West Virginia General Summative Assessment

2021–2022

Volume 2, Part 2 (Science) Test Development

West Virginia DEPARTMENT OF EDUCATION

TABLE OF CONTENTS

1. INTRODUCTION 1
1.1 Claim Structure 2
1.2 Underlying Principles Guiding Item Development 2
1.3 Organization of this Volume 3
2. ITEM DEVELOPMENT PROCESS THAT SUPPORTS VALIDITY OF CLAIMS 3
2.1 Overview 3
2.2 Item Specifications 4
2.3 Selection and Training of Item Writers 7
2.4 Internal Review 7
2.4.1 Preliminary Review 8
2.4.2 Scoring Entry and Review 9
2.4.3 Content Review One 9
2.4.4 Edit Review 9
2.4.5 Senior Review 10
2.5 Review by State Personnel and Stakeholder Committees 10
2.5.1 State Review 10
2.5.2 Content Advisory Committee Reviews 11
2.5.3 Language Accessibility, Bias, and Sensitivity Committee Reviews 12
2.5.4 Markup for Translation and Accessibility Features 13
2.6 Field Testing 13
2.7 Post-Field-Test Review 14
2.7.1 Rubric Validation 14
2.7.2 Data Review 15
3. SCIENCE ITEM BANK SUMMARY 19
3.1 Current Composition of the Science Item Bank 20
3.2 Strategy for Pool Evaluation and Replenishment 25
4. WVGSA SCIENCE TEST CONSTRUCTION 25
4.1 Test Design 25
4.2 Test Blueprints 26
4.3 Test Construction 33
4.4 Paper-Pencil/Braille Accommodation Form Construction 37
5. SIMULATION SUMMARY REPORT 37
5.1 Factors Affecting Simulation Results 38
5.2 Results of Simulated Test Administrations: English 38
5.2.1 Summary of Blueprint Match 38
5.2.2 Item Exposure 38
5.3 Results of Simulated Test Administrations: Spanish 39
5.3.1 Summary of Blueprint Match 39
5.3.2 Item Exposure 39

6. OPERATIONAL TEST ADMINISTRATION SUMMARY REPORT 40
6.1 Blueprint Match 40
6.2 Item Exposure 40
REFERENCES 41

LIST OF TABLES

Table 1: Summary of How Each Step of Development Supports the Validity of Claims 4
Table 2: Sample Science Item Cluster Specifications for Middle School Life Sciences Performance Expectation 6
Table 3: Summary of Content Advisory Committee Meetings 11
Table 4: Summary of Fairness Committee Meetings 12
Table 5: Summary of Data Review Committee Meetings 16
Table 6. Science Interaction Types and Descriptions 20
Table 7. Spring 2022 Shared Science Assessment Operational and Field-Test Item Bank 22
Table 8. Spring 2022 Shared Science Assessment Operational Item Bank 22
Table 9. Spring 2022 Shared Science Assessment Field-Test Item Bank 23
Table 10. Spring 2022 Shared Science Assessment Operational and Field-Test Item Bank by Science Discipline 23
Table 11. Spring 2022 Shared Science Assessment Operational and Field-Test Item Bank by Disciplinary Core Idea (DCI) 24
Table 12. Science Test Blueprint, Grade 5 Science 27
Table 13. Science Test Blueprint, Grade 8 Science 29
Table 14. WVGSA Science 85th Percentile Testing Times by Grade 33
Table 15. Spring 2022 WVGSA Science Operational and Field-Test Item Pool 33
Table 16. Spring 2022 WVGSA Science Operational Item Pool 34
Table 17. Spring 2022 WVGSA Science Field-Test Item Pool 34
Table 18. Spring 2022 WVGSA Science Item Pool by Science Discipline 35
Table 19. Spring 2022 WVGSA Science Item Pool by Disciplinary Core Idea 36
Table 20. Item Exposure Rates by Grade: Percentage of Items by Exposure Rate, Across All English Online Simulation Sessions 39
Table 21. Spring 2022 Spanish Operational Item Pool 39
Table 22. Item Exposure Rates by Grade: Percentage of Items by Exposure Rate, Across All Spanish Simulation Sessions 40
Table 23. Item Exposure Rates by Grade: Percentage of Items by Exposure Rate, Across All Spring 2022 Test Administrations 40

LIST OF APPENDICES

Appendix A: Item Writer Training Materials
Appendix B: Item Review Checklist
Appendix C: Content Advisory Committee Participant Details
Appendix D: Fairness Committee Participant Details
Appendix E: Sample Data Review Training Materials
Appendix F: Data Review Committee Participant Details
Appendix G: Example Item Interactions
Appendix H: Shared Science Assessment Item Bank
Appendix I: Adaptive Algorithm Design
Appendix J: WVGSA Science Assessment Item Pool

1. INTRODUCTION

The West Virginia General Summative Assessment (WVGSA) for science was first administered to students during spring 2018, replacing the WESTEST2 in science. The WVGSA for science was delivered to students in grades 5 and 8 as an online assessment, using an adaptive test design which makes use of several technology-enhanced item types. The content measures the three-dimensional science standards based on the National Research Council's *A Framework for K–12 Science Education* published in 2012.

Additional details on the implementation of the assessments are available in Volume 1, Annual Technical Report.

The interpretation, usage, and validity of test scores rely heavily on the test development process itself. This volume of the technical report describes how the WVGSA science assessment was developed and how that process contributes to the validity of the test scores. Specifically, this volume provides evidence to support the following:

- The Test Item Specifications provided detailed guidance for item writers and reviewers to ensure that the science items were aligned to the performance expectations (PEs) they were intended to measure.
- The item development procedures employed for the WVGSA science assessments was consistent with industry standards.
- The development and maintenance of the item bank, where test items cover the range of measured performance expectations, grade-level difficulties, and levels of cognitive engagement through the use of both item clusters and stand-alone items.
- The Test Design Summary/Blueprint stipulated the range of operational items from each item type and content category required for each test administration. This document was implemented using the item selection algorithm for science.

Note that for the science assessments, as outlined in Volume 1 of this technical report, Cambium Assessment, Inc. (CAI) collaborates with a group of states that share common item development processes. In addition to developing items for each of those states, CAI develops and maintains the Independent College and Career Readiness (ICCR) item bank, which consists of items developed according to the same principles followed when the items owned by each of the collaborator states were created. This volume focuses on general test development activities.

For the WVGSA science test, items are drawn from an item bank that consists of ICCR items, items owned by West Virginia, and items owned by several other states that share a Memorandum of Understanding (MOU) to share content, leadership, and new ideas and methods. Specifically, all items developed under the MOU underwent the same item development process. For the remainder of this volume, the term *item bank* will refer to all items developed under the MOU unless explicitly stated otherwise.

1.1 Claim Structure

The goals, uses, and claims that the Shared Science Assessment Item Bank and subsequent tests would be designed to support were identified in a series of collaborative meetings held August 22–23, 2016. The overarching goal of those meetings was to support the development of statewide summative assessments using science content that measures the three-dimensional science standards based on *A Framework for K–12 Science Education* (National Research Council, 2012).

To that end, CAI invited content and assessment leaders from 10 states as well as four nationally recognized experts who helped author the Next Generation Science Standards (NGSS). Two nationally recognized psychometricians also participated.

CAI staff and the participating states collaborated to develop items and test specifications designed to measure the three-dimensional science standards. The item specifications were generally accompanied by sample item clusters that met those specifications. All specifications and sample item clusters were reviewed by state content experts and committees of educators in at least one of the states.

1.2 Underlying Principles Guiding Item Development

The science item bank was established using a highly structured, evidence-centered design. The process began with detailed item specifications. The specifications, discussed in a later section, described the interaction types that could be used, listed guidelines for targeting the appropriate cognitive engagement, offered suggestions for controlling item difficulty, and provided sample items.

Items were written with the goal that virtually every item would be accessible to all students, either by itself or in conjunction with accessibility tools, such as text-to-speech (TTS), translations, or assistive technologies. This goal was supported by delivering the items via CAI's Test Delivery System (TDS), which has received Web Content Accessibility Guidelines (WCAG) 2.0 AA certification, offers a wide array of accessibility tools, and is compatible with most assistive technologies.

Item development supported the goal of creating high-quality item clusters and stand-alone items through rigorous development processes that were managed and tracked by a content development platform. This platform ensured that every item flowed through the correct sequence of reviews, and it captured every comment and change applied to each item.

CAI sought to ensure that the items measured the PEs in a fair and meaningful way by engaging educators and other stakeholders at each step of the process. Educators evaluated the alignment of items to the PEs and offered guidance and suggestions for improvement. They also reviewed items for fairness and sensitivity. Following item field testing, educators engaged in rubric validation, a process that refines rule-based rubrics upon review of student responses.

Both these principles and the processes that support them were incorporated into an item bank that measures the PEs with fidelity and does so in a manner that minimizes construct-irrelevant variance and barriers to access. The details of these processes are described later in this volume.

1.3 Organization of this Volume

This volume of the technical report is organized into three sections that cover the following topics:

1. An overview of the science item development process that supports the validity of the claims that science assessments are designed to support
2. An overview of the science item bank, the types of assessments the bank is designed to support, and methods for refreshing the bank
3. A description of test construction process for the WVGSA for science, including the blueprint, test design, an evaluation of simulated test sessions, the operational blueprint match results, and item exposure rates

2. Item Development Process That Supports Validity of Claims

2.1 Overview

CAI developed the Shared Science Assessment Item Bank in collaboration with the states that were part of the Memorandum of Understanding (MOU) using a rigorous, structured process that engaged stakeholders at critical junctures. This process was managed by CAI's Item Tracking System (ITS), which is an auditable content-development tool that enforces rigorous workflow and captures each item change comment. Reviewers, including internal CAI reviewers or stakeholders in committee meetings, can review items in ITS as they will appear to the student, along with all accessibility features and tools.

The process begins with the definition of item specifications, and continues with

- selection and training of item writers;
- writing and internal review of items;
- review by state personnel and stakeholder committees;
- markup for translation and accessibility features;
- field testing; and
- post-field-test reviews.

Each step has a role in ensuring that the items support the claims on which they are based. Table 1 describes how each step contributes to this goal and what each step entails.

Table 1: Summary of How Each Step of Development Supports the Validity of Claims

Developmental Steps	Supports Alignment to the Performance Expectations	Reduces Construct-Irrelevant Variance Through Universal Design	Expands Access Through Linguistic and Other Supports
Item specifications	Specifies item interactions, content limits, and guidelines for meeting task demands and levels of cognitive engagement requirements and adjusting difficulty.	Avoids the use of any item interactions with accessibility constraints and provides language guidelines. Allows for multiple response modes to accommodate different styles.	
Selection and training of item writers	Ensures that item writers have the background to understand the PEs and item specifications. Teaches item writers how to select item interactions for measurement and accessibility.	Training in language accessibility, bias, and sensitivity helps item writers avoid unnecessary barriers.	
Writing and internal review of items	Checks content alignment and evaluates and improves overall quality.	Eliminates editorial issues and flags and removes bias and accessibility issues.	
Markup for translation and accessibility features		Adds universal features, such as text-to-speech (TTS) for science, that reduce barriers.	Adds TTS, braille, American Sign Language (ASL), translations and glossaries.
Review by state personnel and stakeholder committees	Checks content and cognitive complexity alignment; evaluates and improves overall quality.	Flags sensitivity issues.	
Field testing	Provides statistical checks on quality and flags issues.	Flags items that appear to function differently for subsequent review to identify issues.	May reveal usability or implementation issues with markup.
Post-field-test reviews	Provides final, more focused checks on flagged items. Rubric validation ensures that scoring reflects PEs.	Provides final, focused review on items flagged for differential item function (DIF).	

2.2 ITEM SPECIFICATIONS

CAI is working with a group of states, psychometricians, and science experts, including the authors of the Next Generation Science Standards (NGSS), to develop powerful innovative solutions to the challenges of measuring three-dimensional science standards based on the National Research Council's *A Framework for K–12 Science Education* published in 2012. Participating states include Connecticut, Hawaii, Idaho, Montana, Oregon, Rhode Island, Utah, Vermont, West Virginia, and Wyoming. New Hampshire, North Dakota, South Dakota, and the US Virgin Islands participate in some activities. This collaboration has yielded item specifications for PEs, sample item clusters for some specification, and hundreds of science item clusters and stand-alone items in various stages of development. Under this collaboration, utilizing guidelines for item specifications proposed by WestEd in collaboration with the Council of Chief State School Officers (CCSSO), state members, and content experts (CCSSO, 2015), states developed item specifications jointly.

Item specifications are documents designed to guide item writers as they craft test questions and stakeholders as they review those items. These specifications are intended to serve as a roadmap for writers to facilitate the creation of items that are properly aligned to the three dimensions comprising each science standard and that together form coherent item clusters and stand-alone items.

The item specifications for science include the following elements:

- **Performance Expectation**. This identifies the PE being assessed.
- **Dimensions**. This identifies the Science and Engineering Practices (SEPs), Crosscutting Concepts (CCCs), and Disciplinary Core Ideas (DCIs) that the PE assesses.
- **Clarifications and Content Limits**. This delineates the specific content that the PE measures and the parameters in which items must be developed to assess the PE accurately, including the lower and upper complexity limits of items. Specifically, content limits refine the intent of the PEs and provide limits of what may be asked of test takers. For example, content limits may identify the specific formulae that students are expected to know or not know.
- **Science Vocabulary**. This section identifies the relevant technical words that students are expected to know, and related words that they are explicitly not expected to know. These categories should not be considered exhaustive, as the boundaries of relevance are ambiguous, and the list is limited by the writers' imagination.
- **Content/Phenomena**. This section provides examples of the types of phenomena that would support the effective items related to the PE in question. In general, these are guideposts, and item writers seek comparable phenomena rather than drawing on those within the documents.
- **Task Demands**. In this section, the PEs and associated evidence statements are broken down into specific task demands aligned to each PE. Task demands denote the specific ways in which students will provide evidence of their understanding of the concept or skill. Specifically, the task demands identify the types of interactions and activities that item writers should employ. Each item should be clearly linked to one or more of the task demands, and the verbs guide the types of interactions writers might employ to elicit the student response.

Table 2 provides a sample of the item specifications developed by content experts for a middle school Life Sciences PE.

Table 2: Sample Science Item Cluster Specifications for Middle School Life Sciences Performance Expectation

<table>
<tr><td>Performance Expectation</td><td colspan="3">MS-LS1-1[a]
Conduct an investigation to provide evidence that living things are made of cells, either one cell or many different numbers and types of cells.</td></tr>
<tr><td>Dimensions</td><td>Planning and Carrying Out Investigations
Conduct an investigation to produce data to serve as the basis for evidence that meets the goals of an investigation.</td><td>LS1.A: Structure and Function
All living things are made up of cells, which is the smallest unit that can be said to be alive. An organism may consist of one single cell (unicellular) or many different numbers and types of cells (multicellular).</td><td>Scale, Proportion, and Quantity
Phenomena that can be observed at one scale may not be observable at another scale.</td></tr>
<tr><td>Clarifications and Content Limits</td><td colspan="3">Clarification Statements
Emphasis is placed on developing evidence that living things are made of cells, distinguishing between living and non-living things, and understanding that living things may be made of one cell or many varying cells.
Content Limits
<u>Students do not need to know the following</u>:
The structures or functions of specific organelles or different proteins
Systems of specialized cells
The mechanisms by which cells are alive
Specifics of DNA and proteins or of cell growth and division
Endosymbiotic theory
Histological procedures</td></tr>
<tr><td>Science Vocabulary Students Are Expected to Know</td><td colspan="3">Multicellular, unicellular, cells, tissues, organ, system, organism hierarchy, bacteria, colony, yeast, prokaryote, eukaryote, magnify, microscope, DNA, nucleus, cell wall, cell membrane, algae, chloroplast(s), chromosomes, cork.</td></tr>
<tr><td>Science Vocabulary Students Are Not Expected to Know</td><td colspan="3">Differentiation, mitosis, meiosis, genetics, cellular respiration, energy transfer, RNA, protozoa, amoeba, histology, protists, archaea, nucleoid, plasmid, diatoms, cyanobacteria.</td></tr>
<tr><td colspan="4">Phenomena</td></tr>
<tr><td>Context/ Phenomena</td><td colspan="3">Some example phenomena for MS-LS1-1:
Plant leaves and roots have tiny box-like structures that can be seen under a microscope.
Small creatures can be seen swimming in samples of pond water viewed through a microscope.
Different parts of a frog’s body (e.g., muscles, skin, tongue) are observed under a microscope, and are seen to be composed of cells.
One-celled organisms (e.g., bacteria, protists) perform the eight necessary functions of life, but nothing smaller has been seen to do this.
Swabs from the human cheek are observed under a microscope. Small cells can be seen.</td></tr>
</table>

This Performance Expectation and associated Evidence Statements support the following Task Demands.
Task Demands
1. Identify from a list the materials/tools, including distractors, needed for an investigation to find the smallest unit of life (cell).
2. Identify the outcome data that should be collected in an investigation of the smallest unit of living things.
3. Evaluate the sufficiency and limitations of data collected to explain that the smallest unit of living things is the cell.
4. Make and/or record observations about whether the sample contains cells.[b]
5. Interpret and/or communicate data from the investigation to determine if a specimen is alive.
6. Construct a statement to describe the overall trend suggested by the observed data.

[a]MS-LS1-1 is the PE code for Middle School Life Sciences 1-1.

[b]Denotes task demands deemed appropriate for use in stand-alone item development.

The specifications help test developers create item clusters and stand-alone items that will support a range of difficulties, furthering the goal of measuring the full range of performance found in the population but remaining at grade level.

2.3 Selection and Training of Item Writers

All item writers developing science items at CAI have at least a bachelor's degree, and many bring teaching experience. All item writers are trained in

- the principles of universal design;
- the appropriate use of item interactions; and
- the science item specifications.

Key materials are shown in Appendix A, Item Writer Training Materials. These include

- CAI's Language Accessibility, Bias, and Sensitivity Guidelines; and
- a training (presented using Microsoft PowerPoint) for the appropriate use of item interactions.

2.4 Internal Review

CAI's test development structure uses highly effective units organized around each content area. Unit directors oversee team leaders who work with team members to ensure item quality and adherence to best practices. All team members, including item writers, are content area experts. Teams include senior content specialists who review the items before the client review and provide training and feedback for all content-area team members.

ICCR and MOU science items undergo a rigorous, multiple-level internal review process before they are sent to external review. Staff members are trained to review items for both content and accessibility throughout the process. A sample of the item review checklist that CAI test developers use is included in Appendix B, Item Review Checklist. The ICCR and MOU science internal review cycle includes the following phases:

- Preliminary Review

- Scoring Entry and Review
- Content Review One
- Edit Review
- Senior Review

2.4.1 Preliminary Review

Team leads or senior content staff conduct Preliminary Review. Sometimes, Preliminary Review is conducted in a group setting, led by a senior test developer. During the Preliminary Review process, team leads or senior content staff analyze items to ensure the following requirements have been met:

- The item aligns with the PE.
- The item matches the item specification for the skills being assessed.
- The item is based on a quality scientific phenomenon (i.e., it assesses something in a reasonable way, and it is a discrete observation that grounds a scenario that allows for the assessment of something worthwhile in a meaningful way).
- The item aligns appropriately with the task demands.
- The vocabulary used in the item is appropriate for the grade and subject matter.
- The item considers language accessibility, bias, and sensitivity.
- The content is accurate and straightforward.
- The graphic and stimulus materials are necessary to answer the question.
- The item follows the approved style guide.
- The stimulus is clear, concise, and succinct (i.e., it contains enough information to convey what is being asked, it is stated positively, and it does not rely on negatives—such as *no, not, none, or never*—unless necessary).

For selected-response item interactions, test developers also check to ensure that the set of response options are

- as succinct and short as possible (without repeating text);
- parallel in structure, grammar, length, and content;
- sufficiently distinct from one another;
- all plausible (but with only one correct option); and
- free of obvious or subtle cueing.

2.4.2 Scoring Entry and Review

During Scoring Entry, the item writer inputs the machine scoring to the team lead or senior staff for review before the Content Review One level. This step is separate from Preliminary Review and allows senior staff to suggest changes to the interaction at Preliminary Review without requiring the writer to overhaul the scoring they have already created. This step also allows senior staff to ensure that the scoring suggested by the writer at the Preliminary Review is appropriate. This ensures that the scoring is entered once, streamlining the process. At this level, the scoring is analyzed to ensure that the following criteria are met:

- The scoring works as intended (i.e., the student receives a point for ALL correct responses and no points for ALL incorrect responses).
- The student receives a point for every unique piece of information they reveal about their understanding through their responses.
- Dependent scoring between and within interactions is captured.
- The way in which the scoring is set up is unambiguous and matches the questions asked (i.e., if students are asked to round to a particular decimal place, they are scored accordingly).

The senior staff approves the intent of the scoring from the Preliminary Review. At the Scoring Entry level, the writer inputs this approved scoring, after which the senior staff checks the functionality of the scoring. Once the scoring is determined to be working correctly, the senior staff signs off on the item and moves it to Content Review One.

2.4.3 Content Review One

Content Review One is conducted by a senior content specialist who was not part of the Preliminary Review. This reviewer carefully examines each item based on all the same criteria identified for Preliminary Review. They also ensure that the revisions made during the Preliminary Review did not introduce errors or content inaccuracies. This reviewer approaches the item from the perspective of potential clients and their own experience in test development.

2.4.4 Edit Review

During Edit Review, editors have four primary tasks:

1. Editors perform basic line editing for correct spelling, punctuation, grammar, and mathematical and scientific notation, ensuring consistency of style across the items.
2. Editors ensure that all items are accurate in content. Editors compare reading passages against the original publications to ensure that all information is internally consistent across stimulus materials and items, including names, facts, or cited lines of text that appear in the item. They ensure that the answer key(s) and all information in the item are correct. For items with mathematical tasks, editors perform all calculations to ensure accuracy.

3. Editors review all material for fairness and language accessibility issues.
4. Editors confirm that the items reflect the accepted guidelines for good item construction. They examine all items for language that is simple, direct, and free of ambiguity with minimal verbal difficulty. Editors confirm that a problem or task and its stem are clearly defined and concisely worded with no unnecessary information. For multiple-choice interactions, editors check that options are parallel in structure and fit logically and grammatically with the stem. They also ensure that the key answers the question posed accurately and correctly, is not inappropriately obvious, and is the only correct answer to an item among the distractors. For constructed-response interactions, editors review the rubrics for appropriate style and grammar.

2.4.5 Senior Review

By the time a science item arrives at Senior Review, both content reviewers and editors have thoroughly vetted it. Senior reviewers (in particular, senior content specialists) look at the item's entire review history, ensuring that all the issues identified in that item have been adequately addressed. Senior reviewers verify the overall content of each item, confirming its accuracy, alignment with the PE, and consistency with expectations for the highest quality. They check whether the scoring is working as intended and ensure that the scoring assertions adequately address the evidence the student provides with each type of response.

2.5 Review by State Personnel and Stakeholder Committees

All science items undergo an exhaustive external review process. Items in the Shared Science Assessment Item Bank were reviewed by content experts in one or several states and reviewed and approved by multiple stakeholder committees that evaluated them for both content and bias/sensitivity.

2.5.1 State Review

After items have been developed for a state participating in the MOU, content experts from the state that owns the item review any eligible items prior to committee review. At this stage in the review process, clients can request edits, such as wording edits, scoring edits, alignment changes, or task demand updates. A CAI science content expert reviews all client-requested edits considering the science item specifications, other clients' requests, and existing items in the bank to determine whether the requested edits will be made. At this stage, clients can either present these items to the committee (based on the edits made) or withhold them from committee review.

ICCR items are reviewed by at least one or two states. The state(s) then provide feedback on the ICCR items, and CAI's science leadership gathers suggestions and makes edits that improve the ICCR items. Not all suggestions are implemented, as these items are owned by CAI. Further, most MOU states accept or reject ICCR and MOU items (as they appear at the time) to be presented to their committees. Some clients skip this step and allow CAI to review all items with their committees before reviewing them. These items can either be set for field testing in a future administration or become a part of the locked operational pool.

2.5.2 Content Advisory Committee Reviews

During the Content Advisory Committee (CAC) Reviews, items are reviewed for content validity, grade-level appropriateness, and alignment with the PE. CAC members are typically grade-level and subject-matter experts. During this review, educators also ensure that the scoring assertions clearly identify what is being scored as correct and give credit where they should (refer to Section 2.7.1, Rubric Validation).

Items developed for each state under the MOU are reviewed by the state that owns the items. ICCR items are reviewed by the CAC of one or more states. In most cases, items are seen by multiple state committees before their field test or operational use.

In 2022, MOU states were all involved in a single CAC process where participants from multiple states reviewed items. The items were edited and returned to their respective owning states for final approval.

A summary of the 2021-2022 committee meetings is presented in Table 3, with additional details about the participants in Appendix C, Content Advisory Committee Participant Details. Appendix C also contains detailed information about the participants of Content Advisory Committee meetings of previous years.

Table 3: Summary of Content Advisory Committee Meetings

State/Item Bank	Meeting	Number of Committee Members	Number of Items Reviewed
Connecticut	July 2021	26	26[c]
	September 2021	27	25
ICCR	July 2021	[a]	141[c]
Idaho	July 2021	12	0[b, c]
	November 2021	11	317
Montana	July 2021	1	36[c]
	October 2021	6	41
Multi-State Science Assessment (Rhode Island and Vermont)	July 2021	7	32[c]
	August 2021	11	93
Oregon	August 2021	14	375
Utah	July 2021	0	55[c]
	August 2021	14	62
West Virginia	July 2021	10	16[c]
Wyoming	June/July 2021	14	39
	July 2021	14	39[c]

Note. [a]Number of Content Advisory Committee Members is not available at the time of writing this report.
[b]Number of science items reviewed by Content Advisory Committees is unavailable at the time of writing this report.

[c]Items were reviewed in a combined multi-state Content Advisory Committee meeting.

2.5.3 Language Accessibility, Bias, and Sensitivity Committee Reviews

During the bias and sensitivity reviews, stakeholders review items to check for issues that might unfairly impact students based on their background. For example, some states include representatives from student populations such as Special Education, low vision, and the hearing impaired. Further, diverse members of this committee represent students of various ethnic and economic backgrounds to ensure that all items are free of bias and sensitivity concerns.

Due to the COVID-19 pandemic during 2020, 2021, and 2022, CAI reviewed items that contained references to virus, vaccine, bacteria, disease, infection, and related words and phrases. CAI content experts reviewed 65 items and rejected one item for sensitivity concerns.

In 2022, MOU states were all involved in a single review process where participants from multiple states reviewed items. The items were edited and returned to their respective owner states for final approval.

A summary of the 2021-2022 committee meetings is presented in Table 4, with additional details about the participants available in Appendix D, Fairness Committee Participant Details. Appendix D also contains detailed information about the participants of Fairness Committee meetings of previous years.

Table 4: Summary of Fairness Committee Meetings

State/Item Bank	Meeting	Number of Committee Members	Number of Items Reviewed	Number of Items Rejected
Connecticut	July 2021	6	20[a]	0
	September 2021	7	111	23
ICCR	July 2021	15	157[a]	1
Idaho	December 2021	21	179	0
Montana	July 2021	3	41[a]	0
Multi-State Science Assessment (Rhode Island and Vermont)	July 2021	3	30[a]	1
	August 2021	3	93	3
Oregon	August 2021	7	353	13
U.S. Virgin Islands	October 2021	6	299	28
Utah	July 2021	11	64[a]	0
	August 2021	6	62	62
West Virginia	July 2021	2	12[a]	1
Wyoming	June/July 2021	6	39	39
	July 2021	4	28[a]	0

Note. [a]Items were reviewed in a combined multi-state Fairness Committee Meeting

2.5.4 Markup for Translation and Accessibility Features

After all approved state- and committee-recommended edits have been applied, the items are considered *locked* and ready for a portion of the accessibility tagging. TTS tagging is applied prior to field testing, while Spanish translation and braille are applied post-field-testing. Accessibility markup is embedded into each item as part of the item development process rather than as a *post-hoc* process applied to completed assessments.

Accessibility markup, whether translations or TTS, follows similar processes. One trained expert enters the markup, and then a second expert reviews the work and recommends changes if necessary. If there is disagreement, a third expert is engaged to resolve the conflict.

Currently, science items are tagged with TTS. Spanish translations, including Spanish TTS and braille, are available for a subset of items.

2.6 FIELD TESTING

A large pool of science field-test items was administered in the following nine states in spring 2018: Connecticut, Hawaii, New Hampshire, Oregon, Rhode Island, Utah, Vermont, West Virginia, and Wyoming. For Hawaii, Oregon, and Wyoming, items were embedded as field-test items in the legacy science test. Connecticut and Rhode Island conducted an independent field test in which all students participated, but no scores were reported. In New Hampshire, Utah, Vermont and West Virginia, an operational field test was administered.

In 2019, a second pool of field-test items was administered in the following nine states: Connecticut, Hawaii, Idaho, New Hampshire, Oregon, Rhode Island, Vermont, West Virginia, and Wyoming. For Hawaii, Idaho (elementary school), and Wyoming, unscored field-test items were added as a separate segment to the operational (scored) legacy science test. An independent field test in which students were administered a full set of items was conducted for a sample of Idaho middle schools. In Connecticut, New Hampshire, Oregon, Rhode Island, Vermont, and West Virginia, field-test items were administered as unscored items embedded within the operational items.

In 2021, a third wave of field-test items was administered in 12 states. An independent field test, in which students were administered a full set of items, was conducted in Idaho and Montana. Unscored field-test items were added as a separate segment to the operational (scored) legacy science test in Wyoming. In the remaining nine states (Connecticut, Hawaii, New Hampshire, North Dakota, Rhode Island, South Dakota, Utah, Vermont, and West Virginia), field-test items were administered as unscored items embedded within the operational items.

In 2022, a fourth wave of field-test items was administered in 13 states and one US territory. In all 13 states and US territories (Connecticut, Hawaii, Idaho, Montana, New Hampshire, North Dakota, Oregon, Rhode Island, South Dakota, Utah, US Virgin Islands, Vermont, West Virginia, and Wyoming), field-test items were administered as unscored items embedded within the operational items. CAI's field-test process is detailed in Section 3.2.1, Field Testing, in Volume 1, Annual Technical Report.

