Science Assessment Item Collaborative

# Item Specifications Guidelines

for the

*Next Generation Science Standards*

September 2015

Developed by WestEd in collaboration with
CCSSO, state members, and content experts.

THE COUNCIL OF CHIEF STATE SCHOOL OFFICERS

The Council of Chief State School Officers (CCSSO) is a nonpartisan, nationwide, nonprofit organization of public officials who head departments of elementary and secondary education in the states, the District of Columbia, the Department of Defense Education Activity, and five U.S. extra-state jurisdictions. CCSSO provides leadership, advocacy, and technical assistance on major educational issues. The Council seeks member consensus on major educational issues and expresses their views to civic and professional organizations, federal agencies, Congress, and the public.

SCIENCE ASSESSMENT ITEM COLLABORATIVE
ITEM SPECIFICATIONS GUIDELINES
FOR THE NEXT GENERATION SCIENCE STANDARDS

COUNCIL OF CHIEF STATE SCHOOL OFFICERS
June Atkinson (North Carolina), President
Chris Minnich, Executive Director

Council of Chief State School Officers
One Massachusetts Avenue, NW, Suite 700
Washington, DC 20001-1431
Phone (202) 336-7000
Fax (202) 408-8072
www.ccsso.org

# TABLE OF CONTENTS

**LIST OF FIGURES AND TABLES** ........ **ii**

**OVERVIEW** ........ **1**

**CHAPTER ONE: INTRODUCTION** ........ **3**
Basic Terminology ........ 4

**CHAPTER TWO: GENERAL ITEM SPECIFICATIONS GUIDELINES** ........ **7**
Cognitive Complexity ........ 7
Universal Design/Vocabulary and Language ........ 7
Scoring Considerations ........ 9
Achievement Level Descriptors and Special Student Populations ........ 9

**CHAPTER THREE: ITEM CLUSTER ALIGNMENTS** ........ **11**
PE Item Specifications ........ 11
Multi-PE Item Cluster Alignments ........ 17
PE Bundling Considerations ........ 17
Example of a Two-PE Item Cluster Alignment ........ 19
From Item Cluster Alignment to Item Cluster ........ 25

**REFERENCES** ........ **26**

**APPENDICES** ........ **29**
**APPENDIX A.** Guidelines for Item Clusters
**APPENDIX B.** Item Types and Subtypes for Item Clusters

# LIST OF FIGURES AND TABLES

**Figure 1.** Simplified flow chart showing the basic outline of a PE item specification .................11

**Figure 2.** An overview of the item cluster development process ..............................................12

**Figure 3.** Sample representation of the relationship of an item cluster aligned to two PEs to its component items, with item-aligned dimension combinations shown..................18

**Figure 4.** Sample representation of the relationship of an item cluster aligned to two PEs to its component items, with sample item subtypes shown..........................................26

**Table 1.** Sample PE item specification at the high school level................................................12

**Table 2.** Sample item cluster alignment for a high school item cluster aligned to two Life Sciences PEs.......................................................................................................................20

# OVERVIEW

The *Science Assessment Item Collaborative (SAIC) Item Specifications Guidelines* document (hereafter referred to as the "Item Specifications Guidelines") has been developed as a companion document to the *Science Assessment Item Collaborative Assessment Framework* (hereafter referred to as the "Assessment Framework"). Together, these documents address the major issues facing state education agencies (SEAs) and other entities that are implementing new science standards by documenting the processes needed to guide the development of assessments for the Next Generation Science Standards (NGSS). Due to the interrelated nature of the documents, elements of the Item Specifications Guidelines that specifically detail the characteristics of the assessments and associated development considerations may also appear in the Assessment Framework.

The Item Specifications Guidelines focus specifically on how the content of the NGSS will be assessed, by articulating the NGSS–to–item cluster correlations that are necessary for the development of NGSS-aligned items, item clusters, and assessments. (See **Appendix A** for a description of item clusters.) In particular, emphasis is placed on developing item pools for summative assessment (i.e., large-scale, evaluative testing done at the end of academic years) of the NGSS. Many states may choose to implement formative assessment (tools and processes) and interim assessment of the NGSS as well. Guidelines for these assessments will vary from state to state, based on variables such as scale, time frame, budgetary constraints, and diagnostic goals, and, as such, are not addressed in this document.

The Item Specifications Guidelines are organized into the following three chapters:

**Chapter One: Introduction** discusses specific assessment-relevant elements of the NGSS and the need for an assessment format that can measure the broad skills and practices that are embedded within each performance expectation (PE) of the NGSS. The *item cluster*, a set of related test items tied to a common stimulus, is introduced in this chapter as the foundational architectural building block for assessment of the NGSS.

**Chapter Two: General Item Specifications Guidelines** provides summaries and information on general issues related to assessment design and development such as cognitive complexity, universal design/vocabulary, and scoring considerations. Connections to the NGSS and associated ancillary materials, as well as practical applications of the presented guidelines, are provided to demonstrate how SEAs can leverage these general guidelines.

**Chapter Three: Item Cluster Alignments** references sample PE-level item specifications to describe how PEs can be articulated into a specification for an item cluster (i.e., an item cluster alignment). The chapter also includes a discussion of a key element of the sample PE item specifications: linkage of the PE's evidence statements with appropriately selected groupings of two or more dimensions. This discussion is illustrated by the inclusion of sample multi-PE item cluster alignments.

Two appendices conclude the Item Specifications Guidelines:

**Appendix A: Guidelines for Item Clusters** delves into the details of crafting specifications for item cluster development. As such, this appendix describes the item cluster requirements that should be addressed in the specifications. A subsection of the appendix is devoted to stimuli for item clusters, reflecting the critical role that this context-setting material has in connecting the items within a cluster and in providing scaffolding to support assessment of all students on the ability continuum.

**Appendix B: Item Types and Subtypes for Item Clusters** includes a discussion of the three main item types to be used in item clusters (selected response, constructed response, and technology enhanced). Several item subtypes exist within each of these three main item types, particularly for technology-enhanced items, and these item subtypes are discussed in more detail with respect to their most effective use in the larger context of the item cluster.

# CHAPTER ONE: INTRODUCTION

The Next Generation Science Standards (NGSS) contain a set of performance expectations (PEs) that form the core of scientific concepts and skills that students are expected to know and perform. While PEs are the foundation of NGSS-aligned assessments, they do not, unto themselves, define a science curriculum, as emphasized in the Executive Summary of the NGSS (NGSS, 2013). A unique aspect of the NGSS is that the standards, as defined by the PEs, do not only include the content that students are expected to know and understand, but also embed related cognitive skills and connections that are the basis of scientific understanding and thinking. These cognitive elements are discussed in the document *A Framework for K–12 Science Education* (NRC, 2012) and are divided into the following categories, collectively referred to as the NGSS “dimensions”:

1. Scientific and Engineering Practices (SEPs)
2. Disciplinary Core Ideas (DCIs)
3. Crosscutting Concepts (CCCs)

The assessable scientific content of each PE is further defined through evidence statements. As outlined in the document *Developing Assessments for the Next Generation Science Standards* (NRC, 2014), the content of a PE in the NGSS is composed of the content of the related dimensions for each PE; therefore, assessment of PEs included in the NGSS must include these elements in order to be considered complete. Thus, given the different elements that must be included in any assessment of the NGSS, there has been considerable discussion about how best to develop items that will effectively assess students’ science understanding and skills at several cognitive levels. From this discussion, much of which is outlined in the Assessment Framework, a consonance has emerged that NGSS-based assessment will require a more complex item scope than that of traditional assessment formats in order to effectively measure students’ mastery of all three dimensions that comprise a PE. The *item cluster*, which utilizes an assessment approach that spans the concepts and dimensions of one or several PEs by scaffolding multiple items through an overarching stimulus, has emerged as an assessment model that satisfies this requirement.

This document provides a methodical and practical guide for item cluster development. It discusses issues pertinent to item clusters and provides a road map for the development of clear, comprehensive specifications for NGSS-aligned item clusters. An important first step in developing NGSS-aligned assessments is to determine how to develop item clusters that will be effective for measuring NGSS-based content across all three dimensions, including those concepts that have been challenging to measure via traditional item types. Using an evidence-based approach (see Chapter One of the Assessment Framework), item clusters must be explicitly linked to particular combinations of scientific learning dimensions within the context of a particular PE or bundle of PEs. Clear links among the NGSS, the measurement model, and the item types must be evident. Developmental appropriateness and accessibility for special student populations (including strategies for differentiating responses) must also be considered.

Guiding questions to consider prior to and during the item cluster development process include the following:

- What information is required in specifications in order to guide developers to create item clusters that can fully measure the multiple dimensions inherent in all NGSS PEs?
- How should the specifications for item clusters be organized? Specifically, what are the linkages between the elements of the NGSS PEs, such as dimensions and evidence statements, and the items composing the item clusters?
- How do general assessment issues translate to the item cluster context, and how should these issues be addressed?
- What is the most desirable combination of item types and subtypes within an item cluster, and how do these item types and subtypes fit the unique measurement goals for each PE?
- How are item clusters organizationally similar or different when dimensional aspects of different PEs within a multi-PE item cluster are mutually aligned?
- What criteria should be used to establish linkages of PEs (for both intra-domain and inter-domain groupings) within item clusters during the development of item clusters?

## Basic Terminology

To provide additional clarity, a glossary of the assessment terminology used throughout this document is provided in this section. Many of these terms are relevant to assessment in general, but may have a specific meaning when referenced in the context of item clusters.

- *Constructed response* (*CR*): An item type in which the response is text or mathematical symbols that are entered into a field.
- *Dimension*: A broad set of expectations with respect to a student's knowledge and skills in the following three areas: Scientific and Engineering Practices (SEPs), Disciplinary Core Ideas (DCIs), and Crosscutting Concepts (CCCs; concepts that unify the study of science and engineering).
- *Domain*: One of the following four disciplinary areas: physical sciences; life sciences; Earth and space sciences; and engineering, technology, and application of science.
- *Evidence statement*: A set of observable features of student performance, encompassing the many aspects of a performance expectation, developed for the NGSS by educators and scientists in a process coordinated by Achieve, Inc.
- *Item*: An individual assessment element, within the structure of an item cluster, that includes item-specific stimulus material (optional), a question/prompt, answer/options or an answer field, scoring criteria, and metadata.

- *Item (or item part) alignment*: The finest grain of alignment, inclusive of alignment to one or more evidence statements and the associated dimensions for that particular evidence statement (and thus for the associated PE).
- *Item cluster*: a set of items (usually between four and six items, with some items having more than one part) that are based on at least one common stimulus (e.g., text, audio, video, animation, simulation, experiment). Administration time for a single item cluster for summative assessment purposes is estimated to be approximately 20 minutes.
- *Item cluster alignment (document)*: The final specification that directs the development of an item cluster targeting specific PE(s); composed of PE item specifications for all PEs selected for the item cluster.
- *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 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*: The most general description of the format of a particular item, divided into three main categories: selected response, constructed response, and technology-enhanced.
- *Multidimensional alignment grouping*: A defined item template, within a PE item specification, that links at least two of the PE's three dimensions with related components from the PE's evidence statement.
- *Performance expectation* (*PE*): An assessable statement of what students should know and be able to do. Performance expectations are the unit of the NGSS and consist of the interweaving of SEPs, DCIs, and CCCs.
- *Performance expectation grouping* (*or bundle*): A selection of 2–3 PEs to be assessed together within an item cluster.
- *Performance expectation item specification*: The distribution of all evidence statements for a PE into the four main two- or three-dimensional grouping categories—DCI/SEP, DCI/CCC, SEP/CCC, and DCI/SEP/CCC—based on an analysis of each evidence statement.
- *Science phenomenon* (*or focus*): The main idea upon which an item cluster focuses. The science phenomenon provides the context necessary to determine which PEs can be bundled together naturally. A phenomenon is an object or aspect known through the senses rather than by thought or intuition. A fact or event of scientific interest susceptible to scientific description and explanation (Moulding, Bybee, & Paulson, 2015).
- *Selected response* (*SR*): An item type in which the response consists of one or more options chosen from a list of options.

- *Stem*: The statement of an item question or prompt to which the student responds.
- *Stimulus*: A component of an item cluster that does not directly require a student response. A stimulus can include one or more of the following: text, audio, video, animation/simulation, experimentation, discussion, activity, and/or demonstration.
- *Target*: Assessable knowledge and skills; for an item or item part in an item cluster, the target consists of the evidence statements and associated dimensions included in the evidence statement for the associated PE.
- *Technology-enhanced item* (*TEI*): A computer-delivered item type in which the response requires specialized computer interaction that is beyond selected-response or constructed-response interactions.

# CHAPTER TWO: GENERAL ITEM SPECIFICATIONS GUIDELINES

The item clusters that are developed for Next Generation Science Standards (NGSS)–based assessments will require specifications that are unique to the item-cluster assessment model. Other guideline specifications will be more generic in the sense that they touch on issues, such as cognitive complexity, accessibility, and scoring considerations, that are important elements of any type of assessment development. This chapter focuses on these more generic aspects of assessment, viewed through the lens of item cluster development.

In addition to these aspects, states should also consider developing a companion style guide that development vendors will use when developing NGSS-aligned item clusters. Style guides function to establish clear expectations with respect to fonts, graphics, units, and other stylistic elements, to ensure that assessment content is developed using styles that are consistent with a state's current assessment system.

## Cognitive Complexity

The NGSS place strong emphasis on student reasoning skills and the application of content knowledge to new contexts. The evidence statements are arranged into numerical categories (1, 2, 3, and sometimes 4) and represent a range in terms of cognitive demand. It is important to note that the categories do not equate with any established cognitive scale (e.g., Webb's Depth of Knowledge or Bloom's Revised Taxonomy), and that the numbered categories do not correlate with or imply degrees of cognitive complexity (e.g., category label 1 does not necessarily imply lower cognitive complexity than category label 2). Developing item clusters that include items aligned to the full range of evidence statements will result in items covering the range of cognitive complexity intended for the performance expectations. It is important to note that the evidence statements, taken in total, are targeting the proficient range. As such, item clusters that do not require students to reason or to utilize the last sequential category would not be considered acceptable assessment of the NGSS. When developing item clusters in this way, states may choose to add another layer of cognitive complexity coding to their metadata expectations. These guidelines assume that states will not add this additional layer of coding to their item cluster alignments.

