item 5

The charts show two proposed solutions for reducing the impacts of desert locust swarms on crops.

Option 1: Spraying Chemical Pesticides.
Airplanes are used to 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 sprayed area. • Animals (including humans) who eat the dead locusts or sprayed plants can become ill. Some do not survive. • Using this method is somewhat expensive.

Option 2: Spraying Biological Pesticides.
A variety of methods are used to spray biological pesticides over affected areas.

Considerations for Option 2
• Effectiveness of this method is unknown and is still being tested. • Takes 1–3 weeks to kill locusts. • There is no impact on mammals or crops. • May kill other insects that are closely related to locusts. • Using this method is very expensive.

Scientists predict that Kenya will experience locust swarms again this year. Use the information provided about Option 1 and Option 2 as well as your understanding of the needs of ecosystems to choose the better solution for Kenya. Use scientific reasoning and at least 2 pieces of evidence to support the solution you choose.

Exhibit A.40. Locusts Constructed Response Scoring Notes

Item	Scoring Notes
3	Student responses should include reasoning that connects the shift to Mode 2, and associated behavioral impacts (they are closer together, reproduce more) as a positive feedback loop that leads to relatively huge population growth. Note that students DO NOT need to use terms like positive feedback loop as long as the concept is present in their response. Students may also colloquially use language like “exponential growth” to indicate the scale at which the population is growing.
4B	Student responses connect the evidence to the provided claim, using reasoning that shows an understanding of the targeted DCs. Note that the emphasis here is on reasoned connection between the evidence and claim, using an accurate understanding of the DC.

Item	Scoring Notes
4C	Students make a claim that human activities will most likely lead to increased swarming. In their argument, they include reasoning that connects provided evidence to the claim, grounded in an understanding of the relationship between human activities, changes to climate/temperature, and impact on locust swarming. While students may include multiple reasons/pieces of evidence, one example connecting evidence to the claim or a generalized response that summarizes across the evidence provided is sufficient here.
5	Students may choose either Option 1 or Option 2 and support their choice with two pieces of evidence provided, along with reasoning grounded in the targeted DC. Note that students may also bring reasoning that reflects their own priorities/values (e.g., monetary cost, impact on animals or environment, impact on peoples' livelihoods, etc.). These are acceptable as elements of student reasoning, as long as they are connected to evidence provided and consistent with scientific reasoning.

Exhibit A.41 is an item set for grade 12 that demonstrates how NAEP DCs, SEPs, and CCCs can be used in the service of sensemaking. The items leverage simple uses of NAEP SEPs and CCCs, allowing a wider range of students to access and engage with a rich context that attends to cultural relevance by highlighting a specific community (in this case, the many, largely rural communities that make up Appalachia) through an asset-based lens. This is particularly important because these specific rural communities are often portrayed and understood through deficit-oriented lenses (e.g., focusing on poverty, lack of educational resources and college degrees, economically less advantageous career options, etc.).

It should be noted that this item set could be expanded to more deeply and comprehensively assess (a) related Earth and Space Sciences DCs (e.g., how the geography of the Appalachian region has contributed to climate resilience over time), (b) the Life Science DCs (e.g., how ecosystem dynamics contribute to resilient biodiversity), or (c) their integrated use (e.g., how geographic features that have evolved over time have led to adaptations and speciation, contributing to the rich biological systems in the area and impacts on human civilizations). These could be explored in further independent items within an item set or in related items within a scenario-based task. NAEP CCC 12.3 could be further explored by examining the mechanisms within any of the subsystems included as part of the inputs of the model.

Exhibit A.41. Human Migration to Appalachia

Item ID: Human Migration to Appalachia
Grade and discipline: Grade 12, Earth and Space Sciences and Life Science
Item set with the following item types: Single-selection multiple choice, extended constructed response
Alignment: This is a **3D item set**, measuring parts of:

	NAEP DC	NAEP SEP	NAEP CCC
Item 1	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. 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. *High complexity*	S12.15: 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. *Low complexity*	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. *Low complexity*
Item 2	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. *Low complexity*	S12.14: Apply or revise algorithms when analyzing data or designing, programming, testing, and revising scientific models, explanations, and design solutions. *Medium complexity*	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. *Medium complexity*

Phenomenon and Engineering Design Problem: Human migration to Appalachia has been greater than predicted by computer models. The model used to make the prediction needs to be revised to better reflect the factors that influence migration into different regions of Appalachia.

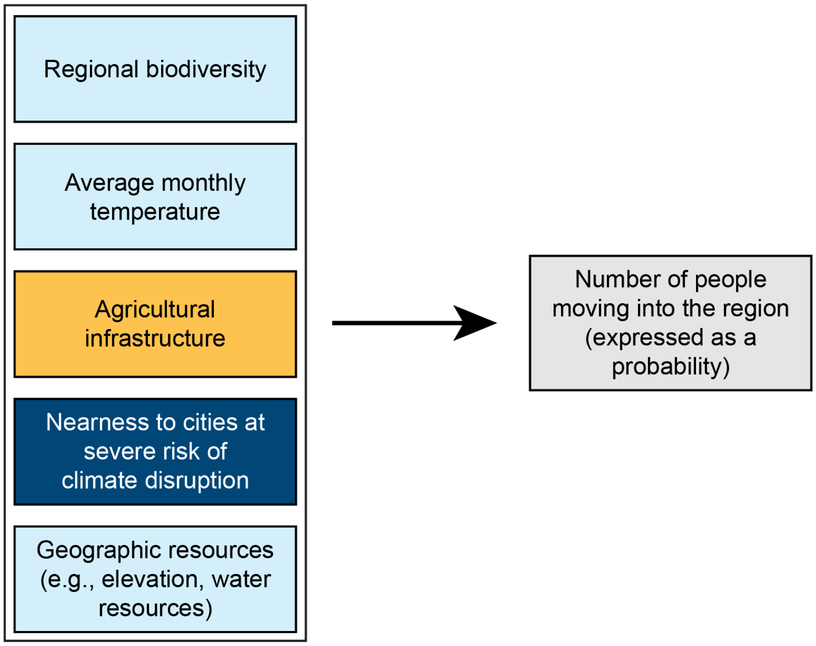
Appalachia is considered “climate resilient.” This means that the area can successfully handle the impacts of changes to climate and can prevent those impacts from growing worse. The green areas in Figure 1 show where Appalachia is located in the United States.

Figure 1. Map of Appalachia

Computational models predict that many people will move into the Appalachian region over the next 20 years as they seek to find places to live that are safer and more stable.

Figure 2 shows one model local leaders are using to predict how many people will move into Appalachia. Blue indicates factors that are expected to increase migration, and orange indicates factors expected to decrease migration. Darker colors indicate more weight on that factor in the model. Agricultural infrastructure includes farms, markets and businesses that support farms, and the transportation and communication systems in the area.

Figure 2. Computational Model for Predicting Human Migration



When this model was tested against recent population growth due to migration into two locations in Appalachia, leaders noticed some differences between what the model predicted and what the data showed. The table shows these differences for the two locations, along with information about how high or low each location is rated on several factors.