2.7 POST-FIELD-TEST REVIEW

Following the field test, items were subjected to a substantial validation process, including rubric validation and data review. That validation process is described in Section 2.7.1, Rubric Validation, and Section 2.7.2, Data Review.

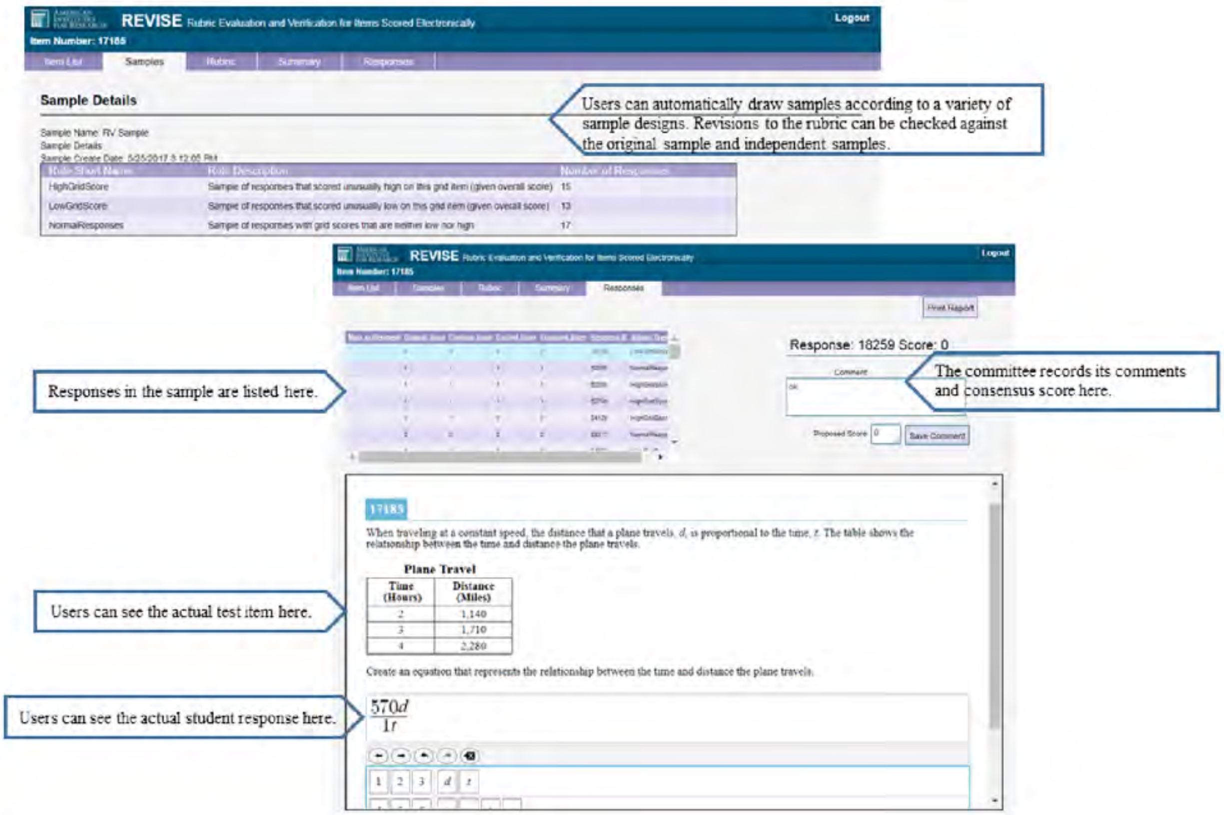
2.7.1 Rubric Validation

The validation process for the field-test items begins with rubric validation to verify and make any necessary revisions to the scoring rubrics. The rubric validation process occurs in two phases. During the first phase, CAI content experts work with the analysis team to prepare for the rubric validation meetings. The CAI content experts use the Rubric Evaluation and Verification for Items Scored Electronically (REVISE) system to generate student responses that are scientifically sampled to overrepresent responses most likely to have been mis-scored. Specifically, the sample overrepresents (1) low-scored responses from otherwise high-scoring students and (2) high-scored responses from otherwise low-scoring students. This process allows CAI to identify any potential scoring concerns before the rubric validation meeting, such as unanticipated (but accurate) responses, equivalent responses that were not originally considered, and responses receiving credit but should not (based on the content and the item rubric). At this point, the rubrics may be adjusted, and responses rescored.

The second phase of rubric validation involves committees of educators in each state. The committees review the response samples generated by CAI to make recommendations to change or confirm each item's rubrics. The committee recommendations are then discussed with the state of ownership to resolve any inconsistencies. The rubric is then edited or confirmed based on this resolution.

Figure 1 illustrates the features provided by the REVISE system.

Figure 1: Features of the REVISE Software



After the rubric validation meetings, CAI staff apply the approved revisions to the rubrics, and any items rejected as part of the process are rejected in ITS. ITS archives critical information regarding the scoring certification completed during the rubric validation process. This includes any rubric changes made during the scoring decision meetings and the sign-off completed by the senior content expert once the rubric has been changed, rescoring the entire sample, and the verification that the final rubric functioned as intended.

Following rubric validation, all items are subject to statistical checks, and flagged items are presented in data review committees.

2.7.2 Data Review

Following rubric validation, all items are rescored, and classical item statistics are computed for the scoring assertions, including item difficulty and item discrimination statistics, testing time, and differential item functioning (DIF) statistics. The states established standards for the statistics. Any items violating these standards are flagged for a second educator review. Even though the scoring assertions are the basic units of analysis used to compute classical item statistics, the business rules to flag items for additional educator review were established at the item level because assertions cannot be reviewed in isolation. A common set of business rules was defined for all the states participating in the field test. The classical item statistics were computed on the data of the students testing in the state that owned the item. For Rhode Island and Vermont, which share their item development, statistics were computed on the combined data of students testing in both states. For

ICCR items, the data from students testing in Connecticut, Idaho, New Hampshire, North Dakota, Oregon, Rhode Island, South Dakota, Utah, Vermont, and West Virginia were combined.

Section 4 of Volume 1, Field-Test Classical Analysis, describes the statistical flags that designate items for data review. The flags are designed to highlight potential content weaknesses, miskeys, or possible bias issues. Committee members are taught to interpret these flags and are given guidelines for examining the items for content or fairness issues.

For each of the states participating in the MOU, flagged items owned by the state were reviewed by a data review committee. The composition of the data review committees generally consisted of content experts from the state's department of education or state educators (in this case, the state educators were science teachers) and were supported by CAI content experts. ICCR field-test items were taken to committee members from several states participating in the MOU. The outcomes were decided by CAI's science content leadership, taking the committees' recommendations into consideration.

At the start of each state-owned item data review meeting, CAI staff leads participants in a training session to familiarize them with the item development process, the purpose of the data review committee and the data review process, and the meaning of the various flags. Committee members are taught to interpret the various flags and are given guidelines for examining the items for content or fairness issues. The training includes a group review of item cards, which detail specific item attributes (including grade level and alignment to the science PEs, the content and rubric of the item, and the various item statistics). A sample of the training materials used for these data review meetings is presented in Appendix E, Sample Data Review Training Materials. Participants use an online environment via laptop computers to review the items, interact with them in a manner similar to that of students, and view the statistics associated with each item.

The items are then reviewed by the participants who are most familiar with the particular grade (band) level and the items' content domain. CAI's content specialists, who are also well versed in item statistics, facilitate the discussion in each room with CAI's psychometricians available to answer questions. At the end of each meeting day, CAI's content specialists meet with the state's content specialists to review the committee recommendations and decide whether to accept or reject the item for inclusion in the operational item pool. Items that were rejected become eligible for potential changes and additional field-test items.

Table 5 summarizes the data review committee meetings. Details, including the composition of each committee, are presented in in Appendix F, Data Review Committee Participant Details.

Table 5: Summary of Data Review Committee Meetings

Owner	Meeting	Number of Committee Members	Item Type	Number of Items Reviewed	Number of Items Rejected
Connecticut	August 2018	29	**Total**	**18**	**11**
			Cluster	7	5
			Stand-Alone	11	6

Owner	Meeting	Number of Committee Members	Item Type	Number of Items Reviewed	Number of Items Rejected
	August 2019	29	**Total**	**53**	**17**
			Cluster	14	6
			Stand-Alone	39	11
	August 2021	19	**Total**	**51**	**12**
			Cluster	8	2
			Stand-Alone	43	10
	August 2022	15	**Total**	**19**	**6**
			Cluster	5	4
			Stand-Alone	14	2
Hawaii	August 2018	18	**Total**	**32**	**3**
			Cluster	7	1
			Stand-Alone	25	2
	August 2019	18	**Total**	**37**	**13**
			Cluster	17	5
			Stand-Alone	20	8
	August 2021	25[d]	**Total**	**26**	**8**
			Cluster	6	0
			Stand-Alone	20	8
	August 2022	12[d]	**Total**	**49**	**8**
			Cluster	11	2
			Stand-Alone	38	6
ICCR	July 2018	18	**Total**	**84**	**8**
			Cluster	33	2
			Stand-Alone	51	6
	August 2019	N/A[c]	**Total**	**43**	**3**
			Cluster	0	1
			Stand-Alone	43	2
	August 2021	25[d]	**Total**	**75**	**6**
			Cluster	11	2
			Stand-Alone	64	4
	August 2022	20[d]	**Total**	**68**	**14**
			Cluster	12	1
			Stand-Alone	56	13
Idaho	August 2019	10	**Total**	**12**	**6**
			Cluster	4	3
			Stand-Alone	8	3
	August 2021	25[d]	**Total**	**60**	**5**
			Cluster	26	1
			Stand-Alone	34	4
	August 2022	8[d]	**Total**	**4**	**0**
			Cluster	3	0

Owner	Meeting	Number of Committee Members	Item Type	Number of Items Reviewed	Number of Items Rejected
			Stand-Alone	1	0
Montana	September 2021	4	**Total**	**17**	**4**
			Cluster	3	2
			Stand-Alone	14	2
	September 2022	5	**Total**	**17**	**3**
			Cluster	5	2
			Stand-Alone	12	1
Multi-State Science Assessment (Rhode Island and Vermont)	August 2018	N/A[a]	**Total**	**9**	**6**
			Cluster	2	0
			Stand-Alone	7	6
	August 2019	N/A[a]	**Total**	**14**	**4**
			Cluster	2	1
			Stand-Alone	12	3
	August 2021	N/A[a]	**Total**	**18**	**9**
			Cluster	4	4
			Stand-Alone	14	5
	September 2022	N/A[a]	**Total**	**11**	**7**
			Cluster	1	1
			Stand-Alone	10	6
Oregon	September 2018	11	**Total**	**44**	**6**
			Cluster	28	5
			Stand-Alone	16	1
	August 2019	4	**Total**	**8**	**7**
			Cluster	1	1
			Stand-Alone	7	6
	August 2022	8[d]	**Total**	**31**	**8**
			Cluster	11	2
			Stand-Alone	20	6
South Dakota	September 2021	N/A[b]	**Total**	**15**	**0**
			Cluster	0	0
			Stand-Alone	15	0
Utah	August 2018	16	**Total**	**40**	**6**
			Cluster	40	6
			Stand-Alone	0	0
	September 2021	6	**Total**	**11**	**3**
			Cluster	11	3
			Stand-Alone	0	0
	September 2022	13	**Total**	**11**	**6**
			Cluster	11	6
			Stand-Alone	0	0
West Virginia	July 2018	4	**Total**	**3**	**1**

Owner	Meeting	Number of Committee Members	Item Type	Number of Items Reviewed	Number of Items Rejected
			Cluster	3	1
			Stand-Alone	0	0
	September 2019	4	**Total**	**7**	**6**
			Cluster	1	1
			Stand-Alone	6	5
	August 2021	25[d]	**Total**	**7**	**3**
			Cluster	1	1
			Stand-Alone	6	2
	August 2022	9[d]	**Total**	**10**	**4**
			Cluster	4	2
			Stand-Alone	6	2
Wyoming	October 2018	19	**Total**	**16**	**6**
			Cluster	6	1
			Stand-Alone	10	5
	August 2019	10	**Total**	**16**	**5**
			Cluster	4	3
			Stand-Alone	12	2
	August 2021	25[d]	**Total**	**16**	**4**
			Cluster	3	1
			Stand-Alone	13	3
	August 2022	12[d]	**Total**	**19**	**3**
			Cluster	2	0
			Stand-Alone	17	3

Note. [a]Conducted by the Rhode Island Department of Education and the Vermont Agency of Education science content experts.
[b]Reviewed by South Dakota Department of Education.
[c]In summer 2019, ICCR field-test items were taken to Connecticut, Hawaii, and Idaho for committee review.
[d]Combined Data Review for multiple states (184 Hawaii, Idaho, West Virginia, Wyoming, and ICCR items in 2021 and 181 Hawaii, Idaho, Oregon, West Virginia, Wyoming, and ICCR items in 2022). There were 25 total participants in 2021 and 38 total participants in 2022. Items are broken out by owning state.

3. Science Item Bank Summary

Tests based on *A Framework for K–12 Science Education* (National Research Council, 2012) adopt a three-dimensional conceptualization of science understanding, including Science and Engineering Practices (SEPs), Crosscutting Concepts (CCCs), and Disciplinary Core Ideas (DCIs). Accordingly, the new science assessments are composed mostly of item clusters representing a series of interrelated student interactions directed towards describing, explaining, and predicting scientific phenomena. Some stand-alone items are added to increase the coverage of the test without increasing the testing time or testing burden.

CAI has built the science item bank in partnership with multiple states. The science item bank is robust and has been constructed to support multiple statewide science assessments. As described earlier, the science items are written to the three-dimensional science standards. The item bank comprises ICCR items and items developed for specific states, which are all shared with MOU partner states. These items follow the same specifications, test development processes, and review processes. In 2018, CAI field tested more than 540 item clusters and stand-alone items, of which 451 (including items from all sources) were accepted and made available as operational items in 2019. In 2019, 347 item clusters and stand-alone items were field tested, of which 268 were accepted and made available as operational items in 2020. In 2021, CAI field tested more than 545 item clusters and stand-alone items, of which 458 have passed rubric validation and item data review. In 2022, CAI field tested 471 item clusters and stand-alone items, of which 403 have passed rubric validation and item data review.

Each state or territory using the Shared Science Assessment Item Bank selects items that are appropriately aligned and have passed required reviews (as described in Section 2, Item Development Process That Supports Validity of Claims) for use on its statewide assessment. The Shared Science Assessment Item Bank continues to grow as participating states and territory continue to field test new items. Participating states and territory collectively share the items and agree to field test new items each year.

3.1 CURRENT COMPOSITION OF THE SCIENCE ITEM BANK

The Shared Science Assessment Item Bank contains item clusters and stand-alone items. Item clusters represent a series of interrelated student interactions directed toward describing, explaining, and predicting scientific phenomena. Item clusters can consist of several item parts requiring the student to interact with the item in various ways. In addition, shorter items (stand-alone items) are included to increase the coverage of the assessments without also increasing testing time or testing burden.

Within each item (item cluster and stand-alone item), a series of explicit assertions is made about the knowledge and skills that a student has demonstrated based on specific features of the student's responses across multiple interactions. For example, a student may correctly graph data points indicating that they can construct a graph showing the relationship between two variables, but they may make an incorrect inference about the relationship between the two variables. In this case, the student's performance would not support the assertion that they can interpret relationships expressed graphically. Table 6 lists and describes the science interaction types. Examples of various interaction types are presented in Appendix G, Example Item Interactions.

Table 6. Science Interaction Types and Descriptions

Interaction Type	Associated Subtypes	Description
Choice	Multiple-Choice	Traditional multiple-choice interaction allows students to select a single option from a list of possible answer options.

Interaction Type	Associated Subtypes	Description
	Multiple-Select	Traditional multiple-select interaction (checkboxes) allows students to select one or more options from a list of possible answer choices.
Text Entry	Simple Text Entry	Students type a response in a text box.
	Embedded Text Entry	Students type their responses in one or more text boxes that are embedded in a section of read-only text.
	Natural Language	Students are directed to provide a short, written response.
	Extended-Response	Students are directed to provide a longer, written response in the form of an essay.
Table	Table Match	Interaction allows students to check a box to indicate if the information from a column header matches information from a row header.
	Table Input	Interaction solicits students to complete tabular data.
Edit Task	Edit Task	Students click a word and replace it with another word that they type to revise a sentence.
	Edit Task with Choice	Students click a word or phrase and select a replacement from several options.
	Edit Task Inline Choice	Drop-down menus are placed in the text, and students select an option to complete the text.
Hot-Text	Selectable	Selectable hot-text interactions require students to select one or more text elements in the response area.
	Re-orderable	Re-orderable hot-text interactions require students to click and drag hot text elements into a different order.
	Drag-from-Palette	Drag-from-Palette hot-text interactions require students to drag elements from a palette into the available blank table cells or “gaps” (text boxes) in the response area.
	Custom	Custom hot-text interactions combine the functionality of the other hot-text interaction sub-types. Students responding to a custom hot-text interaction may need to select text elements, rearrange text elements, and/or drag text elements from a palette to blank table cells or drop targets in the response area.
Equation	N/A	Equation interactions require students to enter a response into input boxes. These boxes may stand alone or be in line with text or embedded in a table. The equation interaction may have an on-screen keypad that might consist of special mathematic characters. Students may also enter their responses via a physical keyboard.
Grid	Grid	Grid interactions require students to enter a response by interacting with a grid area in the answer space. Students may be required to draw a line or shape, plot a point, or create a

Interaction Type	Associated Subtypes	Description
		graph. Students may also drag-and-drop or click selectable hot spots.
	Hot-Spot	Hot-spot interaction sub-types facilitate grid interactions with specific hot-spot functionality. These interactions require students to select hot-spot regions in the grid area.
	Graphic Gap Match	Graphic Gap Match interactions facilitate grid interactions with specific drag-and-drop functionality. These interactions require students to drag image objects from a palette to specified regions (gaps) in the grid area.
Simulation	N/A	Simulation interactions allow students to investigate a phenomenon by selecting variables to get output data. Some simulations are accompanied by animations.

Table 7 through Table 11 present the number of items in the Shared Science Assessment Item Bank available for use in the spring 2022 statewide assessments. Appendix H, Shared Science Assessment Item Bank, presents the items available within the bank by grade band, PE, and origin.

Table 7. Spring 2022 Shared Science Assessment Operational and Field-Test Item Bank

Grade Band and Item Type	ICCR Items	West Virginia Items	MOU Items[a]	Total Bank Items
Elementary School	**148**	**20**	**405**	**573**
Cluster	49	10	235	294
Stand-Alone	99	10	170	279
Middle School	**163**	**18**	**414**	**595**
Cluster	55	8	230	293
Stand-Alone	108	10	184	302
Total	**311**	**38**	**819**	**1168**

Note. [a]Other MOU states include Connecticut, Hawaii, Idaho, Montana, MSSA (Rhode Island and Vermont), Oregon, Utah, and Wyoming.

Table 8. Spring 2022 Shared Science Assessment Operational Item Bank

Grade Band and Item Type	ICCR Operational Items	West Virginia Operational Items	MOU Operational Items[a]	Total Bank Operational Items
Elementary School	**116**	**14**	**273**	**403**
Cluster	40	9	157	206
Stand-Alone	76	5	116	197
Middle School	**101**	**10**	**294**	**405**
Cluster	29	4	172	205
Stand-Alone	72	6	122	200

Grade Band and Item Type	ICCR Operational Items	West Virginia Operational Items	MOU Operational Items[a]	Total Bank Operational Items
Total	**217**	**24**	**567**	**808**

Note. [a]Other MOU operational item states include Connecticut, Hawaii, Idaho, Montana, MSSA (Rhode Island and Vermont), Oregon, Utah, and Wyoming.

Table 9. Spring 2022 Shared Science Assessment Field-Test Item Bank

Grade Band and Item Type	ICCR Field-Test Items	West Virginia Field-Test Items	MOU Field-Test Items[a]	Total Bank Field-Test Items
Elementary School	**32**	**6**	**132**	**170**
Cluster	9	1	78	88
Stand-Alone	23	5	54	82
Middle School	**62**	**8**	**120**	**190**
Cluster	26	4	58	88
Stand-Alone	36	4	62	102
Total	**94**	**14**	**252**	**360**

Note. [a]Other MOU field-test item states include Connecticut, Hawaii, Idaho, Montana, MSSA (Rhode Island and Vermont), Oregon, Utah, and Wyoming.

Table 10. Spring 2022 Shared Science Assessment Operational and Field-Test Item Bank by Science Discipline

Grade Band	Science Discipline	Item Type	ICCR Items	West Virginia Items	MOU Items[a]	Total Bank Items[b]
Elementary School	Earth and Space Sciences	Cluster	18	4	77	99
		Stand-Alone	29	3	56	88
	Life Sciences	Cluster	14	4	64	82
		Stand-Alone	32	3	48	83
	Physical Sciences	Cluster	17	2	94	113
		Stand-Alone	38	4	66	108
Middle School	Earth and Space Sciences	Cluster	16	3	59	78
		Stand-Alone	29	1	54	84
	Life Sciences	Cluster	22	3	88	113
		Stand-Alone	47	3	64	114
	Physical Sciences	Cluster	17	2	76	95
		Stand-Alone	32	6	65	103
Total			**311**	**38**	**811**	**1160**

Note. [a]Other MOU states include Connecticut, Hawaii, Idaho, Montana, MSSA (Rhode Island and Vermont), Oregon, Utah, West Virginia, and Wyoming. [b]Count excludes eight MOU items that do not align to the NGSS.

Table 11. Spring 2022 Shared Science Assessment Operational and Field-Test Item Bank by Disciplinary Core Idea (DCI)

Grade Band	Science Discipline	Disciplinary Core Idea	ICCR Items	West Virginia Items	MOU Items[a]	Total Bank Items[b]
Elementary School	Earth and Space Sciences	ESS1	12	1	37	50
		ESS2	15	4	60	79
		ESS3	20	2	36	58
	Life Sciences	LS1	16	6	42	64
		LS2	6	0	22	28
		LS3	5	1	15	21
		LS4	19	0	33	52
	Physical Sciences	PS1	14	1	40	55
		PS2	15	4	34	53
		PS3	20	1	56	77
		PS4	6	0	30	36
Middle School	Earth and Space Sciences	ESS1	15	1	30	46
		ESS2	16	2	40	58
		ESS3	14	1	43	58
	Life Sciences	LS1	22	3	51	76
		LS2	24	1	42	67
		LS3	5	2	17	24
		LS4	18	0	42	60
	Physical Sciences	PS1	13	2	47	62
		PS2	6	2	41	49
		PS3	19	2	34	55
		PS4	11	2	19	32
Total			**311**	**38**	**811**	**1160**

Note. [a]Other MOU states include Connecticut, Hawaii, Idaho, Montana, MSSA (Rhode Island and Vermont), Oregon, Utah, and Wyoming. [b]Count excludes eight MOU items that do not align to the NGSS.

3.2 STRATEGY FOR POOL EVALUATION AND REPLENISHMENT

Both CAI and the participating MOU states continue to develop items to replenish and grow the science item bank. The general strategy for targeting item development involves gathering information from three sources:

1. Characteristics of released items to be replaced
2. Characteristics of items that are overused
3. Tabulations of content coverage and ranges of difficulty to identify gaps in the bank.

Before a test goes live, simulations are used to fine-tune the parameters of the algorithm that govern the item selection in an adaptive design. Among the many reports from the simulator are items seen by more than 20% of students. The characteristics of these items are the primary targets for development. Overused items become candidates for release in two years once replacements have been introduced into the operational pool.

4. WVGSA SCIENCE TEST CONSTRUCTION

4.1 TEST DESIGN

The West Virginia General Summative Assessment (WVGSA) science assessment was administered online to students in grades 5 and 8 using an adaptive test design. In an adaptive test, operational items are selected on the fly based on a student's performance on past items while ensuring the test blueprint is followed for each student. An advantage of adaptive testing is that it can provide more precise scores for students with lower and higher proficiencies. In contrast, fixed forms and linear-on-the fly tests (LOFTs) are typically targeted to provide the best precision for students with medium proficiencies. Also, as opposed to a fixed form and a LOFT, every student has the potential to see a different set of items that adapt to the student's ability, thus offering a better testing experience.

Items are selected by an item-selection algorithm that is based on the content and information values. At any given point during the test, the content value of an item is determined by its contribution to meeting the blueprint, given the content characteristics of the items that have already been administered. During the test, the content value increases for items that exhibit features that have not met their designated minimum as the end of the approaches. Vice versa, the content value decreases for items with content features for which the minimum has been met. The information value of an item is based on the item information function evaluated at the estimated proficiency. The proficiency estimate is updated throughout the test.

The adaptive item-selection algorithm is the same algorithm CAI uses to deliver English language arts (ELA) and mathematics tests but with some modifications to make it suitable for using item clusters. Specifically, the proficiencies estimated during the test are computed under an item response theory (IRT) model that incorporates cluster effects. To avoid over-selecting items with many scoring assertions, the information of an item at an estimated proficiency level is normalized by the number of assertions in the item (similar to how information is computed for item sets in

ELA and mathematics assessments). Additional details about CAI's adaptive testing algorithm are given in Appendix I, Adaptive Algorithm Design.

A non-segmented test design was used for the WVGSA. Students received items from different disciplines in random order. Compared to a segmented design, in which items are administered by science discipline, a non-segmented test design provides more freedom when selecting items that target the current best estimate of proficiency in an adaptive test. Embedded field-test items were randomly positioned in the test and randomly distributed across students. Every student received either one item cluster or four stand-alone items as embedded field-test items in their test.

4.2 TEST BLUEPRINTS

Test blueprints provide the following guidelines:

- Test length
- Science disciplines to be covered and the acceptable number of items across performance expectations (PEs) within each science discipline and Disciplinary Core Idea (DCI)

The science blueprints for grade 5 and grade 8 are presented in Table 12 and Table 13, respectively.

Table 12. Science Test Blueprint, Grade 5 Science

Grade 5	Min Clusters	Max Clusters	Min Stand-Alone Items	Max Stand-Alone Items	Min Clusters + Min Stand-Alone Items	Max Clusters + Max Stand-Alone Items
Discipline – Physical Science, PE Total = 17	**2**	**2**	**4**	**4**	**6**	**6**
DCI – Motion and Stability: Forces and Interactions	**0**	**1**	**0**	**2**	**0**	**3**
3-PS2-1: Forces – balanced and unbalanced forces	0	1	0	1	0	1
3-PS2-2: Forces – pattern predicts future motion	0	1	0	1	0	1
3-PS2-3: Forces – between objects not in contact	0	1	0	1	0	1
3-PS2-4: Forces – magnets*	0	1	0	1	0	1
5-PS2-1: Space systems	0	1	0	1	0	1
DCI – Energy	**0**	**1**	**0**	**2**	**0**	**3**
4-PS3-1: Energy – relationship between speed and energy of object	0	1	0	1	0	1
4-PS3-2: Energy – transfer of energy	0	1	0	1	0	1
4-PS3-3: Energy – changes in energy when objects collide	0	1	0	1	0	1
4-PS3-4: Energy – converting energy from one form to another*	0	1	0	1	0	1
5-PS3-1: Matter & Energy	0	1	0	1	0	1
DCI – Waves and Their Applications in Technologies for Information Transfer	**0**	**1**	**0**	**2**	**0**	**3**
4-PS4-1: Waves – waves can cause objects to move	0	1	0	1	0	1
4-PS4-2: Structure, function, information processing	0	1	0	1	0	1
4-PS4-3: Waves – using patterns to transfer information*	0	1	0	1	0	1
DCI – Matter and Its Interactions	**0**	**1**	**0**	**2**	**0**	**3**
5-PS1-1: Structure & Properties of Matter	0	1	0	1	0	1
5-PS1-2: Structure & Properties of Matter	0	1	0	1	0	1
5-PS1-3: Structure & Properties of Matter	0	1	0	1	0	1
5-PS1-4: Structure & Properties of Matter	0	1	0	1	0	1
Discipline – Life Science, PE Total = 12	**2**	**2**	**4**	**4**	**6**	**6**
DCI – From Molecules to Organisms: Structure and Function	**0**	**1**	**0**	**2**	**0**	**3**
3-LS1-1: Inheritance	0	1	0	1	0	1

Grade 5	Min Clusters	Max Clusters	Min Stand-Alone Items	Max Stand-Alone Items	Min Clusters + Min Stand-Alone Items	Max Clusters + Max Stand-Alone Items
4-LS1-1: Structure, Function, Information Processing	0	1	0	1	0	1
4-LS1-2: Structure, Function, Information Processing	0	1	0	1	0	1
5-LS1-1: Matter & Energy	0	1	0	1	0	1
DCI – Ecosystems: Interactions, Energy, and Dynamics	**0**	**1**	**0**	**2**	**0**	**3**
3-LS2-1: Ecosystems	0	1	0	1	0	1
5-LS2-1: Matter & Energy	0	1	0	1	0	1
DCI – Inheritance and Variation of Traits	**0**	**1**	**0**	**2**	**0**	**3**
3-LS3-1: Inheritance	0	1	0	1	0	1
3-LS3-2: Inheritance	0	1	0	1	0	1
DCI – Biological Evolution: Unity and Diversity	**0**	**1**	**0**	**2**	**0**	**3**
3-LS4-1: Ecosystems	0	1	0	1	0	1
3-LS4-2: Inheritance	0	1	0	1	0	1
3-LS4-3: Ecosystems	0	1	0	1	0	1
3-LS4-4: Ecosystems*	0	1	0	1	0	1
Discipline – Earth and Space Science, PE Total = 13	**2**	**2**	**4**	**4**	**6**	**6**
DCI – Earth's Systems	**0**	**1**	**0**	**3**[a]	**0**	**3**
3-ESS2-1: Weather & Climate	0	1	0	1	0	1
3-ESS2-2: Weather & Climate	0	1	0	1	0	1
4-ESS2-1: Earth's Systems & Processes	0	1	0	1	0	1
4-ESS2-2: Earth's Systems & Processes	0	1	0	1	0	1
5-ESS2-1: Earth's Systems	0	1	0	1	0	1
5-ESS2-2: Earth's Systems	0	1	0	1	0	1
DCI – Earth and Human Activity	**0**	**1**	**0**	**2**	**0**	**3**
3-ESS3-1: Weather & Climate*	0	1	0	1	0	1
4-ESS3-2: Earth's Systems & Processes*	0	1	0	1	0	1
4-ESS3-1: Energy	0	1	0	1	0	1
5-ESS3-1: Earth's Systems	0	1	0	1	0	1

Grade 5	Min Clusters	Max Clusters	Min Stand-Alone Items	Max Stand-Alone Items	Min Clusters + Min Stand-Alone Items	Max Clusters + Max Stand-Alone Items
DCI – Earth's Place in the Universe	**0**	**1**	**0**	**2**	**0**	**3**
4-ESS1-1: Earth's Systems & Processes	0	1	0	1	0	1
5-ESS1-1: Space Systems	0	1	0	1	0	1
5-ESS1-2: Space Systems	**0**	**1**	**0**	**1**	**0**	**1**
PE Total = 42	**6**	**6**	**12**	**12**	**18**	**18**

Note: Constraints on sampling across grades per discipline (except Grade 3 LS): At most 1 cluster per grade, at most 3 stand-alone items per grade, at most 4 clusters + stand-alone items per grade; for grade 3 LS, at most 2 clusters per grade, at most 3 stand-alone items per grade, at most 4 clusters + stand-alone items per grade.
* These PEs have an engineering component.
[a]Because of the limitation of the item pool in the ESS discipline, the maximum number of stand-alone items allowed was changed from 2 to 3 while keeping the maximum number of items (clusters + stand-alone items) allowed at 3 in ESS2.