## Universal Design/Vocabulary and Language

Universal design for assessment is broadly defined as a set of applied principles that assist in the design of assessments and that minimize and/or mitigate physical, linguistic, cultural, and other barriers to accessibility and threats to test validity. As described in Johnstone, Altman, and Thurlow (2006), universally designed assessments embrace seven basic elements:

1. Inclusive assessment population
2. Precisely defined constructs
3. Accessible, non-biased items
4. Amenable to accommodations
5. Simple, clear, and intuitive instructions and procedures
6. Maximum readability and comprehensibility
7. Maximum legibility

Details and examples of these universal test design elements can be found in Johnstone et al. (2006), and additional information about universal testing is available on the National Center on Educational Outcomes (NCEO)'s Universally Designed Assessments webpage: http://www.cehd.umn.edu/NCEO/TopicAreas/UnivDesign/UnivDesignTopic.htm.

In developing item clusters for NGSS-based assessments, these elements should be viewed as foundational in all stages of item development. Vocabulary and sentence structure should not hinder item accessibility. In science, some disciplinary-specific terminology can and should be used. When a scientific term is to be included in the stem or stimulus but is not mentioned in the PE itself or in the evidence statements, it is generally desirable to use other grade-appropriate words as a substitute for the term or, if that is not practical, to provide a general and grade-appropriate definition for the term in some way (e.g., parenthetical, glossary). Careful use of the grade-band progressions should be considered in determining assessment and grade appropriateness for scientific terminology. Regardless of the policy chosen, all states should consider providing guidance to writers on terminology that may or may not be assumed in writing the item cluster.

The stimuli for item clusters will almost always be richer in terms of content and, in some cases, data, compared to typical stand-alone items in other forms of assessment. For this reason, strong consideration should be given to the ways in which data and ideas are presented, with an emphasis on clarity of both textual presentation and item cluster organization. While excess verbiage and redundancy should always be avoided, item developers should keep in mind that a balance between succinctness and clarity is most desirable. If possible, information that is necessary for a particular item in an item cluster should be provided in the stimulus material immediately preceding that item, rather than in the general stimulus that begins the item cluster. Additionally, considerations with regard to the layout and presentation of the stimulus during an online administration should include the ability for a student to easily and quickly access components of the stimulus throughout administration of the item cluster without unnecessary scrolling or the addition of unnecessary cognitive load (i.e., the user interface of the platform should not hinder a student from accessing the stimulus and navigating to and from it during administration of the entire item cluster).

Item developers should also be cognizant of grade-level appropriateness when choosing contexts for the stimulus material. Even if the basic science content of a context is assumed to be aligned to a PE (a discussion of the intent for PEs to be context-agnostic is provided in **Appendix A**), the context may still be inappropriate for the grade level. For example, chemical species with complex formulae and chemical names should be avoided, as it should not be assumed that students would have had previous exposure to these or similar chemical species at a given grade level. Developers should also avoid using classic textbook examples, contexts, or phenomena, in order to avoid the most common representations of knowledge and the risk of students recalling what they have learned, rather than using their knowledge and skills to demonstrate what they know and are able to do. Providing unique, grade-level-appropriate contexts allows students to demonstrate, in a fair and unbiased way, their ability to purposefully use their knowledge and skills.

## Scoring Considerations

The focus of development of the SAIC NGSS-aligned item pool is assumed to be on computer-administered items, and, as such, item specifications will not cover accommodations for paper-and-pencil or other forms of delivery. (Individual states may choose to develop item specifications that include paper-and-pencil accommodations.) For computer-administered summative assessment, the ability to score items using rule-based machine rubrics based on artificial intelligence (automated scoring) is increasing, though still not fully implemented. The exceptions to this trend are constructed-response (CR) items, which continue to require hand scoring in almost all cases (APT Innovations in Testing, 2015). Most SR items and technology-enhanced items (TEIs) within a cluster will be worth 1 or 2 points; CR items will generally have more points attached to them and will require scoring rubrics to correlate student responses to the levels of full or partial awarding of points (which also supports the structure of a multipart CR).

## Achievement Level Descriptors and Special Student Populations

The Assessment Framework includes content that informs the development of a set of state-specific initial achievement level descriptors (ALDs) that will be aligned with the NGSS for each achievement level and for all tested grades. It is recommended that the development of ALDs occur concurrently with the test development cycle. This shift will allow the ALDs to specifically address student performance expectations that should ultimately inform the ways in which a state’s NGSS-aligned science assessments are conceived and developed.

Effective ALDs break down and make transparent the knowledge, skills, and processes that students are being asked to demonstrate at predetermined levels of achievement (for example, Basic, Proficient, and Advanced, or Levels 1–4). ALDs are often included in student-level score reports as well as in state aggregate reports, and, in order to be effective, ALDs must be able to clearly distinguish the differences among the discrete proficiency levels (that is, what students should know and be able to do at each level) to all stakeholders, including parents, teachers, and state policymakers. The NGSS evidence statements were developed for a single proficiency level (“proficient”), and therefore do not include information on determining multiple levels of achievement beyond a single “proficient” level relevant to the PEs. Developing ALDs for multiple levels of proficiency at the beginning of the test development cycle will aid in the verification of alignment among the assessment targets, and the descriptors will ensure that the assessment content supports the distinctions among the levels. This sequence will ultimately translate into transparent and valuable descriptors for all stakeholders.

For states in which all student populations are tested and alternate assessments are provided for students with the most severe disabilities, there will be a need to provide a “road map” between the skills and practices identified in the evidence statements for the NGSS PEs and the evidence of skills and practices that is provided by the alternate assessments. In particular, it will be useful for educators of severely disabled students to have materials that provide guidance on how the scaffolded competencies exhibited in an NGSS-based item cluster can be mirrored in an alternate assessment, along with guidelines on how these student audiences can progress to higher levels of achievement in NGSS assessment. Because of the emerging nature of this documentation and the reality that states have unique approaches to ALD development and alternate-assessment development, the Item Specifications Guidelines do not further address these aspects of assessment design; states will need to incorporate them into their state-specific item specifications.

# CHAPTER THREE: ITEM CLUSTER ALIGNMENTS

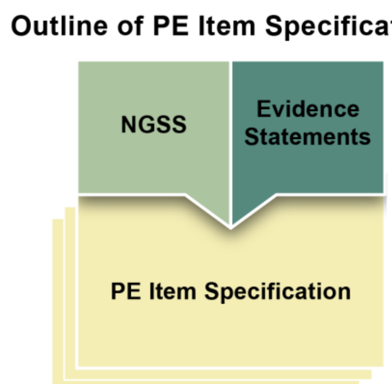
An item cluster alignment is the bridge between the Next Generation Science Standards (NGSS) and NGSS-aligned item clusters. Item cluster alignments will serve as tools to guide developers in their development of item clusters. Item cluster alignments serve the same purpose in the development of item clusters as item specifications serve for the development of individual items.

## PE Item Specifications

The development of an item cluster alignment is a multistep process. The first step in this process is to create a PE item specification by correlating the evidence statements for a single PE into multidimensional groupings (i.e., assigning each evidence statement either to a pair of dimensions or to all three dimensions, thereby sorting the evidence statements into one of the following categories: DCI/SEP, DCI/CCC, SEP/CCC, or DCI/SEP/CCC). This sorting should be done based on a content expert's understanding of the type(s) of items that can be generated for specific evidence statements, and not simply on the dimensional color coding of the evidence statements. It is critical to note that the goal of the NGSS is for students to demonstrate three-dimensional knowledge. Students should be able to demonstrate their ability to use each dimension to explain phenomena or design solutions.

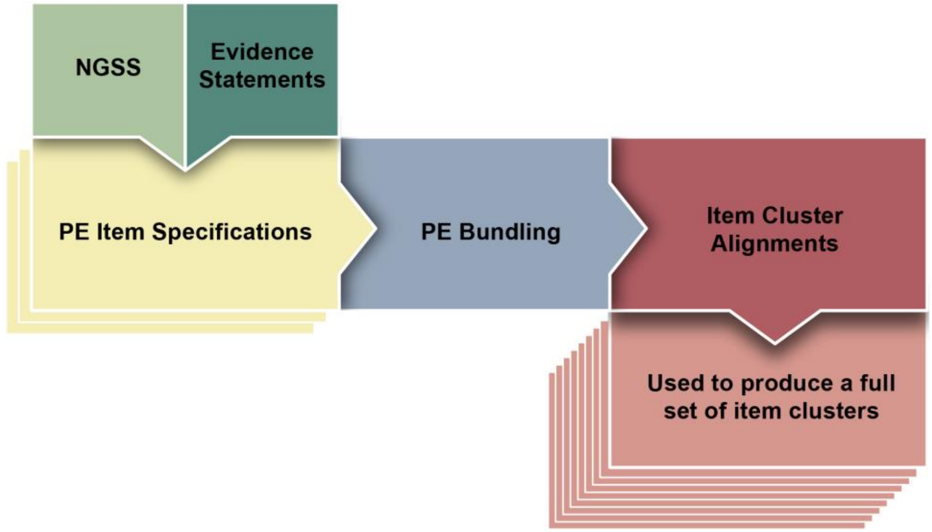
Figure 1 shows how the main components of a PE item specification, the PE itself (labeled "NGSS" in the figure) and the evidence statements that correlate to the PE's dimensions, are combined to form a PE specification.

**Figure 1.** Simplified flow chart showing the basic outline of a PE item specification




As shown in the process overview in Figure 2, the PE item specifications serve as the scaffolding or building-block components to generate item cluster alignments.

**Figure 2.** An overview of the item cluster development process




It is expected that item cluster alignments will be tailored to individual state testing programs, based on state-specific test design decisions (e.g., PE bundling choices, grade or grade-span expectations of tested content). The item cluster alignments built from the PE item specifications then serve as directive guides for generating item clusters to be used in the state's testing program.

Table 1 provides an example of how a PE item specification might be constructed. This particular PE item specification describes a single PE (HS-LS3-2).

**Table 1.** Sample PE item specification at the high school level

| | |
|---|---|
| **Performance Expectations:** | **HS-LS3-2. Make and defend a claim based on evidence that inheritable genetic variations may result from: (1) new genetic combinations through meiosis, (2) viable errors occurring during replication, and/or (3) mutations caused by environmental factors.** |
| **Content Domain:** | *LS3.B: Variation of Traits*<br>• In sexual reproduction, chromosomes can sometimes swap sections during the process of meiosis (cell division), thereby creating new genetic combinations and thus more genetic variation. Although DNA replication is tightly regulated and remarkably accurate, errors do occur and result in mutations, which are also a source of genetic variation. Environmental factors can also cause mutations in genes, and viable mutations are inherited.<br>• Environmental factors also affect expression of traits, and hence affect the probability of occurrences of traits in a population. Thus the variation and distribution of traits observed depends on both genetic and environmental factors. |
| **Target Clarifications:** | Emphasis is on using data to support arguments for the way variation occurs. |
| **Assessment Boundary:** | Assessment does not include the phases of meiosis or the biochemical mechanism of specific steps in the process. |
| **Number of Items in Item Cluster:** <<TBD>> | |
| **Allowable Stimulus Materials:** Graphs, tables, videos, verbal descriptions, simulations, animations, text | |

**Table 1.** (continued)

| | |
|---|---|
| ***Items to DCI and SEP*** | |
| **Disciplinary Core Ideas:** | ***LS3.B: Variation of Traits*** |
| | • In sexual reproduction, chromosomes can sometimes swap sections during the process of meiosis (cell division), thereby creating new genetic combinations and thus more genetic variation. Although DNA replication is tightly regulated and remarkably accurate, errors do occur and result in mutations, which are also a source of genetic variation. Environmental factors can also cause mutations in genes, and viable mutations are inherited.<br>• Environmental factors also affect expression of traits, and hence affect the probability of occurrences of traits in a population. Thus the variation and distribution of traits observed depends on both genetic and environmental factors. |
| **Science and Engineering Practices:** | ***Engaging in Argument from Evidence*** |
| | Engaging in argument from evidence in 9-12 builds on K-8 experiences and progresses to using appropriate and sufficient evidence and scientific reasoning to defend and critique claims and explanations about the natural and designed world(s). Arguments may also come from current scientific or historical episodes in science.<br>• Make and defend a claim based on evidence about the natural world that reflects scientific knowledge and student-generated evidence. |
| **Evidence Statements:** | (2) Identifying scientific evidence. (a) Students identify and describe evidence that supports the claim, including: (ii) Genetic mutations can occur due to: (a) errors during replication; and/or (b) environmental factors. |
| | (2) Identifying scientific evidence. (a) Students identify and describe evidence that supports the claim, including: (iii) Genetic material is inheritable. |
| | (2) Identifying scientific evidence. (b) Students use scientific knowledge, literature, student-generated data, simulations and/or other sources for evidence.* |
| | (3) Evaluating and critiquing evidence. (a) Students identify the following strengths and weaknesses of the evidence used to support the claim: (i) Types and numbers of sources.* |
| | (3) Evaluating and critiquing evidence. (a) Students identify the following strengths and weaknesses of the evidence used to support the claim: (iii) Validity and reliability of the evidence.* |
| **Recommended Item Types:** | SR and TE |
| **Item Point Total:** | SR: 1 point, TE: 1-2 points |
| **Allowable Stimulus Materials:** | simulations, animations, graphs, tables, videos, text |

**Table 1.** (continued)

| | |
|---|---|
| ***Items to DCI and CCC*** | |
| **Disciplinary Core Ideas:** | ***LS3.B: Variation of Traits*** |
| | • In sexual reproduction, chromosomes can sometimes swap sections during the process of meiosis (cell division), thereby creating new genetic combinations and thus more genetic variation. Although DNA replication is tightly regulated and remarkably accurate, errors do occur and result in mutations, which are also a source of genetic variation. Environmental factors can also cause mutations in genes, and viable mutations are inherited. <br>• Environmental factors also affect expression of traits, and hence affect the probability of occurrences of traits in a population. Thus the variation and distribution of traits observed depends on both genetic and environmental factors. |
| **Crosscutting Concepts:** | ***Cause and Effect*** |
| | • Empirical evidence is required to differentiate between cause and correlation and make claims about specific causes and effects. |
| **Evidence Statements:** | *(None of the given Evidence Statements contain only these two dimensions. Items aligning to only these two dimensions can be developed to Evidence Statements that align to all three dimensions.)* |
| **Allowable Item Types:** | SR and TE |
| **Item Point Total:** | SR: 1 point, TE: 1-2 points |
| **Allowable Stimulus Materials:** | simulations, animations, graphs, tables, videos, text |
| ***Items to SEP and CCC*** | |
| **Science and Engineering Practices:** | ***Engaging in Argument from Evidence*** |
| | Engaging in argument from evidence in 9-12 builds on K-8 experiences and progresses to using appropriate and sufficient evidence and scientific reasoning to defend and critique claims and explanations about the natural and designed world(s). Arguments may also come from current scientific or historical episodes in science. <br>• Make and defend a claim based on evidence about the natural world that reflects scientific knowledge and student-generated evidence. |
| **Crosscutting Concepts:** | ***Cause and Effect*** |
| | • Empirical evidence is required to differentiate between cause and correlation and make claims about specific causes and effects. |
| **Evidence Statements:** | (3) Evaluating and critiquing evidence. (a) Students identify the following strengths and weaknesses of the evidence used to support the claim: (ii) Sufficiency to make and defend the claim, and to distinguish between causal and correlational relationships. |
| | (4) Reasoning and synthesis. (c) Students defend a claim against counter-claims and critique by evaluating counter-claims and by describing the connections between the relevant and appropriate evidence and the strongest claim. |
| **Allowable Item Types:** | SR and TE |
| **Item Point Total:** | SR: 1 point, TE: 1-2 points |
| **Allowable Stimulus Materials:** | simulations, animations, graphs, tables, videos, text |