Predicted and Actual Population Growth in Pittsburgh and the Shenandoah Valley

Location	Pittsburgh	Shenandoah Valley
Predicted population growth	high	low
Actual population growth	low	high
Relative biodiversity	low	high
Average monthly temperature range	29–73° F	32–74° F
Relative agricultural infrastructure	low	medium
Nearness to climate-impacted urban centers	high	high
Access to usable water	medium	high

Item 1
Based on your understanding of what human societies need to be successful, which idea **best** explains why the model prediction was different from the observed data?

A. Pittsburgh has highly limited access to water. The model did not account for how this shortage would limit the growth of local businesses.
B. Human societies require reasonable annual temperature ranges. The model incorrectly assumed that average monthly temperature was a less important factor.
C. Human societies require access to food and water. The model did not account for how important access to food and water would be for human migration to Appalachia.
D. The Shenandoah Valley has very little access to fresh fruits and vegetables. The model incorrectly assumed that agricultural infrastructure would not promote migration.

Key: C

Item 2
Human migration to areas such as Appalachia can result in rural gentrification. People who are currently living in low-cost, natural resource–rich areas are forced to leave, resulting in the loss of access to the resources and communities they have actively contributed to developing and maintaining. As a result, they can no longer enjoy the benefits of these regions.

Describe **one** way you could revise or build on the computational model to better understand how migration into Appalachia could impact current residents' access to natural resources. Be sure to explain how the change you describe will provide a better understanding of how migration into Appalachia could impact current residents' access to natural resources.

Exhibit A.42. Human Migration to Appalachia Item 2 Constructed Response Scoring Notes

Emphasis here is on an understanding of how to revise the model parameters to better understand more nuanced population-environment dynamics. Student ideas can be wide-reaching but should be justifiable as at least one of the following: (a) updating the model to better understand who moves into and out of the region, and/or (b) determining the feedback impacts of population growth on biodiversity, and natural resources such as water. This item specifically focuses on the computational reasoning aspect of this SEP and can include either quantitative or qualitative reasoning from students. Appropriate lines of reasoning here can include, but are not limited to, the following:

- updating the model to account for demographic subgroups
- considering housing costs / other metrics for socioeconomic status as part of the inputs and/or outputs, as a mechanism to better understand the relative wealth/characteristics of who lives in the area
- calculations of migration out of Appalachian regions
- feedback mechanisms that influence biodiversity and natural resource availability (Note: This could be specific to particular resources, or general at the level of the categories included in the model.)
- relative factors for scaling variables (quantitative or qualitative)

APPENDIX B: Special Studies

On occasion, NAEP carries out special studies to inform future assessments. Four special study proposals ranked as top priority for the NAEP Science Assessment:

- cultural relevance of science scenarios intended to tap into specific cultural features
- factors increasing the accessibility of items to students with lower science achievement
- dimensionality of assessment with three distinct science constructs with engineering interwoven
- exploring the use of process data as evidence of science achievement

Each study contributes to a further understanding of science assessment.

Cultural Relevance. Guidance in the Framework encourages item writers to include culturally relevant context in the item stems. One of the sample items, for instance, focuses on a type of seaweed native to the waters in Hawaiʻi. When items are tagged as culturally relevant, they should be researched to verify that the content is, in fact, relevant to the intended culture, does not touch on negative stereotypes, and is not unduly burdensome to anyone outside of that culture.

To determine how the inclusion of culturally relevant context affects students both in and outside of the targeted culture, this study could be conducted using cognitive labs during item development or through surveys during item tryouts. Students could be probed about their reactions to scenarios that were designed to be culturally relevant to a particular demographic group. Research questions include the following:

- Did you find [this scenario] interesting?
- Was there anything about [this item/task] that made you especially interested in figuring out the answer?
- Was there anything about [this item/task] that was hard to understand or that made the science more difficult?
- Have you ever encountered a situation like the scenario described in [this item/task]?

The outcome of this study would influence the inclusion of additional items that highlight culturally relevant context and help us better understand the features that engage students versus those that create unintended barriers.

Item Accessibility. Another special study should be conducted to ensure students at the lower end of the scale can answer items as intended. Focusing on students who score near the lowest cut score in 4th and 8th grade science on their state assessments and administering the NAEP items intended to measure skills identified in the *Basic* achievement level would help determine the degree to which items tap into the knowledge and skills of students at the lower end of the achievement continuum. Listening to students think aloud while answering these items would provide insights into how they are tapping into their understanding of science.

The most important element of this work is understanding what makes a science item accessible to lower achieving students. The complexity framework indicates that the amount of cueing provided should be related to difficulty, with more cueing leading to higher p-values. On the other hand, cueing could increase the reading load. The ALDs were written to demonstrate how phenomena can also be more or less complex, thereby also influencing difficulty. And the three dimensions of disciplinary concepts, science and engineering practices, and crosscutting concepts can also vary in complexity in ways that should influence difficulty.

With all of these factors potentially increasing or limiting access to items, it will be important to further tease out the relationship between these layers of complexity and item difficulty. Multiple statistical analyses will help determine elements that seem most related to difficulty, but ultimately, this question will need to be explored through student interviews. Using cognitive lab protocols, after the think-aloud portion, students can be probed on elements of items and task sets that they found challenging or that helped them answer the item. These results, in turn, can be used to further refine item specifications.

Likewise, better understanding how students are responding to the items and whether the test should allow for multiple representations of responses, particularly in constructed response items might result in more culturally responsive rubrics. For example, if different students are showing the same knowledge in different formats (e.g., written response versus a drawing of the model), they should receive equal credit. Understanding how students approach the task will help test developers be appropriately inclusive in the rubric development and scoring process.

Engineering. The decision was made early in the Framework development to incorporate Engineering into the three primary science domains where appropriate. However, a special study should be conducted to determine if engineering might actually be a separate construct. Once sufficient data are collected through a test administration, a confirmatory factor analysis (CFA) should be run to ensure the three factors load as intended on the physical sciences, life sciences, and earth and space sciences constructs.

Conversely, the CFA could specify the four constructs to see if engineering appears. If all items load most heavily on the first three, that would provide evidence that engineering should not be reported separately. If, however, there is clear evidence of a fourth factor, additional simulations could be run to determine how scores might change if engineering was weighted more heavily to balance it with the other three factors and report that domain separately.

Process Data. Process data reflect all the interactions between the student and the assessment and may provide relevant evidence about whether the student can use a practice that is part of an assessment target. Thus, process data can be captured, measured, and interpreted to generate a score. Clickstream data, activity logs, text, and transcribed voice responses are among the ways to capture the state of student activity as they work through a problem. These types of data hold the potential power to measure student interactivity in carrying out investigations, modeling systems, and solving problems.

An opportunity for future NAEP Science Assessments built on this framework is to develop validated measures from process data, such as time on task, number of times a simulation is run, the frequency of returning to a task. Comparing these aspects to total score will provide insights into student engagement. Additionally, process data that looks at the actual simulation the student runs and how many they run before answering the test questions could be used eventually to model student understanding. Further research on incorporating process data into the score is needed to determine whether ability can be calculated with greater reliability and accuracy through including process data in the modeling.

APPENDIX C: Achievement Level Descriptions

The NAEP Achievement Level Descriptions (ALDs) in this appendix provide examples of what students performing at the *NAEP Basic*, *NAEP Proficient*, and *NAEP Advanced* achievement levels should know and be able to do in terms of the science disciplinary content, science and engineering practices, and crosscutting concepts identified in the framework.