Table 13. Science Test Blueprint, Grade 8 Science

Grade 8	Min Clusters	Max Clusters	Min Stand-Alone Items	Max Stand-Alone Items	Min Clusters + Min Stand-Alone Items	Max Clusters + Max Stand-Alone Items
Discipline – Physical Science, PE Total = 19	**2**	**2**	**4**	**4**	**6**	**6**
DCI – Matter and Its Interactions	**0**	**1**	**0**	**2**	**0**	**3**
8-MS-PS1-1: Structure & Properties of Matter	0	1	0	1	0	1
8-MS-PS1-2: Chemical Reactions	0	1	0	1	0	1
8-MS-PS1-3: Structure & Properties of Matter	0	1	0	1	0	1
8-MS-PS1-4: Structure & Properties of Matter	0	1	0	1	0	1
8-MS-PS1-5: Chemical Reactions	0	1	0	1	0	1
8-MS-PS1-6: Chemical Reactions*	0	1	0	1	0	1
DCI – Motion and Stability: Forces and Interactions	**0**	**1**	**0**	**2**	**0**	**3**
7-MS-PS2-1: Forces & Interactions*	0	1	0	1	0	1
7-MS-PS2-2: Forces & Interactions	0	1	0	1	0	1
7-MS-PS2-3: Forces & Interactions	0	1	0	1	0	1

Grade 8	Min Clusters	Max Clusters	Min Stand-Alone Items	Max Stand-Alone Items	Min Clusters + Min Stand-Alone Items	Max Clusters + Max Stand-Alone Items
7-MS-PS2-4: Forces & Interactions	0	1	0	1	0	1
7-MS-PS2-5: Forces & Interactions	0	1	0	1	0	1
DCI – Energy	**0**	**1**	**0**	**2**	**0**	**3**
7-MS-PS3-1: Energy	0	1	0	1	0	1
7-MS-PS3-2: Energy	0	1	0	1	0	1
7-MS-PS3-3: Energy*	0	1	0	1	0	1
7-MS-PS3-4: Energy	0	1	0	1	0	1
7-MS-PS3-5: Energy	0	1	0	1	0	1
DCI – Waves and Their Applications in Technologies for Information Transfer	**0**	**1**	**0**	**2**	**0**	**3**
6-MS-PS4-1: Waves & Electromagnetic Radiation	0	1	0	1	0	1
6-MS-PS4-2: Waves & Electromagnetic Radiation	0	1	0	1	0	1
6-MS-PS4-3: Waves & Electromagnetic Radiation	0	1	0	1	0	1
Discipline – Life Science, PE Total = 21	**2**	**2**	**4**	**4**	**6**	**6**
DCI – From Molecules to Organisms: Structures and Processes	**0**	**1**	**0**	**2**	**0**	**3**
7-MS-LS1-1: Structure, Function, Information Processing	0	1	0	1	0	1
7-MS-LS1-2: Structure, Function, Information Processing	0	1	0	1	0	1
7-MS-LS1-3: Structure, Function, Information Processing	0	1	0	1	0	1
8-MS-LS1-4: Growth, Development, Reproduction	0	1	0	1	0	1
8-MS-LS1-5: Growth, Development, Reproduction	0	1	0	1	0	1
6-MS-LS1-6: Matter & Energy	0	1	0	1	0	1
6-MS-LS1-7: Matter & Energy	0	1	0	1	0	1
7-MS-LS1-8: Structure, Function, Information Processing	0	1	0	1	0	1
DCI – Ecosystems: Interactions, Energy, and Dynamics	**0**	**1**	**0**	**2**	**0**	**3**
6-MS-LS2-1: Matter & Energy	0	1	0	1	0	1
6-MS-LS2-2: Interdependent Relationships in Ecosystems	0	1	0	1	0	1
6-MS-LS2-3: Matter & Energy	0	1	0	1	0	1

Grade 8	Min Clusters	Max Clusters	Min Stand-Alone Items	Max Stand-Alone Items	Min Clusters + Min Stand-Alone Items	Max Clusters + Max Stand-Alone Items
6-MS-LS2-4: Matter & Energy	0	1	0	1	0	1
6-MS-LS2-5: Interdependent Relationships in Ecosystems*	0	1	0	1	0	1
DCI – Hereditary: Inheritance and Variation of Traits	**0**	**1**	**0**	**2**	**0**	**3**
8-MS-LS3-1: Growth, Development, Reproduction	0	1	0	1	0	1
8-MS-LS3-2: Growth, Development, Reproduction	0	1	0	1	0	1
DCI – Biological Evolution: Unity and Diversity	**0**	**1**	**0**	**2**	**0**	**3**
8-MS-LS4-1: Natural Selection & Adaptation	0	1	0	1	0	1
8-MS-LS4-2: Natural Selection & Adaptation	0	1	0	1	0	1
8-MS-LS4-3: Natural Selection & Adaptation	0	1	0	1	0	1
8-MS-LS4-4: Natural Selection & Adaptation	0	1	0	1	0	1
8-MS-LS4-5: Growth, Development, Reproduction	0	1	0	1	0	1
8-MS-LS4-6: Natural Selection & Adaptation	0	1	0	1	0	1
Discipline – Earth and Space Science, PE Total = 15	**2**	**2**	**4**	**4**	**6**	**6**
DCI – Earth's Place in the Universe	**0**	**1**	**0**	**2**	**0**	**3**
6-MS-ESS1-1: Space Systems	0	1	0	1	0	1
6-MS-ESS1-2: Space Systems	0	1	0	1	0	1
6-MS-ESS1-3: Space Systems	0	1	0	1	0	1
7-MS-ESS1-4: History of Earth	0	1	0	1	0	1
DCI – Earth's Systems	**0**	**1**	**0**	**2**	**0**	**3**
7-MS-ESS2-1: Earth's Systems	0	1	0	1	0	1
7-MS-ESS2-2: History of Earth	0	1	0	1	0	1
7-MS-ESS2-3: History of Earth	0	1	0	1	0	1
7-MS-ESS2-4: Earth's Systems	0	1	0	1	0	1
6-MS-ESS2-5: Weather & Climate	0	1	0	1	0	1
6-MS-ESS2-6: Weather & Climate	0	1	0	1	0	1
DCI – Earth and Human Activity	**0**	**1**	**0**	**2**	**0**	**3**
7-MS-ESS3-1: Earth's Systems	0	1	0	1	0	1

Grade 8	Min Clusters	Max Clusters	Min Stand-Alone Items	Max Stand-Alone Items	Min Clusters + Min Stand-Alone Items	Max Clusters + Max Stand-Alone Items
6-MS-ESS3-2: Human Impacts	0	1	0	1	0	1
7-MS-ESS3-3: Human Impacts*	0	1	0	1	0	1
8-MS-ESS3-4: Human Impacts	0	1	0	1	0	1
6-MS-ESS3-5: Weather & Climate	0	1	0	1	0	1
Total PE= 55	**6**	**6**	**12**	**12**	**18**	**18**

Note: Constraints on sampling across grades per discipline: at most 1 cluster per grade, at most 3 stand-alone items per grade, at most 4 clusters + stand-alone items per grade.
* These PEs have an engineering component.

The main characteristics of the blueprint were that any PE could be tested only once (indicated by the values of 0 and 1 for the minimum and maximum values of the individual PEs in Table 12 and Table 13). In general, no more than one item cluster or two stand-alone items could be sampled from the same DCI; and no more than three total items could be sampled from the same DCI (as indicated by the minimum and maximum values in the rows representing DCIs). In addition, based on item data from the most recent year, a few stand-alone items, which took more than 3.5 minutes for 80% of students to complete, and cluster items, which took more than 14 minutes for 80% students to complete, were identified as items that require long testing time. At most one of such cluster items and two of such stand-alone items could be sampled in a test session. Furthermore, at most, one item cluster and three stand-alone items could be sampled from a given grade within the grade band.

While tests are not timed, the West Virginia Department of Education (WVDE) published estimated testing times for the WVGSA science assessment. The 85th percentile of the testing times by grade is presented in Table 14.

Table 14. WVGSA Science 85th Percentile Testing Times by Grade

Subject	Grade	85th Percentile Testing (min)
Science	5	100.92
	8	76.03

4.3 Test Construction

During fall 2021, CAI psychometricians and content experts worked with WVDE content specialists and leadership to build item pools for the spring 2022 administration. The WVGSA test construction used a structured test construction plan, explicit blueprints, and active collaborative participation from all parties.

The 2022 WVGSA science test item pools were built by CAI test developers to match items exactly to the detailed test blueprints. Operational items were selected from nine item banks (ICCR, Connecticut, Hawaii, Idaho, MSSA, Oregon, Utah, West Virginia, and Wyoming) to fulfill the blueprint for each grade. Table 15 through Table 19 summarize the 2022 WVGSA science item pool. Appendix J, WVGSA Science Assessment Item Pool, outlines the 2022 WVGSA science item pool by grade band, PE, and origin.

Table 15. Spring 2022 WVGSA Science Operational and Field-Test Item Pool

Grade and Item Type	ICCR Items	West Virginia Items	MOU Items[a]	Total Pool Items
Grade 5	**96**	**19**	**143**	**258**
Cluster	38	10	82	130
Stand-Alone	58	9	61	128

Grade and Item Type	ICCR Items	West Virginia Items	MOU Items[a]	Total Pool Items
Grade 8	**106**	**18**	**171**	**295**
Cluster	28	8	115	151
Stand-Alone	78	10	56	144
Total	**202**	**37**	**314**	**553**

Note. [a]Other MOU state items administered included items from Connecticut, Hawaii, Idaho, MSSA (Rhode Island and Vermont), Oregon, Utah, and Wyoming.

Table 16. Spring 2022 WVGSA Science Operational Item Pool

Grade and Item Type	ICCR Operational Items	West Virginia Operational Items	MOU Operational Items[a]	Total Operational Pool Items
Grade 5	**96**	**13**	**130**	**239**
Cluster	38	9	72	119
Stand-Alone	58	4	58	120
Grade 8	**82**	**10**	**170**	**262**
Cluster	27	4	115	146
Stand-Alone	55	6	55	116
Total	**178**	**23**	**300**	**501**

Note. [a]Other MOU state operational items administered included items from Connecticut, Hawaii, Idaho, MSSA (Rhode Island and Vermont), Oregon, Utah, and Wyoming.

Table 17. Spring 2022 WVGSA Science Field-Test Item Pool

Grade and Item Type	ICCR Field-Test Items	West Virginia Field-Test Items	MOU Field-Test Items[a]	Total Field-Test Pool Items
Grade 5	**0**	**6**	**13**	**19**
Cluster	0	1	10	11
Stand-Alone	0	5	3	8
Grade 8	**24**	**8**	**1**	**33**
Cluster	1	4	0	5
Stand-Alone	23	4	1	28
Total	**24**	**14**	**14**	**52**

Note. [a]Other MOU states field-test items administered include Hawaii, MSSA (Rhode Island and Vermont), and Utah.

Table 18. Spring 2022 WVGSA Science Item Pool by Science Discipline

Grade	Science Discipline	Item Type	ICCR Items	West Virginia Items	MOU Items[a]	Total Pool Items
Grade 5	Earth and Space Sciences	Cluster	12	4	25	41
		Stand-Alone	17	3	20	40
	Life Sciences	Cluster	14	4	27	45
		Stand-Alone	21	2	18	41
	Physical Sciences	Cluster	12	2	30	44
		Stand-Alone	20	4	23	47
Grade 8	Earth and Space Sciences	Cluster	9	3	30	42
		Stand-Alone	22	1	17	40
	Life Sciences	Cluster	8	3	49	60
		Stand-Alone	34	3	19	56
	Physical Sciences	Cluster	11	2	36	49
		Stand-Alone	22	6	20	48
Total			**202**	**37**	**314**	**553**

Note. [a]Other MOU state items administered included items from Connecticut, Hawaii, Idaho, MSSA (Rhode Island and Vermont), Oregon, Utah, and Wyoming.

Table 19. Spring 2022 WVGSA Science Item Pool by Disciplinary Core Idea

Grade	Science Discipline	Disciplinary Core Idea	ICCR Items	West Virginia Items	MOU Items[a]	Total Pool Items
Grade 5	Earth and Space Sciences	ESS1	8	1	14	23
		ESS2	11	4	19	36
		ESS3	10	2	12	24
	Life Sciences	LS1	12	5	15	32
		LS2	5	0	9	14
		LS3	3	1	6	10
		LS4	15	0	15	30
	Physical Sciences	PS1	9	1	14	24
		PS2	7	4	16	27
		PS3	13	1	14	28
		PS4	3	0	9	12
Grade 8	Earth and Space Sciences	ESS1	12	1	10	23
		ESS2	11	2	24	36
		ESS3	8	1	15	23
	Life Sciences	LS1	12	3	23	38
		LS2	15	1	22	38
		LS3	3	2	8	13
		LS4	12	0	15	27
	Physical Sciences	PS1	7	2	23	32
		PS2	5	2	15	22
		PS3	12	2	10	24
		PS4	9	2	8	19
Total			**202**	**37**	**314**	**553**

Note. [a]Other MOU state items administered included items from Connecticut, Hawaii, Idaho, MSSA (Rhode Island and Vermont), Oregon, Utah, and Wyoming.

Additional information on *p*-values, biserial correlations, and item response theory (IRT) parameters is available in Volume 1, Annual Technical Report. The details on calibration, equating, and scoring of the WVGSA can also be found in Volume 1.

4.4 Paper-Pencil/Braille Accommodation Form Construction

Student scores should not depend upon the mode of administration or type of test form. In spring 2022, the WVGSA was primarily administered in an online test delivery system in spring 2022, with only seven students in grade 5 taking the paper-pencil form and one student in grade 5 taking the braille form. Scores obtained via alternate modes of administration must be established as comparable to scores obtained through online testing. This section outlines the overall test development plans that ensured the comparability between the online and paper-pencil/braille tests.

To build paper-pencil/braille forms, content specialists began with the online pool and removed any items that could not render on paper or be transcribed in braille. Next, content specialists constructed fixed forms adhering to the test blueprint. In Spring 2022, the paper-pencil/braille forms met all blueprint requirements.

5. Simulation Summary Report

This section describes the results of the simulated test administrations used to configure and evaluate the adequacy of the item-selection algorithm used to administer the 2021–2022 West Virginia General Summative Assessment (WVGSA) for grades 5 and 8 science. Simulations were conducted to configure the algorithm's settings and evaluate whether individual tests adhered to the test blueprint.

Some important settings included "Select Candidate Set 1" (cset1) and "Select Candidate Set 2" (cset2), which represent subsets of the item pool that were eligible for item selection. Refer to Appendix I, Adaptive Algorithm Design, for more details about the current item selection algorithm. In spring 2022, cset1 and cset2 values were set to 5 and 1. Psychometricians reviewed the simulation results and configured settings based on some key diagnostics, including:

- *Match-to-Test Blueprint.* This diagnostic determines whether the tests have the correct number of test items overall and the appropriate proportion by content categories at each level of the content hierarchy, as specified in the test blueprints for each science grade.
- *Item Exposure Rate.* This diagnostic evaluates the utility of the item pools and identifies overexposed and underexposed items.

These diagnostics are interrelated. For example, if the test pool for a particular content category is limited (i.e., there are only a few test items available), achieving a 100% match to the blueprint for this content level will lead to a high item exposure rate, which means that a large number of students will see the same items. The software system that performs the simulation allows adjustments to the setting parameters to attain the best possible balance among these diagnostics.

The simulation involves an iterative process that reviews the initial results, adjusts the system parameters, runs new simulations, reviews the new results, and repeats the exercise until an optimal balance is achieved. The final setting would then be applied for the operational assessments.

5.1 Factors Affecting Simulation Results

Several factors may influence the simulation results in an adaptive administration. These factors include

1. *The proportional relationship between the pool and the constraints to be met.* Proportionally distributed pools tend to make better use of the item pool (i.e., more uniform item exposure) and make it easier to meet blueprint and other constraints. For example, if the specifications call for at least one item cluster per Disciplinary Core Idea (DCI), but the pool has no item cluster for some DCIs, it may be impossible to meet this constraint.
2. *The correlational structure between constraints.* It is easier to satisfy a constraint if there are instances of the constraint at all levels of another constraint. For example, if stand-alone items within a discipline are associated only with a specific DCI, it may be difficult to meet both the desired distribution of content and the desired distribution of item type.
3. *Whether there is a "strict maximum" on a given constraint.* This means that the requirement must be met exactly in each test administration.

5.2 Results of Simulated Test Administrations: English

This section presents the simulation results for the English online tests, which is the assessment taken by almost all students (99.9%). Simulations were evaluated for all content areas using 5,000 simulated cases per grade.

5.2.1 Summary of Blueprint Match

The simulation results showed no blueprint violations at all content levels for both grades.

5.2.2 Item Exposure

The simulator output also reports the degree to which the constraints outlined in the blueprints may yield greater exposure of items to students. This is reported by examining the percentage of test administrations in which an item appears. For instance, in a fixed paper-pencil form, 100% of the items appear on 100% of the test administrations because every test taker sees the same form. In an adaptive test or a LOFT with a sufficiently large item pool, it is expected that most of the items would appear on a relatively small percentage of the test administrations only.

When this condition holds, it suggests that test administrations between students are more or less unique. Therefore, the item exposure rate was calculated for each item by dividing the total number of test administrations in which an item appears by the total number of tests administered. Then the distribution of the item exposure rate (r) was shown in eight bins. The bins $r = 0\%$ (unused), $0\% < r \leq 1\%$, $1\% < r \leq 5\%$, $5\% < r \leq 20\%$, $20\% < r \leq 40\%$, $40\% < r \leq 60\%$, $60\% < r \leq 80\%$, and $80\% < r \leq 100\%$. If global item exposure is minimal, it is expected that the largest proportion of items would appear in the bins of 0% < r ≤ 20%, indicating that most of the items appeared on a very small percentage of the test forms.

Table 20 presents the percentage of items that fell into each exposure bin for both grades. Most test items were administered in 1%–20% of the test administrations. The minimum exposure rate was 0.3% in grade 5.

Table 20. Item Exposure Rates by Grade: Percentage of Items by Exposure Rate, Across All English Online Simulation Sessions

Grade	Total Items	[0,0]%	(0,1]%	(1,5]%	(5,20]%	(20,40]%	(40,60]%	(60,80]%	(80,100]%
5	239	0	23.01	27.62	39.75	9.62	0	0	0
8	262	0	10.69	43.13	40.08	4.96	1.15	0	0

5.3 Results of Simulated Test Administrations: Spanish

This section presents the simulation results for the Spanish tests. The Spanish item pool consisted of a subset of ICCR items and some MOU items that had Spanish translations available. Table 21 presents the numbers of items available for the Spanish tests in spring 2022.

Table 21. Spring 2022 Spanish Operational Item Pool

Grade	Item Type	Total Number of Items
5	Cluster	19
	Stand-Alone	37
8	Cluster	16
	Stand-Alone	37
Total		**109**

Simulations were evaluated for all content areas using 1,000 simulated cases per grade.

5.3.1 Summary of Blueprint Match

The simulation results showed no blueprint violations at all content levels for both grades.

5.3.2 Item Exposure

Table 22 presents the percentage of items that fell into each exposure bin for all grades. Most items were administered in more than 20% of the test administrations. A few items had an exposure rate of 100% because of the limited Spanish item pool. Only those items were available to satisfy the blueprint constraints.

Table 22. Item Exposure Rates by Grade: Percentage of Items by Exposure Rate, Across All Spanish Simulation Sessions

Grade	Total Items	[0,0]%	(0,1]%	(1,5]%	(5,20]%	(20,40]%	(40,60]%	(60,80]%	(80,100]%
5	56	0	0	10.71	19.64	39.29	19.64	7.14	3.57
8	53	0	0	5.66	16.98	37.74	26.42	13.21	0

6. Operational Test Administration Summary Report

This section presents the blueprint match reports and item exposure rates for the spring 2022 operational test administrations.

6.1 Blueprint Match

The simulation results showed no blueprint violations at all content levels for both English and Spanish tests at both grades.

6.2 Item Exposure

Table 23 presents the item exposure rates for the spring 2022 test administration. The exposure rates were very similar to the simulation results described in Section 5.2.2, Item Exposure, of this volume for the English test administrations. The item exposure rate for field-test items ranged from 6.97% to 8.27% for grade 5 and 7.84% to 8.70% for grade 8. More items had an exposure rate between 20% to 100% in the Spanish tests compared to the English assessments due to a smaller item pool. In addition, the operational exposure rates were somewhat different from the simulation results because of the small population sizes in both grades. In spring 2022, only 2 students in grade 5 and 24 students in grade 8 took the Spanish version of the assessments.

Table 23. Item Exposure Rates by Grade: Percentage of Items by Exposure Rate, Across All Spring 2022 Test Administrations

Grade	Total Items	[0,0]%	(0,1]%	(1,5]%	(5,20]%	(20,40]%	(40,60]%	(60,80]%	(80,100]%
					English				
5	239	0	23.43	25.52	42.68	7.53	0.84	0	0
8	262	0	13.74	44.27	35.11	4.96	1.91	0	0
					Spanish				
5	27	51.79	0	0	0	0	32.14	0	16.07
8	48	9.43	0	5.66	26.42	24.53	13.21	11.32	9.43

References

Council of Chief State School Officers (CCSSO). (2015). *Science Assessment Item Collaborative (SAIC) Assessment Framework for the Next Generation Science Standards*. Washington, DC: Council of Chief State School Officers. Retrieved from https://ccsso.org/sites/default/files/2017-12/SAICAssessmentFramework_FINAL.pdf

National Research Council. (2012). *A framework for K–12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: The National Academies Press.

Appendix A

Item Writer Training Materials

Exhibit 1: LABS Guidelines

Cambium Assessment

LABS GUIDELINES

1. Stereotyping

Testing materials should not present persons stereotyped according to the following characteristics:

- Age
- Disability
- Gender
- Race/Ethnicity
- Sexual orientation

2. Sensitive or Controversial Subjects

Controversial or potentially distressing subjects should be avoided or treated sensitively. For example, a passage discussing the historical importance of a battle is acceptable whereas a graphic description of a battle would not be. Controversial subjects include:

a. Death and Disease
b. Gambling*
c. Politics (Current)
d. Race relations
e. Religion
f. Sexuality
g. Superstition
h. War

**References to gambling should be avoided in mathematics items related to probability.*

3. Advice

Testing materials should not advocate specific lifestyles or behaviors except in the most general or universally agreed-upon ways. For example, a recipe for a healthful fruit snack is acceptable but a passage recommending a specific diet is not. The following categories of advice should be avoided:

- Religion
- Sexual preference
- Exercise
- Diet

4. Dangerous Activity

Tests should not contain content that portrays people engaged in or explains how to engage in dangerous activities. Examples of dangerous activities include:

- Deep-sea diving
- Stunts
- Parachuting
- Smoking
- Drinking

5. Population Diversity and Ethnocentrism

Testing materials should:

- Reflect the diversity of the testing population
- Use stimulus materials (such as works of literature) produced by members of minority communities
- Use personal names from different ethnic origin communities
- Use pictures of people from different ethnic origin communities
- Avoid ethnocentrism (the attitude that all people should share a particular group's language, beliefs, culture, or religion)

6. Differential Familiarity and Elitism

Specialized concepts and terminology extraneous to the core content of test questions should be avoided. This caveat applies to terminology from the fields of:

- Construction
- Finance
- Sports
- Law
- Machinery
- Military topics
- Politics
- Science
- Technology
- Agriculture

7. Language Use

Language should be as inclusive as possible.

- Avoid masculine-coded words like mankind, manmade, and the generic "he"
- Use equal pairs such as husband and wife rather than man and wife

8. Language Accessibility

The grammar and vocabulary should be clear, concise, and appropriate for the intended grade level. The following should be avoided or used with care:

- Passive constructions
- Idioms
- Multiple subordinate clauses
- Pronouns with unclear antecedents
- Multiple-meaning words
- Nonstandard grammar
- Dialect
- Jargon

9. Illustrations and Graphics

Illustrations and graphics should embody all of the previously referenced LABS Guidelines.

Exhibit 2: LABS Checklist

Cambium Assessment

LABS – Checklist

Stereotyping Considerations

- ☐ Does the material negatively represent or stereotype people based on gender or sexual preference?
- ☐ Does the material portray one or more people with disabilities in a negative or stereotypical manner?
- ☐ Does the material portray one or more religious groups as aggressive or violent?
- ☐ Does the material romanticize or demean people based on socioeconomic status?
- ☐ Does the material portray one or more ethnic groups or cultures participating in certain stereotypical activities or occupations?
- ☐ Does the material portray one or more age groups in a negative or stereotypical manner?

Sensitive / Controversial Material Considerations

- ☐ Does the material require a student to take a position that challenges authority?
- ☐ Does the material present war or violence in an overly graphic manner?
- ☐ Does the material present sensitive or highly controversial subjects, such as death, war, abortion, euthanasia, or natural disasters, except where they are needed to meet State Content Standards?
- ☐ Does the material require examinees to disclose values that they would rather hold confidential?
- ☐ Does the material present sexual innuendoes?
- ☐ Does the material trivialize significant or tragic human experiences?
- ☐ Does the material require the parent, teacher, or examinee to support a position that is contrary to their religious beliefs?

Advice Considerations

- ☐ Does the material contain advice pertaining to health and well-being about which there is not universal agreement?

Population Diversity

- Is the material written by members of diverse groups?
- Does the material reflect the experiences of diverse groups?
- Does the material portray people in positive nontraditional roles?
- Does test material represent the racial and ethnic composition of the testing population?
- Does the material reflect ethnocentrism?
- Does the material refer to population subgroups accurately?
- Does test material reflect diversity through the use of names, cultural references, pictures, and roles?

Differential Familiarity / Elitism

- Does the material contain phrases, concepts, and beliefs that are irrelevant to testing domain and are likely to be more familiar to specific groups that others?
- Does the material require knowledge of individuals, events, or groups that is not familiar to all groups of students?
- Does the material suggest that affluence is related to merit or intelligence?
- Does the material suggest that poverty is related to increased negative behaviors in society?
- Does the material use language, content, or context that is offensive to people of a particular economic status?
- Does success with the material assume that the examinee has experience with a certain type of family structure?
- Does the material favor one socioeconomic group over another?
- Does the material assume values not shared by all test takers?