**Table 1.** (continued)

| | |
|---|---|
| ***Items to DCI, SEP, and CCC*** | |
| **Disciplinary Core Ideas:** | ***LS3.B: Variation of Traits*** |
| | • In sexual reproduction, chromosomes can sometimes swap sections during the process of meiosis (cell division), thereby creating new genetic combinations and thus more genetic variation. Although DNA replication is tightly regulated and remarkably accurate, errors do occur and result in mutations, which are also a source of genetic variation. Environmental factors can also cause mutations in genes, and viable mutations are inherited.<br>• Environmental factors also affect expression of traits, and hence affect the probability of occurrences of traits in a population. Thus the variation and distribution of traits observed depends on both genetic and environmental factors. |
| **Science and Engineering Practices:** | ***Engaging in Argument from Evidence*** |
| | Engaging in argument from evidence in 9-12 builds on K-8 experiences and progresses to using appropriate and sufficient evidence and scientific reasoning to defend and critique claims and explanations about the natural and designed world(s). Arguments may also come from current scientific or historical episodes in science.<br>• Make and defend a claim based on evidence about the natural world that reflects scientific knowledge and student-generated evidence. |
| **Crosscutting Concepts:** | ***Cause and Effect*** |
| | • Empirical evidence is required to differentiate between cause and correlation and make claims about specific causes and effects. |
| **Evidence Statements:** | (1) Developing a claim. (a) Students make a claim that includes the idea that inheritable genetic variations may result from: (i) New genetic combinations through meiosis. |
| | (1) Developing a claim. (a) Students make a claim that includes the idea that inheritable genetic variations may result from: (ii) Viable errors occurring during replication. |
| | (1) Developing a claim. (a) Students make a claim that includes the idea that inheritable genetic variations may result from: (iii) Mutations caused by environmental factors. |
| | (2) Identifying scientific evidence. (a) Students identify and describe evidence that supports the claim, including: (i) Variations in genetic material naturally result during meiosis when corresponding sections of chromosome pairs exchange places. |
| | (4) Reasoning and synthesis. (a) Students use reasoning to describe links between the evidence and claim, such as: (i) Genetic mutations produce genetic variations between cells or organisms. |
| | (4) Reasoning and synthesis. (a) Students use reasoning to describe links between the evidence and claim, such as: (ii) Genetic variations produced by mutation and meiosis can be inherited. |
| | (4) Reasoning and synthesis. (b) Students use reasoning and valid evidence to describe that new combinations of DNA can arise from several sources, including meiosis, errors during replication, and mutations caused by environmental factors. |
| **Allowable Item Types:** | TE |
| **Item Point Total:** | 1 to 3 points |
| **Allowable Stimulus Materials:** | simulations, animations, graphs, tables, videos, text, equations, |

Source: Nevada Department of Education (n.d.)

The sample item specification begins with an introductory section identifying the requirements of, and information concerning, the PE as a whole. In this section, the relevant PE, content domain, target clarifications (if any), and assessment boundary (if any) are identified. The number of items in the final cluster and the allowable stimulus materials are also included. If desired, this section could include the number of items that can be multipart, primarily as an indirect means of controlling the overall length of the item cluster. This sample specification allows for virtually all stimulus types, although it is anticipated that, in practice, the list of allowable stimulus types would likely be significantly shorter due to budget and administration constraints. (It is assumed that each individual state's PE item specifications will include some subset of the requirements and information shown in Table 1, and possibly some additional information at the state's discretion.)

Following the introductory section are four multidimensional alignment groupings organized by the grouped dimensions intended for assessment. The evidence statements have been color coded to show the alignment of the evidence statement wording to the PE's dimensions (blue = SEP, orange = DCI, green = CCC). This color coding was developed by Achieve, Inc., as a tool to help demonstrate how each dimension is represented in a given evidence statement. (Note: The color coding in Table 1 was created by WestEd prior to the Achieve color coding being completed, but based on WestEd's understanding of the process that Achieve was using, and may vary to some extent from the color coding subsequently created by Achieve. At the time of this printing, the color coding was in final content review at Achieve.) In general, all parts of an evidence statement will have some degree of alignment to a DCI, which reflects the scientific content of the PE, as well as the SEP, given the structure and format of the evidence statements (i.e., organized by SEP). In the sample item specification, the DCI/CCC grouping does not show unique alignment to any part of the evidence statement. However, for this multidimensional alignment grouping, parts of the evidence statement from the more inclusive SEP/DCI/CCC multidimensional alignment grouping can be used to develop items for an item cluster assessing this PE.

Although there are only four groupings listed in the sample, it is expected that some of these item alignments will be used for alignment for more than one item within a cluster. For example, based on its related evidence statements, the "Items to DCI, SEP, and CCC" grouping has sufficient breadth and depth to support more than one item. In each item alignment grouping, the allowable item types and subtypes, item point totals, and allowable stimulus materials are identified for items within the item cluster.

Consistent with the idea of having the item specifications at the PE level and the item cluster alignments be "context agnostic" (see **Appendix A**), the sample item cluster alignments do not attempt to offer possible scenarios or structural frameworks (e.g., "simulated laboratory investigation"). The main requirement for item developers is that they align the science phenomenon or focus of the item cluster (e.g., context) with the specified content domain (drawn from the relevant sections of the DCIs that align to the PE[s]) and with the requirements laid out for the item cluster as a whole and for the individual items in the cluster.

## Multi-PE Item Cluster Alignments

As previously stated, item clusters should focus on a particular science phenomenon and/or engineering problem. In order to fully support the item cluster, two or more PEs should be bundled (or grouped) together, with engineering PEs always being assessed together with science content. Ultimately, the nature of the test design will impact the total number and focus of item clusters, and, by extension, the final PE bundles (groupings). An inherent balance between breadth and depth will also influence the grouping of PEs. (Please refer to the discussions of test design and reporting in Appendix A of the Assessment Framework.) The process of bundling PEs should take into account the distribution of points across domains and dimensions and the degree of overlap among the dimensions. This should be done while maintaining a focus on the coherence of the PEs with respect to the DCI content knowledge associated with the PE (shared science phenomenon and context).

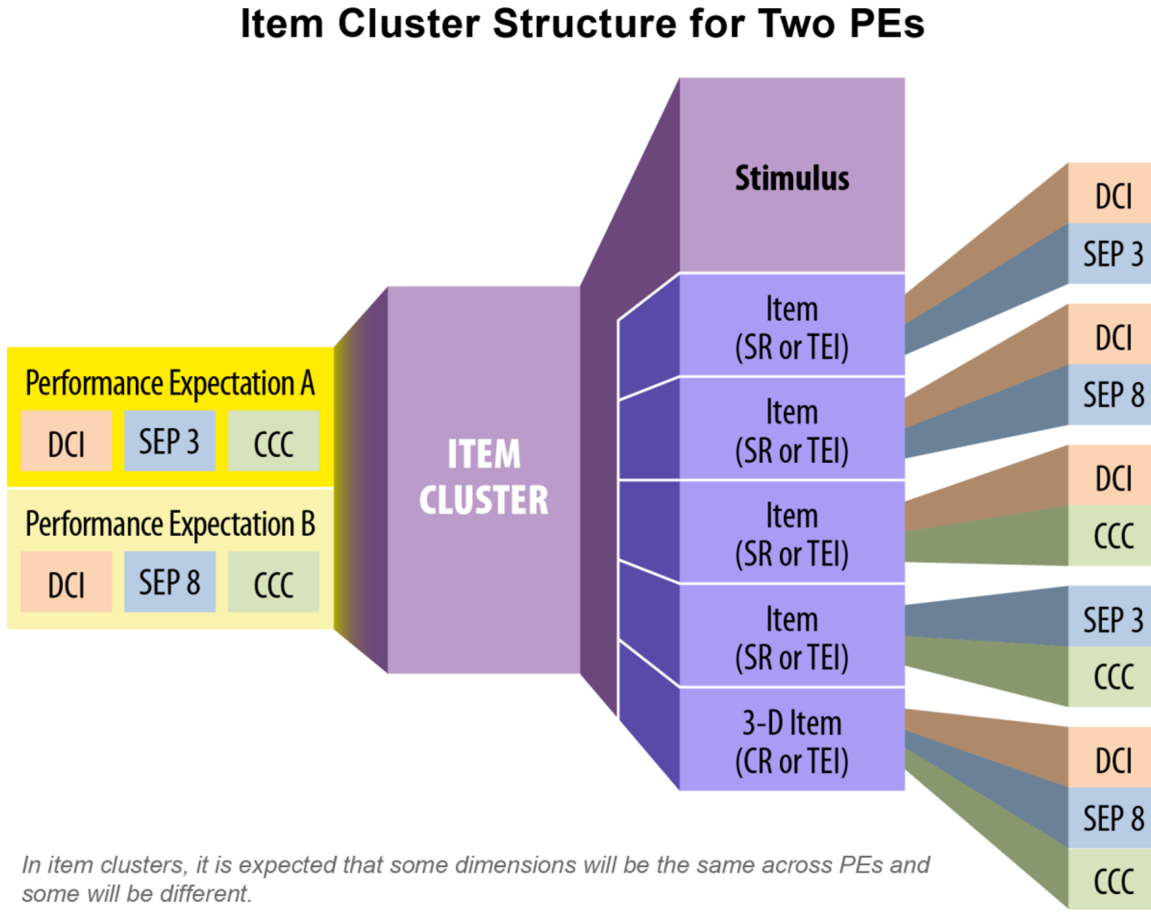
### *PE Bundling Considerations*

Multi-PE item clusters function to help ensure that a high number of PEs across an assessment are assessed in a meaningful way. Bundling decisions must be carefully considered, to ensure that PEs are grouped together in a way that supports the assessment of all targeted aspects of each PE in a meaningful way. Several PEs may lend themselves to a natural grouping when a particular science phenomenon is considered, and the science phenomenon chosen may even support the bundling of PEs in many different combinations.

Item clusters that encompass more than one PE require more specification of the individual item alignments. This is due, in large part, to the multiple dimensions (i.e., SEPs, DCIs, and CCCs) contained within the individual PEs. If the dimensions of the PEs within a bundle do not contain overlap, it becomes more challenging to develop item clusters that address an acceptable breadth of each PE's components in a natural manner. If the dimensions of the PEs do contain overlap of dimensions, then care must be taken to not develop item clusters that are narrowly focused and do not reflect the nature of student understanding and the NGSS.

Every individual item within the cluster should assess at least two of the dimensions from any particular PE. Dimensional alignment should be considered as an emphasis on a specific dimension, but not exclusive of other dimensions. Individual dimensions are not intended to be understood or practiced in isolation, so the assessment of the dimensions should not artificially isolate the dimensions. As Figure 3 shows, if PE A has SEP 3 as one of its dimensions and PE B has SEP 8 as one of its dimensions, then items aligning to PE A should emphasize SEP 3 (and not SEP 8).

**Figure 3.** Sample representation of the relationship of an item cluster aligned to two PEs to its component items, with item-aligned dimension combinations shown




Given the number of permutations possible, PE bundling will be a challenging process, but some general rules-of-thumb for bundling PEs will help winnow the possibilities. The following list is a set of guidelines for consideration when selecting PEs to bundle in a multi-PE item cluster:

1. PEs should be bundled in a way that naturally supports the assessment of a science phenomenon (that spans PEs and that, in some cases, can span domains and/or interdisciplinary contexts).
2. To ensure that the breadth of dimensions can be assessed, two to three PEs should be bundled in an item cluster, although single-PE item clusters may be preferable in situations in which natural groupings between a PE and other PEs do not exist. In most cases, it is recommended that no more than three PEs be bundled together in an item cluster.
3. PEs from the Engineering Design DCI should always be bundled together with PEs from one of the science disciplines.

4. An evidence statement that is used from one PE should be grouped with comparable evidence statements within and/or across the PEs in the cluster, to help ensure that items measure the correct aspects of the PEs. It is possible that an individual item will align to a single evidence statement (since the evidence statements incorporate multiple dimensions, this does not violate multidimensional alignment).
5. States may choose to bundle PEs across domains and/or across grade levels. Further, some states may choose to develop numerous permutations for their PE bundling, while other states may choose to bundle PEs such that each PE appears in only a single bundle (i.e., each PE is used once). Since each PE bundle can support many different stimuli (contexts), all bundling approaches should support a rich and varied item cluster pool.

### ***Example of a Two-PE Item Cluster Alignment***

Table 2 provides a sample item cluster alignment for high school and illustrates how the degrees of overlap among the dimensions and evidence statements can be organized to support multi-PE item clusters through the process of combining PE item specifications to compose item cluster alignments. The two PEs bundled together in the sample item cluster alignment include both a natural conceptual connection and a dimensional alignment (CCC: Cause and Effect). The arrangement of evidence statements next to each other in an item cluster alignment does not denote any specific relationship.