The ALDs in the 2028 NAEP Science Framework have changed, relative to ALDs presented in previous frameworks. The differences reflect not only changes to the science knowledge, skills, and abilities assessed (science disciplinary content, science and engineering practices, and crosscutting concepts) but also an effort to develop ALDs that provide explicit guidance for item developers. Specifically, across grade levels, the 2028 NAEP Science Framework ALDs have changed in the following ways:

- Updates to the grade-level objectives in Chapter 2 of the framework are reflected in the content foci described in each grade-level ALD.
- The science practices from previous science frameworks have been expanded to Science and Engineering Practices (SEPs) for the 2028 NAEP Science Framework, and Crosscutting Concepts (CCCs) have been added. A new paragraph was developed to show the progression of application of the SEPs and CCCs to make sense of science phenomena.
- To provide specific and unambiguous guidance to item developers, these ALDs provide more explicit elaborations of the knowledge and skills students should demonstrate and the actions they should perform at each grade level and within each achievement level. In addition to the overall section and the section on SEPs and CCCs, the ALDs continue to be broken out by science domain: Life Science, Physical Science, and Earth and Space Sciences. The ALDs provide samples of how the SEPs and CCCs can be applied to specific concepts within each domain.

Within each grade level, the shifts from one achievement level to the next have commonalities, and the content of each achievement level can be described generally. Descriptions at each achievement level, for all grade levels, are as follows:

- Descriptions at the *NAEP Basic* achievement level focus on partial understanding of grade-appropriate concepts and simple applications of SEPs and CCCs to that content to make sense of real-world situations and common phenomena.
- Descriptions at the *NAEP Proficient* achievement level focus on solid understanding of grade-appropriate concepts and skillful application of SEPs and CCCs to that content to reason with and interpret real-world situations and phenomena.
- Descriptions at the *NAEP Advanced* achievement level focus on superior understanding of grade-appropriate concepts and expert engagement with SEPs and CCCs to that content to interpret, explain, and predict real-world situations and phenomena.

Text that elaborates on these statements is included within the ALD tables. The ALDs are organized into three sections. The first demonstrates how the DCs progress in depth of understanding from *NAEP Basic* to *NAEP Proficient* to *NAEP Advanced*. The second section illustrates the same progression for the SEPs and CCCs. However, the intent is that none of these dimensions is assessed in isolation. Therefore, the third section provides examples of the intersection of the DCs with the SEPs and CCCs. The third section is subdivided by discipline, with four examples per discipline. Although it wasn't possible to write ALDs for every possible crossing of DCs with SEPs and CCCs, the four selected represent big ideas that can also be shown to progress from grade 4 to grade 8 to grade 12.

To add clarity and specificity, the NAEP Science Assessment and Item Specifications include example performance expectations targeting each achievement level within each grade level. In Appendix A of the Assessment and Item Specifications, three items (one each for grades 4, 8, and 12), along with annotations that describe items across the achievement levels, illustrate the knowledge and skills required at different NAEP achievement levels.

Exhibit C.1. NAEP Grade 4 Science Achievement Level Descriptions

<table>
<tr><th>NAEP Basic Level</th><th>NAEP Proficient Level</th><th>NAEP Advanced Level</th></tr>
<tr><td>Students at this level should be able to demonstrate partial mastery and competency in making sense of common phenomena or designing solutions using science and engineering practices and/or crosscutting concepts together with disciplinary concepts such as:
A. different types of matter (materials) have different properties,
B. a force acting on an object at rest can move the object,
C. water and light are needed for a plant's survival,
D. the location of rocks and fossils can be used to establish Earth's history,
E. Earth surface features can be changed by natural processes, such as wind or water or living organisms,
F. humans can impact the land, water, and air where they live.</td><td>Students at this level should be able to demonstrate solid academic performance and competency in making sense of phenomena or designing solutions using science and engineering practices and/or crosscutting concepts together with disciplinary concepts such as:
A. matter (materials) can be classified based on its properties,
B. a change in motion requires unequal forces acting on an object,
C. varying amounts of water and light may affect a plant's growth,
D. fossils can provide evidence for the nature of an environment where organisms lived long ago,
E. some changes to Earth's surface features by wind, water, and living organisms can be observable,
F. humans can impact the land, water, and air around the world.</td><td>Students at this level should be able to demonstrate superior performance and competency in making sense of complex phenomena or designing solutions using science and engineering practices and/or crosscutting concepts together with disciplinary concepts such as:
A. different types of matter (materials) have multiple different properties; it is therefore necessary to consider pros and cons when selecting a material for a specific purpose,
B. two objects can each exert a force on the other; the change in motion of either object depends on all the forces that act on it,
C. some animals obtain the matter they need for growth and survival from plants or from other animals,
D. the location of fossils within rock layers can be used to show the changes that occurred to Earth and life on Earth over time,
E. small changes to Earth's surface features by water, wind, or organisms can result in large changes over time,
F. humans can impact the land, water, and air around the world in positive and negative ways.</td></tr>
</table>

NAEP Basic Level	NAEP Proficient Level	NAEP Advanced Level
Working with the disciplinary concepts, students require a well-defined set of actions to be able to apply science and engineering practices and crosscutting concepts such as: A. asking a simple question about an observation, B. using models to describe a phenomenon or describe a design proposal, C. using a diagram to identify one way that changes might affect a phenomenon, D. identifying an evidence-based argument, E. describing quantitative evidence related to a design problem, F. using evidence to support the solution to a design problem while considering the criteria that the solution should meet, G. developing an evidence-based description of a phenomenon.	Working with the disciplinary concepts, students require some cueing to be able to apply science and engineering practices and crosscutting concepts such as: A. asking questions about observed data to aid in interpretation, B. describing how the parts and relationships in a model represent a phenomenon or proposed design solution, C. describing observations or measurements that can be used as evidence to explain a phenomenon, D. evaluating the merits of an evidence-based argument, E. organizing data to reveal patterns that can be used to solve a design problem, F. making a claim about the solution to a design problem using evidence while considering criteria and constraints, G. using patterns in information to develop an evidence-based explanation of a phenomenon.	Working with the disciplinary concepts, students require limited cueing to be able to apply science and engineering practices and crosscutting concepts such as: A. using questions as a tool to clarify an argument or investigate a problem, B. identifying the limitations of a model used to represent a phenomenon or proposed design solution, C. predicting the outcome of an experiment designed to explore changes to a phenomenon, D. comparing evidence-based arguments about the changes of a system based on the evidence or the reasoning the arguments include, E. estimating or predicting data points using patterns in recorded data to solve a design problem, F. proposing a solution to a design problem using evidence to help ensure it will meet criteria and constraints, G. developing an evidence-based explanation of a phenomenon supported by reasoning about cause-and-effect relationships.