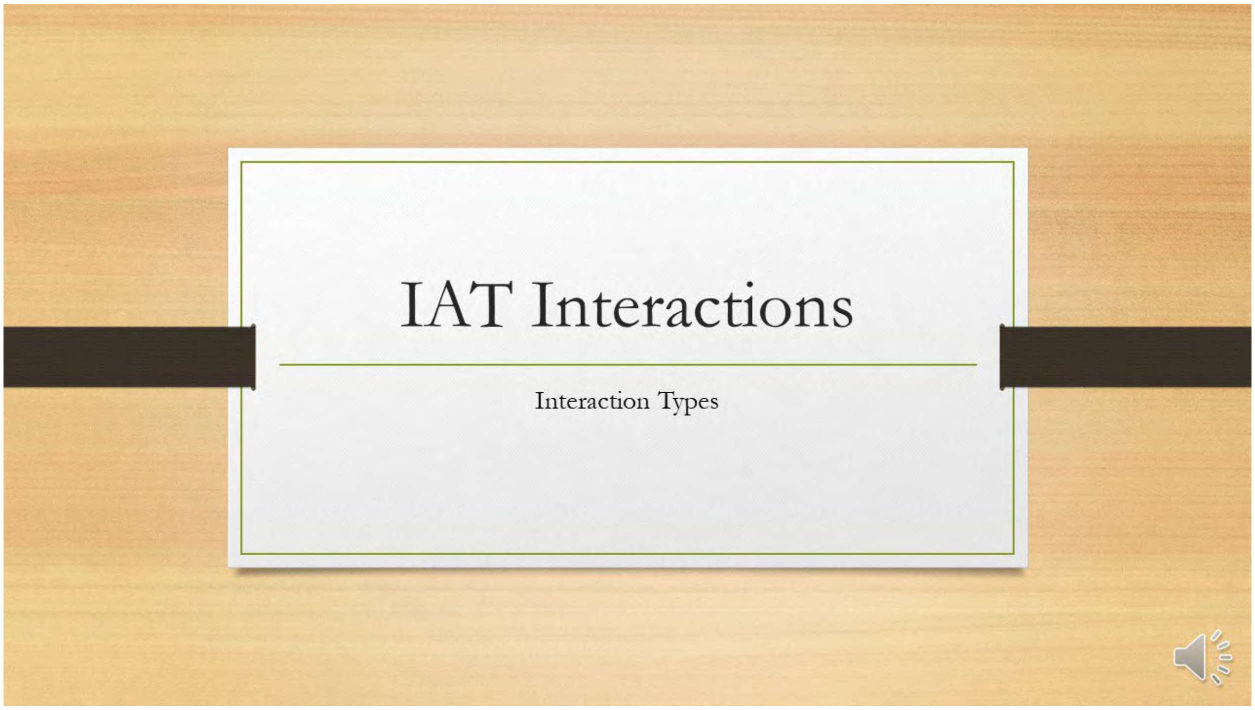
Linguistic Features / Language Accessibility/Graphics

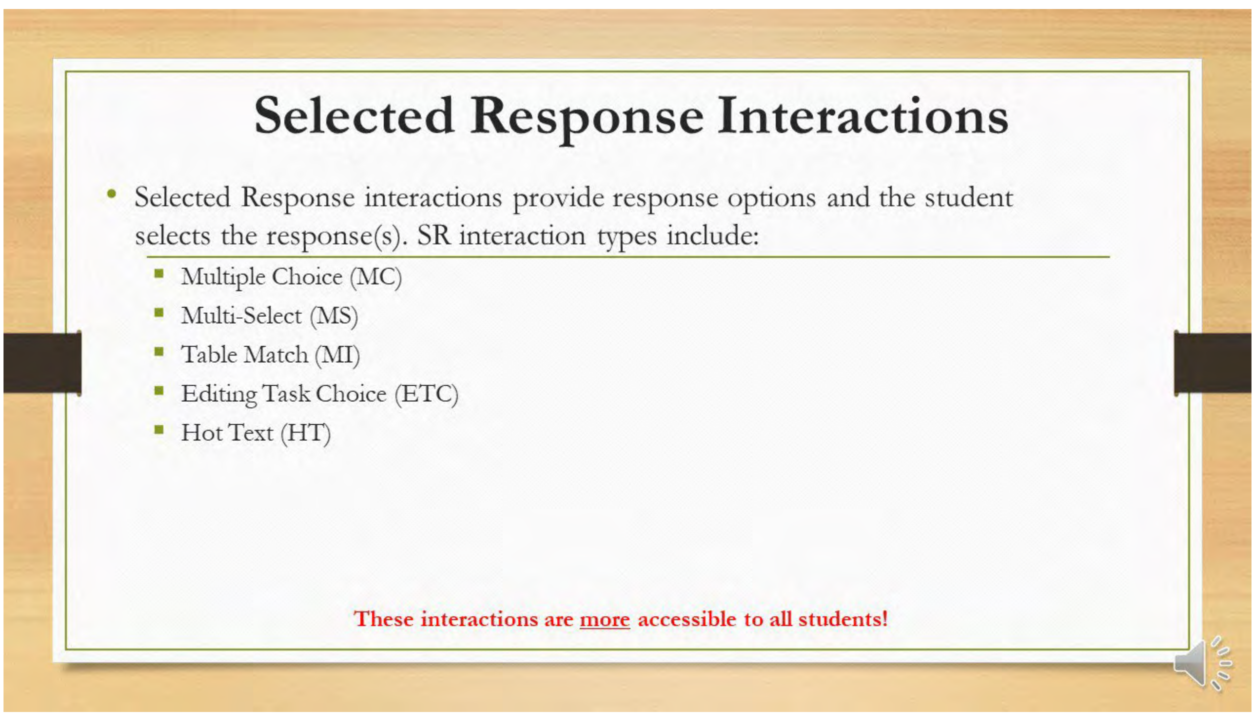
- Is grammar and vocabulary used in the items clear, concise and appropriate for the intended grade level?
- Are passages at a difficulty level that is appropriate for the intended grade level?
- Do the illustrations and graphics embody all of the previously referenced LABS Guidelines?

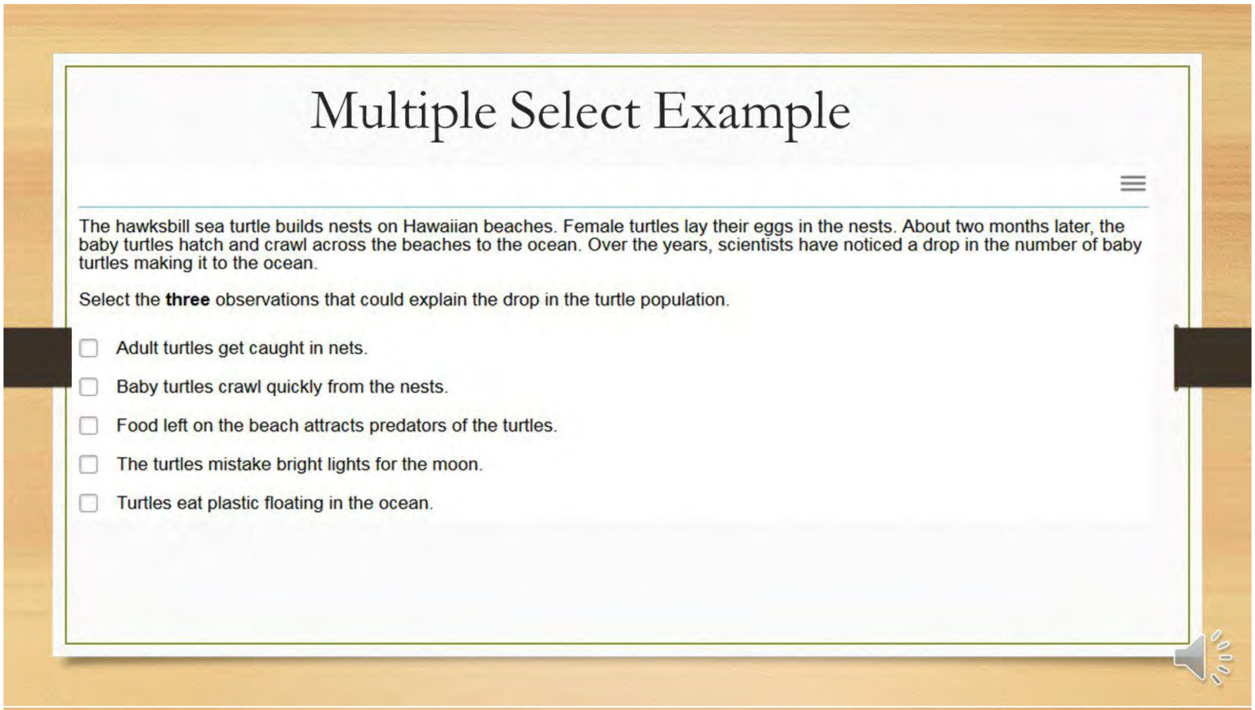
Other questions to consider

- Does the material favor one age group over others except in a context where experience or maturation is relevant?
- Does the material use language, content, or context that is not accessible to one or more of the age groups tested?
- Does the material contain language or content that contradicts values held by a certain culture?
- Does the material favor one racial or ethnic group over others?
- Does the material degrade people based on physical appearance or any physical, cognitive, or emotional challenge?
- Does the material focus only on a person’s disability rather than portraying the whole person?
- Does the material favor one religion and/or demean others?

Exhibit 3: An Overview of Interaction Types


IAT Interactions
Interaction Types


Selected Response Interactions
• Selected Response interactions provide response options and the student selects the response(s). SR interaction types include:
▪ Multiple Choice (MC)
▪ Multi-Select (MS)
▪ Table Match (MI)
▪ Editing Task Choice (ETC)
▪ Hot Text (HT)
These interactions are more accessible to all students!


Multiple Select Example
The hawksbill sea turtle builds nests on Hawaiian beaches. Female turtles lay their eggs in the nests. About two months later, the baby turtles hatch and crawl across the beaches to the ocean. Over the years, scientists have noticed a drop in the number of baby turtles making it to the ocean.
Select the three observations that could explain the drop in the turtle population.
Adult turtles get caught in nets.
Baby turtles crawl quickly from the nests.
Food left on the beach attracts predators of the turtles.
The turtles mistake bright lights for the moon.
Turtles eat plastic floating in the ocean.

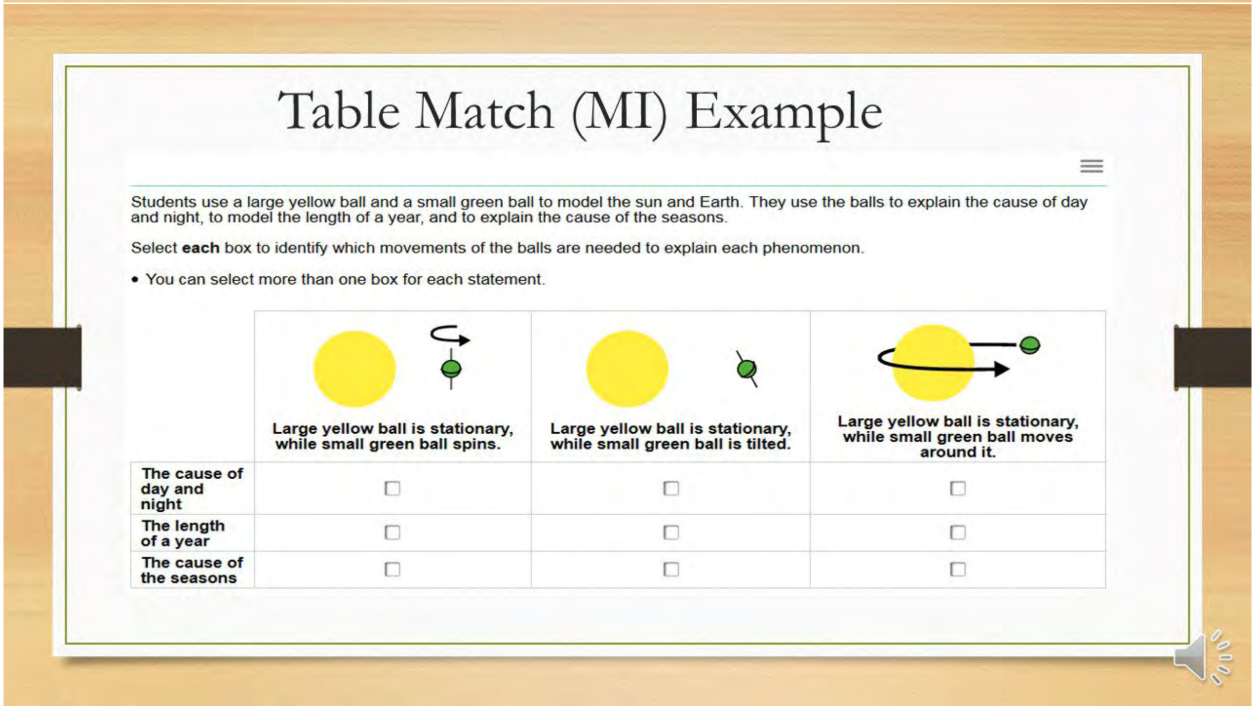
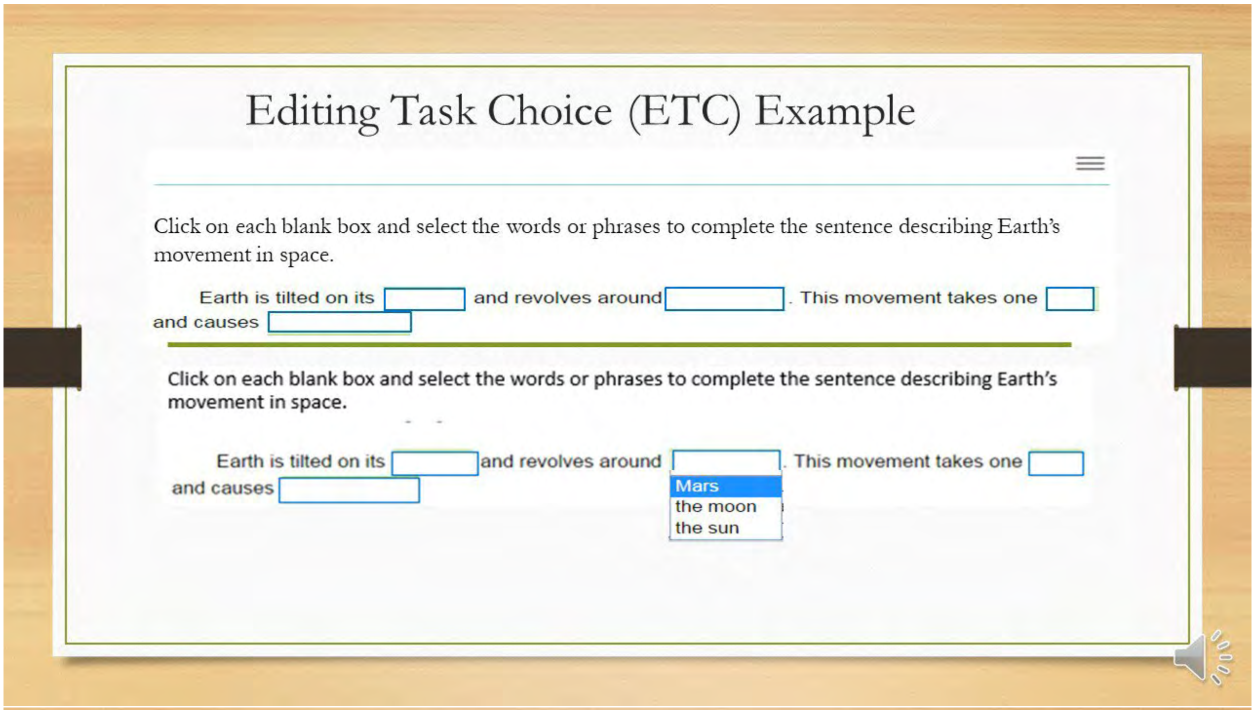
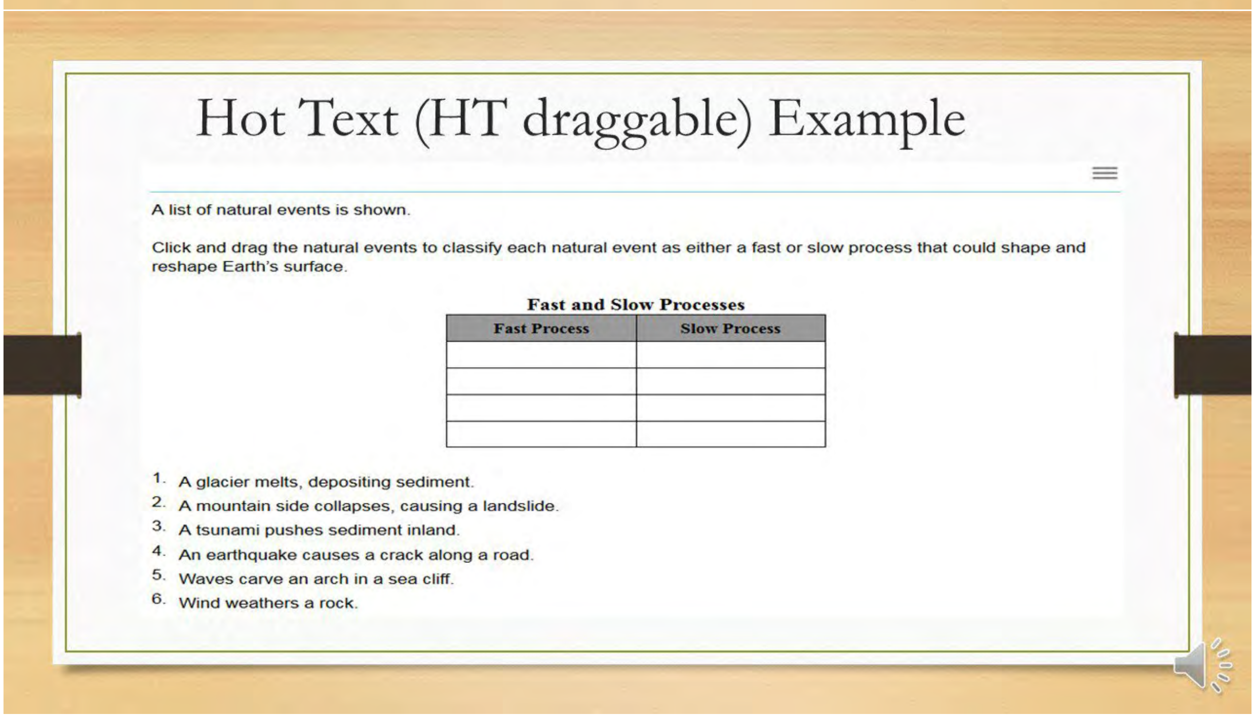
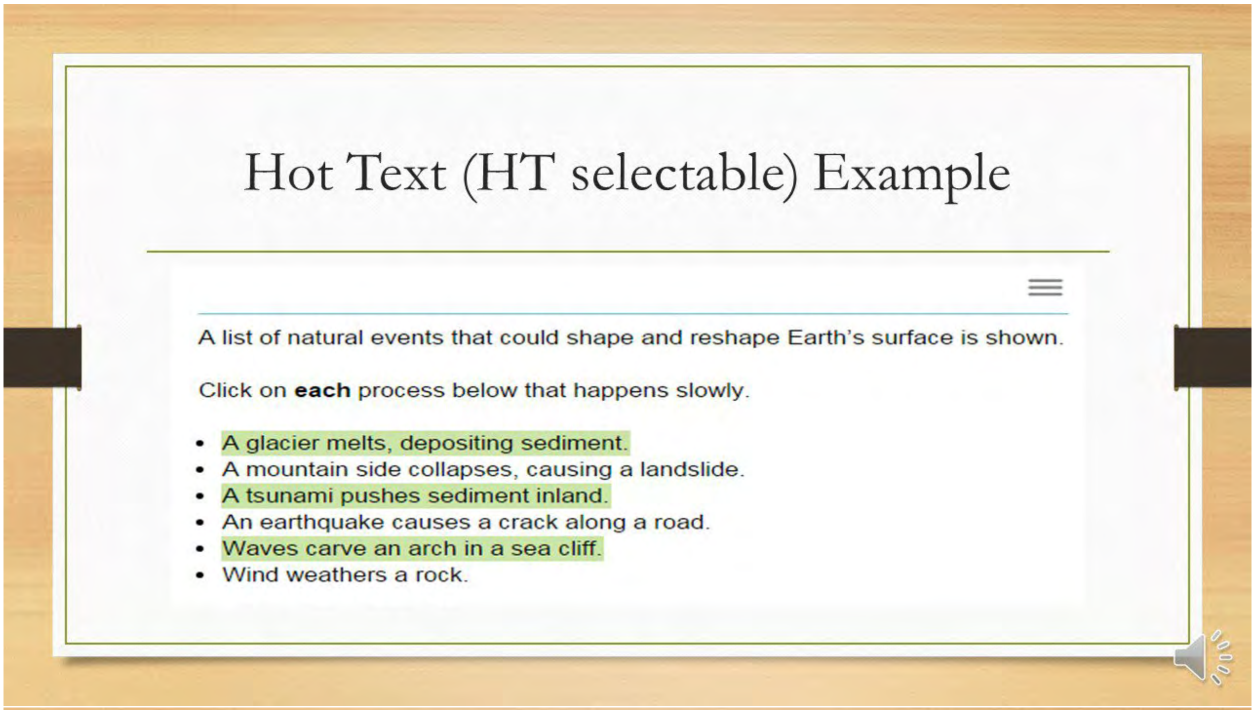
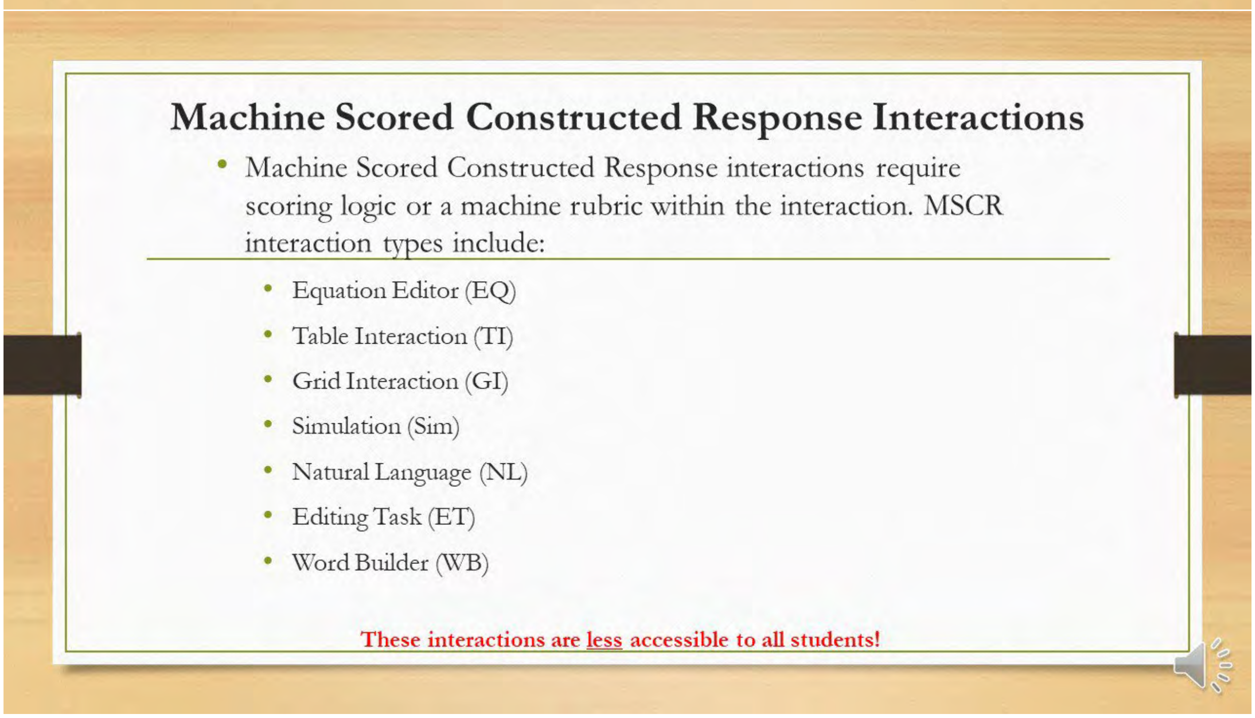


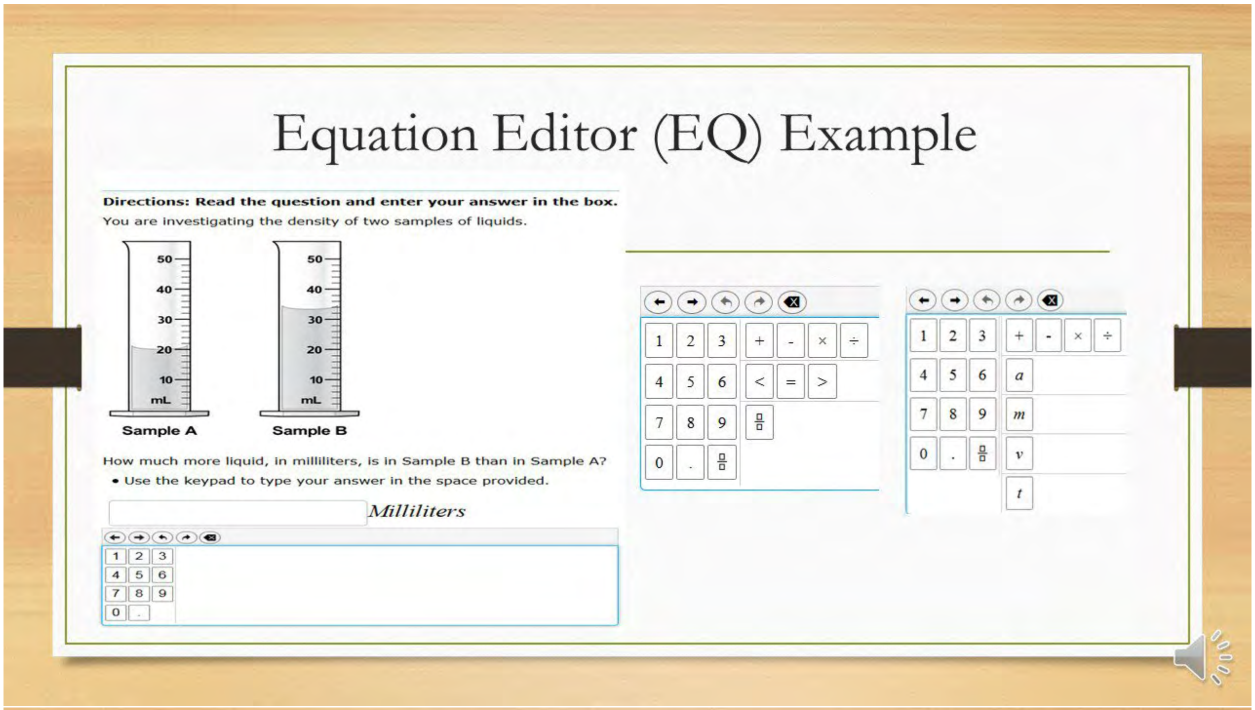
Table Match (MI) Example
Students use a large yellow ball and a small green ball to model the sun and Earth. They use the balls to explain the cause of day and night, to model the length of a year, and to explain the cause of the seasons.
Select each box to identify which movements of the balls are needed to explain each phenomenon.
• You can select more than one box for each statement.
Large yellow ball is stationary, while small green ball spins.
Large yellow ball is stationary, while small green ball is tilted.
Large yellow ball is stationary, while small green ball moves around it.
The cause of day and night
The length of a year
The cause of the seasons


Editing Task Choice (ETC) Example
Click on each blank box and select the words or phrases to complete the sentence describing Earth's movement in space.
Earth is tilted on its and revolves around . This movement takes one
and causes
Click on each blank box and select the words or phrases to complete the sentence describing Earth's movement in space.
Earth is tilted on its and revolves around . This movement takes one
and causes
Mars
the moon
the sun


Hot Text (HT draggable) Example
A list of natural events is shown.
Click and drag the natural events to classify each natural event as either a fast or slow process that could shape and reshape Earth's surface.
Fast and Slow Processes
Fast Process
Slow Process
1. A glacier melts, depositing sediment.
2. A mountain side collapses, causing a landslide.
3. A tsunami pushes sediment inland.
4. An earthquake causes a crack along a road.
5. Waves carve an arch in a sea cliff.
6. Wind weathers a rock.


Hot Text (HT selectable) Example
A list of natural events that could shape and reshape Earth's surface is shown.
Click on **each** process below that happens slowly.
• A glacier melts, depositing sediment.
• A mountain side collapses, causing a landslide.
• A tsunami pushes sediment inland.
• An earthquake causes a crack along a road.
• Waves carve an arch in a sea cliff.
• Wind weathers a rock.


Machine Scored Constructed Response Interactions
• Machine Scored Constructed Response interactions require scoring logic or a machine rubric within the interaction. MSCR interaction types include:
• Equation Editor (EQ)
• Table Interaction (TI)
• Grid Interaction (GI)
• Simulation (Sim)
• Natural Language (NL)
• Editing Task (ET)
• Word Builder (WB)
These interactions are less accessible to all students!