**Table 2.** Sample item cluster alignment for a high school item cluster aligned to two Life Sciences PEs

| | | |
|---|---|---|
| **Performance Expectations:** | **HS-LS3-1. Ask questions to clarify relationships about the role of DNA and chromosomes in coding the instructions for characteristic traits passed from parents to offspring.** | **HS-LS3-2. Make and defend a claim based on evidence that inheritable genetic variations may result from: (1) new genetic combinations through meiosis, (2) viable errors occurring during replication, and/or (3) mutations caused by environmental factors.** |
| **Content Domain:** | **LS1.A: Structure and Function** | **LS3.B: Variation of Traits** |
| | • All cells contain genetic information in the form of DNA molecules. Genes are regions in the DNA that contain the instructions that code for the formation of proteins. *(secondary)* | • In sexual reproduction, chromosomes can sometimes swap sections during the process of meiosis (cell division), thereby creating new genetic combinations and thus more genetic variation. Although DNA replication is tightly regulated and remarkably accurate, errors do occur and result in mutations, which are also a source of genetic variation. Environmental factors can also cause mutations in genes, and viable mutations are inherited.<br>• Environmental factors also affect expression of traits, and hence affect the probability of occurrences of traits in a population. Thus the variation and distribution of traits observed depends on both genetic and environmental factors. |
| | **LS3.A: Inheritance of Traits** | |
| | • Each chromosome consists of a single very long DNA molecule, and each gene on the chromosome is a particular segment of that DNA. The instructions for forming species' characteristics are carried in DNA. All cells in an organism have the same genetic content, but the genes used (expressed) by the cell may be regulated in different ways. Not all DNA codes for a protein; some segments of DNA are involved in regulatory or structural functions, and some have no as-yet known function. | |
| **Target Clarifications:** | No target clarifications are specified in the standards. | Emphasis is on using data to support arguments for the way variation occurs. |
| **Assessment Boundary:** | Assessment does not include the phases of meiosis or the biochemical mechanism of specific steps in the process. | Assessment does not include the phases of meiosis or the biochemical mechanism of specific steps in the process. |
| **Number of Items in Item Cluster:** <<TBD>> | | |
| **Allowable Stimulus Materials:** Graphs, tables, videos, verbal descriptions, simulations, animations, text | | |

**Table 2.** (continued)

| | | |
|---|---|---|
| | ***Items to DCI and SEP*** | |
| **Disciplinary Core Ideas:** | ***LS1.A: Structure and Function***<br>• All cells contain genetic information in the form of DNA molecules. Genes are regions in the DNA that contain the instructions that code for the formation of proteins. *(secondary)*<br>***LS3.A: Inheritance of Traits***<br>• Each chromosome consists of a single very long DNA molecule, and each gene on the chromosome is a particular segment of that DNA. The instructions for forming species' characteristics are carried in DNA. All cells in an organism have the same genetic content, but the genes used (expressed) by the cell may be regulated in different ways. Not all DNA codes for a protein; some segments of DNA are involved in regulatory or structural functions, and some have no as-yet known function. | ***LS3.B: Variation of Traits***<br>• In sexual reproduction, chromosomes can sometimes swap sections during the process of meiosis (cell division), thereby creating new genetic combinations and thus more genetic variation. Although DNA replication is tightly regulated and remarkably accurate, errors do occur and result in mutations, which are also a source of genetic variation. Environmental factors can also cause mutations in genes, and viable mutations are inherited.<br>• Environmental factors also affect expression of traits, and hence affect the probability of occurrences of traits in a population. Thus the variation and distribution of traits observed depends on both genetic and environmental factors. |
| **Science and Engineering Practices:** | ***Asking Questions and Defining Problems***<br>Asking questions and defining problems in 9-12 builds on K-8 experiences and progresses to formulating, refining and evaluating empirically testable questions and design problems using models and simulations.<br>• Ask questions that arise from examining models or a theory to clarify relationships | ***Engaging in Argument from Evidence***<br>Engaging in argument from evidence in 9-12 builds on K-8 experiences and progresses to using appropriate and sufficient evidence and scientific reasoning to defend and critique claims and explanations about the natural and designed world(s). Arguments may also come from current scientific or historical episodes in science.<br>• Make and defend a claim based on evidence about the natural world that reflects scientific knowledge and student-generated evidence. |
| **Evidence Statements:** | (1) Addressing phenomena or scientific theories. (a) Students use models of DNA to formulate questions, the answers to which would clarify: (ii) That the DNA and chromosomes that are used by the cell can be regulated in multiple ways. | (2) Identifying scientific evidence. (a) Students identify and describe evidence that supports the claim, including: (ii) Genetic mutations can occur due to: (a) errors during replication; and/or (b) environmental factors. |
| | (1) Addressing phenomena or scientific theories. (a) Students use models of DNA to formulate questions, the answers to which would clarify: (iii) The relationship between the non-protein coding sections of DNA and their functions (e.g., regulatory functions) in an organism. | (2) Identifying scientific evidence. (a) Students identify and describe evidence that supports the claim, including: (iii) Genetic material is inheritable. |
| | | (2) Identifying scientific evidence. (b) Students use scientific knowledge, literature, student-generated data, simulations and/or other sources for evidence.* |
| | | (3) Evaluating and critiquing evidence. (a) Students identify the following strengths and weaknesses of the evidence used to support the claim: (i) Types and numbers of sources.* |

**Table 2.** (continued)

| | | |
|---|---|---|
| **Evidence Statements (continued):** | | (3) Evaluating and critiquing evidence. (a) Students identify the following strengths and weaknesses of the evidence used to support the claim: (iii) Validity and reliability of the evidence.* |
| **Recommended Item Types:** | SR and TE | SR and TE |
| **Item Point Total:** | SR: 1 point, TE: 1-2 points | SR: 1 point, TE: 1-2 points |
| **Allowable Stimulus Materials:** | simulations, animations, graphs, tables, videos, text | simulations, animations, graphs, tables, videos, text |
| | ***Items to DCI and CCC*** | |
| **Disciplinary Core Ideas:** | ***LS1.A: Structure and Function***<br>• All cells contain genetic information in the form of DNA molecules. Genes are regions in the DNA that contain the instructions that code for the formation of proteins. *(secondary)*<br>***LS3.A: Inheritance of Traits***<br>• Each chromosome consists of a single very long DNA molecule, and each gene on the chromosome is a particular segment of that DNA. The instructions for forming species' characteristics are carried in DNA. All cells in an organism have the same genetic content, but the genes used (expressed) by the cell may be regulated in different ways. Not all DNA codes for a protein; some segments of DNA are involved in regulatory or structural functions, and some have no as-yet known function. | ***LS3.B: Variation of Traits***<br>• In sexual reproduction, chromosomes can sometimes swap sections during the process of meiosis (cell division), thereby creating new genetic combinations and thus more genetic variation. Although DNA replication is tightly regulated and remarkably accurate, errors do occur and result in mutations, which are also a source of genetic variation. Environmental factors can also cause mutations in genes, and viable mutations are inherited.<br>• Environmental factors also affect expression of traits, and hence affect the probability of occurrences of traits in a population. Thus the variation and distribution of traits observed depends on both genetic and environmental factors. |
| **Crosscutting Concepts:** | ***Cause and Effect***<br>• Empirical evidence is required to differentiate between cause and correlation and make claims about specific causes and effects. | ***Cause and Effect***<br>• Empirical evidence is required to differentiate between cause and correlation and make claims about specific causes and effects. |
| **Evidence Statements:** | *(None of the given Evidence Statements contain only these two dimensions. Items aligning to only these two dimensions can be developed to Evidence Statements that align to all three dimensions.)* | *(None of the given Evidence Statements contain only these two dimensions. Items aligning to only these two dimensions can be developed to Evidence Statements that align to all three dimensions.)* |
| **Recommended Item Types:** | SR and TE | SR and TE |
| **Item Point Total:** | SR: 1 point, TE: 1-2 points | SR: 1 point, TE: 1-2 points |
| **Allowable Stimulus Materials:** | simulations, animations, graphs, tables, videos, text | simulations, animations, graphs, tables, videos, text |

**Table 2.** (continued)

| | | |
|---|---|---|
| | ***Items to SEP and CCC*** | |
| **Science and Engineering Practices:** | ***Asking Questions and Defining Problems*** | ***Engaging in Argument from Evidence*** |
| | Asking questions and defining problems in 9-12 builds on K-8 experiences and progresses to formulating, refining and evaluating empirically testable questions and design problems using models and simulations.<br>• Ask questions that arise from examining models or a theory to clarify relationships | Engaging in argument from evidence in 9-12 builds on K-8 experiences and progresses to using appropriate and sufficient evidence and scientific reasoning to defend and critique claims and explanations about the natural and designed world(s). Arguments may also come from current scientific or historical episodes in science.<br>• Make and defend a claim based on evidence about the natural world that reflects scientific knowledge and student-generated evidence. |
| **Crosscutting Concepts:** | ***Cause and Effect*** | ***Cause and Effect*** |
| | • Empirical evidence is required to differentiate between cause and correlation and make claims about specific causes and effects. | • Empirical evidence is required to differentiate between cause and correlation and make claims about specific causes and effects. |
| **Evidence Statements:** | (2) Evaluating empirical testability. (a) Students' questions are empirically testable by scientists. | (3) Evaluating and critiquing evidence. (a) Students identify the following strengths and weaknesses of the evidence used to support the claim: (ii) Sufficiency to make and defend the claim, and to distinguish between causal and correlational relationships. |
| | | (4) Reasoning and synthesis. (c) Students defend a claim against counter-claims and critique by evaluating counter-claims and by describing the connections between the relevant and appropriate evidence and the strongest claim. |
| **Recommended Item Types:** | SR and TE | SR and TE |
| **Item Point Total:** | SR: 1 point, TE: 1-2 points | SR: 1 point, TE: 1-2 points |
| **Allowable Stimulus Materials:** | simulations, animations, graphs, tables, videos, text | simulations, animations, graphs, tables, videos, text |

**Table 2.** (continued)

| | | |
|---|---|---|
| [redacted] | | |
| ***Items to DCI, SEP, and CCC*** | | |
| **Disciplinary Core Ideas:** | ***LS1.A: Structure and Function*** | ***LS3.B: Variation of Traits*** |
| | • All cells contain genetic information in the form of DNA molecules. Genes are regions in the DNA that contain the instructions that code for the formation of proteins. *(secondary)* | • In sexual reproduction, chromosomes can sometimes swap sections during the process of meiosis (cell division), thereby creating new genetic combinations and thus more genetic variation. Although DNA replication is tightly regulated and remarkably accurate, errors do occur and result in mutations, which are also a source of genetic variation. Environmental factors can also cause mutations in genes, and viable mutations are inherited.<br>• Environmental factors also affect expression of traits, and hence affect the probability of occurrences of traits in a population. Thus the variation and distribution of traits observed depends on both genetic and environmental factors. |
| | ***LS3.A: Inheritance of Traits*** | |
| | • Each chromosome consists of a single very long DNA molecule, and each gene on the chromosome is a particular segment of that DNA. The instructions for forming species' characteristics are carried in DNA. All cells in an organism have the same genetic content, but the genes used (expressed) by the cell may be regulated in different ways. Not all DNA codes for a protein; some segments of DNA are involved in regulatory or structural functions, and some have no as-yet known function. | |
| **Science and Engineering Practices:** | ***Asking Questions and Defining Problems*** | ***Engaging in Argument from Evidence*** |
| | Asking questions and defining problems in 9-12 builds on K-8 experiences and progresses to formulating, refining and evaluating empirically testable questions and design problems using models and simulations.<br>• Ask questions that arise from examining models or a theory to clarify relationships | Engaging in argument from evidence in 9-12 builds on K-8 experiences and progresses to using appropriate and sufficient evidence and scientific reasoning to defend and critique claims and explanations about the natural and designed world(s). Arguments may also come from current scientific or historical episodes in science.<br>• Make and defend a claim based on evidence about the natural world that reflects scientific knowledge and student-generated evidence. |
| **Crosscutting Concepts:** | ***Cause and Effect*** | ***Cause and Effect*** |
| | • Empirical evidence is required to differentiate between cause and correlation and make claims about specific causes and effects. | • Empirical evidence is required to differentiate between cause and correlation and make claims about specific causes and effects. |
| **Evidence Statements:** | (1) Addressing phenomena or scientific theories. (a) Students use models of DNA to formulate questions, the answers to which would clarify: (i) The cause and effect relationships (including distinguishing between causal and correlational relationships) between DNA, the proteins it codes for, and the resulting traits observed in an organism. | (1) Developing a claim. (a) Students make a claim that includes the idea that inheritable genetic variations may result from: (i) New genetic combinations through meiosis. |
| | | (1) Developing a claim. (a) Students make a claim that includes the idea that inheritable genetic variations may result from: (ii) Viable errors occurring during replication. |
| | | (1) Developing a claim. (a) Students make a claim that includes the idea that inheritable genetic variations may result from: (iii) Mutations caused by environmental factors. |

**Table 2.** (continued)

| | | |
|---|---|---|
| **Evidence Statements (continued):** | | (2) Identifying scientific evidence. (a) Students identify and describe evidence that supports the claim, including: (i) Variations in genetic material naturally result during meiosis when corresponding sections of chromosome pairs exchange places. |
| | | (4) Reasoning and synthesis. (a) Students use reasoning to describe links between the evidence and claim, such as: (i) Genetic mutations produce genetic variations between cells or organisms. |
| | | (4) Reasoning and synthesis. (a) Students use reasoning to describe links between the evidence and claim, such as: (ii) Genetic variations produced by mutation and meiosis can be inherited. |
| | | (4) Reasoning and synthesis. (b) Students use reasoning and valid evidence to describe that new combinations of DNA can arise from several sources, including meiosis, errors during replication, and mutations caused by environmental factors. |
| **Recommended Item Types:** | TE | TE |
| **Item Point Total:** | 1 to 3 points | 1 to 3 points |
| **Allowable Stimulus Materials:** | simulations, animations, graphs, tables, videos, text, equations, | simulations, animations, graphs, tables, videos, text, equations, |

Source: Nevada Department of Education (n.d.)

Note: The recommended item types listed in this item cluster alignment sample are specific to the state's intended use. Due to cost constraints, the NDE did not intend to use CR item types for this assessment. In relation to the NGSS, CR item types are seen as important for fully assessing the intent of the standards.