NAEP Basic Level	NAEP Proficient Level	NAEP Advanced Level
In **Physical Science**, students should be able to integrate disciplinary concepts, science and engineering practices, and crosscutting concepts to engage in the *simple sensemaking* of common phenomena when provided *a well-defined set of actions* to perform tasks such as: A. using evidence to describe that temperature affects the physical state of a material (solid vs. liquid), B. proposing appropriate variables and tests when planning an investigation to determine whether objects exert a force on each other when they collide, C. organizing data to identify a pattern between how fast an object moves and its energy, D. asking questions to clarify the relationship between the parts of a simple model used to represent that an object can be seen when light produced by the object or reflected from its surface enters the eyes.	In **Physical Science**, students should be able to integrate disciplinary concepts, science and engineering practices, and crosscutting concepts to engage in the *moderate degree of sensemaking* of phenomena when provided *some cues* to perform tasks such as: A. making a claim using data about the relationship between temperature and the physical state of a material (solid vs. liquid), B. planning an investigation considering the variables to control and/or the number of trials to conduct to produce data about whether objects exert forces on each other when they are touching or colliding, C. describing patterns in data to support the claim that the motion of an object is related to its energy, D. asking questions to evaluate a model that represents that objects can be seen only when light produced by the object or reflected from its surface enters the eyes.	In **Physical Science**, students should be able to integrate disciplinary concepts, science and engineering practices, and crosscutting concepts to engage in the *high degree of sensemaking* of complex phenomena when provided *limited cues* to perform tasks such as: A. evaluating a claim based on the evidence or reasoning it includes about how a material might function differently at a different temperature due to the physical state of the material (solid vs. liquid) at that temperature, B. evaluating an investigation plan designed to explore a scientific question about the way that objects exert forces on each other when they touch or collide, C. analyzing patterns in data gathered from the motion of two different objects to support claims about the relationship between the motion and speed of objects and their energy, D. asking questions to identify the limitations of a model that represents the conditions under which various objects and materials can be seen.

NAEP Basic Level	NAEP Proficient Level	NAEP Advanced Level
In **Life Science**, students should be able to integrate disciplinary concepts, science and engineering practices, and crosscutting concepts to engage in the *simple sensemaking* of common phenomena when provided *a well-defined set of actions* to perform tasks such as: A. evaluating textual information to determine whether it is related to given data about the similarities and differences between the life cycles of different animals, B. using data to describe that some of the matter that an organism needs to survive comes from other organisms, C. supporting a claim that inherited characteristics can affect what an organism looks like, D. identifying evidence for an argument about characteristics of an individual providing advantages in surviving.	In **Life Science**, students should be able to integrate disciplinary concepts, science and engineering practices, and crosscutting concepts to engage in the *moderate degree of sensemaking* of phenomena when provided *some cues* to perform tasks such as: A. evaluating whether textual information accurately summarizes data about the diverse life cycles of plants or animals, B. analyzing data that can be used to support a claim that much of the matter organisms need to grow and survive comes from other organisms, C. making an evidence-based claim about the relationship between organisms' characteristics and their look and function, D. supporting an argument with evidence about different characteristics of individuals providing advantages in surviving and finding mates.	In **Life Science**, students should be able to integrate disciplinary concepts, science and engineering practices, and crosscutting concepts to engage in the *high degree of sensemaking* of complex phenomena when provided *limited cues* to perform tasks such as: A. evaluating whether textual information summarizing a table of data comparing the life cycles of different plants or animals accurately reflects a claim about the essential nature of reproduction for all organisms, B. analyzing two different sets of data to determine which one can be used to support a claim that much of the matter organisms need to grow and survive comes from other organisms and that same matter is used again later by other organisms, C. evaluating the evidence to support or reject various claims about whether the characteristics of an organism are inherited, result from interactions with the environment, or both, D. evaluating evidence and reasoning of arguments about whether different characteristics of individuals can provide advantages in surviving and finding mates when a habitat changes.

NAEP Basic Level	NAEP Proficient Level	NAEP Advanced Level
In **Earth and Space Sciences**, students should be able to integrate disciplinary concepts, science and engineering practices, and crosscutting concepts to engage in the *simple sensemaking* of common phenomena when provided *a well-defined set of actions* to perform tasks such as: A. using a model to describe that objects in the sky are not always visible from Earth, B. identifying evidence to support arguments that the Earth's surface was different in the past, C. analyzing patterns in data to describe a weather event that occurred in a region, D. identify evidence that can be used to support a claim about how natural processes can cause hazards for humans in some areas.	In **Earth and Space Sciences**, students should be able to integrate disciplinary concepts, science and engineering practices, and crosscutting concepts to engage in the *moderate degree of sensemaking* of phenomena when provided *some cues* to perform tasks such as: A. developing a model to describe that objects in the sky are not always visible due to Earth's rotation, B. making arguments based on evidence about differences of the surface of the Earth between the present and in the past, C. analyzing patterns in data to describe the kind of weather expected in a region, D. supporting an argument with evidence for how natural processes can cause hazards for humans in some areas but not others.	In **Earth and Space Sciences**, students should be able to integrate disciplinary concepts, science and engineering practices, and crosscutting concepts to engage in the *high degree of sensemaking* of complex phenomena when provided *limited cues* to perform tasks such as: A. developing and using a model to predict which objects in the sky may not be visible at certain points of Earth's rotation, B. evaluating an argument based on the evidence or reasoning it includes about changes to the surface of the Earth and life on Earth over time, C. analyzing patterns in data to determine whether the data support or refute predictions about the kind of weather expected in a region, D. constructing an argument with evidence about a cause-and-effect relationship between natural and human-designed processes and hazards that occur for humans in some areas.

Exhibit C.2. NAEP Grade 8 Science Achievement Level Descriptions

NAEP Basic Level	NAEP Proficient Level	NAEP Advanced Level
Students performing at this level should be able to demonstrate partial mastery and competency in making sense of common phenomena or designing solutions using science and engineering practices and/or crosscutting concepts together with disciplinary concepts such as: A. in solids atoms are close together, in liquids atoms are close together but are moving relative to one another, and in gasses the atoms are relatively far apart, B. an object at rest will remain at rest if the net force on it is zero, C. photosynthetic organisms use energy from light to make food, D. the fossil record documents the existence and extinction of many life-forms throughout Earth's history of life, E. the movement of water within the water cycle is a function of phase changes and downhill movement, F. human activities have caused changes in the areas where they live, bringing about major changes in the land, water, and air.	Students performing at this level should be able to demonstrate solid academic performance and competency in making sense of phenomena or designing solutions using science and engineering practices and/or crosscutting concepts together with disciplinary concepts such as: A. in solids atoms are close together, in liquids atoms are close together but are moving relative to one another and, in gasses the atoms are relatively far apart and move freely and movement of atoms are influenced by temperature and move freely in gasses, B. the change in motion of an object is determined by the sum of the forces acting on it, C. photosynthetic organisms use energy from light, carbon dioxide and water to make sugars and release oxygen, D. the fossil record documents the existence, diversity, extinction, and change of many life-forms throughout the history of the Earth, supporting inferences of lines of evolutionary descent, E. the movement of water within the water cycle is a function of phase changes and is driven by gravity,	Students performing at this level should be able to demonstrate superior performance and competency in making sense of complex phenomena or designing solutions using science and engineering practices and/or crosscutting concepts together with disciplinary concepts such as: A. in solids atoms are close together vibrating in place, in liquids atoms are close together but are moving relative to one another, and in gasses, the atoms are relatively far apart and move rapidly and freely and the higher the temperature, the more rapid the movement, B. the greater the mass of the object, the greater the force needed to achieve the same change in motion, C. through a series of chemical reactions that release energy photosynthetic organisms use energy from light, carbon dioxide, and water to form new molecules, sugars to support growth and oxygen which is released to the atmosphere, D. the changes in life forms documented through the fossil record can be used to reconstruct an evolutionary history for organisms. E. the movement of water within the water cycle is driven by gravity and