Equation Editor (EQ) Example
Directions: Read the question and enter your answer in the box.
You are investigating the density of two samples of liquids.
Sample A
Sample B
How much more liquid, in milliliters, is in Sample B than in Sample A?
• Use the keypad to type your answer in the space provided.
Milliliters

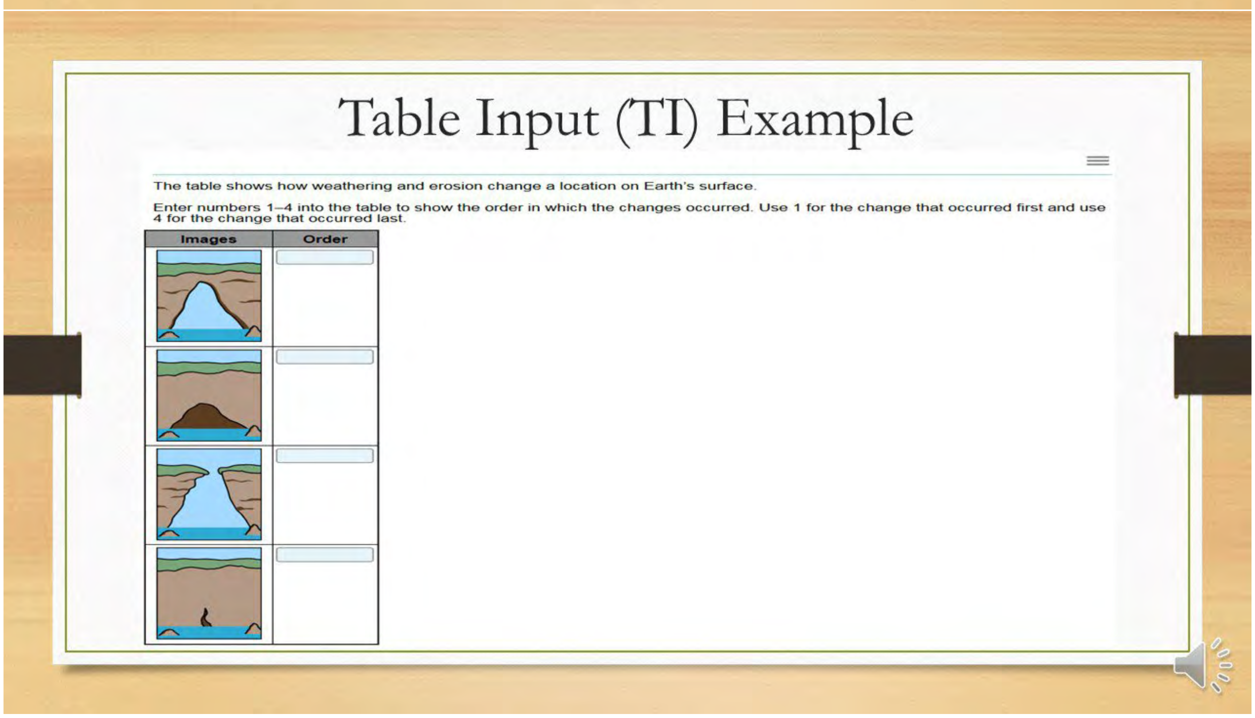
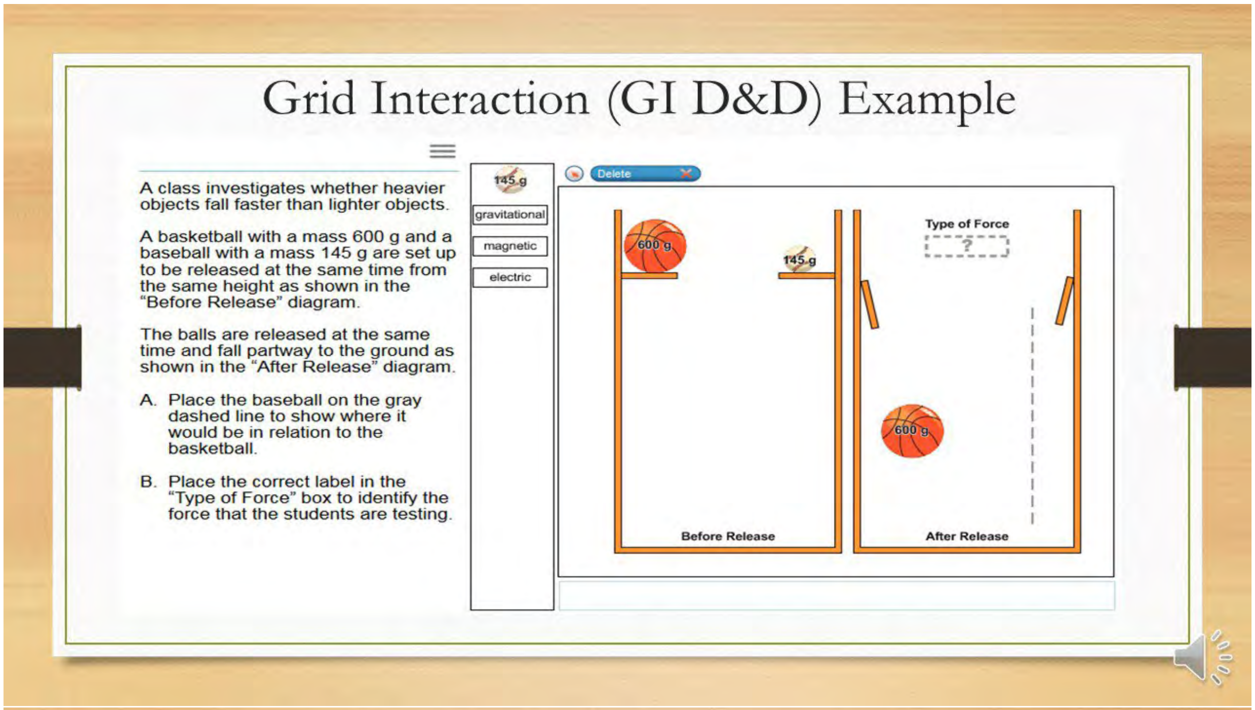
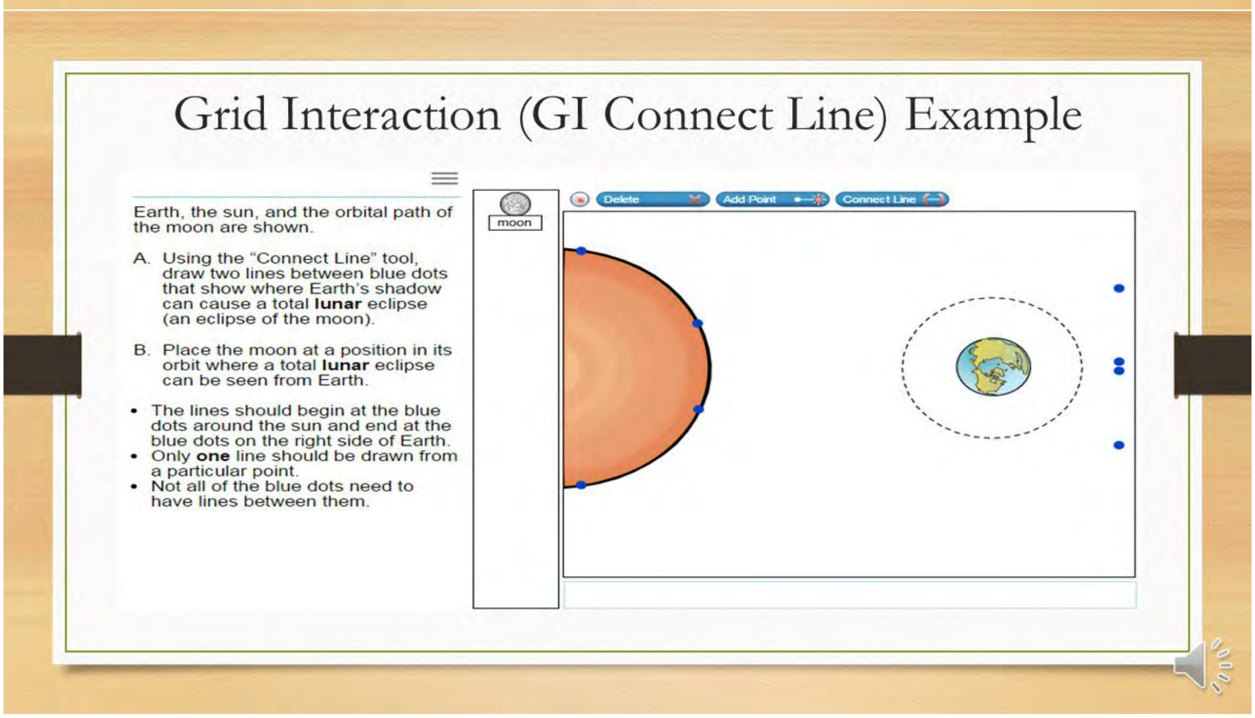
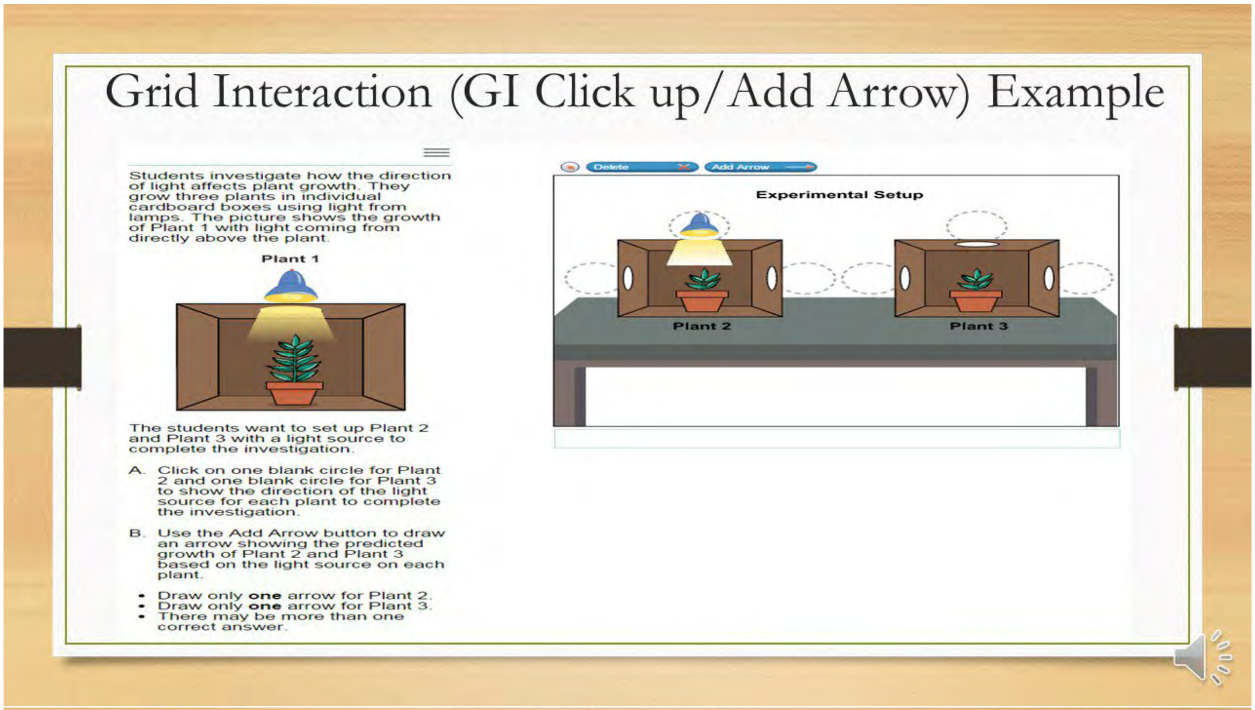
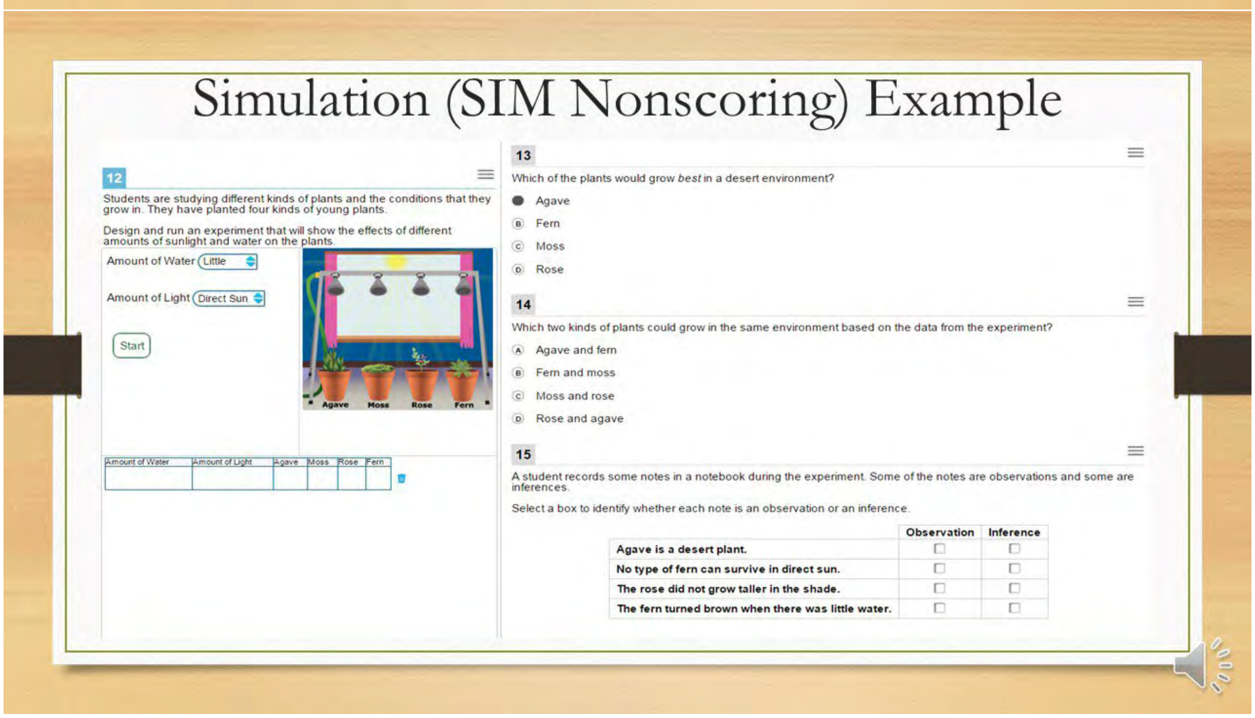


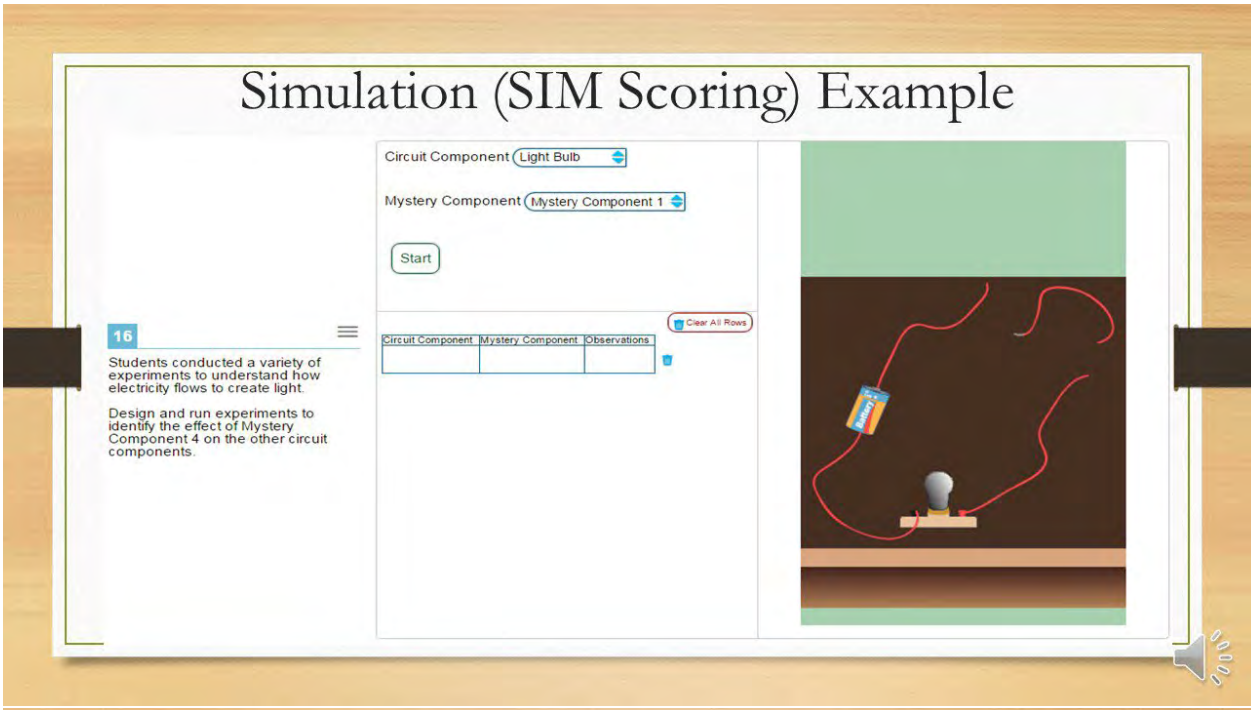
Table Input (TI) Example
The table shows how weathering and erosion change a location on Earth's surface.
Enter numbers 1–4 into the table to show the order in which the changes occurred. Use 1 for the change that occurred first and use 4 for the change that occurred last.
Images
Order

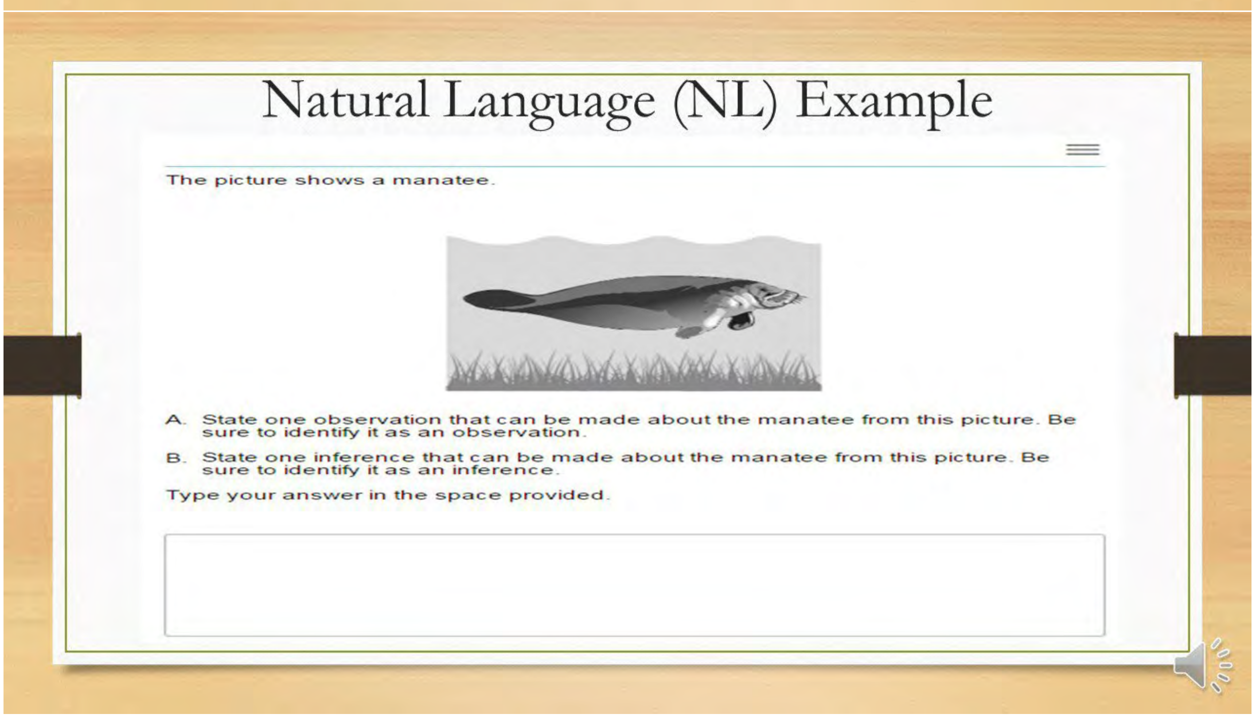

Grid Interaction (GI D&D) Example
A class investigates whether heavier objects fall faster than lighter objects.
A basketball with a mass 600 g and a baseball with a mass 145 g are set up to be released at the same time from the same height as shown in the "Before Release" diagram.
The balls are released at the same time and fall partway to the ground as shown in the "After Release" diagram.
A. Place the baseball on the gray dashed line to show where it would be in relation to the basketball.
B. Place the correct label in the "Type of Force" box to identify the force that the students are testing.
145 g
gravitational
magnetic
electric
Delete
600 g
145 g
Type of Force
?
600 g
Before Release
After Release

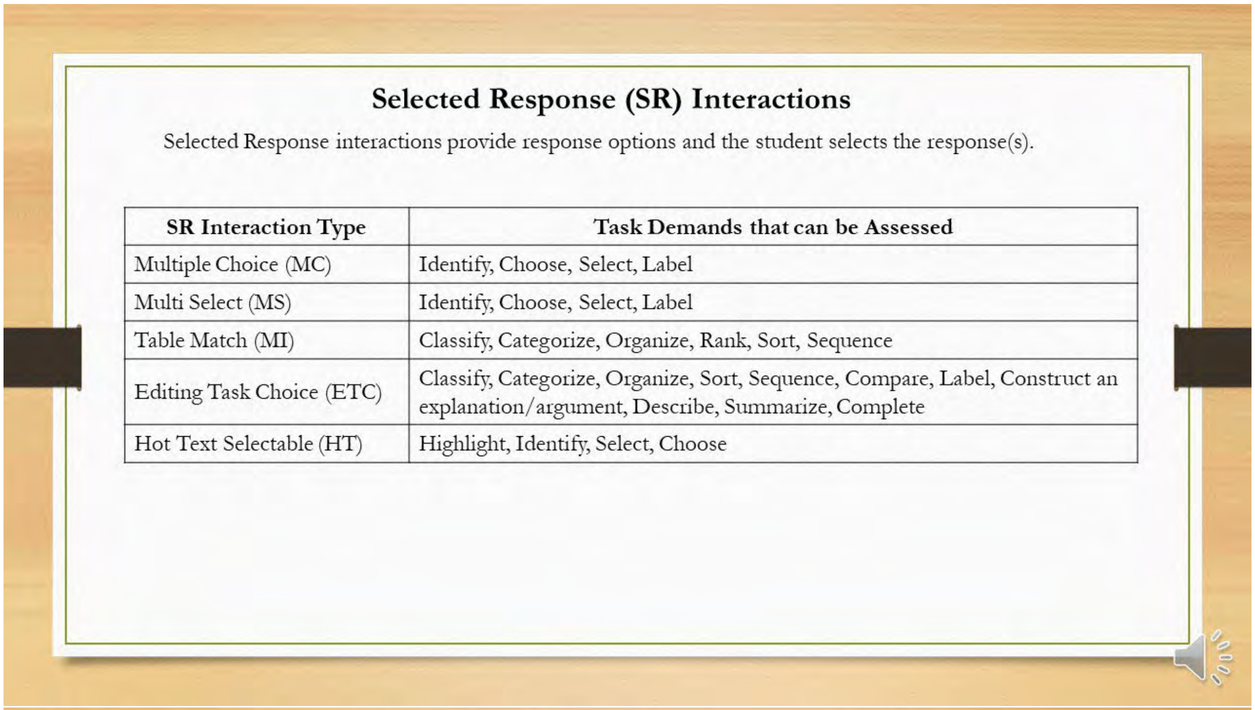

Grid Interaction (GI Connect Line) Example
Earth, the sun, and the orbital path of the moon are shown.
A. Using the "Connect Line" tool, draw two lines between blue dots that show where Earth's shadow can cause a total lunar eclipse (an eclipse of the moon).
B. Place the moon at a position in its orbit where a total lunar eclipse can be seen from Earth.
• The lines should begin at the blue dots around the sun and end at the blue dots on the right side of Earth.
• Only one line should be drawn from a particular point.
• Not all of the blue dots need to have lines between them.
moon
Delete
Add Point
Connect Line


Grid Interaction (GI Click up/Add Arrow) Example
Students investigate how the direction of light affects plant growth. They grow three plants in individual cardboard boxes using light from lamps. The picture shows the growth of Plant 1 with light coming from directly above the plant.
Plant 1
The students want to set up Plant 2 and Plant 3 with a light source to complete the investigation.
A. Click on one blank circle for Plant 2 and one blank circle for Plant 3 to show the direction of the light source for each plant to complete the investigation.
B. Use the Add Arrow button to draw an arrow showing the predicted growth of Plant 2 and Plant 3 based on the light source on each plant.
• Draw only one arrow for Plant 2.
• Draw only one arrow for Plant 3.
• There may be more than one correct answer.
Delete
Add Arrow
Experimental Setup
Plant 2
Plant 3


Simulation (SIM Nonscoring) Example
12
Students are studying different kinds of plants and the conditions that they grow in. They have planted four kinds of young plants.
Design and run an experiment that will show the effects of different amounts of sunlight and water on the plants.
Amount of Water Little
Amount of Light Direct Sun
Start
Agave Moss Rose Fern
Amount of Water Amount of Light Agave Moss Rose Fern
13
Which of the plants would grow best in a desert environment?
Agave
Fern
Moss
Rose
14
Which two kinds of plants could grow in the same environment based on the data from the experiment?
Agave and fern
Fern and moss
Moss and rose
Rose and agave
15
A student records some notes in a notebook during the experiment. Some of the notes are observations and some are inferences.
Select a box to identify whether each note is an observation or an inference.
Observation Inference
Agave is a desert plant.
No type of fern can survive in direct sun.
The rose did not grow taller in the shade.
The fern turned brown when there was little water.


Simulation (SIM Scoring) Example
16
Students conducted a variety of experiments to understand how electricity flows to create light.
Design and run experiments to identify the effect of Mystery Component 4 on the other circuit components.
Circuit Component
Light Bulb
Mystery Component
Mystery Component 1
Start
Clear All Rows
Circuit Component
Mystery Component
Observations
Battery

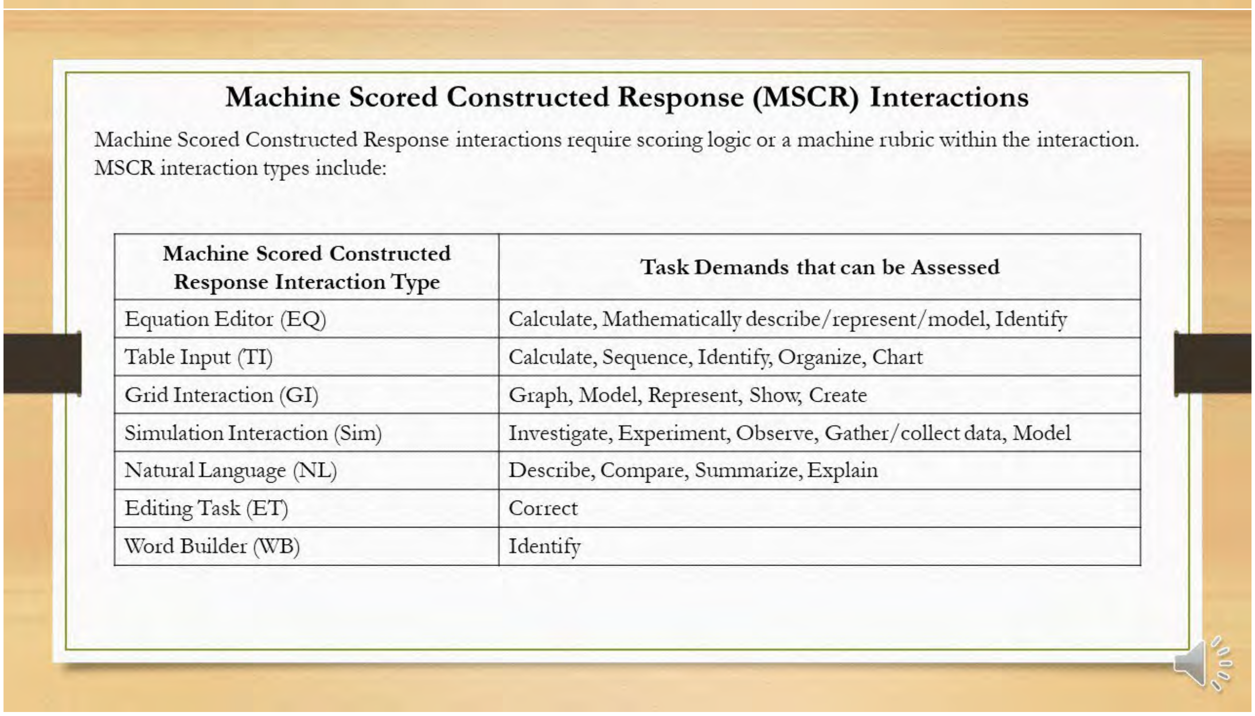

Natural Language (NL) Example
The picture shows a manatee.
A. State one observation that can be made about the manatee from this picture. Be sure to identify it as an observation.
B. State one inference that can be made about the manatee from this picture. Be sure to identify it as an inference.
Type your answer in the space provided.



Selected Response (SR) Interactions

Selected Response interactions provide response options and the student selects the response(s).

SR Interaction Type	Task Demands that can be Assessed
Multiple Choice (MC)	Identify, Choose, Select, Label
Multi Select (MS)	Identify, Choose, Select, Label
Table Match (MI)	Classify, Categorize, Organize, Rank, Sort, Sequence
Editing Task Choice (ETC)	Classify, Categorize, Organize, Sort, Sequence, Compare, Label, Construct an explanation/argument, Describe, Summarize, Complete
Hot Text Selectable (HT)	Highlight, Identify, Select, Choose



Machine Scored Constructed Response (MSCR) Interactions

Machine Scored Constructed Response interactions require scoring logic or a machine rubric within the interaction. MSCR interaction types include:

Machine Scored Constructed Response Interaction Type	Task Demands that can be Assessed
Equation Editor (EQ)	Calculate, Mathematically describe/represent/model, Identify
Table Input (TI)	Calculate, Sequence, Identify, Organize, Chart
Grid Interaction (GI)	Graph, Model, Represent, Show, Create
Simulation Interaction (Sim)	Investigate, Experiment, Observe, Gather/collect data, Model
Natural Language (NL)	Describe, Compare, Summarize, Explain
Editing Task (ET)	Correct
Word Builder (WB)	Identify

Appendix B

Item Review Checklist

Item Review Checklist

Tier 1 – The elements in this Tier are critical. If there are:

<u>Sufficiency/Appropriateness of the Phenomenon to assess the Performance Expectation</u>

1. Is the phenomenon based on a specific real-world scenario and focused enough to get the student to investigate what the Performance Expectation (PE) intends for them to investigate (i.e., the students' application of the Practice in the context of the Disciplinary Core Idea [DCI] and Crosscutting Concepts [CCC] as intended by the PE is sufficient to make sense of the phenomena)?
2. Is there an appropriate science-related activity that is puzzling and/or intriguing for students to engage in? Is the scenario focused on real-world observations that students can connect with or have direct experience with?
3. Is the context and complexity of the phenomenon grade-appropriate?
4. Cluster Task Statement: Does the 'call to action' reflect the end goal of the interactions to be answered? Does the statement make sense? Is this an engaging and reasonable outcome to work towards?
5. Is the phenomenon presented in way(s) that all students can access and comprehend it based on information provided (including text, graphics, data, images, animations, etc.)? Is the phenomenon free of cultural bias, insensitivity or depreciation of unsafe situations?

Tier 2 – Review of Specific Elements by Component

<u>Stimulus</u>

Reading Load/Readability/Style

1. Is the reading load appropriate for the grade (i.e., the amount of text minimized to reduce cognitive load)?
2. Is the language and vocabulary appropriate for the grade?
    a. Non-specific vocabulary should be one grade level lower than the tested grade.
    b. Science vocabulary should be part of the "Science Vocabulary Students Are Expected to Know" in the item specifications.
3. Is all of the information in the stimulus necessary for the student to complete the item interactions?
4. Is language consistent throughout the cluster (i.e., does not switch between steam and vapor)?
5. Is everything in active voice (i.e., avoids unnecessary and unclear passive construction)?

Measurement/Units

1. Are the data in SI units? Check style guide for exceptions.
2. Are units of measure introduced or defined before they are used in graphs/tables?
3. Are the dependent/independent variables on the correct axes or in the correct columns?
4. Are the graphs/tables/pictures free of extraneous information and appropriate for the grade level?

5. Is there information included on the graph/picture/table that is not necessary and can be removed?
6. Do the graphs/tables/pictures depend on color? Is there another way to represent the difference in the data other than by color (e.g., using patterns.)?

Data Source and Scientific Reference

1. Is content both accurate and appropriate in its context?
2. Are the data sources appropriate for the subject and grade and from reliable academic sources?
3. Does the item use the most up-to-date explanation?