### *From Item Cluster Alignment to Item Cluster*

With the completion of the item cluster alignment (see the example of a completed item cluster alignment in Table 2), the basic elements for item cluster development are set in place. The development of the item cluster will involve two major additional elements not prescribed by the item cluster alignment:

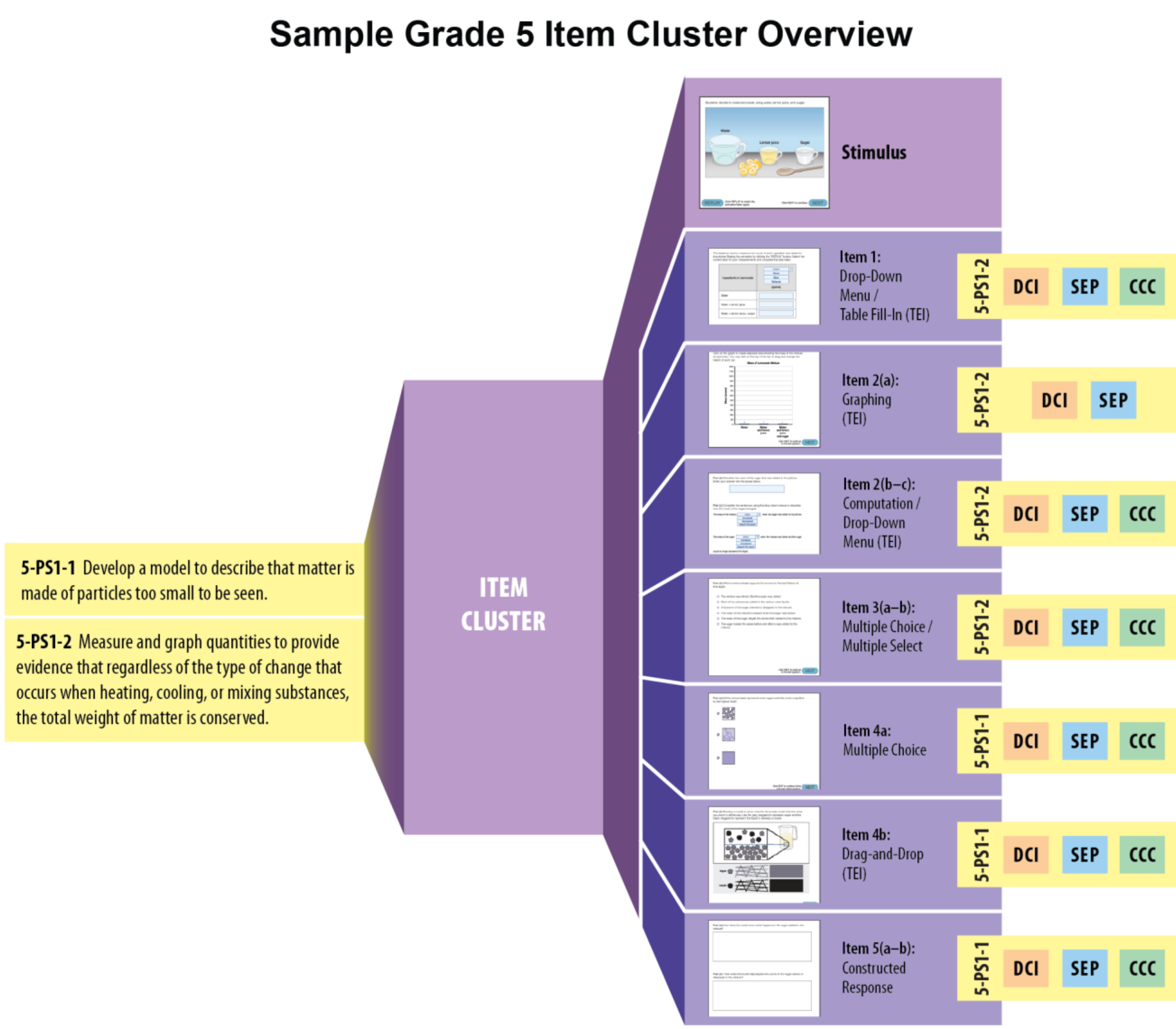
1. **The context for the item cluster**. This will drive the creation of both the stimulus and the overall scaffolding of the item cluster. It is assumed that, in most cases, the item cluster developer will choose the context, although states may want to add specifications regarding how to determine which contexts may or may not be used.

2. **The item type structure of the item cluster**. This decision can be handled in one of two ways: a particular number of items and selection of item types can be included in the

specifications section of the item cluster, and the item cluster developer can then create the item cluster prototype within the confines of the chosen context and the given specifications; or the item cluster developer can create a prototype for the item cluster, specifying the context, the number of items, and the specific item types. It is anticipated that both of these processes would involve some iterative review stages between the state and the item cluster developer.

A visual overview of how a final item cluster alignment might translate into the schematics of the final item cluster is shown in Figure 4.

**Figure 4.** Sample representation of the relationship of an item cluster aligned to two PEs to its component items, with sample item subtypes shown



# REFERENCES


APT Innovations in Testing. (2015). *Implementing automatic scoring in K-12 assessments: Development and scoring considerations*. Retrieved from http://www.innovationsintesting.org/program-webcast-implementing-automated-scoring.aspx

Johnstone, C. J., Altman, J., & Thurlow, M. (2006). *A state guide to the development of universally designed assessments.* Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved from http://www.cehd.umn.edu/nceo/OnlinePubs/StateGuideUD/default.htm

National Assessment Governing Board. (2007). *Science assessment and item specifications for the 2009 National Assessment of Educational Progress*. Retrieved from http://www.nagb.org/content/nagb/assets/documents/publications/frameworks/science/2009-science-specification.pdf

National Assessment of Educational Progress. (n.d.) *NAEP – 2009 science interactive computer tasks.* Retrieved from http://www.nationsreportcard.gov/science_2009/ict_tasks.asp

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

National Research Council (NRC). (2014). *Developing assessments for the Next Generation Science Standards*. Committee on Developing Assessments of Science Proficiency in K–12. Board on Testing and Assessment and Board on Science Education, J. W. Pellegrino, M. R. Wilson, J. A. Koenig, & A. S. Beatty (Eds.). Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Next Generation Science Standards (NGSS). (2013). *NGSS front matter executive summary.* Retrieved from http://www.nextgenscience.org/sites/ngss/files/Final%20Release%20NGSS%20Front%20Matter%20-%206.17.13%20Update_0.pdf

Partnership for Assessment of Readiness for College and Careers (PARCC). (2013). *PARCC item development technical guide (Table 5.3)*. Retrieved from https://parccsharepoint.org/Public_Access/Diagnostic%20and%20K-1%20Assessment%20ITN%20-%20June%202013%20-%20Reference%20Documentation/June%202013%20DRAFT_PARCC%20Item%20Development%20Technical%20Guide%2020130627.pdf

Smarter Balanced Assessment Consortium (Smarter Balanced). (2014). *Getting students ready for the field test: Information on item and response types*. Retrieved from http://sbac.portal.airast.org/wp-content/uploads/2014/03/Student-Item-and-Response-Types.pdf

Moulding, B.D., Bybee, R.W., & Paulson, N. (2015). *A Vision and Plan for Science Teaching and Learning.*

Webb, N. L. (2002). *Depth-of-knowledge levels for four content areas*. Retrieved from http://www.hed.state.nm.us/uploads/files/ABE/Policies/depth_of_knowledge_guide_for_all_subject_areas.pdf

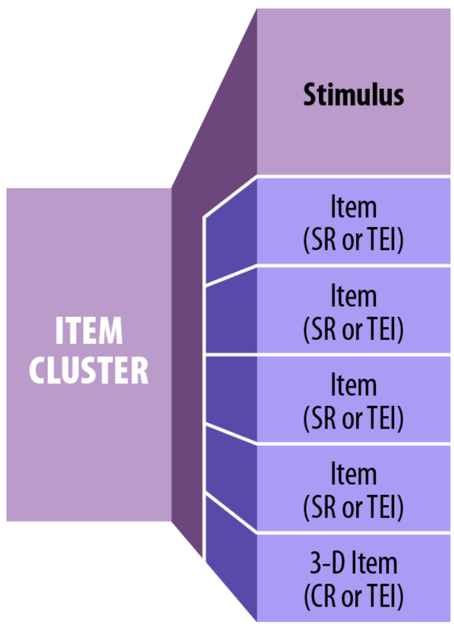
# APPENDICES

## APPENDIX A. GUIDELINES FOR ITEM CLUSTERS

The basic premises of the discussion in this appendix are as follows:

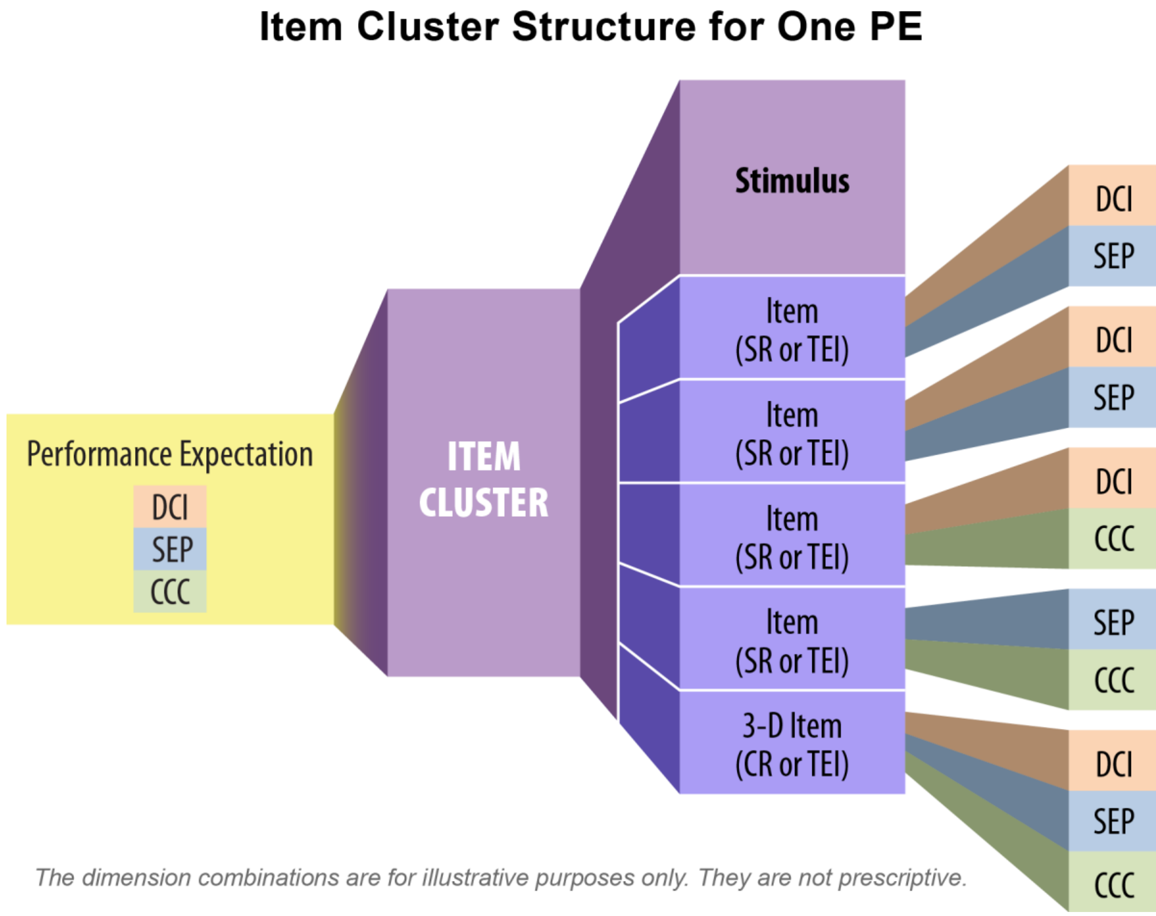
- Item clusters, not individual items, are the base unit for SAIC test development. That is, individual items are intentionally developed to be situated within the context of an item cluster and not to be used as stand-alone items. The basic organization of a typical item cluster is shown in Figure A-1.

**Figure A-1.** Sample representation of the relationship of an item cluster to its component items




- Item clusters are the primary focus for developers in terms of alignment to the NGSS. That is, each item cluster ***must*** demonstrate strong three-dimensional alignment to the NGSS.
- To meet NGSS alignment expectations, item clusters must be inclusive of all three dimensions of the NGSS that are inherent in the associated PE(s) (i.e., DCI, SEP, and CCC).
- Each individual item within the cluster must align with at least two dimensions of the NGSS (i.e., DCI, SEP, and/or CCC) to qualify for inclusion in an item cluster. As an example, Figure A-2 shows an elaboration of Figure A-1, with the dimensions of each item in a simplified single-PE cluster included.

**Figure A-2.** Sample representation of the relationship of an item cluster aligned to a single PE to its component items, with item-aligned dimension combinations shown




- It should be noted that all items will exhibit some degree of alignment to the disciplinary context of the DCI, as all items are inextricably linked to the context, which was selected to align to the discipline(s) associated with the PEs. Therefore, every item in an item cluster will naturally fall within the content limits of the DCI, but not every item may truly call for the assessment of understanding of the content put forth in the DCI. Thus, items that only align to SEPs/CCCs are not intended to be viewed as devoid of a disciplinary context, but, rather, are intended to be viewed as items that place relatively greater emphasis on assessing an associated SEP and/or CCC than they do on assessing the underlying DCI content. In fact, each SEP and CCC has its own knowledge that is most relevant in context of a DCI.
- If an evidence statement appears to align to a single SEP or CCC dimension, it is recommended that the evidence statement be grouped with the DCI in order to prevent an item writer from developing an item to a single dimension in isolation (e.g., attempting to assess a science practice in isolation without tying the item to the context and/or the DCI).
- At least one item should be aligned to all three dimensions (as shown in Figure A-2), as this is the overall vision of the NGSS.

- Each item is inextricably linked to the stimulus and to the other items within the item cluster. This means that student exposure to the stimulus is considered essential in order for the student to respond correctly to any individual item, and that the cluster of items must be constructed in such a way that individual performance on each item is adversely affected if an item is responded to without the context of the other items in the cluster. (See the following “Item Cluster Stimuli” subsection for more information on stimuli for item clusters.)
- Testing time for each item cluster will be content dependent, but an estimate of 20 minutes of testing time per item cluster is assumed for summative assessment purposes. This estimate will be further refined as prototypes are completed.
- Each item cluster will have items tied to evidence statement selections for one or more PEs. These evidence statement selections are the fundamental component of item alignment with scientific content. Item clusters aligned to more than one PE could be from the same domain (i.e., Physical Sciences, Life Sciences, Earth and Space Sciences), but could also be from related, but different, content areas (e.g., photosynthesis and chemical reactions). PEs can also be from different domains. PEs from the domain of Engineering, Technology, and Applications of Science should always be bundled with PEs from one of the science disciplines.