<table>
<tr><th>NAEP Basic Level</th><th>NAEP Proficient Level</th><th>NAEP Advanced Level</th></tr>
<tr><td></td><td>F. human activities have significantly altered the biosphere, atmosphere, and geosphere, sometimes causing the extinction of many organisms.</td><td>energy from the sun and continually cycles, changing the land on and below the Earth’s surface,
F. human activity can significantly alter the biosphere, atmosphere, and geosphere, and human choices can increase or decrease harm to organisms and the environment.</td></tr>
<tr><td>Working with the disciplinary concepts, students require a well-defined set of actions to be able to apply science and engineering practices and crosscutting concepts such as:
A. asking questions that arise from observations of phenomena,
B. using a model to describe a phenomenon,
C. evaluating whether a simple experimental design would meet the goals of an investigation,
D. identifying evidence to support an argument for a proposed model or explanation of a phenomenon,
E. applying simple mathematical concepts (such as basic operations and simple computations) to scientific questions or designed problems,
F. using graphical displays of data to identify relationships between variables,</td><td>Working with the disciplinary concepts, students require some cueing to be able to apply science and engineering practices and crosscutting concepts such as:
A. asking questions to clarify or refine the explanation for a phenomenon,
B. developing and using a model to explain a phenomenon by identifying relationships among parts and or quantities in a system,
C. planning an experimental design to produce data that can be used as evidence that meets the goals of an investigation,
D. using evidence to support an argument for a proposed model or explanation of a phenomenon,
E. applying simple mathematical concepts (such as ratios or proportional thinking) to scientific questions or designed problems,
F. constructing graphical displays to identify relationships (linear vs. nonlinear) between variables,</td><td>Working with the disciplinary concepts, students require limited cueing to be able to apply science and engineering practices and crosscutting concepts such as:
A. asking questions that can be answered with empirical evidence to refine an explanation of cause-and-effect relationships in phenomena,
B. revising a model to explain a phenomenon by identifying relationships among parts and or quantities in a system,
C. revising an experimental design to help ensure it produces data that can be used as evidence that meets the goals of an investigation,
D. revising an argument with evidence for a proposed model or explanation of a phenomenon,
E. applying mathematical concepts (such as ratios, rates, or percent) to scientific questions or designed problems,</td></tr>
</table>

NAEP Basic Level	NAEP Proficient Level	NAEP Advanced Level
G. describing a solution to a design problem while considering prioritized criteria, H. identifying flaws in a science-related argument in text (poor assumptions).	G. evaluating a solution to a design problem while considering criteria and constraints, H. identifying and critiquing flaws in science-related arguments in text (faulty explanations or overgeneralizations from limited data).	F. evaluating the limitation of data presented in graphical displays of data to identify causal versus correlational relationships between variables, G. evaluating the merits of a solution to a design problem using evidence while considering criteria and constraints, H. identifying and critiquing flaws in science-related arguments in text (cause vs. correlation).
In **Physical Science**, students should be able to integrate disciplinary concepts, science and engineering practices, and crosscutting concepts to engage in *simple sensemaking* of common phenomena when provided *a well-defined set of actions* to perform tasks such as: A. using a simple model to describe patterns associated with the position and movement of atoms relative to one another in solids, liquids, or gasses. B. evaluating whether a simple experimental design provides evidence that the net force on an object is zero when an object is at rest, C. using graphical displays of data on a pair of interacting objects to identify the relationship between the force	In **Physical Science**, students should be able to integrate disciplinary concepts, science and engineering practices, and crosscutting concepts to engage in the *moderate degree of sensemaking* of phenomena when provided *some cues* to perform tasks such as: A. developing a model to describe the patterns associated with the position and motion of atoms relative to one another in solids, liquids, or gasses and how temperature influences the position and movement of atoms, B. planning an experimental design to produce data that can be used as evidence that the change in motion of an object is determined by the sum of the forces on it, C. constructing graphical displays of data to showcase a relationship	In **Physical Science**, students should be able to integrate disciplinary concepts, science and engineering practices, and crosscutting concepts to engage in the *high degree of sensemaking* of complex phenomena when provided *limited cues* to perform tasks such as: A. revising a model-based explanation to describe the patterns associated with the position and motion of atoms relative to one another in solids, liquids, or gasses, including the cause-and-effect relationship between temperature and states of matter, B. revising an experimental design to produce data that can serve as evidence that the greater the mass of the object, the greater the force

NAEP Basic Level	NAEP Proficient Level	NAEP Advanced Level
exerted by the first object on the second object and the force exerted by the second object on the first object, D. 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.	between the forces between any pair of interacting objects, D. asking questions to refine an explanation for how the material an object is made of influences whether light shining on the object is reflected or transmitted.	needed to achieve the same change in motion, C. constructing graphical displays of data to showcase the causal or correlational relationship between changes in motion of any pair of interacting objects due to the fact the force exerted by the first object on the second and that exerted by the second on the first are second object are equal in strength but in the opposite direction, D. asking questions that can elicit empirical evidence to refine an explanation for how the material an object is made of, or the frequency (color) of the light, influences the reflection, absorption, or transmission of the light shining on the object.
In **Life Science**, students should be able to integrate disciplinary concepts, science and engineering practices, and crosscutting concepts to engage in *simple sensemaking* of common phenomena when provided *a well-defined set of actions* to perform tasks such as: A. identifying weak assumptions in an argument about the species involved in predatory interactions in an ecosystem,	In **Life Science**, students should be able to integrate disciplinary concepts, science and engineering practices, and crosscutting concepts to engage in the *moderate degree of sensemaking* of phenomena when provided *some cues* to perform tasks such as: A. identifying and critiquing faulty explanations or overgeneralizations from limited data in an argument about the impact of predatory actions on a population of organisms,	In **Life Science**, students should be able to integrate disciplinary concepts, science and engineering practices, and crosscutting concepts to engage in the *high degree of sensemaking* of complex phenomena when provided *limited cues* to perform tasks such as: A. identifying and critiquing flaws in scientific arguments about patterns of predatory interactions across various ecosystems,

NAEP Basic Level	NAEP Proficient Level	NAEP Advanced Level
B. using a simple model to explain that matter is transferred between producers, consumers, and decomposers, C. identifying evidence to support an argument about how variations in inherited traits between parent and offspring arise from the subset of genes inherited, D. using evidence to support an explanation about the changes a population undergoes over time in response to a change in physical components of an ecosystem.	B. developing and using a model to explain that matter and energy are transferred between producers, consumers, and decomposers, C. revising an argument using new evidence about how variations in inherited traits between parent and offspring arise from the subset of genes inherited, D. constructing an explanation that uses a chain of cause-and-effect associations between the changes a population undergoes over time and changes in physical or biological components of that population's ecosystem.	B. revising a proposed model to explain that the atoms that make up the organisms in an ecosystem are cycled repeatedly between the living and nonliving parts of the ecosystem, C. revising an argument using new evidence to support or reject an explanation that genetic mutations may result in changes in the structure and function of the proteins encoded by genes, D. revising an explanation that uses a chain of cause-and-effect associations between the changes a population undergoes over time in response to changes in physical or biological components of an ecosystem, therefore helping or hurting the health of the ecosystem, including biodiversity.
In **Earth and Space Sciences**, students should be able to integrate disciplinary concepts, science and engineering practices, and crosscutting concepts to engage in *simple sensemaking* of common phenomena when provided *a well-defined set of actions* to perform tasks such as: A. using a simple model to describe an observable pattern in the motion of an object in the sky relative to Earth, B. making a claim about the relative timing of major events in Earth's	In **Earth and Space Sciences**, students should be able to integrate disciplinary concepts, science and engineering practices, and crosscutting concepts to engage in the *moderate degree of sensemaking* of phenomena when provided *some cues* to perform tasks such as: A. developing and using a model to explain the observable patterns in the motion of objects in the sky relative to Earth,	In **Earth and Space Sciences**, students should be able to integrate disciplinary concepts, science and engineering practices, and crosscutting concepts to engage in the *high degree of sensemaking* of complex phenomena when provided *limited cues* to perform tasks such as: A. 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,