Formatting

1. Is everything presented within the browser dimensions (1024x768) without horizontal scrolling?
2. Are the tables/graphs/etc. laid out in a way that is easy to read?
3. Are details and text in animations easy to see? Are labels in diagrams easy to read?
4. Is the average file size appropriate for test delivery (approximately 100KB, 250KB maximum)?

Item

Interaction and Alignment to Specifications

1. Does the item make sense if you are responding to the interactions as if you are the student in the intended grade-level?
2. Does the interaction require the student to demonstrate the science practice and/or content that the PE is assessing them on?
3. Are the interactions grade level/developmentally appropriate and do they follow a logical progression? Do the interactions use appropriate scaffolding to guide students in making sense of the phenomena?
4. Do the interactions align with the task demands?
5. Do the interactions avoid redundancy? Do the student interactions follow a coherent progression?
6. Do the student interactions follow a coherent progression? Does the order of the interactions allow students to make sense of the phenomenon or problem?
7. Is the item stem worded in a way that makes the intent of the interaction clear to the student?
8. Is it clear to the student what they will be scored on in the interaction?
9. Is the language (words and phrases) consistent throughout the stimulus and items?

Grade Appropriate

1. Is the content within the item accurate and grade appropriate?
2. Are the correct units used? Are the units grade appropriate? Where necessary, are the abbreviations of the units introduced?
3. Are the number of item parts/scoring assertions appropriate for the grade level?
4. Is the mathematics level appropriate for the grade being tested?

Formatting

1. Is everything presented within the browser frame without horizontal scrolling?
2. Are the tables/graphs/etc. easy to read? Are the images created in an appropriate color palette per the Style Guide?
3. Are details and text in animations easy to see?

Tier 3 – Review of the Scoring and Assertion(s)

Scoring Accuracy

1. Do the interactions/task provide clear guidance on how student responses will be scored/interpreted?
2. Are scores assigned appropriately as correct or incorrect?
3. Are the dependencies logical?
4. Are any of the scoring assertions exclusive (i.e., the student can only get one assertion correct and not another at any given time)?
5. Is the correct answer clear and distinct from the distractors?
6. Does the scoring result in an appropriate distribution of points?

Scoring Assertions

1. Is the appropriate wording used for each scoring assertion (e.g., <What the student did as a response> provides evidence of an understanding of/ability to <inference about student's ability relative to the PE being measured>)?
2. Does the inference follow from the data?
3. Are the assertions specific to the individual interactions (i.e., does not just repeat the PE)?
4. Are the scoring assertions in the same order as the interactions?
5. Does the wording of the scoring assertion make it very clear which interaction and action it refers to?

Strategies for Editing Text to Produce Plain Language

- Reduce excessive length.
- Use common words.
- Avoid ambiguous words.
- Limit irregularly spelled words.
- Avoid inconsistent naming and graphic conventions.
- Avoid multiple terms for the same concept.
- Limit the use of embedded clauses and phrases.
- Avoid the passive voice.

Appendix C

Content Advisory Committee Participant Details

Content Advisory Committee Participant Details

Table C-1. Content Advisory Committee Participants, Science

State/Item Bank	Date	Location	Number of Committee Members	Committee Member Demographic Summary	Number of Items Reviewed	Number of Items Approved by Committees
Connecticut	February 2017	Cromwell, Connecticut	41	**Gender:** 78% Female, 22% Male	45	31
	May 2017	New Britain, Connecticut	42	**Gender:** 74% Female, 26% Male	40	38
	October 2017	New Britain, Connecticut	41	**Gender:** 80% Female, 20% Male	75	64
	November 2017	New Britain, Connecticut	35	**Gender:** 83% Female, 17% Male	41	32
	January/ February 2018	New Britain, Connecticut	33	**Gender:** 82% Female, 18% Male	42	25
	October 2018	New Britain, Connecticut	45	**Gender:** 84% Female, 16% Male	84	54
	November 2018	New Britain, Connecticut	49	**Gender:** 86% Female, 14% Male	235	200
	December 2018	New Britain, Connecticut	32	**Gender:** 81% Female, 19% Male	56	55
	January 2019	New Britain, Connecticut	44	**Gender:** 82% Female, 18% Male	65	59
	September 2019	Rocky Hill, Connecticut	50	**Gender:** 82% Female, 18% Male	60	57
	November 2019	Cromwell, Connecticut	44	**Gender:** 80% Female, 20% Male **Ethnicity:** 5% Hispanic, 93% White, 2% Preferred Not to Answer **Region:** 14% Rural, 59% Suburban, 16% Urban, 11% Not Applicable **Teaching Experience:** 2% None, 9% 1 to 5 years, 9% 6 to 10 years, 30% 11 to 15 years, 25% 16 to 20 years, 25% More than 20 years	171	153
	January 2020	Cromwell, Connecticut	57	**Gender:** 75% Female, 25% Male **Ethnicity:** 5% Black or African American, 2% Franco-American, 5% Hispanic, 88% White	190	161

State/Item Bank	Date	Location	Number of Committee Members	Committee Member Demographic Summary	Number of Items Reviewed	Number of Items Approved by Committees
				Region: 14% Rural, 63% Suburban, 19% Urban, 4% Not Applicable **Teaching Experience:** 12% 1 to 5 years, 14% 6 to 10 years, 25% 11 to 15 years, 21% 16 to 20 years, 28% More than 20 years		
	July 2020[e]	Virtual	23[e]	**Gender:** 83% Female, 17% Male **Ethnicity:** 4% Black or African American, 91% White, 4% Prefer Not to Say	48	44
	July 2021[e]	Virtual	68[e]	**State:** 22% Connecticut, 4% Hawaii, 9% Idaho, 4% Montana, 4% Oregon, 4% Rhode Island, 13% South Dakota, 4% Utah, 3% Vermont, 12% West Virginia, 13% Wyoming, 4% Did Not Respond **Teaching Experience:** 9% 1 to 5 years, 18% 6 to 10 years, 29% 11 to 15 years, 19% 16 to 20 years, 25% More than 20 years	26	26
	September 2021	Virtual	27	**Gender:** 74% Female, 26% Male **Ethnicity:** 4% Black or African American, 96% White or Caucasian **Region:** 41% Suburban, 11% Urban, 7% Not Applicable, 41% Did Not Respond **Teaching Experience:** 7% 1 to 5 years, 15% 6 to 10 years, 15% 11 to 15 years, 15% 16 to 20 years, 48% More than 20 years	149	120
Hawaii	July 2017	Honolulu, Hawaii	22	**Gender:** 64% Female, 36% Male **Ethnicity:** 5% Black, 5% Chinese and White, 9% Filipino, 14% Hawaiian, 9% Hispanic, 14% Japanese, 41% White, 5% Did Not Respond **Teaching Experience:** 64% General Education, 5% General Education with SPED Certification, 5% SPED Teacher, 23% Other, 5% Did Not Respond	25	[d]
	September 2017	Honolulu, Hawaii	20	**Gender:** 75% Female, 25% Male **Ethnicity:** 5% Black, 10% Filipino, 10% Hispanic, 15% Japanese, 50% White, 10% Did Not Respond **Teaching Experience:** 65% General Education, 15% General Education with SPED Certification, 20% Other	65	[d]
	October 2018	Honolulu, Hawaii	28	**Gender:** 83% Female, 17% Male	85	79

State/Item Bank	Date	Location	Number of Committee Members	Committee Member Demographic Summary	Number of Items Reviewed	Number of Items Approved by Committees
				Ethnicity: 31% Asian, 7% Asian Pacific Islander, 3% Hawaiian, 10% Hispanic, 10% Not Applicable, 10% Two or More, 28% White **Teaching Experience:** 83% General Education, 24% Other		
	February/ March 2019	Honolulu, Hawaii	21	**Gender:** 80% Female, 20% Male **Ethnicity:** 50% Asian, 35% White, 15% Two or More **Teaching Experience:** 65% General Education, 5% General Education with SPED Certification, 5% SPED Teacher, 25% Other	44	44
	June/ July 2020	Virtual	17	**Gender:** 18% Female, 12% Male, 70% Did Not Respond **Ethnicity**: 18% White or Caucasian, 6% Asian or Pacific Islander, 6% Multiracial or Biracial, 70% Did Not Respond **Region:** 12% Rural, 12% Suburban, 76% Did Not Respond **Teaching Experience:** 6% 6 to 10 years, 12% 11 to 15 years, 12% More than 20 years, 70% Did Not Respond	344	324
	July 2020[e]	Virtual	28[e]	**State:** 14% Connecticut, 4% Hawaii, 14% Idaho, 14% Montana, 7% Oregon, 4% Rhode Island, 4% Utah, 7% Vermont, 11% West Virginia, 7% Wyoming, 14% Did Not Respond **Gender:** 86% Female, 14% Male **Ethnicity:** 46% White or Caucasian, 4% Black or African American, 50% Did Not Respond **Region:** 7% Rural, 14% Suburban, 14% Urban, 64% Did Not Respond **Teaching Experience:** 3% 6 to 10 years, 11% 11 to 15 years, 14% 16 to 20 years, 18% More than 20 years, 54% Did Not Respond	90	90
ICCR	March 2018	Virtual	38	**State:** 45% Connecticut, 5% Hawaii, 3% Indiana, 3% Maryland, 8% Oregon, 8% Utah, 26% West Virginia, 3% Wyoming **Gender:** 74% Female, 26% Male	152	[d]
	July 2020[e]	Virtual	6[e]	**State:** 17% Connecticut, 17% Idaho, 17% Oregon, 17% Rhode Island, 33% Did Not Respond **Gender:** 83% Female, 17% Male **Ethnicity**: 33% White or Caucasian, 67% Did Not Respond **Region:** 17% Suburban, 83% Did Not Respond	57	56

State/Item Bank	Date	Location	Number of Committee Members	Committee Member Demographic Summary	Number of Items Reviewed	Number of Items Approved by Committees
				Teaching Experience: 33% 16 to 20 years, 67% Did Not Respond		
	July 2021[e]	Virtual	68[e]	**State:** 22% Connecticut, 4% Hawaii, 9% Idaho, 4% Montana, 4% Oregon, 4% Rhode Island, 13% South Dakota, 4% Utah, 3% Vermont, 12% West Virginia, 13% Wyoming, 4% Did Not Respond **Teaching Experience:** 9% 1 to 5 years, 18% 6 to 10 years, 29% 11 to 15 years, 19% 16 to 20 years, 25% More than 20 years	141	141
	July 2021	Virtual	45	**State:** 33% Connecticut, 9% Hawaii, 4% Idaho, 2% Montana, 7% North Dakota, 4% Oregon, 18% South Dakota, 2% Vermont, 4% West Virginia, 13% Wyoming **Gender:** 80% Female, 18% Male, 2% Did Not Respond **Ethnicity:** 4% Asian or Pacific Islander, 2% Black or African American, 4% Hispanic or Latino, 87% White or Caucasian, 2% Did Not Respond **Region:** 36% Rural, 24% Suburban, 20% Urban, 20% Did Not Respond **Teaching Experience:** 2% None, 2% Less than 1 year, 11% 1 to 5 years, 33% 6 to 10 years, 16% 11 to 15 years, 9% 16 to 20 years, 24% More than 20 years, 2% Did Not Respond	163	158
Idaho	December 2018	Boise, Idaho	21	Not Collected	241	230
	October 2019	Boise, Idaho	18	**Gender:** 83% Female, 11% Male, 6% Did Not Respond **Ethnicity:** 100% White **Region:** 50% Rural, 17% Suburban, 22% Urban, 11% Not Applicable **Teaching Experience:** 11% 1 to 5 years, 22% 6 to 10 years, 17% 11 to 15 years, 11% 16 to 20 years, 28% 21+ years, 11% Did Not Respond	231	211
	July 2020[e]	Virtual	2[e]	**State:** 100% Hawaii **Gender:** 100% Female	12	12
	October 2020	[a]	[b]	Not Collected	14	14
	July 2021[e]	Virtual	8[e]	**Gender:** 88% Female, 13% Male **Ethnicity:** 100% White or Caucasian **Region:** 25% Rural, 25% Suburban, 50% Did Not Respond	[d]	[d]

State/Item Bank	Date	Location	Number of Committee Members	Committee Member Demographic Summary	Number of Items Reviewed	Number of Items Approved by Committees
				Teaching Experience: 38% 6 to 10 years, 38% 11 to 15 years, 13% 16 to 20 years, 13% More than 20 years		
	November 2021	Virtual	11	**Gender:** 91% Female, 9% Male **Ethnicity**: 73% White or Caucasian, 17% Did Not Respond **Region:** 18% Rural, 18% Suburban, 64% Did Not Respond **Teaching Experience:** 9% 1 to 5 years, 46% 6 to 10 years, 9% 11 to 15 years, 9% More than 20 years, 27% Did Not Respond	317	286
Montana	January 2020	Helena, Montana	15	Not Collected	149	139
	July 2020[e]	Virtual	4[e]	**State:** 25% Hawaii, 25% Idaho, 25% Oregon, 25% Rhode Island **Gender:** 75% Female, 25% Male **Ethnicity**: 50% White or Caucasian, 50% Did Not Respond **Region:** 50% Urban, 50% Did Not Respond **Teaching Experience:** 50% More than 20 years, 50% Did Not Respond	9	9
	October 2020	Virtual	8	**Gender:** 13% Female, 88% Did Not Respond **Ethnicity**:13% White or Caucasian, 88% Did Not Respond **Region:** 13% Rural, 88% Did Not Respond **Teaching Experience:** 13% 16 to 20 years, 88% Did Not Respond	156	140
	July 2021[e]	Virtual	68[e]	**State:** 22% Connecticut, 4% Hawaii, 9% Idaho, 4% Montana, 4% Oregon, 4% Rhode Island, 13% South Dakota, 4% Utah, 3% Vermont, 12% West Virginia, 13% Wyoming, 4% Did Not Respond **Teaching Experience:** 9% 1 to 5 years, 18% 6 to 10 years, 29% 11 to 15 years, 19% 16 to 20 years, 25% More than 20 years	36	36
	October 2021	Virtual	6	**Gender:** 83% Female, 17% Did Not Respond **Ethnicity**: 67% White or Caucasian, 17% Hispanic or Latino, 17% Did Not Respond **Region:** 67% Rural, 33% Did Not Respond **Teaching Experience:** 33% 6 to 10 years, 33% 11 to 15 years, 17% More than 20 years, 17% Did Not Respond	41	39
Multi-State Science	January 2018	Providence, Rhode Island	42	**State:** 90% Rhode Island, 10% Vermont	73	58

State/Item Bank	Date	Location	Number of Committee Members	Committee Member Demographic Summary	Number of Items Reviewed	Number of Items Approved by Committees
Assessment (Rhode Island and Vermont)				**Teaching Experience:** 69% General Education, 2% Bilingual Education, 14% Science Coordinator, 14% Other		
	March 2018	Providence, Rhode Island	34	**State:** 25% Rhode Island, 75% Vermont	107	90
	January 2019	Concord, New Hampshire	21	**Gender:** 74% Female, 26% Male **Teaching Experience:** 69% General Education, 3% Special Education, 29% Other, 6% Not Applicable	116	97
	November 2019	Fairlee, Vermont	17	**State:** 29% Rhode Island, 6% Vermont, 65% Did Not Respond **Gender:** 23% Female, 12% Male, 65% Did Not Respond **Ethnicity**: 35% White or Caucasian, 65% Did Not Respond **Region:** 6% Rural, 17% Suburban, 77% Did Not Respond **Teaching Experience:** 6% 11 to 15 years, 17% 16 to 20 years, 12% More than 20 years, 65% Did Not Respond	136	118
	July 2020[e]	Virtual	8[e]	**State:** 12.5% Connecticut, 12.5% Hawaii, 12.5% Montana, 12.5% Oregon, 25% Oregon, 25% Did Not Respond **Gender:** 87.5% Female, 12.5% Male **Ethnicity**: 37.5% White or Caucasian, 62.5% Did Not Respond **Region:** 12.5% Suburban, 87.5% Did Not Respond **Teaching Experience:** 12.5% 6 to 10 years, 12.5% 11 to 15 years, 12.5% More than 20 years, 62.5% Did Not Respond	27	27
	July 2021[e]	Virtual	68[e]	**State:** 22% Connecticut, 4% Hawaii, 9% Idaho, 4% Montana, 4% Oregon, 4% Rhode Island, 13% South Dakota, 4% Utah, 3% Vermont, 12% West Virginia, 13% Wyoming, 4% Did Not Respond **Teaching Experience:** 9% 1 to 5 years, 18% 6 to 10 years, 29% 11 to 15 years, 19% 16 to 20 years, 25% More than 20 years	32	31
	August 2021	Virtual	11	**State:** 45% Rhode Island, 55% Vermont **Gender:** 73% Female, 27% Male **Ethnicity:** 100% White or Caucasian **Region:** 36% Rural, 18% Suburban, 9% Urban, 36% Did Not Respond **Teaching Experience:** 27% 6 to 10 years, 18% 11 to 15 years, 27% 16 to 20 years, 27% More than 20 years	93	91

State/Item Bank	Date	Location	Number of Committee Members	Committee Member Demographic Summary	Number of Items Reviewed	Number of Items Approved by Committees
Oregon	August 2017	Salem, Oregon	10	**Gender:** 90% Female, 10% Male **Region:** 50% Urban, 50% Rural **Teaching Experience:** 100% General Education, 10% Bilingual Education, 10% Special Education, 20% Administration	235	142
	August 2018	Salem, Oregon	20	**Gender:** 80% Female, 20% Male **Ethnicity:** 95% White, 5% Other **Region:** 56% Urban, 44% Rural **Teaching Experience:** 65% Bilingual Education, 65% Special Education, 55% Other	257	200
	October 2018	Salem, Oregon	11	**Gender:** 100% Female **Ethnicity:** 91% White, 9% Other **Region:** 55% Urban, 45% Rural **Teaching Experience:** 18% General Education, 91% Bilingual Education, 45% Special Education, 55% Other	60	30
	December 2018	Virtual	16	**Gender:** 63% Female, 38% Male **Ethnicity:** 6% Asian, 94% White **Region:** 50% Urban, 50% Suburban **Teaching Experience:** 38% General Education, 63% Bilingual Education, 25% Special Education	62	48
	October 2019	Salem, Oregon	17	**Gender:** 76% Female, 24% Male **Ethnicity:** 88% White, 6% Asian, 6% Other **Region:** 71% Urban, 29% Rural **Teaching Experience:** 82% General Education, 29% Bilingual Education, 18% Special Education	255	221
	July 2020[e]	Virtual	9[e]	**State:** 22% Idaho, 11% Vermont, 22% West Virginia, 11% Wyoming, 33% Did Not Respond **Gender:** 78% Female, 22% Male **Ethnicity**: 44% White or Caucasian, 56% Did Not Respond **Region:** 11% Rural, 11% Suburban, 78% Did Not Respond **Teaching Experience:** 11% 11 to 15 years, 11% 16 to 20 years, 22% More than 20 years, 56% Did Not Respond	22	20
	August 2020	Virtual	21	**Gender:** 71% Female, 29% Male **Ethnicity:** 90% White, 5% Hispanic, 5% Native American **Region:** 5% Urban, 43% Suburban, 52% Rural	159	134

State/Item Bank	Date	Location	Number of Committee Members	Committee Member Demographic Summary	Number of Items Reviewed	Number of Items Approved by Committees
				Teaching Experience: 86% General Education, 81% Bilingual Education, 81% Special Education, 14% Administration, 5% Other		
	August 2021	Virtual	14	**Gender:** 86% Female, 14% Male **Ethnicity:** 86% White, 7% Asian and/or Pacific Islander, 7% Hispanic **Region:** 14% Urban, 72% Suburban, 14% Rural **Teaching Experience:** 64% General Education, 7% Bilingual Education, 7% Special Education, 22% Other	375	308
South Dakota	October 2019	Pierre, South Dakota	26	**Gender:** 81% Female, 19% Male **Ethnicity:** 4% American Indian or Alaska Native, 4% Asian, 92% White **Region:** 65% Rural, 15% Suburban, 15% Urban, 4% Not Applicable **Teaching Experience:** 12% 1 to 5 years, 12% 6 to 10 years, 19% 11 to 15 years, 19% 16 to 20 years, 38% More than 20 years	235	222
Utah	July 2017	Park City, Utah	18	**Gender:** 74% Female, 26% Male **Ethnicity:** 4% Native American, 91% White, 4% Other **Teaching Experience:** 100% General Education, 4% Special Education, 4% Other	55	51
	December 2017	Salt Lake City, Utah	36	**Gender:** 84% Female, 16% Male **Ethnicity:** 3% American Indian/Alaska Native and White, 94% White, 3% Other **Teaching Experience:** 87% General Education, 10% General Education and Other, 3% General Education and ESOL	64	62
	October 2019	Provo, Utah	16	**Gender:** 25% Female, 75% Did Not Respond **Ethnicity**: 25% White or Caucasian, 75% Did Not Respond **Region:** 25% Suburban, 75% Did Not Respond **Teaching Experience:** 6% 6 to 10 years, 6% 16 to 20 years, 13% More than 20 years, 75% Did Not Respond	91	44
	July 2020[e]	Virtual	17[e]	**State:** 6% Connecticut, 12% Hawaii, 18% Idaho, 12% Montana, 12% Oregon, 12% Rhode Island, 6% Vermont, 6% West Virginia, 12% Wyoming, 6% Did Not Respond **Gender:** 82% Female, 18% Male **Ethnicity**: 47% White or Caucasian, 6% Other, 47% Did Not Respond	44	44

State/Item Bank	Date	Location	Number of Committee Members	Committee Member Demographic Summary	Number of Items Reviewed	Number of Items Approved by Committees
				Region: 6% Rural, 12% Suburban, 6% Urban, 76% Did Not Respond **Teaching Experience:** 12% 6 to 10 years, 6% 11 to 15, 18% 16 to 20 years, 18% More than 20 years, 47% Did Not Respond		
	July 2020	Virtual	16	**Gender:** 31% Female, 6% Male, 63% Did Not Respond **Ethnicity**: 12.5% White or Caucasian, 6% Asian or Pacific Islander, 19% Hispanic or Latino, 62.5% Did Not Respond **Region:** 19% Urban, 6% Suburban, 75% Did Not Respond **Teaching Experience:** 6% 6 to 10 years, 12.5% 11 to 15 years, 12.5% 16 to 20 years, 6% More than 20 years, 63% Did Not Respond	82	76
	December 2020	Virtual	6	**Gender:** 50% Female, 50% Did Not Respond **Ethnicity**: 33% White or Caucasian, 17% Hispanic or Latino, 50% Did Not Respond **Region:** 17% Suburban, 83% Did Not Respond **Teaching Experience:** 17% 1 to 5 years, 33% 16 to 20 years, 50% Did Not Respond	[d]	[d]
	July 2021[e]	Virtual	68[e]	**State:** 22% Connecticut, 4% Hawaii, 9% Idaho, 4% Montana, 4% Oregon, 4% Rhode Island, 13% South Dakota, 4% Utah, 3% Vermont, 12% West Virginia, 13% Wyoming, 4% Did Not Respond **Teaching Experience:** 9% 1 to 5 years, 18% 6 to 10 years, 29% 11 to 15 years, 19% 16 to 20 years, 25% More than 20 years	55	53
	August 2021	Virtual	14	**Gender:** 86% Female, 14% Male **Ethnicity:** 7% Asian or Pacific Islander, 21% Hispanic or Latino, 71% White or Caucasian **Region:** 14% Rural, 36% Suburban, 43% Urban, 7% Did Not Respond **Teaching Experience:** 7% 1 to 5 years, 21% 6 to 10 years, 21% 11 to 15 years, 29% 16 to 20 years, 21% More than 20 years	62	62
West Virginia	January 2017	Charleston, West Virginia	28[c]	Not Collected	39	[d]
	October 2018	Charleston, West Virginia	10	**Gender:** 89% Female, 11% Male **Ethnicity:** 11% Black, 89% White **Region:** 100% Rural	191	[d]

State/Item Bank	Date	Location	Number of Committee Members	Committee Member Demographic Summary	Number of Items Reviewed	Number of Items Approved by Committees
				Teaching Experience: 100% General Education		
	January 2019	Charleston, West Virginia	9	**Gender:** 89% Female, 11% Male **Ethnicity:** 11% Black, 89% White **Region:** 100% Rural	71	67
	July 2019	Charleston, West Virginia	12	**Gender:** 87% Female, 13% Male **Ethnicity:** 4% Asian, 4% Black, 87% White, 4% Not Applicable **Region:** 70% Rural, 30% Urban, 4% Not Applicable **Teaching Experience:** 72% General Education, 4% Special Education, 13% Other, 13% Not Applicable	50	[d]
	July 2020[e]	Virtual	8[e]	**State:** 12.5% Connecticut, 37.5% Idaho, 12.5% Oregon, 12.5% Wyoming, 25% Did Not Respond **Gender:** 100% Female **Ethnicity**: 37.5% White or Caucasian, 62.5% Did Not Respond **Region:** 12.5% Suburban, 12.5% Rural, 75% Did Not Respond **Teaching Experience:** 12.5% 6 to 10 years, 25% More than 20 years, 62.5% Did Not Respond	102	102
	July 2021[e]	Virtual	68[e]	**State:** 22% Connecticut, 4% Hawaii, 9% Idaho, 4% Montana, 4% Oregon, 4% Rhode Island, 13% South Dakota, 4% Utah, 3% Vermont, 12% West Virginia, 13% Wyoming, 4% Did Not Respond **Teaching Experience:** 9% 1 to 5 years, 18% 6 to 10 years, 29% 11 to 15 years, 19% 16 to 20 years, 25% More than 20 years	16	16
Wyoming	December 2017	Cheyenne, Wyoming	18	Not Collected	32	30
	October 2018	Cheyenne, Wyoming	19	**Gender:** 79% Female, 21% Male **Teaching Experience:** 5% 3 to 5 years, 21% 6 to 10 years, 42% 11 to 20 years, 32% 21+ years	39	36
	November 2019	Cheyenne, Wyoming	22	**Gender:** 91% Female, 9% Male **Teaching Experience:** 9% 3 to 5 years, 23% 6 to 10 years, 18% 11 to 20 years, 50% 21+ years	44	43

State/Item Bank	Date	Location	Number of Committee Members	Committee Member Demographic Summary	Number of Items Reviewed	Number of Items Approved by Committees
	July 2020[e]	Virtual	13[e]	**State:** 8% Connecticut, 15% Hawaii, 8% Montana, 15% Oregon, 8% Rhode Island, 23% West Virginia, 23% Did Not Respond **Gender:** 77% Female, 23% Male **Ethnicity**: 23% White or Caucasian, 8% Asian or Pacific Islander, 8% Other, 61% Did Not Respond **Region:** 8% Suburban, 15% Urban, 77% Did Not Respond **Teaching Experience:** 15.4% 6 to 10 years, 23% 11 to 15 years, 15.4% 16 to 20 years, 46.2% Did Not Respond	37	37
	August 2020	Virtual	14	**Gender:** 29% Female, 7% Male, 64% Did Not Respond **Ethnicity**: 36% White or Caucasian, 64% Did Not Respond **Region:** 22% Rural, 78% Did Not Respond **Teaching Experience:** 7% 11 to 15 years, 7% 16 to 20 years, 22% More than 20 years, 64% Did Not Respond	37	36
	June/ July 2021	Virtual	14	**Gender:** 43% Female, 7% Male, 50% Did Not Respond **Ethnicity:** 50% White or Caucasian, 50% Did Not Respond **Region:** 14% Rural,7% Suburban, 7% Urban, 71% Did Not Respond **Teaching Experience:** 14% 11 to 15 years, 36% More than 20 years, 50% Did Not Respond	39	39
	July 2021[e]	Virtual	68[e]	**State:** 22% Connecticut, 4% Hawaii, 9% Idaho, 4% Montana, 4% Oregon, 4% Rhode Island, 13% South Dakota, 4% Utah, 3% Vermont, 12% West Virginia, 13% Wyoming, 4% Did Not Respond **Teaching Experience:** 9% 1 to 5 years, 18% 6 to 10 years, 29% 11 to 15 years, 19% 16 to 20 years, 25% More than 20 years	39	38

Note. [a]Location of Content Advisory Committee Meeting is unavailable at the time of writing this report.
[b]Number of Committee Members is not available at the time of writing this report.
[c]The number of Committee Members includes total committee members for ELA, math, and science. The number for science only committee members is not available.
[d]Number of science items reviewed and/or approved by Content Advisory Committee is unavailable at the time of writing this report.
[e]Items were reviewed in a combined Content Advisory Committee Meeting that included all MOU state items.