  The rationale for correlating the parts of a PE evidence statement with two or more of the PE’s dimensions is that such a correlation provides a building block for item construction when the PE is bundled with one or more other PEs in an item cluster. Looking at the entirety of the dimensions and evidence statements for two or more PEs in an item cluster can be somewhat overwhelming in terms of the amount of information provided in relation to assessment goals. By structuring the PE and evidence statement components into natural dimensional/evidence-statement relationships that might form the basis of an item in an item cluster, the item cluster developer can better perceive how all of these PE elements fit together and how they might be used, along with the multidimensional alignment groupings for other PEs in an item cluster, to form a balanced, conceptually cohesive item cluster.
- While it may be possible to develop items within a single cluster that are collectively sufficient to assess the entirety of the evidence statement for a single PE, this is not preferable and will not be possible in many, if not most, cases. For item clusters inclusive of more than one PE, it is not expected that a single item cluster will be able to fully assess the complete set of evidence statements for each PE, and thus, PEs may appear in other clusters in the assessment. For example, PEs HS-PS1-3 and HS-ESS2-5 might be combined in a single item cluster. The evidence statements for these two PEs are shown in Table A-1 and Table A-2, respectively.

**Table A-1.** Evidence statements for HS-PS1-3

| Observable features of the student performance by the end of the course: | | |
|---|---|---|
| **1** | **Identifying the phenomenon to be investigated** | |
| | a | Students describe the phenomenon under investigation, which includes the following idea: the relationship between the measurable properties (e.g., melting point, boiling point, vapor pressure, surface tension) of a substance and the strength of the electrical forces between the particles of the substance. |
| **2** | **Identifying the evidence to answer this question** | |
| | a | Students develop an investigation plan and describe the data that will be collected and the evidence to be derived from the data, including bulk properties of a substance (e.g., melting point and boiling point, volatility, surface tension) that would allow inferences to be made about the strength of electrical forces between particles. |
| | b | Students describe why the data about bulk properties would provide information about strength of the electrical forces between the particles of the chosen substances, including the following descriptions: |
| | | i. The spacing of the particles of the chosen substances can change as a result of the experimental procedure even if the identity of the particles does not change (e.g., when water is boiled the molecules are still present but further apart). |
| | | ii. Thermal (kinetic) energy has an effect on the ability of the electrical attraction between particles to keep the particles close together. Thus, as more energy is added to the system, the forces of attraction between the particles can no longer keep the particles close together. |
| | | iii. The patterns of interactions between particles at the molecular scale are reflected in the patterns of behavior at the macroscopic scale. |
| | | iv. Together, patterns observed at multiple scales can provide evidence of the causal relationships between the strength of the electrical forces between particles and the structure of substances at the bulk scale. |
| **3** | **Planning for the investigation** | |
| | a | In the investigation plan, students include: |
| | | i. A rationale for the choice of substances to compare and a description of the composition of those substances at the atomic molecular scale. |
| | | ii. A description of how the data will be collected, the number of trials, and the experimental set up and equipment required. |
| | b | Students describe how the data will be collected, the number of trials, the experimental set up, and the equipment required. |
| **4** | **Collecting the data** | |
| | a | Students collect and record data — quantitative and/or qualitative — on the bulk properties of substances. |
| **5** | **Refining the design** | |
| | a | Students evaluate their investigation, including evaluation of: |
| | | i. Assessing the accuracy and precision of the data collected, as well as the limitations of the investigation; and |
| | | ii. The ability of the data to provide the evidence required. |
| | b | If necessary, students refine the plan to produce more accurate, precise, and useful data. |

**Table A-2.** Evidence statements for HS-ESS2-5

| Observable features of the student performance by the end of the course: | | |
|---|---|---|
| **1** | **Identifying the phenomenon to be investigated** | |
| | a | Students describe the phenomenon under investigation, which includes the following idea: a connection between the properties of water and its effects on Earth materials and surface processes. |
| **2** | **Identifying the evidence to answer this question** | |
| | a | Students develop an investigation plan and describe the data that will be collected and the evidence to be derived from the data, including: |
| | | i. Properties of water, including: |
| | | a) The heat capacity of water; |
| | | b) The density of water in its solid and liquid states; and |
| | | c) The polar nature of the water molecule due to its molecular structure. |
| | | ii. The effect of the properties of water on energy transfer that causes the patterns of temperature, the movement of air, and the movement and availability of water at Earth's surface. |
| | | iii. Mechanical effects of water on Earth materials that can be used to infer the effect of water on Earth's surface processes. Examples can include: |
| | | a) Stream transportation and deposition using a stream table, which can be used to infer the ability of water to transport and deposit materials; |
| | | b) Erosion using variations in soil moisture content, which can be used to infer the ability of water to prevent or facilitate movement of Earth materials; and |
| | | c) The expansion of water as it freezes, which can be used to infer the ability of water to break rocks into smaller pieces. |
| | | iv. Chemical effects of water on Earth materials that can be used to infer the effect of water on Earth's surface processes. Examples can include: |
| | | a) The solubility of different materials in water, which can be used to infer chemical weathering and recrystallization; |
| | | b) The reaction of iron to rust in water, which can be used to infer the role of water in chemical weathering; |
| | | c) Data illustrating that water lowers the melting temperature of most solids, which can be used to infer melt generation; and |
| | | d) Data illustrating that water decreases the viscosity of melted rock, affecting the movement of magma and volcanic eruptions. |
| | b | In their investigation plan, students describe how the data collected will be relevant to determining the effect of water on Earth materials and surface processes. |
| **3** | **Planning for the investigation** | |
| | a | In their investigation plan, students include a means to indicate or measure the predicted effect of water on Earth's materials or surface processes. Examples include: |
| | | i. The role of the heat capacity of water to affect the temperature, movement of air and movement of water at the Earth's surface; |
| | | ii. The role of flowing water to pick up, move and deposit sediment; |
| | | iii. The role of the polarity of water (through cohesion) to prevent or facilitate erosion; |
| | | iv. The role of the changing density of water (depending on physical state) to facilitate the breakdown of rock; |
| | | v. The role of the polarity of water in facilitating the dissolution of Earth materials; |
| | | vi. Water as a component in chemical reactions that change Earth materials; and |
| | | vii. The role of the polarity of water in changing the melting temperature and viscosity of rocks. |
| | b | In the plan, students state whether the investigation will be conducted individually or collaboratively. |
| **4** | **Collecting the data** | |
| | a | Students collect and record measurements or indications of the predicted effect of a property of water on Earth's materials or surface. |
| **5** | **Refining the design** | |
| | a | Students evaluate the accuracy and precision of the collected data. |
| | b | Students evaluate whether the data can be used to infer the effect of water on processes in the natural world. |
| | c | If necessary, students refine the plan to produce more accurate and precise data. |

It is clear that a single item cluster could not fully address the combined evidence statements of the two PEs shown in Tables A-1 and A-2, and, in fact, it is very unlikely that a single item cluster could fully address the evidence statements for even one of these two PEs. The specifications for a particular item cluster would only include a subset of these two sets of evidence statements, focusing on the most apparent points of intersection between the two.

It is important to note, however, that the NGSS emphasize that the PEs, and therefore their related evidence statements are intentionally written to be “context agnostic” (i.e., the ideas and relationships addressed in the PE need not be assessed within a specific context), and that, rather, the *application* of the content and skills in any PE should be the focus of assessment. It is envisioned that numerous contexts can be used to assess all or part of any given PE. In some cases, the context might be indirectly steered in a certain direction by the nature of the dimensions for the PE, but there will always be some latitude for the item cluster developer to choose how to best frame a context to align with the PEs. Thus, the item cluster alignment should also be carefully crafted in a manner that does not delimit the boundaries of potential contexts that could be used for a particular item cluster. The item cluster evidence statements should be chosen so as to allow for flexibility in terms of the specific context that is applied.

Beyond the identification of the pertinent PE(s) and the related components of the supporting evidence statements, the specifications for an item cluster may be further specified by individual states, but should contain several other elements, including the following:

1. The number (or range) of items to be included in the item cluster.
2. The number of points (or range of points) to be assigned to each item in the cluster.
3. The SEPs, DCIs, and/or CCCs (collectively referred to as the dimensions) that apply to each item in the item cluster. Each item should align to at least two dimensions (e.g., an SEP and a DCI or a DCI and a CCC); for some items in a cluster, including any CR items, it is anticipated that all three dimensions could align to the item. Item parts may align to two dimensions with one part fully aligned to all three dimensions.
4. The type of each item (e.g., SR, TEI, or CR). In many cases, a choice of item types and/or subtypes, rather than a specific item type, may be identified for a particular item. For example, for the first three items in an item cluster, the specifications might require that the first item be a TEI, the second item be either an SR or a TEI (with allowable SR and TEI subtypes indicated), and the third item be either a TEI or a CR. (Note that CR items and item parts may appear in any position in the item cluster, either in a sequence of CR item parts or interspersed throughout the item cluster.)
5. The assessment boundaries for the item cluster, based on the PE(s).
6. Any target clarifications associated with the PE(s).
7. Guidance on the gradual and purposeful building of cognitive complexity and difficulty (which may be related, but may differ with respect to item development).
8. Guidance on the inclusion of within-cluster scaffolding (for example, sample student data may be provided/introduced to prevent a student from using erroneous or incorrectly

determined data from a previous item that will impact his or her score on a subsequent item).

### *Item Cluster Stimuli*

Each item cluster will have a common stimulus upon which all items in the cluster are dependent. For example, the stimulus may be a text-based description that includes a description of experimental data or an experimental setup, or may be a fully interactive computer simulation in which students can control for different variables, run multiple trials, and collect data needed to address the full depth and breadth of one or more PEs. The majority of the stimulus may appear at the beginning of the item cluster, with additional stimulus material interspersed throughout the item cluster (often to support scaffolding). Items that can be answered without referring to the stimulus are not appropriate for an NGSS-based assessment.

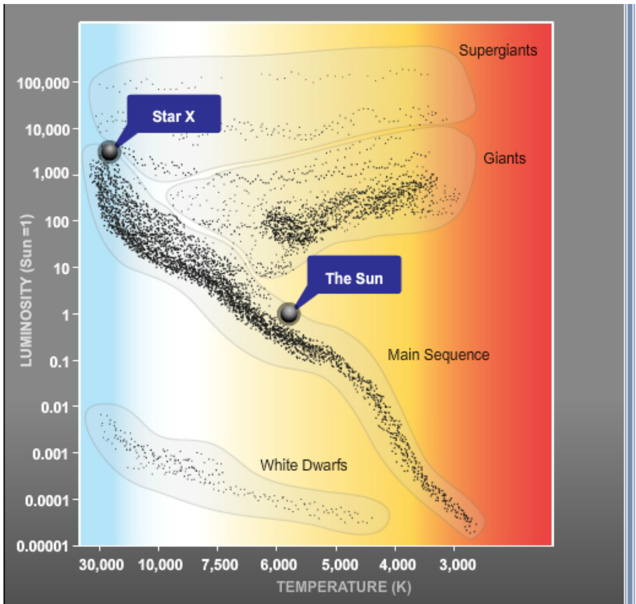
Since the stimuli will be identified or developed for use on a large-scale summative assessment, it may be assumed that the large majority of stimuli will be text-based. However, developers may propose creative solutions and should not allow current challenges of administration to constrain their thinking.

The stimulus for an item cluster must be broad enough in content to support all of the items in the cluster yet flexible enough for students to exhibit their ability to demonstrate their capabilities to apply SEPs. Because many item clusters may require students to demonstrate capacity to develop and use models, plan and carry out an investigation, and subsequently interpret the data and construct explanations based on the data, the stimulus may need to include information that is extraneous or tangential to the overall goal in order for students to demonstrate their capacity to identify appropriate or pertinent information or data from a stimulus.

The stimulus that is common to all items in the cluster should always precede the first item in the item cluster. It may be necessary to add follow-up information and related stimulus material prior to other items in the cluster, in order to effectively scaffold to higher-level thinking tasks. Additionally, providing timely information and data only at the point(s) when needed for items other than the first item in the cluster helps to avoid the problem of information overload or of unnecessarily adding to the student's cognitive load (see the "Universal Design/Vocabulary and Language" section in **Chapter Two**).

As an example, Figure A-3 shows an SR item that occurs in the middle of an item cluster. In this example, the student is given follow-up information about the Sun and asked to predict future stages of the Sun's evolution, based on these data and the student's understanding of what the Hertzsprung-Russell diagram says about stellar evolution. (it is important to note that this item is not necessarily aligned to the NGSS, it is for illustrative purposes only.  NGSS does not specify for students to know the name "Hertzsprung-Russell diagram").

**Figure A-3.** Example of stimulus and task for a mid-cluster Item




| | Temperature | Mass | Radius | Luminosity |
|---|---|---|---|---|
| Star X | 22,000 K | 10 M | 5 R | 5,800 L |
| The Sun | 5,800 K | 1 M | 1 R | 1 L |

Your next task is to examine data on two stars, Star X and our Sun, which are currently on the main sequence.

The stars are plotted on the H-R diagram and data about them are shown in the table below the H-R diagram.

Use the information on the left and your knowledge of processes that occur in stars to predict what stages you expect each star to go through as it ages.

What future stages will the Sun likely go through? Select all that apply.

- ☐ **A** Giant
- ☐ **B** Supergiant
- ☐ **C** White dwarf
- ☐ **D** Planetary nebula
- ☐ **E** Supernova

Click "NEXT" to continue.

Source: National Assessment of Educational Progress (n.d.).

## APPENDIX B. ITEM TYPES AND SUBTYPES FOR ITEM CLUSTERS

This section discusses the types of items that can be used to populate each item cluster developed for online administration. The typical item cluster will consist of four to six single-part or multipart items. Table B-1 summarizes the known item types that are frequently used on large-scale state and multistate summative assessments. Following the table, each item type and its associated item subtypes are discussed in greater detail.