NAEP Basic Level	NAEP Proficient Level	NAEP Advanced Level
history based on the sequence of rock strata, C. interpreting graphical displays of data to identify a relationship between the sunlight, the ocean, and the weather patterns in a given location, D. identifying evidence that predicts future natural hazards through the patterns that precede those hazardous events.	B. identifying evidence that can be used to refute a claim about the relative timing of major events in Earth's history based on the sequence of rock strata and fossil records, C. constructing graphical displays of data to describe relationships between sunlight, the ocean, the atmosphere, ice, or landforms and the weather patterns in a given location, D. constructing an argument using evidence for how observable phenomena that precede the occurrence of some natural hazards can help forecast future events.	B. evaluating evidence used to refute a claim about the relative timing of major events in Earth's history based on fossil records and rock types, C. constructing graphical displays of data to explain relationships between sunlight, the ocean, the atmosphere, ice, and landforms and the weather patterns in a given location, D. evaluating arguments about minimizing the risk from hazards using evidence from observable phenomena that precede the occurrence of some natural hazards.

Exhibit C.3. NAEP Grade 12 Science Achievement Level Descriptions

NAEP Basic Level	NAEP Proficient Level	NAEP Advanced Level
Students performing at this level should be able to demonstrate partial mastery and competency in making sense of common phenomena or designing solutions using science and engineering practices and/or crosscutting concepts together with disciplinary concepts such as: A. all matter is made of atoms, B. the mass and speed of a moving object determine its momentum, C. photosynthesis converts light energy to stored chemical energy, D. DNA sequences vary among species, but there are many sequence similarities, E. the decay of radioactive isotopes in rocks provides a way to date rock formations, F. water's unique properties include expanding upon freezing, G. humans can mitigate the negative impacts on the environment resulting from the use of Earth's resources by applying engineering solutions.	Students performing at this level should be able to demonstrate solid academic performance and competency in making sense of phenomena or designing solutions using science and engineering practices and/or crosscutting concepts together with disciplinary concepts such as: A. all matter is made of atoms that contain protons, neutrons, and electrons, B. momentum is always conserved, C. photosynthesis converts light energy to stored chemical energy by converting carbon dioxide plus water into sugars plus released oxygen, D. DNA sequences vary among species, but there are many overlaps, providing evidence of evolution, E. 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 early history, F. water's unique properties include expanding upon freezing, dissolving and transporting solid material, and separating different chemical elements, G. humans can mitigate the negative impacts on the environment resulting from the use of Earth's resources and	Students performing at this level should be able to demonstrate superior performance and competency in making sense of complex phenomena or designing solutions using science and engineering practices and/or crosscutting concepts together with disciplinary concepts such as: A. the electrostatic forces between subatomic particles explain both the structure of isolated atoms and why atoms combine to form molecules, compounds, and extended materials, B. 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, C. photosynthesis converts light energy to stored chemical energy by converting carbon dioxide plus water into sugars, which have more chemical bonds than does carbon dioxide, plus released oxygen, D. evidence of evolution includes overlaps of DNA sequences among species, similarities/differences in amino acid sequences, and anatomical and embryological evidence, E. the decay of radioactive isotopes in rocks from Earth, moon rocks, and

NAEP Basic Level	NAEP Proficient Level	NAEP Advanced Level
	waste disposal by applying engineering and design solutions.	meteorites provides a way to date rock formations that can be used as evidence for Earth's formation and early history, F. interactions between the atmosphere, hydrosphere, and geosphere are influenced by water's unique properties, including expanding upon freezing, dissolving and transporting solid materials, and separating different chemical elements, G. humans can mitigate negative impacts on the environment from resource extraction and waste disposal by applying engineering and design solutions, but when the sources of such problems are not well understood, some actions could magnify the problems.
Working with the disciplinary concepts, students require a well-defined set of actions to be able to apply science and engineering practices and crosscutting concepts such as: A. asking a question that arises from examining a model to clarify the model, B. using a simple model of a system that includes a mathematical relationship to describe phenomena, C. planning an investigation that will produce data that can support the	Working with the disciplinary concepts, students require some cueing to be able to apply science and engineering practices and crosscutting concepts such as: A. asking a question that arises from examining a model to identify additional needed information, B. developing a simple model of a system that includes scale, proportion, and other mathematical relationships to describe phenomena,	Working with the disciplinary concepts, students require limited cueing to be able to apply science and engineering practices and crosscutting concepts such as: A. asking multiple questions that arise from examining a model to identify all needed additional information, B. revising and using a model of a system that includes scale, proportion, and other mathematical relationships to explain phenomena,

NAEP Basic Level	NAEP Proficient Level	NAEP Advanced Level
scientific explanation of a phenomenon, D. using evidence to support an argument about a proposed explanation for structure-function relationships in a system, E. applying simple statistical reasoning to represent design problems, F. analyzing data to identify evidence that could support a model, G. identifying an explanation of a phenomenon that uses a chain of evidence-based associations between variables, H. identifying a flaw in a science-related argument.	C. 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, D. constructing an argument with evidence and scientific reasoning to support a proposed explanation for structure-function relationships in a system, E. applying statistical reasoning to solve design problems, F. analyzing patterns in data to provide evidence to support or reject a model, G. constructing an explanation of a phenomenon that uses a chain of evidence-based associations between variables, H. identifying and critiquing a flaw related to overgeneralization from limited data in a science-related argument.	C. evaluating the design of an investigation intended to produce data that can be used as evidence for cause-and-effect relationships in a phenomenon, considering possible confounding variables, D. revising an argument to support or reject a proposed explanation for structure-function relationships in a system, addressing new evidence, E. applying statistical reasoning in the context of complicated measurement problems to represent and solve design problems, F. analyzing patterns in data to provide evidence of cause-and-effect relationships that could support or reject a model, G. revising an explanation of a phenomenon that uses a chain of cause-and-effect associations between factors to account for relationships between variables, H. critiquing science-related arguments by identifying multiple flaws, including flaws related to overgeneralizations from limited data.
In **Physical Science**, students should be able to integrate disciplinary concepts, science and engineering practices, and crosscutting concepts to engage in *simple sensemaking* of	In **Physical Science**, students should be able to integrate disciplinary concepts, science and engineering practices, and crosscutting concepts to engage in the *moderate degree*	In **Physical Science**, students should be able to integrate disciplinary concepts, science and engineering practices, and crosscutting concepts to engage in the *high degree of*