Appendix D

Fairness Committee Participant Details

Fairness Committee Participant Details

Table D-1. Fairness Committee Participants, Science

State/Item Bank	Date	Location	Number of Committee Members	Committee Member Demographic Summary	Number of Items Reviewed	Number of Items Rejected by Committees
Connecticut	February 2017	Cromwell, Connecticut	6	**Gender:** 83% Female, 17% Male	45	1
	December 2017	New Britain, Connecticut	9	**Gender:** 78% Female, 22% Male	75	[c]
	December 2017	Cromwell, Connecticut	10	**Gender:** 70% Female, 30% Male	41	[c]
	February 2018	New Britain, Connecticut	3	**Gender:** 67% Female, 33% Male	42	1
	November 2018	New Britain, Connecticut	11	**Gender:** 91% Female, 9% Male	319	38
	December 2018	New Britain, Connecticut	10	**Gender:** 80% Female, 20% Male	56	[b]
	January 2019	New Britain, Connecticut	9	**Gender:** 78% Female, 22% Male	65	1
	September 2019	Cromwell, Connecticut	9	**Gender:** 89% Female, 11% Male	48	0
	November 2019	Cromwell, Connecticut	10	**Gender:** 80% Female, 20% Male **Ethnicity:** 100% White **Region:** 10% Rural, 70% Suburban, 20% Urban **Teaching Experience:** 10% 6 to 10 years, 20% 11 to 15 years, 10% 16 to 20 years, 60% More than 20 years	52	1
	July 2020[c]	Virtual	8[c]	**Gender:** 88% Female, 13% Male **Ethnicity:** 13% Hispanic, 75% White, 13% Prefer Not to Say	43	0
	July 2021[c]	Virtual	6[c]	**State:** 22% Connecticut, 4% Hawaii, 9% Idaho, 4% Montana, 4% Oregon, 4% Rhode Island, 13% South Dakota, 4% Utah, 3% Vermont, 12% West Virginia, 13% Wyoming, 4% Did Not Respond **Teaching Experience:** 9% 1 to 5 years, 18% 6 to 10 years, 29% 11 to 15 years, 19% 16 to 20 years, 25% More than 20 years	20	0

State/Item Bank	Date	Location	Number of Committee Members	Committee Member Demographic Summary	Number of Items Reviewed	Number of Items Rejected by Committees
	September 2021	Virtual	7	**Gender:** 43% Female, 57% Male **Ethnicity:** 100% White or Caucasian **Region:** 29% Suburban, 29% Urban, 43% Did Not Respond **Teaching Experience:** 14% 6 to 10 years, 29% 11 to 15 years, 14% 16 to 20 years, 43% More than 20 years	111	23
Hawaii	July 2017	Honolulu, Hawaii	22	**Gender:** 64% Female, 36% Male **Ethnicity:** 5% Black, 5% Chinese and White, 9% Filipino, 14% Hawaiian, 9% Hispanic, 14% Japanese, 41% White, 5% Did Not Respond **Teaching Experience:** 64% General Education, 5% General Education with SPED Certification, 5% SPED Teacher, 23% Other, 5% Did Not Respond	25	2
	September 2017	Honolulu, Hawaii	20	**Gender:** 75% Female, 25% Male **Ethnicity:** 5% Black, 10% Filipino, 10% Hispanic, 15% Japanese, 50% White, 10% Did Not Respond **Teaching Experience:** 65% General Education, 15% General Education with SPED Certification, 20% Other	65	13
	October 2018	Honolulu, Hawaii	29	**Gender:** 79% Female, 21% Male **Ethnicity:** 7% Asian, 3% Hawaiian, 7% Asian Pacific Islander, 7% Chinese, 3% Filipino, 10% Hispanic, 10% Japanese, 28% White, 14% Multi-Racial/Ethnic, 10% Not Applicable	85	6
	February/ March 2019	Honolulu, Hawaii	21	**Gender:** 80% Female, 20% Male **Ethnicity:** 50% Asian, 35% White, 15% Two or More **Teaching Experience:** 65% General Education, 5% General Education with SPED Certification, 5% SPED Teacher, 25% Other	44	0
	June/ July 2020	Virtual	17	**Gender:** 18% Female, 12% Male, 70% Did Not Respond **Ethnicity**: 18% White or Caucasian, 6% Asian or Pacific Islander, 6% Multiracial or Biracial, 70% Did Not Respond	344	324

State/Item Bank	Date	Location	Number of Committee Members	Committee Member Demographic Summary	Number of Items Reviewed	Number of Items Rejected by Committees
				Region: 12% Rural, 12% Suburban, 76% Did Not Respond **Teaching Experience:** 6% 6 to 10 years, 12% 11 to 15 years, 12% More than 20 years, 70% Did Not Respond		
	July 2020[c]	Virtual	4[c]	**State:** 25% Connecticut, 50% Rhode Island, 25%Utah **Gender:** 100% Female **Ethnicity**: 25% White or Caucasian, 25% Hispanic or Latino, 50% Did Not Respond **Region:** 25% Urban, 75% Did Not Respond **Teaching Experience:** 25% 6 to 10 years, 25% 16 to 20 years, 50% Did Not Respond	55	8
ICCR	March 2018	Virtual	13	**State:** 46% Connecticut, 8% Indiana, 15% Utah, 23% West Virginia, 8% Wyoming **Gender:** 85% Female, 15% Male	152	7
	July 2020[c]	Virtual	5[c]	**State:** 20% Connecticut, 40% Rhode Island, 20% Utah, 20% Vermont **Gender:** 100% Female **Ethnicity**: 60% White or Caucasian, 20% Hispanic or Latino, 20% Did Not Respond **Region:** 40% Rural, 20% Suburban, 20% Urban, 20% Did Not Respond **Teaching Experience:** 20% 6 to 10 years, 20% 11 to 15 years, 20% 16 to 20 years, 20% More than 20 years, 20% Did Not Respond	57	0
	July 2021[c]	Virtual	15[c]	**State:** 22% Connecticut, 4% Hawaii, 9% Idaho, 4% Montana, 4% Oregon, 4% Rhode Island, 13% South Dakota, 4% Utah, 3% Vermont, 12% West Virginia, 13% Wyoming, 4% Did Not Respond **Teaching Experience:** 9% 1 to 5 years, 18% 6 to 10 years, 29% 11 to 15 years, 19% 16 to 20 years, 25% More than 20 years	157	1
Idaho	December 2018	Boise, Idaho	15	Not Collected	111	1
	December 2021	Boise, Idaho	21	**Gender:** 81% Female, 19% Male **Ethnicity**: 95% White or Caucasian, 5% Hispanic or Latino	179	0

State/Item Bank	Date	Location	Number of Committee Members	Committee Member Demographic Summary	Number of Items Reviewed	Number of Items Rejected by Committees
				Region: 33% Rural, 19% Suburban, 5% Urban, 43% Did Not Respond **Teaching Experience:** 19% None, 5% Less than 1 year, 5% 1 to 5 years, 19% 6 to 10 years, 5% 11 to 15 years, 14% 16 to 20 years, 33% More than 20 years		
Montana	January 2020	Helena, Montana	15	Not Collected	[b]	[b]
	July 2021[c]	Virtual	3[c]	**State:** 22% Connecticut, 4% Hawaii, 9% Idaho, 4% Montana, 4% Oregon, 4% Rhode Island, 13% South Dakota, 4% Utah, 3% Vermont, 12% West Virginia, 13% Wyoming, 4% Did Not Respond **Teaching Experience:** 9% 1 to 5 years, 18% 6 to 10 years, 29% 11 to 15 years, 19% 16 to 20 years, 25% More than 20 years	41	0
Multi-State Science Assessment (Rhode Island and Vermont)	January 2018	Providence, Rhode Island	21	**State:** 100% Rhode Island **Teaching Experience:** 67% General Education, 14% Bilingual Education, 5% Special Education, 5% Science Coordinator, 10% Other	73	14
	March 2018	Providence, Rhode Island	11	**State:** 55% Rhode Island, 45% Vermont	100	24
	January 2019	Concord, New Hampshire	14	**Gender:** 63% Female, 23% Male **Teaching Experience:** 69% General Education, 3% Special Education, 11% Coach, 17% Other	116	18
	November 2019	Fairlee, Vermont	17	**State:** 29% Rhode Island, 6% Vermont, 65% Did Not Respond **Gender:** 23% Female, 12% Male, 65% Did Not Respond **Ethnicity**: 35% White or Caucasian, 65% Did Not Respond **Region:** 6% Rural, 17% Suburban, 77% Did Not Respond **Teaching Experience:** 6% 11 to 15 years, 17% 16 to 20 years, 12% More than 20 years, 65% Did Not Respond	66	0
	July 2020[c]	Virtual	2[c]	**State:** 50% Utah, 50% Vermont **Gender:** 100% Female	27	0

State/Item Bank	Date	Location	Number of Committee Members	Committee Member Demographic Summary	Number of Items Reviewed	Number of Items Rejected by Committees
				Ethnicity: 50% Hispanic or Latino, 50% White or Caucasian **Region:** 50% Rural, 50% Did Not Respond **Teaching Experience:** 50% 6 to 10 years, 50% More than 20 years		
	July 2021[c]	Virtual	3[c]	**State:** 22% Connecticut, 4% Hawaii, 9% Idaho, 4% Montana, 4% Oregon, 4% Rhode Island, 13% South Dakota, 4% Utah, 3% Vermont, 12% West Virginia, 13% Wyoming, 4% Did Not Respond **Teaching Experience:** 9% 1 to 5 years, 18% 6 to 10 years, 29% 11 to 15 years, 19% 16 to 20 years, 25% More than 20 years	30	1
	August 2021	Virtual	3	**State:** 100% Rhode Island **Gender:** 100% Female **Ethnicity:** 100% White or Caucasian **Region:** 33% Suburban, 67% Urban **Teaching Experience:** 33% 6 to 10 years, 67% More than 20 years	93	3
Oregon	September 2017	Salem, Oregon	5	**Gender:** 100% Female **Region:** 80% Urban, 20% Suburban **Teaching Experience:** 40% General Education, 20% Bilingual Education, 20% Special Education, 60% Administration, 20% Other	235	114
	August 2018	Salem, Oregon	39	**Gender:** 74% Female, 26% Male **Ethnicity:** 3% Asian, 8% Hispanic, 3% Native American, 82% White, 10% Other **Region:** 56% Urban, 44% Rural **Teaching Experience:** 15% General Education, 72% Bilingual Education, 33% Special Education, 33% Other	257	8
	October 2018	Salem, Oregon	8	**Gender:** 100% Female **Ethnicity:** 80% White, 20% Other **Region:** 80% Urban, 20% Rural **Teaching Experience:** 88% Bilingual Education, 50% Special Education, 63% Other	60	12
	December 2018	Virtual	11	**Gender:** 91% Female, 9% Male **Ethnicity:** 9% Hispanic, 91% White **Region:** 55% Urban, 45% Rural	62	14

State/Item Bank	Date	Location	Number of Committee Members	Committee Member Demographic Summary	Number of Items Reviewed	Number of Items Rejected by Committees
				Teaching Experience: 27% General Education, 64% Bilingual Education, 18% Special Education, 9% Administration, 64% Other		
	October 2019	Salem, Oregon	9	**Gender:** 78% Female, 22% Male **Ethnicity:** 89% White, 11% Native American **Region:** 44% Urban, 56% Rural **Teaching Experience:** 89% General Education, 67% Bilingual Education, 44% Special Education	246	23
	January 2020	Salem, Oregon	11	**Gender:** 55% Female, 45% Male **Ethnicity:** 100% White **Region:** 45% Urban, 45% Suburban, 9% Rural **Teaching Experience:** 100% General Education, 90% Bilingual Education, 81% Special Education, 81% Other	262	33
	July 2020[c]	Virtual	2[c]	**State:** 50% Connecticut, 50% Utah **Gender:** 100% Female **Ethnicity**: 50% Hispanic or Latino, 50% Did Not Respond **Region:** 100% Did Not Respond **Teaching Experience:** 50% 6 to 10 years, 50% Did Not Respond	22	3
	August 2020	Virtual	7	**Gender:** 72% Female, 14% Male, 14% Nonbinary **Ethnicity:** 14% Asian, 43% African American, 29% Hispanic, 14% Native American **Region:** 14% Urban, 72% Suburban, 14% Rural **Teaching Experience:** 57% General Education, 57% Bilingual Education, 29% Special Education, 29% Administration	86	7
	August 2021	Virtual	7	**Gender:** 100% Female **Ethnicity:** 100% White or Caucasian **Region:** 14% Urban, 29% Suburban, 57% Rural **Teaching Experience:** 43% General Education, 14% Bilingual Education, 14% Administration, 29% Other	353	13
South Dakota	October 2019	Pierre, South Dakota	26	**Gender:** 81% Female, 19% Male **Ethnicity:** 4% American Indian or Alaska Native, 4% Asian, 92% White	[b]	[b]

State/Item Bank	Date	Location	Number of Committee Members	Committee Member Demographic Summary	Number of Items Reviewed	Number of Items Rejected by Committees
				Region: 65% Rural, 15% Suburban, 15% Urban, 4% Not Applicable **Teaching Experience:** 12% 1 to 5 years, 12% 6 to 10 years, 19% 11 to 15 years, 19% 16 to 20 years, 38% More than 20 years		
U.S. Virgin Islands	October 2021	Virtual	18	**Gender:** 72% Female, 28% Male **Ethnicity**: 6% Asian, 88% Black, 6% White **Region:** 17% Rural, 17% Urban, 11% Suburban, 17% Not Applicable, 38% Did Not Respond **Teaching Experience:** 22% 1 to 5 years, 5% 6 to 10 years, 17% 11 to 15 years, 17% 16 to 20 years, 39% More than 20 years	299	28
Utah	July 2017	Park City, Utah	6	**Gender:** 100% Female **Ethnicity:** 33% American Indian or Alaska Native, 33% Hispanic, 33% White **Region:** 17% Rural, 83% Did Not Respond **Teaching Experience:** 17% General Education, 17% Special Education, 33% Administration, 33% Other	44	2
	December 2017	Salt Lake City, Utah	6	**Gender:** 83% Female, 17% Make **Ethnicity:** 33% Black, 17% Hispanic, 33% Native American, 17% Not Applicable **Teaching Experience:** 33% Administration, 83% Other	48	1
	October 2019	Provo, Utah	11	**Gender:** 27% Female, 73% Did Not Respond **Ethnicity**: 9% Hispanic or Latino, 18% White or Caucasian, 73% Did Not Respond **Region:** 9% Urban, 91% Did Not Respond **Teaching Experience:** 9% 1 to 5 years, 18% More than 20 years, 73% Did Not Respond	31	0
	July 2020[c]	Virtual	9[c]	**Gender:** 22% Female, 78% Did Not Respond **Ethnicity**: 22% Hispanic or Latino, 78% Did Not Respond **Region:** 100% Did Not Respond **Teaching Experience:** 11% None, 11% 6 to 10 years, 78% Did Not Respond	38	1
	December 2020	Virtual	6	**Gender:** 50% Female, 50% Did Not Respond **Ethnicity**: 17% Hispanic or Latino, 33% White or Caucasian, 50% Did Not Respond	14	0

State/Item Bank	Date	Location	Number of Committee Members	Committee Member Demographic Summary	Number of Items Reviewed	Number of Items Rejected by Committees
				Region: 17% Suburban, 83% Did Not Respond **Teaching Experience:** 17% 1 to 5 years, 33% 16 to 20 years, 50% Did Not Respond		
	July 2021[c]	Virtual	11[c]	**State:** 22% Connecticut, 4% Hawaii, 9% Idaho, 4% Montana, 4% Oregon, 4% Rhode Island, 13% South Dakota, 4% Utah, 3% Vermont, 12% West Virginia, 13% Wyoming, 4% Did Not Respond **Teaching Experience:** 9% 1 to 5 years, 18% 6 to 10 years, 29% 11 to 15 years, 19% 16 to 20 years, 25% More than 20 years	64	0
	August 2021	Virtual	6	**Gender:** 100% Female **Ethnicity:** 17% Hispanic or Latino, Native American or Alaskan American, White or Caucasian, Multiracial or Biracial, 83% White or Caucasian **Region:** 33% Rural, 33% Suburban, 17% Urban, 17% Did Not Respond **Teaching Experience:** 17% Less than 1 year, 33% 1 to 5 years, 50% More than 20 years	67	1
West Virginia	January 2017	Charleston, West Virginia	28[a]	Not Collected	34	[b]
	January 2019	Charleston, West Virginia	10	**Gender:** 89% Female, 11% Male **Ethnicity:** 11% Black, 89% White **Region:** 100% Rural **Teaching Experience:** 100% General Education	191	[b]
	July 2021[c]	Virtual	2[c]	**State:** 22% Connecticut, 4% Hawaii, 9% Idaho, 4% Montana, 4% Oregon, 4% Rhode Island, 13% South Dakota, 4% Utah, 3% Vermont, 12% West Virginia, 13% Wyoming, 4% Did Not Respond **Teaching Experience:** 9% 1 to 5 years, 18% 6 to 10 years, 29% 11 to 15 years, 19% 16 to 20 years, 25% More than 20 years	12	1
Wyoming	December 2017	Cheyenne, Wyoming	5	Not Collected	32	3
	October 2018	Cheyenne, Wyoming	5	Not Collected	39	0

State/Item Bank	Date	Location	Number of Committee Members	Committee Member Demographic Summary	Number of Items Reviewed	Number of Items Rejected by Committees
	November 2019	Cheyenne, Wyoming	7	**Gender:** 14% Female, 86% Male **Teaching Experience:** 14% 6 to 10 years, 57% 11 to 20 years, 29% 21+ years	44	1
	August 2020	Virtual	14	**Gender:** 29% Female, 7% Male, 64% Did Not Respond **Ethnicity**: 36% White or Caucasian, 64% Did Not Respond **Region:** 22% Rural, 78% Did Not Respond **Teaching Experience:** 7% 11 to 15 years, 7% 16 to 20 years, 22% More than 20 years, 64% Did Not Respond	37	1
	June/ July 2021	Virtual	6	**Gender:** 67% Female, 17% Male, 17% Did Not Respond **Ethnicity:** 83% White or Caucasian, 17% Did Not Respond **Region:** 50% Rural, 17% Suburban, 33% Did Not Respond **Teaching Experience:** 17% 6 to 10 years, 50% 11 to 15 years, 17% More than 20 years, 17% Did Not Respond	39	39
	July 2021[c]	Virtual	4[c]	**State:** 22% Connecticut, 4% Hawaii, 9% Idaho, 4% Montana, 4% Oregon, 4% Rhode Island, 13% South Dakota, 4% Utah, 3% Vermont, 12% West Virginia, 13% Wyoming, 4% Did Not Respond **Teaching Experience:** 9% 1 to 5 years, 18% 6 to 10 years, 29% 11 to 15 years, 19% 16 to 20 years, 25% More than 20 years	28	0

Note. [a]The number of Committee Members includes total committee members for ELA, math, and science. The number for science only committee members is not available.
[b]Number of science items reviewed and/or rejected by Fairness Committees is unavailable at the time of writing this report.
[c]Items were reviewed in a combined Fairness Committee Meeting that included all MOU state items.

Appendix E

Sample Data Review Training Materials

Sample Data Review Training Materials

Item Data Review

- Item Data Review is the final step before items move to the operational pool
- For every state: data review is carried out for the items owned by the state
- Decision to send an item to data review is based on empirical data
- Statistics are computed at the assertion-level
- Inclusion in data review is decided at the item level, not at the assertion level
  - Inclusion in data review is based on statistical flags that rely on assertion level statistics but are evaluated for the entire item

Cambium Assessment, Inc. 2

Item Data Review: Flagging Rules

- Common flagging rules across states
- Flagging is based on business rules related to
  - Difficulty of the item
  - Relation between the score on item and the overall student's score
  - Response time of the item
  - Statistical flags for differential item functioning

Cambium Assessment, Inc. 3

Flagging Rules: *P*-value

- The *p*-value is the proportion of students for which the assertion is TRUE
- Corresponds to the difficulty of an item in a traditional assessment
- Across an item bank, we want to see assertions with *p*-values across the full range to be able to precisely measure proficiency across all proficiency levels
  - A low *p*-value is not bad per se
- However, we want to make sure the low *p*-value is not a result of an item being misleading

Cambium Assessment, Inc. 4

Flagging Rules: *P*-value

- Criteria for clusters:
  - average *p*-value < .30 (across the assertions within a cluster)
  - average *p*-value > .85 (across the assertions within a cluster)
- Criteria for stand-alone items (typically has 1-3 assertions):
  - average *p*-value < . 15 (across the assertions within a stand-alone item)
  - average *p*-value > .95 (across the assertions within a stand-alone item)

Cambium Assessment, Inc. 5

Flagging Rules: Item-Total Correlation

- We expect students who do well on the test overall to have a higher probability of doing well on individual assertions
- The item-total correlation describes that relation
- Criteria
  - Average item-total (biserial) correlation < .25
  - One or more assertions with an item-total correlation < 0.05

Cambium Assessment, Inc. 6

Flagging Rules: Differential Item Functioning

- Fair items behave similar across groups
- Probability of answering correctly is the same for all students of similar ability regardless of group membership
- Groups are defined by
  - Gender
  - Ethnicity
  - Economically disadvantaged vs. not
  - ELL vs. not ELL
  - Special Education vs. not

Cambium Assessment, Inc. 7

Flagging Rules: Differential Item Functioning

- Method: Mantel-Haenszel (Holland & Thayer, 1988)
  - Compares performance on an item (i.e., assertion for assertion-based scoring) between reference and focal groups conditional on overall performance
  - Using the theta estimate as stratification variable

Cambium Assessment, Inc.

8

Flagging Rules: Differential Item Functioning

- Severity of possible bias based on significance testing and effect size
  - "A" No statistical evidence of DIF
  - "B" Evidence for potential mild DIF
  - "C" Evidence for potential severe DIF
- Direction of possible bias
  - "–" assertion favors reference groups (e.g., whites/male/non ELLs)
  - "+" assertion favors focal group

Cambium Assessment, Inc. 9

Flagging Rules: Differential Item Functioning

- DIF Criterion
  - For clusters: 2 or more assertions show ‘C’ DIF in the same direction
  - For stand-alone items: 1 or more assertions show ‘C’ DIF in the same direction

Cambium Assessment, Inc. 10

Flagging Rules: Timing

- We want a good balance between the amount of information an item provides, and the time students spend on the item
- Criteria
  - For clusters: percentile 80 > 15 minutes
    - A percentile 80 of *x* minutes: 80% of the students spent *x* minutes or less on the item
  - For stand-alone items: percentile 80 > 3 minutes
  - Assertions per minute < .5 for clusters and stand-alone items

Cambium Assessment, Inc. 11

Item Data Review: Process

- Facilitator presents the item
- Item is presented with information on
  - Grade
  - Discipline
  - Disciplinary Core Idea
  - Performance Expectation
- Statistics on the assertions of the item are presented
  - Including the reason for flagging

Item Data Review: Process

- Evaluation of item (stimulus, interactions, assertions)
- For every item, one of the following decisions is made
  - Reject
  - Accept as is

Cambium Assessment, Inc. 13

Appendix F

Data Review Committee Participant Details

Data Review Committee Participant Details

Table F-1. Data Review Committee Participants, Science

State/Item Bank	Date	Location	Number of Committee Members	Committee Member Demographic Summary	Number of Items Reviewed
Connecticut	August 2018	New Britain, Connecticut	29	**Gender:** 88% Female, 12% Male	18
	August 2019	Cromwell, Connecticut	29	**Gender:** 83% Female, 17% Male	53
	August 2021	Virtual	19	**Gender:** 63% Female, 21% Male, 16% Did Not Respond **Ethnicity:** 84% White or Caucasian, 16% Did Not Respond **Region:** 21% Suburban, 21% Urban, 58% Did Not Respond **Teaching Experience:** 5% 1 to 5 years, 5% 6 to 10 years, 16% 11 to 15 years, 21% 16 to 20 years, 37% More than 20 years, 16% Did Not Respond	51
	August 2022	Virtual	15	**Gender:** 73% Female, 20% Male, 7% Did Not Respond **Ethnicity:** 7% Hispanic or Latino, 87% White or Caucasian, 7% Did Not Respond **Region:** 60% Suburban, 7% Urban, 33% Did Not Respond **Teaching Experience:** 13% 6 to 10 years, 27% 16 to 20 years, 53% More than 20 years, 7% Did Not Respond	19
Hawaii	August 2018	Honolulu, Hawaii	18	Not Collected	32
	August 2019	Honolulu, Hawaii	18	**Gender:** 71% Female, 29% Male **Ethnicity:** 12% American Indian and White, 41% Asian, 6% Asian and White, 12% Hispanic and White, 18% Native Hawaiian or Pacific Islander, 12% White **Teaching Experience:** 53% General Education, 6% General Education with SPED Certification, 12% Special Education, 29% Other	37
	August 2021[e]	Virtual	7[e]	**State:** 14% Connecticut, 29% Hawaii, 14% Idaho, 29% West Virginia, 14% Wyoming **Gender:** 86% Female, 14% Male **Ethnicity**: 86% White or Caucasian, 14% Did Not Respond **Region:** 14% Rural, 29% Suburban, 57% Did Not Respond **Teaching Experience:** 29% 11 to 15 years, 14% 16 to 20 years, 29% More than 20 years, 14% Did Not Respond	26

State/Item Bank	Date	Location	Number of Committee Members	Committee Member Demographic Summary	Number of Items Reviewed
	August 2022[e]	Virtual	12[e]	**State:** 17% Connecticut, 17% Hawaii, 8% Idaho, 25% Oregon, 33% Wyoming **Gender:** 75% Female, 25% Male **Ethnicity:** 8% Asian or Pacific Islander, 82% White or Caucasian **Region:** 50% Rural, 42% Suburban, 8% Did Not Respond **Island:** 8% Not Applicable, 8% Oahu, 75% Did Not Respond **Teaching Experience:** 33% 6 to 10 years, 8% 16 to 20 years, 58% More than 20 years	49
ICCR	July 2018	Virtual	18	Not Collected	84
	August 2019[d]	Virtual	N/A[d]	N/A[d]	43
	August 2021[e]	Virtual	11[e]	**State:** 27% Connecticut, 9% Hawaii, 18% Idaho, 36% West Virginia, 9% Wyoming **Gender:** 82% Female, 18% Male **Ethnicity:** 54% White or Caucasian, 46% Did Not Respond **Region:** 9% Rural, 27% Suburban, 64% Did Not Respond **Teaching Experience:** 9% 6 to 10 years, 9% 11 to 15 years, 36% More than 20 years, 46% Did Not Respond	75
	August 2022[e]	Virtual	20[e]	**State:** 15% Connecticut, 20% Idaho, 5% North Dakota, 35% Oregon, 5% South Dakota, 20% Wyoming **Gender:** 85% Female, 15% Male **Ethnicity:** 5% Asian or Pacific Islander, 95% White or Caucasian **Region:** 30% Rural, 25% Suburban, 15% Urban, 30% Did Not Respond **Teaching Experience:** 10% 1 to 5 years, 35% 6 to 10 years, 15% 16 to 20 years, 40% More than 20 years	68
Idaho	August 2019	[a]	10	**Gender:** 70% Female, 20% Male, 1% Did Not Respond **Ethnicity:** 100% White **Region:** 60% Rural, 40% Suburban **Teaching Experience:** 60% General Education, 2% Administration, 2% Coach	12
	August 2021[e]	Virtual	9[e]	**State:** 11% Hawaii, 56% Idaho, 11% West Virginia, 22% Wyoming **Gender:** 89% Female, 11% Male **Ethnicity:** 89% White or Caucasian, 11% Did Not Respond **Region:** 11% Rural, 22% Suburban, 67% Did Not Respond	60

State/Item Bank	Date	Location	Number of Committee Members	Committee Member Demographic Summary	Number of Items Reviewed
				Teaching Experience: 22% 6 to 10 years, 22% 11 to 15 years, 11% 16 to 20 years, 33% More than 20 years, 11% Did Not Respond	
	August 2022[e]	Virtual	8[e]	**State:** 25% Connecticut, 13% Idaho, 25% Oregon, 38% Wyoming **Gender:** 63% Female, 38% Male **Ethnicity:** 13% Hispanic or Latino, 88% White or Caucasian **Region:** 38% Rural, 50% Suburban, 13% Did Not Respond **Teaching Experience:** 13% 1 to 5 years, 13% 6 to 10 years, 25% 11 to 15 years, 13% 16 to 20 years, 38% More than 20 years	4
Montana	September 2021	Virtual	4	**Gender:** 50% Female, 50% Did Not Respond **Ethnicity:** 50% White or Caucasian, 50% Did Not Respond **Region:** 50% Rural, 50% Did Not Respond **Teaching Experience:** 25% 6 to 10 years, 25% 16 to 20 years, 50% Did Not Respond	17
	September 2022	Virtual	5	**Gender:** 100% Female **Ethnicity:** 100% White or Caucasian **Region:** 60% Rural, 40% Suburban **Teaching Experience:** 40% 6 to 10 years, 20% 16 to 20 years, 40% More than 20 years	17
Multi-State Science Assessment (Rhode Island and Vermont)	August 2018	Virtual	2[b]	N/A[b]	9
	August 2019	Virtual	2[b]	N/A[b]	14
	August 2021	Virtual	2[b]	N/A[b]	18
	September 2022	Virtual	2[b]	N/A[b]	11
Oregon	September 2018	Salem, Oregon	11	**Gender:** 82% Female, 18% Male **Ethnicity:** 100% White **Region:** 27% Urban, 73% Rural **Teaching Experience:** 64% General Education, 55% Bilingual Education, 36% Special Education, 18% Administration, 18% Other	44
	August 2019	Virtual	4	**Gender:** 50% Female, 50% Male **Ethnicity:** 100% White **Region:** 50% Urban, 50% Rural **Teaching Experience:** 50% General Education, 25% Bilingual Education, 25% Special Education, 25% Administration, 75% Other	8

State/Item Bank	Date	Location	Number of Committee Members	Committee Member Demographic Summary	Number of Items Reviewed
	August 2022[e]	Virtual	8[e]	**State:** 38% Connecticut, 38% Idaho, 13% Wyoming, 13% Did Not Respond **Gender:** 75% Female, 13% Male, 13% Did Not Respond **Ethnicity:** 88% White or Caucasian, 13% Did Not Respond **Region:** 25% Rural, 13% Suburban, 25% Urban, 38% Did Not Respond **Teaching Experience:** 25% 6 to 10 years, 13% 11 to 15 years, 13% 16 to 20 years, 38% More than 20 years, 13% Did Not Respond	31
South Dakota	September 2021[d]	Virtual	N/A[c]	N/A[c]	15
Utah	August 2018	Salt Lake City, Utah	16	**Gender:** 93% Female, 7% Male **Ethnicity:** 87% White, 13% Did Not Respond **Region:** 13% Suburban, 27% Rural, 60% Did Not Respond **Teaching Experience:** 100% General Education	40
	September 2021	Virtual	6	**Gender:** 63% Female, 38% Male **Ethnicity:** 13% Native Hawaiian or Pacific Islander, 88% White or Caucasian **Region:** 50% Rural, 13% Suburban, 38% Urban **Teaching Experience:** 38% 6 to 10 years, 38% 11 to 15 years, 25% More than 20 years	11
	September 2022	Salt Lake City, Utah	13	**Gender:** 77% Female, 15% Male, 8% Did Not Respond **Ethnicity:** 8% Asian, 8% Hispanic, 8% Mixed, 77% White **Region:** 15% Rural, 38% Suburban, 38% Urban, 8% Not Applicable **Teaching Experience:** 23% 6 to 10 years, 46% 11 to 15 years, 15% More than 20 years, 15% Not Applicable	11
West Virginia	July 2018	[a]	4	Not Collected	3
	September 2019	[a]	4	Not Collected	7
	August 2021[e]	Virtual	4[e]	**State:** 25% Hawaii, 50% West Virginia, 25% Wyoming **Gender:** 100% Female **Ethnicity**: 75% White or Caucasian, 25% Did Not Respond **Region:** 25% Rural, 25% Suburban, 50% Did Not Respond **Teaching Experience:** 50% 11 to 15 years, 25% More than 20 years, 25% Did Not Respond	7
	August 2022[e]	Virtual	9[e]	**State:** 22% Connecticut, 33% Idaho, 11% Oregon, 33% Wyoming **Gender:** 89% Female, 11% Male **Ethnicity:** 100% White or Caucasian	10

State/Item Bank	Date	Location	Number of Committee Members	Committee Member Demographic Summary	Number of Items Reviewed
				Region: 56% Rural, 11% Suburban, 11% Urban, 22% Did Not Respond **Teaching Experience:** 11% 1 to 5 years, 22% 6 to 10 years, 11% 11 to 15 years, 11% 16 to 20 years, 44% More than 20 years	
Wyoming	October 2018	Cheyenne, Wyoming	19	Not Collected	16
	August 2019	Cheyenne, Wyoming	10	**Gender:** 90% Female, 10% Male **Region:** 40% Suburban, 60% Rural **Teaching Experience:** 90% General Education, 10% Administration	16
	August 2021[e]	Virtual	8[e]	**State:** 37.5% Connecticut, 12.5% Hawaii, 12.5% West Virginia, 37.5% Wyoming **Gender:** 75% Female, 25% Male **Ethnicity**: 75% White or Caucasian, 25% Did Not Respond **Region:** 12.5% Rural, 25% Suburban, 12.5% Urban, 50% Did Not Respond **Teaching Experience:** 12.5% 11 to 15 years, 62.5% More than 20 years, 25% Did Not Respond	16
	August 2022[e]	Virtual	12[e]	**State:** 17% Connecticut, 8% Hawaii, 17% Idaho, 17% Oregon, 42% Wyoming **Gender:** 67% Female, 33% Male **Ethnicity:** 8% Asian or Pacific Islander, 8% Hispanic or Latino, 83% White or Caucasian **Region:** 42% Rural, 50% Suburban, 8% Did Not Respond **Teaching Experience:** 8% 1 to 5 years, 25% 6 to 10 years, 17% 11 to 15 years, 8% 16 to 20 years, 42% More than 20 years	19

Note. [a]Location of Data Review Committee Meeting is unavailable at the time of writing this report.
[b]Conducted by the Rhode Island Department of Education and the Vermont Agency of Education science content experts.
[c]Reviewed by South Dakota Department of Education.
[d]In summer 2019, ICCR field-test items were taken to Connecticut, Hawaii, and Idaho for committee review.
[e]Combined Data Review for multiple states (184 Hawaii, Idaho, West Virginia, Wyoming, and ICCR items in 2021 and 181 Hawaii, Idaho, Oregon, West Virginia, Wyoming, and ICCR items in 2022). There were 25 total participants in 2021 and 38 total participants in 2022. Items are broken out by owning state.