**Table B-1.** Frequently used item types

| Item Type | Item Subtype and Structure | Response Behavior | Sample Task/Purpose |
|---|---|---|---|
| **Selected response (SR)** | Multiple choice, single correct response (MC) | Select an option by clicking on a radio button or anywhere in the text; generally four options | Identify an appropriate rationale to explain a scientific phenomenon; select an appropriate solution to an engineering design problem |
| | Multiple choice, multiple correct responses (multiple select) (MS) | Select among multiple options by marking a checkbox or clicking anywhere in the text; generally five or more options | Identify a plausible explanation for a phenomenon and the appropriate rationale; select statements that support a claim of a scientific phenomenon |
| | Matching tables (with True/False or Yes/No) (MT) | Select among multiple statements by marking an option in a table cell for each row. | Identify appropriate data; identify appropriate statements given constraints |
| | Inline choice (IC) | Select an option by clicking on a drop-down menu; four options | Identify evidence that would support a claim or explanation |
| | Hot spot (HS) | Select text or objects in a response area; may include more than four options; each option should be a salient feature | Identify aspects of a model that support a given claim |

**Table B-1.** *(continued)*

| Item Type | Item Subtype and Structure | Response Behavior | Sample Task/Purpose |
|---|---|---|---|
| **Constructed response (CR)** | Short text (ST) | Enter text into a multiline text box. | Generate a hypothesis; describe a possible engineering problem |
| | Equation or numeric entry; edit equations (EQ) | Enter mathematical symbols and/or numbers; may include selecting special symbols from an on-screen table or menu | Use a mathematical model to represent a scientific phenomenon; determine a solution to an engineering problem |
| | Cloze text (CT) | Enter text into a text box within a sentence | Construct a description or simple explanation of a scientific phenomenon or solution to an engineering problem |
| | Table text (TT) | Enter text into a table or chart | Design an investigation or make predictions |
| | Constructed response (essay) (CR) | Use keyboard to enter text into a multiline text box; may include text formatting tools | Construct a detailed explanation of a scientific phenomenon or solution to an engineering problem |
| **Technology-enhanced items (TEIs): Data selection** | Slider (SL) | Select a value on a scale by clicking on a slider and dragging it to the appropriate location on the scale | Select values for variables to design and conduct an investigation |
| | Data inspector (DI) | Select a value on a graph by clicking on a slider attached to a vertical line and dragging it to the appropriate location on the *x*-axis | Select data as evidence to support an explanation or the solution to an engineering problem |

**Table B-1.** *(continued)*

| Item Type | Item Subtype and Structure | Response Behavior | Sample Task/Purpose |
|---|---|---|---|
| **Technology-enhanced items (TEIs): Data display** | Graphing: plot points or line graphs (G) | Click in the question response area to create a point or start a line; click and drag to complete the line and to add additional data points | Create a model; analyze data |
| | Function graph (FG) | Click on an icon to select the type of graph; drag two points to the correct position | Create a model; analyze data |
| | Composite graph (CMG) | Click in the question response area to create composite displays including two or more of the following: points, lines, curves, or shaded areas | Analyze data by fitting a line or curve to a set of points; represent possible solutions to an engineering problem, using shaded areas |
| | Bar graph; histogram (BG) | Drag bars to display data in a bar graph or histogram | Create a model; analyze data |
| | Fraction model (circle graph) (FM) | Click on the edge of a circle to create a new division, and/or drag division lines to the appropriate location within a circle | Create a model; analyze data |
| | Interactive number line (INL) | Click in the question response area to create a point or start a line; click and drag to complete the line | Create a model; analyze data |
| | Zoom number line (ZNL) | Present graphical data by zooming in on a number line to graph one point (often used for fractions) | Analyze data; demonstrate how an outcome may be affected by a unique context |

**Table B-1.** *(continued)*

| Item Type | Item Subtype and Structure | Response Behavior | Sample Task/Purpose |
|---|---|---|---|
| **Technology-enhanced items (TEIs): Drag and drop** | Drag and drop single or multiple elements (DD) | Select an object by clicking on it; then drag and drop it into an appropriate location within the response area (including tables and art) | Modify a model to better fit a new constraint |
| | Hot text: select and order text (HT) | Select text by clicking on it; then drag and drop it into an appropriate location within the response area (including tables and art) | Reorder steps/stages into an appropriate sequence, given a context or scenario |
| | Text extraction (EXT) | Select text from a sentence or equation by clicking on it; then drag and drop it into an appropriate location within the response area (including tables and art) | Select parts of a description of a scientific phenomenon that support an explanation or solution to a problem |
| **Multi-component** | Two-part multiple choice, with evidence-based response (EB MC) | Part 1: Select an option by clicking on a radio button or anywhere in the text<br>Part 2: Select a response to support the response to Part 1 | Identify a response/claim and the appropriate rationale to support the response/claim |
| | Other | Any combination of two functionalities within a single item | Generate and test models; display, analyze, and interpret data; design and conduct investigations or solutions and explain results |

Source: Based on Smarter Balanced (2014) and PARCC (2013).

## Selected-Response Items

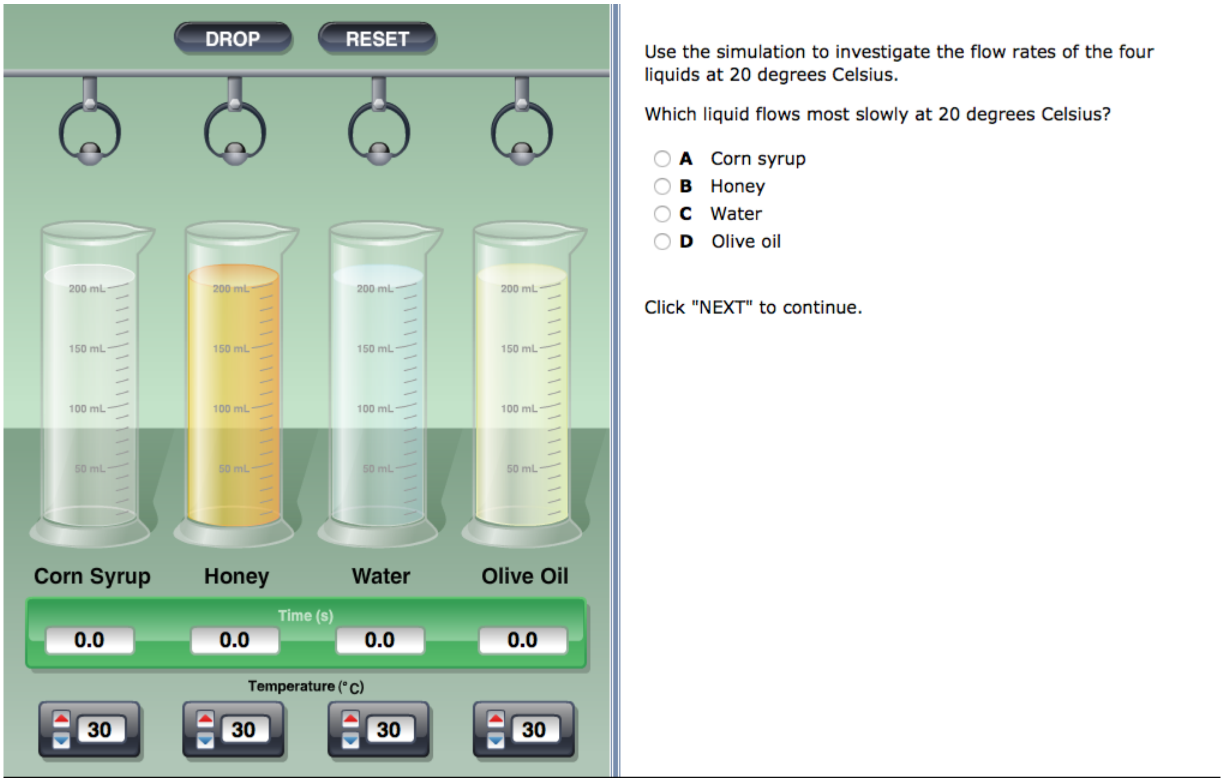
Selected-response (SR) items have long been one of the most common item types used in summative assessments. Historically, the most commonly used SR item subtype is multiple choice (MC). Typically, an MC item has a stimulus, a stem, and four or five answer options, of which only one option is correct. (The stimulus may occur prior to an MC item in an item cluster.) Items in which a larger number of answer options and correct answers are possible are referred to as multiple-select (MS) items. Another variation of the traditional MC includes multiple “Yes-No” or “True-False” questions. The inclusion of just four such questions increases the number of possible answer options to 16, thereby increasing the difficulty level. This item type is referred to as matching tables (MT). Each of these item types is discussed later in this section.

SR items offer the opportunity to leverage automated scoring, as discrete student interactions can be easily tracked and tallied through machine scoring. Partial credit is also possible through automated scoring of selected-response item subtypes such as MS and MT, although this is not typical.

### ***Multiple Choice***

While MC questions are generally not considered technology-enhanced items (TEIs), they can be designed to correlate with technology-enhanced scenarios and simulations. For example, Figure B-1 shows an MC item that is tailored to a simulation investigating liquid flow rates.

**Figure B-1.** Example of an MC item with simulation




Source: National Assessment of Educational Progress (n.d.).

While this MC item itself does not directly align to an SEP, the combination of the simulation and the follow-up question asked in the MC partially aligns to two SEPs (i.e., SEP 3—*Planning and carrying out investigations* and SEP 4—*Analyzing and interpreting data*).

A well-crafted MC item will include carefully chosen distracter answers that closely align to common student errors or misconceptions. The avoidance of cuing is also a concern when creating distracters, and answer options should avoid distinctive wording or visual appearance that might signal that one answer option is more likely to be correct (or, in some cases, incorrect) than the other options. In most cases, outlier distracters, as well as distracters that are a subset of another distracter (thereby logically disqualifying both distracters as possible correct answers), should be avoided. The options should be ordered to follow some logical pattern (e.g., least to greatest, shortest to longest) that filters out any unwitting bias in choosing the correct answer option.

MC items embedded within item clusters can be used as scaffolding to other items that require students to extend their understanding to new contexts or to more challenging concepts. Because an MC item, by definition, provides a student with a correct answer that the student must only distinguish from incorrect answers, its efficacy for use in items requiring higher-level cognitive abilities may be limited in most cases. For such items, the following SR item subtypes may prove to be more appropriate for use in item development.

### *Multiple Select*

MS items are less susceptible to guessing and process-of-elimination techniques, owing to their greater numbers of answers and options. An MS item usually has between five and eight options, of which at least two options form the key. In lower grade bands, the number of options will typically be fewer than in higher grade bands, and in the lowest grades, the number of correct responses is often supplied to the student. Because they include several correct options, MS items are useful for items in which students must identify several characteristics or properties of a system or material, or for items in which several supporting arguments must be distinguished from distracter arguments.

### *Matching Tables*

MT items offer a series of two-option selections of the yes/no or true/false variety. This format lends itself to items in which students are called upon to sort properties or data into the correct categories or identify relevant variables or characteristics from irrelevant ones. Because a matching table with just four choices offers 16 possible answers, guessing strategies are generally ineffective with this subtype.

### *Two-Part Multiple Choice*

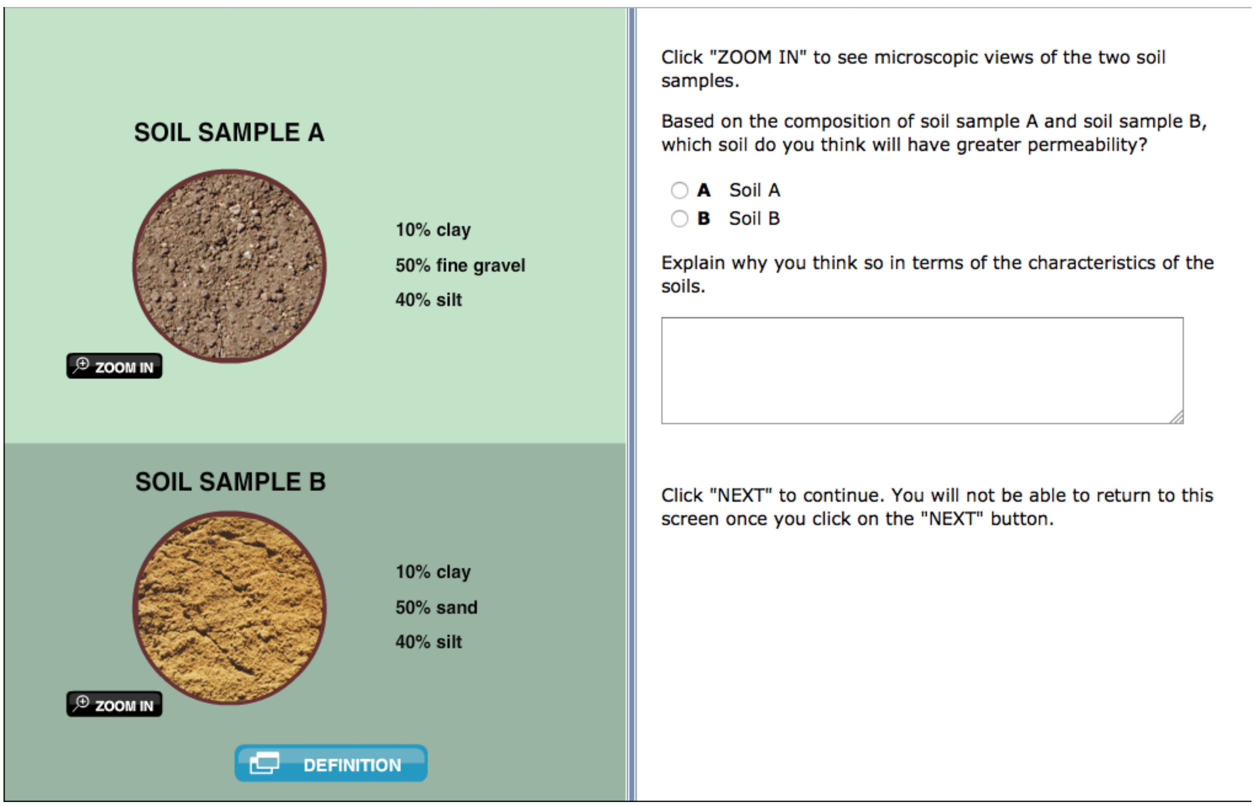
A special type of SR item is the two-part multiple choice subtype, in which the option selected in the second multiple-choice question provides the evidence to support the option chosen in the first multiple-choice question. For this reason, this subtype is referred to as “two-part multiple choice, with evidence-based response” (EB MC) in Table B-1, where it is listed as a multi-component item type. Such an item format is particularly appropriate for assessing scientific reasoning. Generally, credit is given for EB MC items only if both parts are answered correctly (because the reasoning is linked to the evidence), which, like the MT subtype, provides students very little opportunity to apply guessing strategies.

## Constructed-Response Items

CR items generally require students to provide an explanation, and, therefore, are most appropriate for items requiring unambiguous evidence of analytical thinking. These types of items can be divided into two main categories: short entry (short text [ST], Cloze text [CT], and table text [TT]) and long entry (essay [CR]). (A third CR item subtype category is equation/numeric entry [EQ items].) The two primary categories are distinguished primarily by the level of sophistication and textual elaboration required to fully answer a prompt. The length of time needed to answer a CR item can be as short as less than a minute for some short-entry items, to upwards of 15 minutes for scaffolded CR items that have multiple parts.