NAEP Basic Level	NAEP Proficient Level	NAEP Advanced Level
common phenomena when provided *a well-defined set of actions* to perform tasks such as: A. evaluating whether data could support a relationship between temperature and pressure in a gas, at fixed volume, B. identifying relevant and irrelevant variables in the design of an investigation about the relationship between the forces acting on an object and the change in motion of the object, C. identifying scientific questions that arise from examining an explanation of the relationship between the energy within a system and the motion and interactions of matter and radiation within that system, D. using models to compare visible light with X-rays.	*of sensemaking* of phenomena when provided *some cues* to perform tasks such as: A. evaluating whether data are sufficient in quantity to support a qualitative statement of the relationship between temperature and pressure of a gas at fixed volume, B. planning an investigation that will produce data to serve as the basis for evidence about the relationship between the sum of two aligned forces acting on an object and the change in motion of the object, C. asking questions that arise from examining an explanation related to the way the energy available within a system depends on the motion and interactions of matter and radiation within that system, D. developing models of the mathematical relationships between visible light and X-rays.	*sensemaking* of complex phenomena when provided *limited cues* to perform tasks such as: A. evaluating whether data are sufficient in quantity, accuracy, and reliability to support or reject a claim that there is a relationship between temperature and pressure or between temperature and the rate of chemical reactions in gasses, at fixed volume, B. 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, considering possible confounding variables, C. evaluating questions that arise from examining a model that illustrates that the quantity of energy available for processes within a system depends on the motion and interactions of matter and radiation within that system, D. revising models of the mathematical relationships between visible light, X-rays, and radio waves.

NAEP Basic Level	NAEP Proficient Level	NAEP Advanced Level
In **Life Science**, students should be able to integrate disciplinary concepts, science and engineering practices, and crosscutting concepts to engage in *simple sensemaking* of common phenomena when provided *a well-defined set of actions* to perform tasks such as: A. describing the merits of a simulation of an ecosystem that illustrates that not all of the matter consumed at one level of a food web is transferred to other levels in a food web, B. revising an argument that genes code for the formation of proteins that determine traits, C. evaluating whether the quantity of the data are sufficient to support an explanation about whether natural selection can result from competition for resources, D. identifying the merits and limitations of a model of cell division.	In **Life Science**, students should be able to integrate disciplinary concepts, science and engineering practices, and crosscutting concepts to engage in the *moderate degree of sensemaking* of phenomena when provided *some cues* to perform tasks such as: A. evaluating a simulation of an ecosystem that illustrates that only a small fraction of the matter consumed at lower levels of a food web is transferred to upper levels in a food web, B. revising an argument that genes code for the formation of proteins, which determine traits, addressing new evidence about DNA sequences that do not code for a protein, C. evaluating whether the quantity and accuracy of the data used as evidence are sufficient to support an explanation about the four factors that can cause natural selection, D. evaluating the merits and limitations of two different models of the reproduction of genetic information in mitosis to select a model that best fits the evidence.	In **Life Science**, students should be able to integrate disciplinary concepts, science and engineering practices, and crosscutting concepts to engage in the *high degree of sensemaking* of complex phenomena when provided *limited cues* to perform tasks such as: A. evaluating simulations of an ecosystem by comparing the outcomes of the simulation with what is known about the problems in the ecosystem that result from the fact that only a small fraction of the matter consumed at lower trophic levels is transferred to upper trophic levels in a food web to produce growth and release energy at the higher level, B. revising an argument to support or reject an explanation about the various functions of genes and whether all genes code for the formation of proteins that determine traits, C. evaluating whether the quantity, accuracy, and reliability of the data used as evidence are sufficient to support an explanation about the interaction of the four factors that can cause natural selection, D. evaluating the merits and limitations of two different models of the role of mitosis in the growth of an organism

NAEP Basic Level	NAEP Proficient Level	NAEP Advanced Level
		to revise the model that best fits the evidence.
In **Earth and Space Science**s, students should be able to integrate disciplinary concepts, science and engineering practices, and crosscutting concepts to engage in *simple sensemaking* of common phenomena when provided *a well-defined set of actions* to perform tasks such as: A. analyzing data to identify evidence that changes in the orientation of Earth's axis of rotation affects the amount of sunlight falling on the planet, B. identifying evidence that the decay of radioactive isotopes in rocks provides a way to date rock formations, C. using models to describe ways that ocean and atmospheric circulations influence climate, D. identifying evidence that the size of a human population has been affected by a natural hazard.	In **Earth and Space Sciences**, students should be able to integrate disciplinary concepts, science and engineering practices, and crosscutting concepts to engage in the *moderate degree of sensemaking* of phenomena when provided *some cues* to perform tasks such as: A. analyzing data to support an explanation for changes in the shape of Earth's orbit and the orientation of its axis of rotation driving changes in the amount of sunlight falling on the planet, B. evaluating evidence that uses measurements of the decay of radioactive elements in minerals and rocks to provide evidence of Earth's early history, C. using system models that include mathematical relationships to describe how ocean and atmospheric circulations influence temporal and spatial patterns on Earth's climate, D. developing an argument that the size and location of a human population has been affected by natural hazards.	In **Earth and Space Sciences**, students should be able to integrate disciplinary concepts, science and engineering practices, and crosscutting concepts to engage in the *high degree of sensemaking* of complex phenomena when provided *limited cues* to perform tasks such as: A. analyzing global temperature data to reveal patterns that can support an explanation for changes in the shape of Earth's orbit and the orientation of its axis of rotation altering the intensity and distribution of sunlight falling on the planet, B. evaluating evidence from multiple sources using measurements of the decay of radioactive elements in minerals and rocks to provide evidence for Earth's formation and early history, C. developing system models that employ mathematical relationships to describe ways that the absorption, reflection, storage, and redistribution of energy from the sun lead to temporal and spatial patterns in Earth's climate system, D. evaluating an argument that feedback mechanisms could magnify or mitigate the intensity of the effect

NAEP Basic Level	NAEP Proficient Level	NAEP Advanced Level
		on a human population by a natural hazard.

APPENDIX D: 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*.

achievement level descriptions (ALDs): Descriptions of students' performance at achievement levels (*NAEP Basic*, *NAEP Proficient*, and *NAEP Advanced*), detailing what students should know and be able to do in terms of the science disciplinary concepts, science and engineering practices, and crosscutting concepts on the NAEP assessment.

alignment: The coordination of goals, instruction, and assessment in a mutually reinforcing educational system.

argument: A process of evaluating competing claims, models, or explanations of a phenomenon based on available evidence (in science) or to a process of evaluating prospective designs based on specifications and constraints (in engineering).

construct: An image, idea, or theory, especially a complex one formed from a number of simpler elements and often embedded in a web of related ideas.

constructed response (CR): An open-ended, text-based response. Every constructed response item has a scoring guide that defines the criteria used to evaluate students' responses.

Constructed response item types that may be used on the 2028 NAEP Science Assessment are listed below.

- **Short constructed response** – Students respond by giving a short response, from a single word or number to a few sentences.
- **Extended constructed response** – Students respond by giving a description or explanation that requires more than a few words.

context: 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.

contextual variables: Students, teacher, administrator, and school factors that shape students' opportunities to learn, including instructional time, content, strategies, and resources.

crosscutting concepts: 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.

digital tools: Any technology that stores and transmits data electronically. In a digitally based environment, students require tools to make and enter calculations, build models, run simulations, and to create and modify graphical representations of data.

dimensions: 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): 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): A single, standalone item.

English Learners: Active learners of the English language who may benefit from various types of language support programs; students from a diverse set of backgrounds who often come from non-English-speaking homes and backgrounds and who typically require specialized or modified instruction in both the English language and in their academic courses.

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

exhibit: Tables and figures in this framework. 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.

item: 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: A group of independent items that uses common stimulus material. 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 sets should include at least one item that is three-dimensional.

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: 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.

multidimensional: Items that integrate two or all three dimensions.

multipart item (MPI): An item that includes multiple 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.

opportunity to learn: Inputs and processes that enable student achievement of intended outcomes.

phenomena: Observable events that occur in either natural or human-designed systems.

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.

process data: Information collected as students navigate the digital assessment, including the time taken to engage in item stem or stimulus and respond to questions, how often they return to the stem or stimulus to answer questions, and their use of digital tools. Process data reflect interactions in which the student engages and may provide relevant evidence about whether the student possesses a skill that is an assessment target. Thus, process data can be captured, measured, and interpreted to generate a score.

scenario-based task (SBT): 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. Scenario-based tasks should include at least one item that is three-dimensional.

science achievement expectation: An assessable statement of what students should know and be able to do. Formulating a multidimensional science achievement expectation—expressed as a performance—is often the starting point in developing an assessment item.

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

selected response (SR): Assessment responses that involve a student selecting one or more response options from a given, limited set of choices.

Different types of selected response items that may be used on the 2028 NAEP Science Assessment include the following.

- **Grid** – Students evaluate statements, such as claims or explanations, or classify components of a system based on their properties or interactions. The answer is entered by selecting cells in a table.
- **Inline choice** – Students respond by selecting one option from one or more drop-down menus that may appear in various sections of an item.
- **Matching** – Students respond by inserting (i.e., dragging and dropping) one or more source elements (e.g., an image) into target fields (e.g., a table).
- **Multiple-selection multiple choice** – Students respond by selecting two or more choices that meet the condition stated in the stem of the item.
- **Single-selection multiple choice** – Students respond by selecting a single choice from a set of given choices.
- **Zones** – Students respond by selecting one or more regions on a graphic stimulus.

sensemaking: Actively applying disciplinary concepts, science and engineering practices, and crosscutting concepts to figure out a phenomenon or address a real-world problem. The degree and nature of sensemaking required by students is determined by the complexity of each item, and the degree to which each dimension contributes to that complexity in each item.

sophisticated/sophistication: For the purpose of this framework, increasing sophistication refers to a student expression of understanding that is more thorough, more precise, more accurate, and more coherent throughout.

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

stimulus: Written descriptions, images, videos, simulations, long-form texts, infographics, data tables, graphs, and all other information provided to students in a NAEP test question. Item sets and scenario-based tasks include stimuli that are shared among multiple items.

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.

REFERENCES

Abbott, I. A., & Shimazu, C. (1985). The geographic origin of the plants most commonly used for medicine by Hawaiians. *Journal of Ethnopharmacology*, *14*(2-3), 213-222.

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

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

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.

Anderson, C. W., de los Santos, E. X., Bodbyl, S., Covitt, B., Edwards, K. D., & Hancock, J. (2018). Designing educational systems to support enactment of the Next Generation Science Standards. *Journal of Research in Science Teaching, 55*(7), 1026–1052.

Brown, B. A. (2019). *Science in the city: Culturally relevant STEM education.* Harvard Education Press.

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.

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.

The Nation's Report Card. (2023). NAEP Data Explorer. Retrieved from https://www.nationsreportcard.gov/ndecore/landing

National Academies of Sciences, Engineering, and Medicine. (2017). *Seeing students learn science: Integrating assessment and instruction in the classroom.* 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.* National Academies Press. https://doi.org/10.17226/24783

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 Assessment Governing Board (Governing Board). (n.d.). *The nation's report card*. https://nces.ed.gov/nationsreportcard/about/accom_table.aspx

National Assessment Governing Board. (2002). *Item development and review policy statement*. Author, adopted May 18, 2002. https://www.nagb.gov/content/dam/nagb/en/documents/policies/Item%20Development%20and%20Review.pdf

National Assessment Governing Board. (2014). *NAEP testing and reporting on students with disabilities and English language learners policy statement*. Author, adopted March 6, 2010. Updated August 2, 2014. https://www.nagb.gov/content/dam/nagb/en/documents/policies/naep_testandreport_studentswithdisabilities.pdf

National Assessment Governing Board. (2018). *Developing student achievement levels for the national assessment of educational progress policy statement*. Author, adopted November 17, 2018. https://www.nagb.gov/content/dam/nagb/en/documents/policies/ALS-revised-policy-statement-11-17-18.pdf

National Assessment Governing Board. (2022). *NAEP assessment tutorials*. https://nces.ed.gov/nationsreportcard/experience

National Assessment Governing Board. (2023a). *NAEP accommodations*. https://nces.ed.gov/nationsreportcard/about/accom_table.aspx

National Assessment Governing Board. (2023b). *Trial Urban District Assessment (TUDA).* https://nces.ed.gov/nationsreportcard/tuda/

National Assessment of Educational Progress Authorization Act, Pub. L. 107-279, 3 U.S.C. §§ 301–305 (2002). https://www.govinfo.gov/content/pkg/PLAW-107publ279/pdf/PLAW-107publ279.pdf

National Research Council. (1996). *National Science Education Standards.* National Academies Press.

National Research Council. (2005). *How students learn: Science in the classroom.* 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*. National Academies Press. https://doi.org/10.17226/11625

National Research Council. (2012). *A framework for K–12 science education: Practices, crosscutting concepts.* National Academies Press.

National Research Council. (2015). *Guide to implementing the next generation science standards*. Washington, DC: National Academies Press.

Penuel, W. R., Harris, C. J., & DeBarger, A. H. (2015). Implementing the next generation science standards. *Phi Delta Kappan, 96*(6), 45–49.

Pourciau, B. (2023, June 2). *Softball pitching speed chart by age*. TopVelocity. https://www.topvelocity.net/2023/06/02/softball-pitching-speed-chart-by-age/

Sánchez González, J. D. J., Ruiz Corral, J. A., García, G. M., Ojeda, G. R., Larios, L. D. L. C., Holland, J. B., …& García Romero, G. E. (2018). Ecogeography of teosinte. *PLoS One*, *13*(2), e0192676.

Solano-Flores, G., Shade, C., & Chrzanowski, A. (2014). Item accessibility and language variation conceptual framework. Submitted to the Smarter Balanced Assessment Consortium. P. 34.

Steele, C. M., Spencer, S. J., & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. In M. P. Zanna (Ed.), *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.

Thompson, S. J., Johnstone, C. J., Anderson, M. E., & Miller, N. A. (2005). *Considerations for the development and review of universally designed assessments*. National Center on Educational Outcomes, University of Minnesota.

U.S. Department of the Interior. (2015). *History & culture*. National Parks Service. https://www.nps.gov/chcu/learn/historyculture/index.htm

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