Figure B-2 shows an ST CR item (for illustrative purposes; inclusion here does not denote NGSS alignment) in which the student is first asked to choose which of two soil samples is the most permeable in a two-option MC question, and is then asked to provide a short explanation for the choice in terms of the soil characteristics. The salient ideas for answering the second part are relatively narrow in scope and do not require an extended analysis.

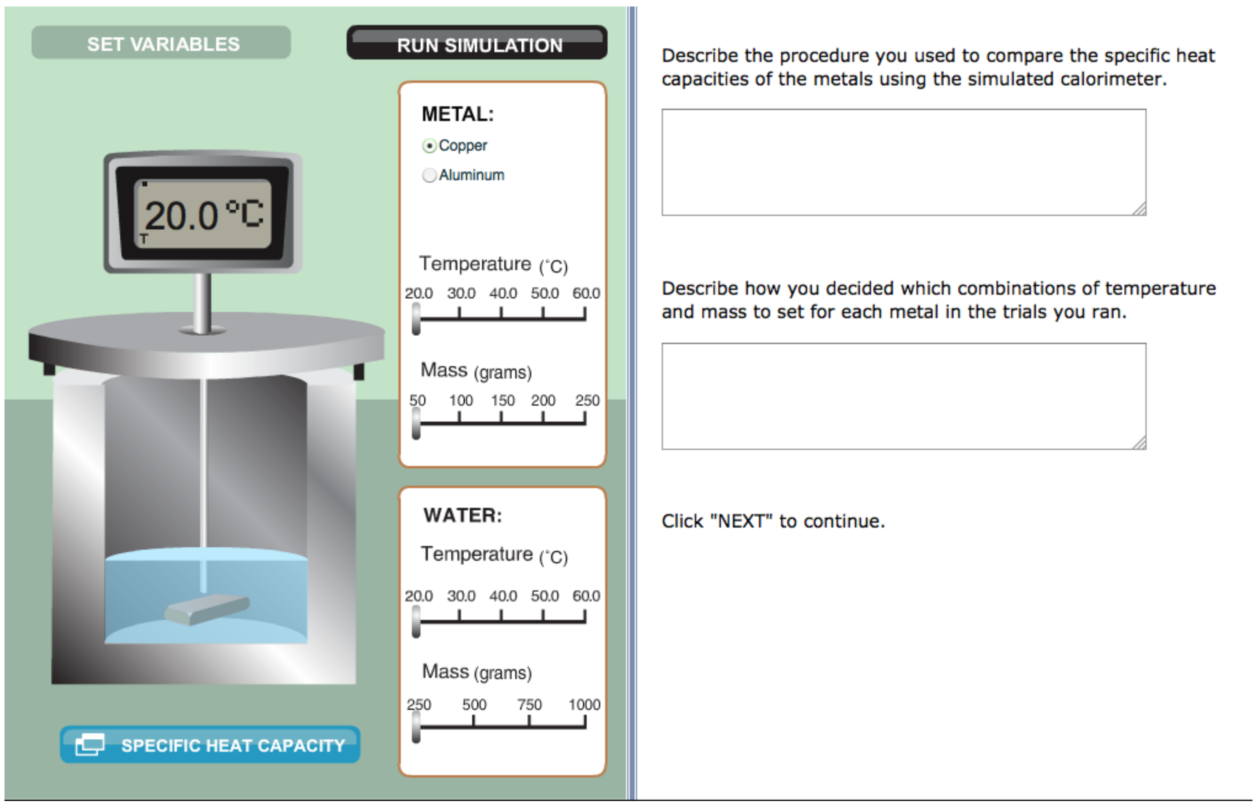
**Figure B-2.** Example of an ST CR item




Source: National Assessment of Educational Progress (n.d.).

In contrast to an ST CR item, an essay item might include two or more parts that build upon one another. These parts might be answered in a single CR box or spread out across more than one box, as in Figure B-3. The example shown in Figure B-3 (for illustrative purposes; inclusion here does not denote NGSS alignment) focuses on identifying the rationales for planning an investigation; other essays might instead concentrate on offering explanations of observed phenomena and data supported by reasoning.

**Figure B-3.** Example of an essay item in which several boxes are used to complete the analysis




Source: National Assessment of Educational Progress (n.d.).

### ***Item Construction***

It is anticipated that all essay CRs and most short-entry CRs will need to be hand scored, although artificial-intelligence capabilities might lend themselves to the scoring of simpler short-entry CRs. In all cases, essay CRs need to be supported with a complete key, describing necessary elements of a correct response, and a scoring rubric, detailing how various levels of completeness in responses translate into score points. Depending upon how open ended a CR item is, additional correct responses may be identified through the review of student work during the benchmarking process. Prompts for CRs must be very specific about what is expected for a complete response, and may require a limited amount of cuing to alert students as to what general ideas and topics need to be considered in their responses. All stems should be robust and sufficiently detailed to provide students with the data and background information that are critically necessary to generate the desired response.

## Technology-Enhanced Items

The TEI item type spans a large class of item subtypes that have been employed in summative assessment of the Common Core State Standards (CCSS) by the Smarter Balanced Assessment Consortium (Smarter Balanced) and the Partnership for Assessment of Readiness for College and Careers (PARCC). These item subtypes can be used within a text-based framework or can be employed with multimedia stimulus materials, such as videos or simulation tools.

These items ideally should only be used for assessment of content understanding and skills that cannot be measured as effectively using SR subtypes, even when existing template subtypes exist, due to the added cost of creating, scoring, and calibrating TEIs. In particular, TEIs should never be used in situations in which the TEI is logically equivalent to an SR subtype. For example, a drag-and-drop TEI in which students are asked to drag each of seven statements or properties into one of two boxes (labeled True and False) is cognitively equivalent to an MS item and should be either revised or converted to the MS format.

Despite the incremental costs of developing TEIs in comparison to developing traditional SR items, TEIs do have some clear advantages over traditional SR formats. Foremost among these is the ability to create items that are of higher-order cognitive complexity but can be machine scored, as are SR items. As is the case for some of the more sophisticated SR item subtypes, such as matching tables, the number of available options for answering some TEIs effectively rules out guessing strategies and thus ensures that a student is demonstrating true understanding when entering a response.

As previously noted, there are a large number of TEI subtypes, but many of them have similar functionalities or primary assessment purposes that can be used as a basis for grouping them into TEI classes. Each TEI class tends to lend itself best to assessment of specific SEPs, although it should be noted that this is a general conclusion and is not meant to imply that a given TEI subtype cannot be used with other SEPs or with CCCs when a natural connection exists. In Table B-1, the listed TEIs have been placed into three classes of TEIs, each of which is briefly discussed in the following sections.

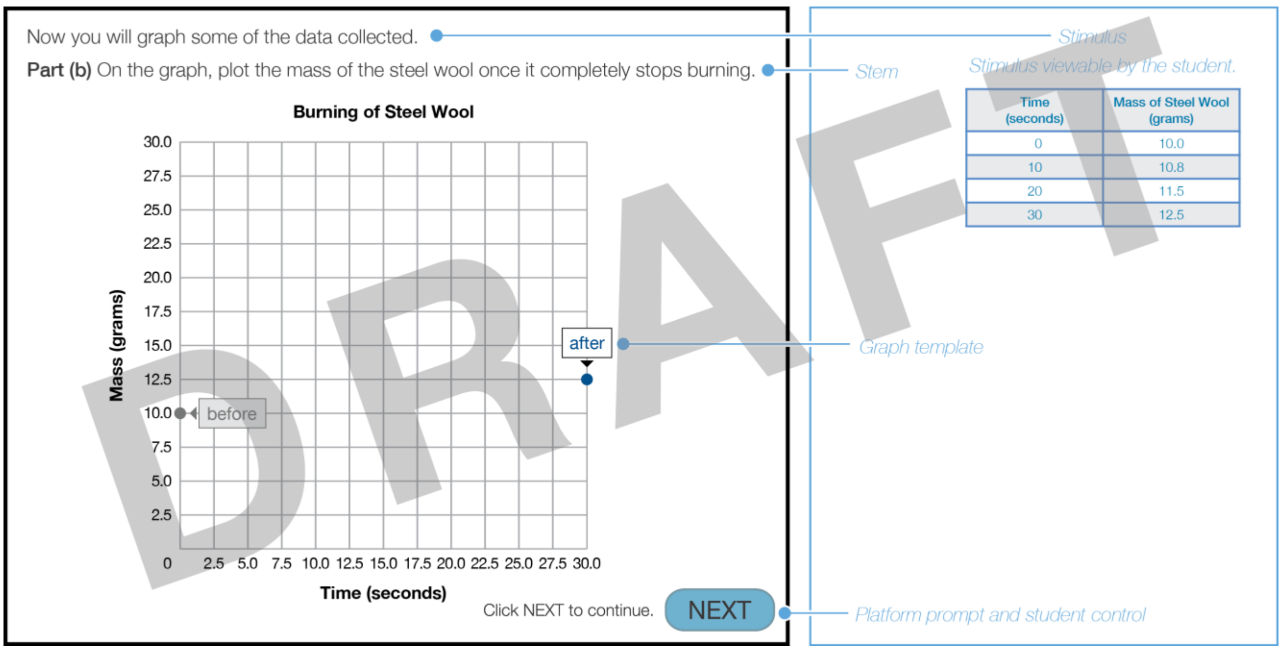
### ***Data Selection***

This class of TEIs, which includes the SI and DI subtypes, fits well with SEP 4 (*Analyzing and interpreting data*) and can also be used to select variable values for investigations (SEP 3—*Planning and carrying out investigations*) and numerical values on a graph that support a conclusion or provide evidence for an explanation (SEP 7—*Engaging in argument from evidence*).

## *Data Display*

Most of the graphing and data plotting tools occur in this class of TEIs. As a result, TEI data display subtypes tend to align well with SEP 2 (*Developing and using models*), SEP 4 (*Analyzing and interpreting data*), SEP 5 (*Using mathematics and computational thinking*), and SEP 8 (*Obtaining, evaluating, and communicating information*). Figure B-4 shows a sample data display (graphing) item prototype (within the context of an item cluster). (Please note that SAIC NGSS-aligned prototypes are still in development, and the sample item shown in Figure B-4 will undergo further revisions during the development process, as it is currently only aligned to one dimension of the NGSS.)

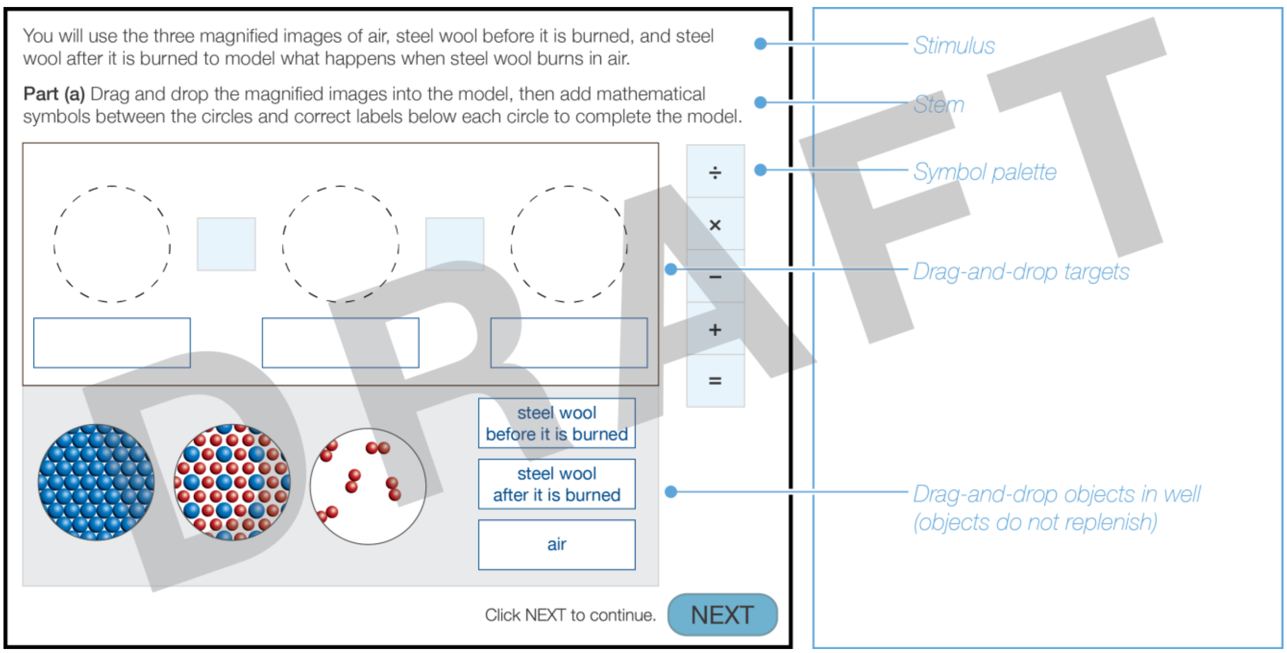
**Figure B-4.** Sample data display (graphing) item prototype (draft version)



### Drag-and-Drop

The drag-and-drop TEI subtypes are useful for categorizing and ordering statements, data, or properties. The ordering capabilities of the HT subtype are particularly useful in aligning to SEP 3 (*Planning and carrying out investigations*) and SEP 6 (*Constructing explanations and designing solutions*), as well as SEP 2 (*Developing and using models*), by allowing students to sequence steps in an investigation, identify missing elements of a scientific explanation, or organize objects to illustrate the functions of a system. Figure B-5 shows a sample drag-and-drop item prototype (within the context of an item cluster). (Please note that SAIC NGSS-aligned prototypes are still in development, and the sample item shown in Figure B-5 may undergo further revisions during the development process.)

**Figure B-5.** Sample drag-and-drop item prototype (draft version)



## Hybrid Approach to TEIs

While much work has been done by the CCSS assessment consortia and by platform vendors, individual states, and organizations to define the functionality for the TEI subtypes listed in Table B-1 and discussed earlier in this appendix, these subtypes were originally developed to assess English language arts and mathematics content and associated skills. It is generally accepted that much of their functionality will cross over effectively to assess science, but it is also prudent to assume that additional functionality may be necessary to target specific science skills and concepts identified in the NGSS. Accordingly, a hybrid approach, in which aspects of different TEIs are layered or fused together to create new and unique item types that more effectively assess the NGSS, should be considered.

Additionally, richer interactivity in stimuli may prove to be a necessary component of an NGSS-aligned assessment. States are encouraged to explore the development of novel item types and functionalities, and to consider the technical requirements for online delivery of the novel item types and richer stimuli when developing item specifications and when selecting a development platform and delivery vendor.