Appendix G

Example Item Interactions

Item Interactions Available in the West Virginia GSA for Science

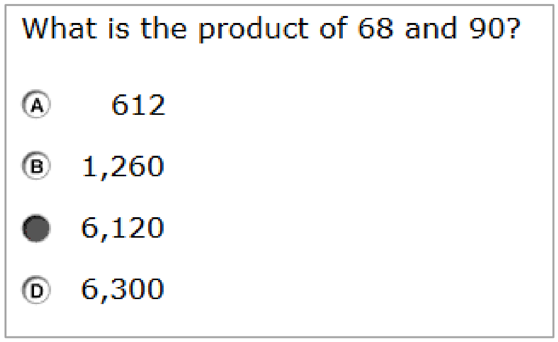
Review of Different Interaction Types

Interaction Type	Associated Sub-Types	Legacy Item Types Supported
Choice	Multiple Choice	MC
	Multiple Select	MS
	Scaffolding	ASI2, ASI3
Text Entry	Simple Text Entry	EA, ECR, LA, OE, SA, SR, WCR, RW, SCR
	Embedded Text Entry	CL, FI
	Natural Language	NL
	Extended Response	ER
Table	Table Match	MI
	Table Input	TI
	Column Match	MI
Edit Task	Edit Task	ET
	Edit Task with Choice	ETC
	Edit Task Inline Choice	ETC
Hot Text	Selectable	HTQ
	Re-orderable	HT
	Drag-from-Palette	DnD
	Custom	HTQ, HT, DnD
Equation	N/A	EQN
Grid	Grid	GI
	Hot Spot	GI
	Graphic Gap Match	GI
Simulation*	N/A	SIM

**Note: the abbreviations correlate to the attributes used in CAI's Item Tracking System*

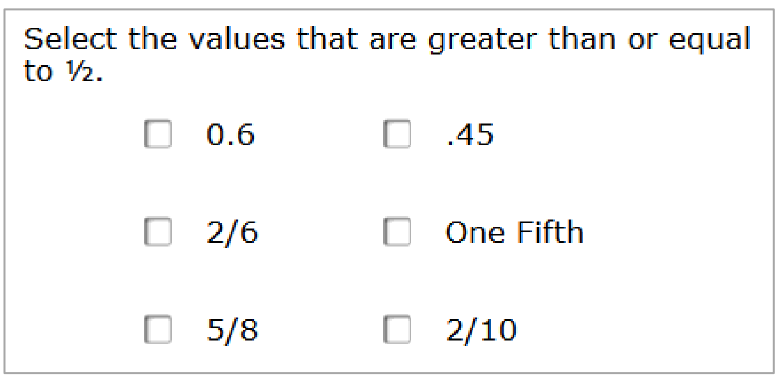
Multiple-Choice Interactions

Multiple-Choice (MC) interactions require students to select a single option from a list of possible answer options. The number and orientation of answer options in a multiple-choice interaction are configurable. Answer options may appear vertically, horizontally, vertically-stacked (in a specified number of columns), or horizontally-stacked (in a specified number of rows).



Multiple-Select Interactions

Multiple-Select interactions require students to select one or more options from a list of possible answer options. The number and orientation of answer options in a multiple-select interaction are configurable. Answer options may appear vertically, horizontally, horizontally-stacked (in a specified number of rows), or vertically-stacked (in a specified number of columns).



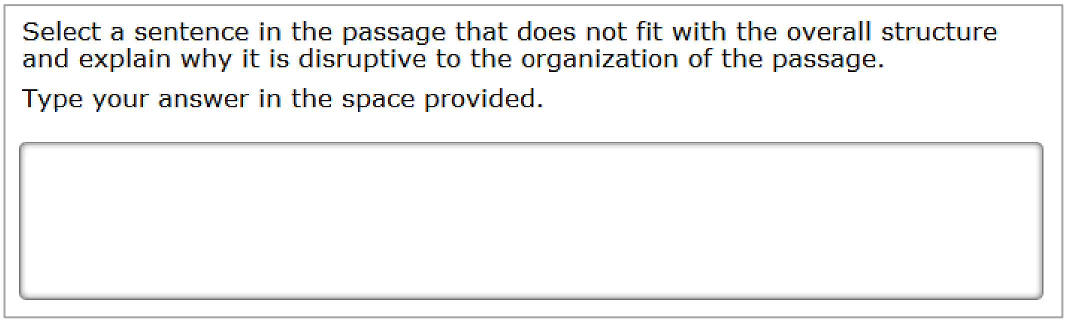
Text Entry Interactions

The Text Entry Interaction Editor allows you to create content for the following interaction types:

- Simple Text Entry Interactions
- Embedded Text Entry Interactions
- Natural Language Interactions
- Extended Response Interactions

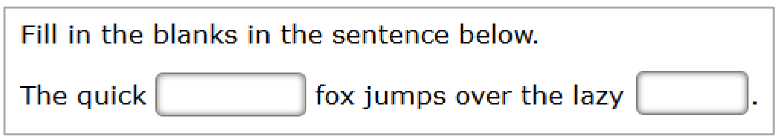
Simple Text Entry Interactions

Simple Text Entry interactions require students to type a response in a text box. For Simple Text Entry interactions, we can allow you to specify the maximum response length for the text box and the type of text editor available to students.



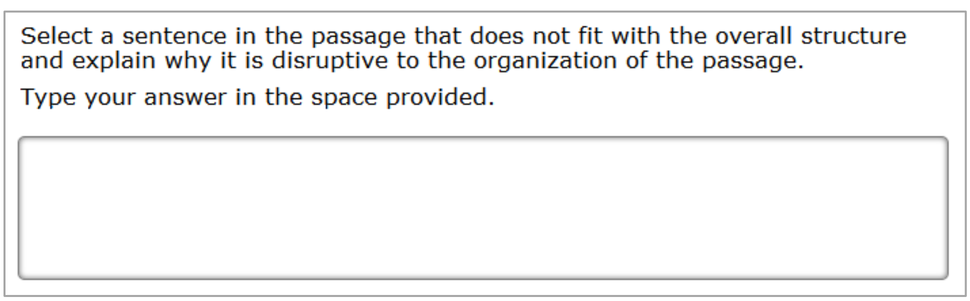
Embedded Text Entry Interactions

Embedded Text Entry interactions require students to type their response in one or more text boxes that are embedded in a section of read-only text.



Extended Response Interactions

Extended Response interactions require students to type a response in a text box. Extended Response interactions are scored by an uploaded essay scoring model that analyzes the student's response to identify variations of acceptable key words and phrases. For Extended Text Entry interactions, we can allow you to specify the maximum response length for the text box and the type of text editor available to students.



Alert: Extended Response interactions cannot be combined with any other interactions in the item.

Table Entry Interaction

The Table Entry Interaction Editor allows you to create content for the following interaction types:

- Table Match Interactions
- Table Input Interactions
- Column Match Interactions

Table Match Interactions

Table Match interactions arrange two sets of match options in a table, with one set listed in columns and the other set listed in rows. Students match options in the columns to options in the rows by marking checkboxes in the cells where the columns and rows intersect.

For each number listed in the rows of the table, mark the checkboxes for each column that describes that number.

	Perfect Square	**Prime Number**	**Odd Number**	**Even Number**
5	☐	☐	☐	☐
12	☐	☐	☐	☐
9	☐	☐	☐	☐

Table Match interactions allow you to customize the number of match options in each set and enter the content for each match option. You can also set restrictions on the number of matches students can make. By default, the panel includes a basic table consisting of three rows and columns (including the row header and column header).

Table Input Interactions

Table Input interactions provide students with a table that includes one or more blank cells. Each blank cell displays a text box in which students can type their response.

Enter a stage direction that you might give to each theater technician listed in the table below.

The first one has been done for you.

Theater technicians	Stage direction
Set designer	A circular bench around a small obelisk
Props manager	
Sound technician	
Lighting technician	

Table Input interactions allows you to customize the number of rows and columns in the table, specify which cells display text boxes, and enter content for the read-only cells. By default, the panel includes a basic table consisting of three rows and columns (including the row header and column header).

Alert: If a table does not include row headers, then it must include column headers. If a table does not include column headers, then it must include row headers.

Column Match Interactions

Column Match interactions provide students with two columns that each contain a set of match options. Students respond to the interaction by selecting a match option in the left column and then selecting the corresponding match option in the right column. A match option in one set may have one, multiple, or no matches in the other set.

Match the words in the left column with their synonyms in the right column.

Happy	Despondent
Sad	Famished
Angry	Elated
Hungry	Weary
Tired	Irate

Column Match interactions allows you to customize the number of match options in each set and enter the content for each match option. By default, the panel includes two single-column tables, each of which includes two match options. You can also set restrictions on the number of matches students can make.

Edit Task Interactions

The Edit Task Interaction Editor allows you to create content for the following interaction types:

- Edit Task Interactions
- Edit Task with Choice Interactions
- Edit Task Inline Choice Interactions

Edit Task Interactions

Edit Task interactions provide students with a sentence or paragraph containing one or more tagged text elements. Tagged elements usually contain an error, such as improper spelling or grammar. To respond to these interactions, students click a tagged element and enter corrected text in an editing window. The entered text replaces the original tagged text.

> The sentence below contains several grammatical mistakes. Click the highlighted words to correct the grammar.
>
> The quick foxes jumps over the lazy, dogs.

Edit Task interactions allow you to enter the text that appears in the response area and tag elements within the text that students can edit.

Warning: You cannot include hand-scored and machine-scored interactions in the same item.

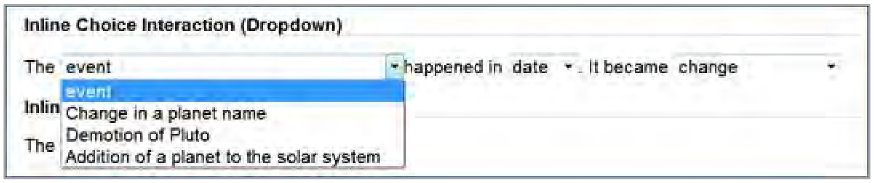
Edit Task with Choice Interactions

Edit Task with Choice interactions are similar to Edit Task interactions. The only difference is that when responding to Edit Task with Choice interactions, students replace the tagged text elements with options selected from a drop-down list.

Edit Task with Choice interactions allow you to enter the text that appears in the response area and tag elements within the text that students can edit.

Edit Task Inline Choice Interactions

Edit Task Inline Choice interactions are similar to Edit Task with Choice interactions. The only difference is that students select replacement options from a drop-down list embedded within the read-only text, rather than accessing the drop-down list via a pop-up window.



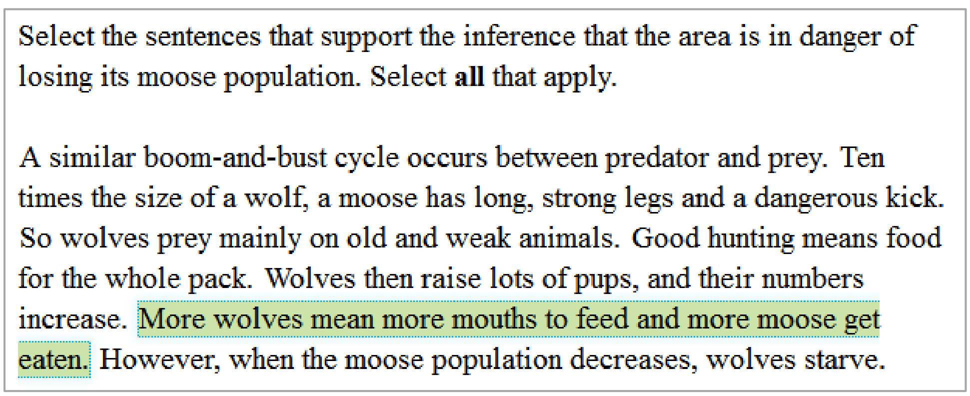
Hot Text Interactions

The Hot Text Interaction Editor allows you to create content for the following interaction types:

- Selectable Hot Text Interactions
- Re-orderable Hot Text Interactions
- Drag-from-Palette Hot Text Interactions
- Custom Hot Text Interactions

Selectable Hot Text Interactions

Selectable Hot Text interactions require students to select one or more text elements in the response area.


Select the sentences that support the inference that the area is in danger of losing its moose population. Select **all** that apply.

A similar boom-and-bust cycle occurs between predator and prey. Ten times the size of a wolf, a moose has long, strong legs and a dangerous kick. So wolves prey mainly on old and weak animals. Good hunting means food for the whole pack. Wolves then raise lots of pups, and their numbers increase. More wolves mean more mouths to feed and more moose get eaten. However, when the moose population decreases, wolves starve.

Selectable Hot Text interactions allows you to set the minimum and maximum number of elements students can select, enter the text that appears in the response area, and tag the text elements that will be selectable.

Re-orderable Hot Text Interactions

Re-orderable Hot Text interactions require students to click and drag hot text elements into a different order.

Place the following sentences in the correct order.

Hey Jude. And make it better. Don't be afraid. Take a sad song.

Re-orderable Hot Text interactions allow you to enter the re-orderable text elements in the response area. You can specify the elements' orientation and set them to appear in random order to students.

Drag-from-Palette Hot Text Interactions a.k.a. Hot Text Gap Match

Drag-from-Palette Hot Text interactions require students to drag elements from a palette into the available blank table cells or "gaps" (text boxes) in the response area. Palette elements may consist of text and/or images. Students may be able to drag the same palette element into multiple gaps, depending on the interaction's configuration.

Drag and drop the characteristics into the appropriate table cells below.

Fortunato's character	**Montressor's character**

Sinister and calculating
Cowardly and irreverent
Egotistical and rude
Lazy and inconsiderate

Drag-from-Palette Hot Text interactions allow you to enter the elements that appear in the palette, enter static text for the response area, and create the gap targets where students can drag the text elements. You can enter all of the elements in a single text box or enter each segment in its own text box.

- Can set a minimum/maximum number of times a student is required/allowed to use a specific palette object
- Only supports drag-and-drop of palette items (images or plain text) onto pre-defined drop targets (“gaps” or “blanks”) in the body text
    - These palette items are always confined to a special palette region (no “preplacing” them)
    - There is some control over palette placement
    - The items can only be placed in predefined “target” regions

Custom Hot Text Interactions

Custom Hot Text interactions combine the functionality of the other Hot Text interaction sub-types. Students responding to a Custom Hot Text interaction may need to select text elements, rearrange text elements, and/or drag text elements from a palette to blank table cells or drop targets in the response area. In many ways, this is the grid of the text-interaction world. In practice, it is typically used to do drag-and-drop with text, but it can technically do more:

- Supports dragging and dropping text elements onto drop target areas
    - Text elements can originally be placed anywhere in the interaction (there's no dedicated palette)
    - Multiple elements can be dropped onto a target
        - this constitutes a "group"
        - much like grid hotspots, you can set constraints on the group
        - Supports selectable text elements
    - Like grid hotspots, these too can be grouped

> Use the word bank to fill in the blank in the sentence below. Then, select all the words in the sentence that are nouns.
>
> **Word bank**:
>
> *young dull good rich*
>
> **Sentence**:
>
> *All work and no play makes Jack a _________ boy.*

Custom Hot Text interactions allow you to create groups of text elements, as well as the drop targets and static text that appear in the response area. When you create a group of text elements, you must assign a Hot Text functionality to that group. The following functionalities are available:

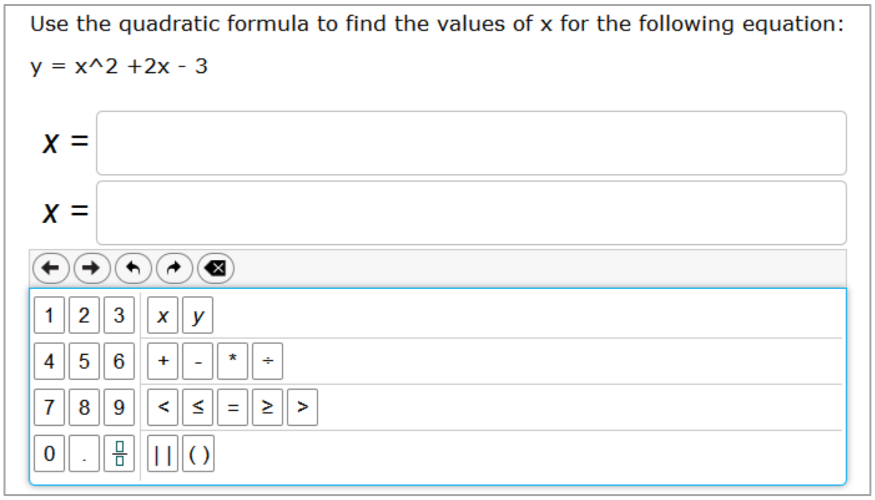
- **Selectable**: When you assign this functionality to a group, the text elements in the group behave like elements in a Selectable Hot Text interaction. You cannot add drop target elements to this kind of group.

- **Draggable**: When you assign this functionality to a group, the text elements in the group behave like elements in a Re-Orderable Hot Text interaction. If you assign this functionality to a group and also add drop targets to the group, the text elements in the group behave like elements in a Drag-from-Palette Hot Text interaction.

You can create as many groups as you wish, but you can only assign one Hot Text functionality to each group.

Equation Interaction Editor

The Equation Interaction Editor allows you to create content for Equation interactions only. Equation interactions require students to enter a response into input boxes using an on-screen keypad, which may consist of special mathematics characters. Students can also enter their response via a physical keyboard, but they cannot enter any characters that are not included in the on-screen keyboard.



Equation interactions allow you to select the buttons to include in the on-screen keypad, enter static text in the response area, and specify the number of input boxes to include in the response area. When selecting buttons to include in the keypad, you can add individual buttons or an entire row or tab of buttons.

Grid Interactions

The Grid Interaction Editor allows you to create content for the following interaction types:

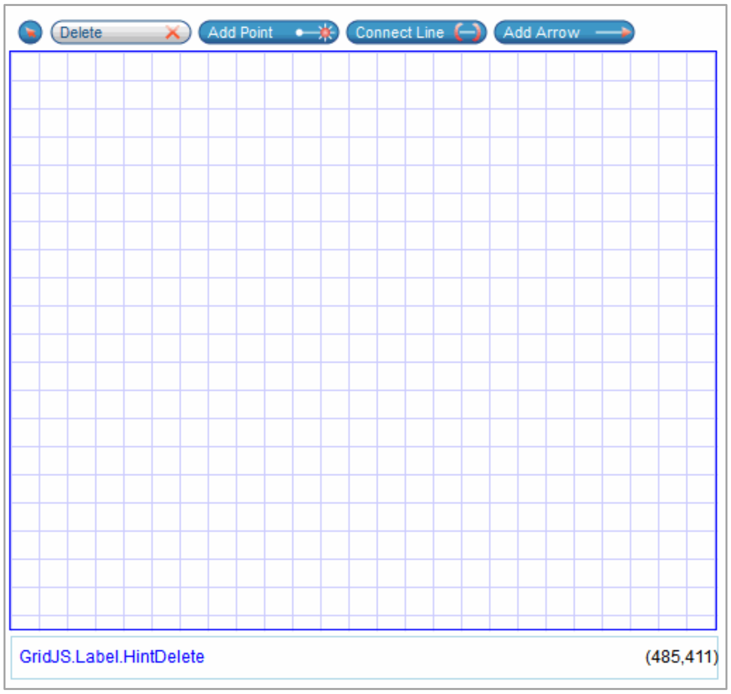
- Grid Interactions
- Hot Spot Interactions
- Graphic Gap Match Interactions

Note: Although there are three options available in the **Interaction Type** drop-down list, the generic **Grid** option allows you to create interactions with functionality similar to Hot Spot and Graphic Gap Match sub-types.

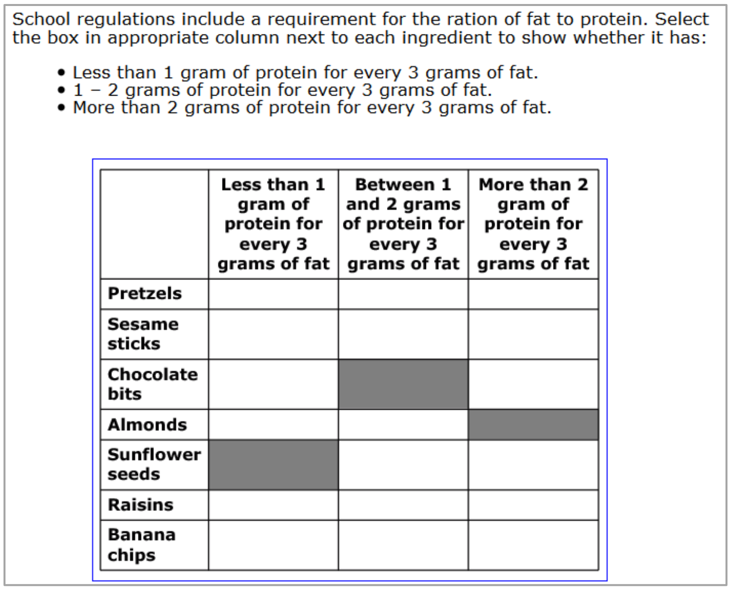
Grid Interactions Types

Grid interactions require students to enter a response by interacting with a grid area in the answer space. There are three general ways in which students can interact with the grid area.

- **Graphing Functionality**: Students can use various tool buttons to add points, lines, and other geometric shapes to the grid area. Only the Grid interaction sub-type allows you to create interactions with this functionality.



- **Hot Spot Functionality**: Students can click or hover over interactive regions in the grid area (hot spots) in order to activate them. Activated hot spots become highlighted, become outlined, or display an image. The Grid and Hot Spot interaction sub-types allow you to create interactions with this functionality.
    - Hotspots can be defined in groups, each of which can have its own selection constraints
    - These regions support events so clicking a hotspot might change the appearance of the interaction by showing/hiding other images, for example



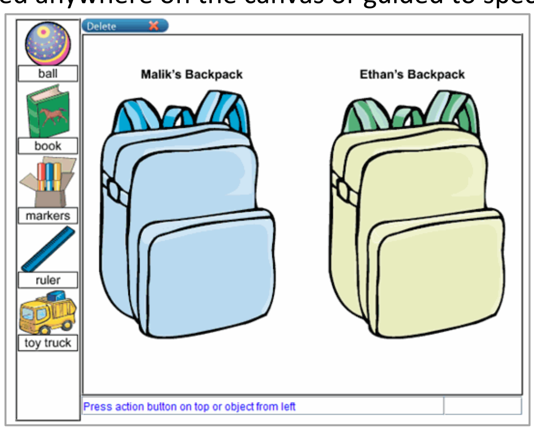
School regulations include a requirement for the ration of fat to protein. Select the box in appropriate column next to each ingredient to show whether it has:

- Less than 1 gram of protein for every 3 grams of fat.
- 1 – 2 grams of protein for every 3 grams of fat.
- More than 2 grams of protein for every 3 grams of fat.

	Less than 1 gram of protein for every 3 grams of fat	Between 1 and 2 grams of protein for every 3 grams of fat	More than 2 gram of protein for every 3 grams of fat
Pretzels			
Sesame sticks			
Chocolate bits			
Almonds			
Sunflower seeds			
Raisins			
Banana chips			

- **Drag-and-Drop Functionality**: Students can click image or text objects and drag them into various locations in the grid area. The objects for these interactions are either provided in a palette beside the grid area or pre-placed within the grid area itself. The Grid and Graphic Gap Match interaction sub-types allow you to create interactions with this functionality; however, only Graphic Gap Match interactions allow text objects.

  - o These palette items can be "preplaced" on the canvas or listed in a separate palette
  - o The items can be placed anywhere on the canvas or guided to specific regions with snap points



Note: The functionalities of these interaction types are not mutually exclusive. A single Grid interaction may require students to select hot spots and place objects, or graph lines and select hot spots, and so on. However, a Grid interaction cannot include preplaced objects if it also includes the **Delete** tool button above the grid area.

Grid Hot Spot Interactions

Hot Spot interaction sub-types allow you to create Grid interactions with hot spot functionality. These interactions require students to select hot spot regions in the grid area.

- o Only supports click-to-select "hotspots"

- No visual side-effect events are supported
- No hotspot groups are supported

Grid Graphic Gap Match Interactions

Graphic Gap Match interactions allow you to create Grid interactions with both hot spot and drag-and-drop functionality. These interactions require students to drag image objects from a palette to hot spot regions (gaps) in the grid area.

o Only supports drag-and-drop of palette items (images or plain text) onto the canvas/background

- These palette items are always confined to a special palette region (no “preplacing” them on the canvas)
- The items can only be placed in predefined “target” regions

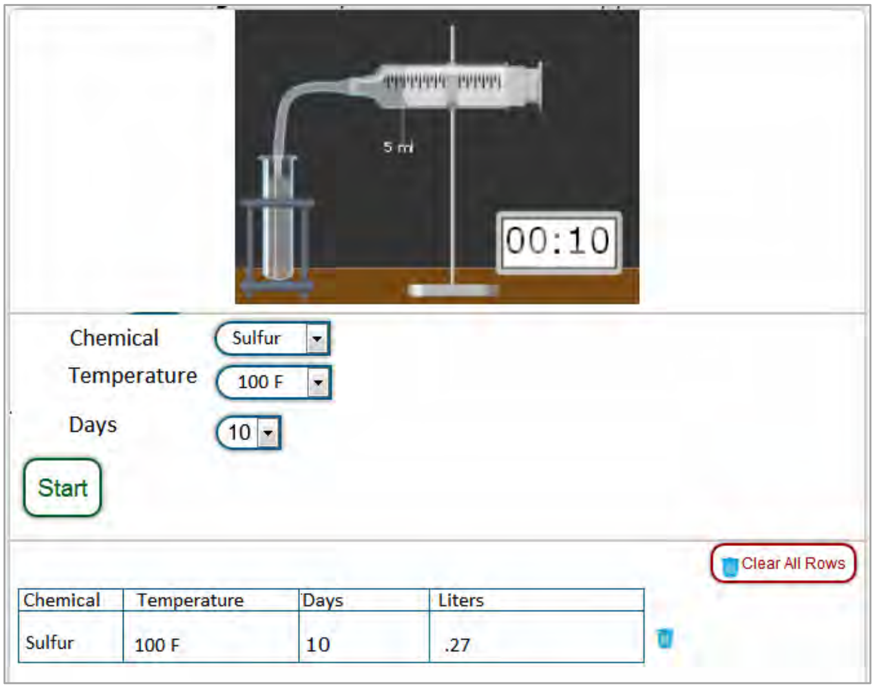
Alert: Graphic Gap Match interactions do not allow you to enable Snap-to-Point or Snap-to-Grid mode. You cannot pre-place image or text objects in the grid area with Graphic Gap Match Interactions.

Basically, graphic gap match and hotspot are dedicated interactions that don’t support all the features of a grid. The trade-off here is:

- Graphic gap match and hotspot interactions are rendered differently (more simplistically)
- In some ways, graphic gap match and hotspot are easier to author and maintain
- Grid interactions need to use the “grid rubric tool,” which is quite complicated

Simulation Interaction Editor

The Simulation Interaction Editor allows you to create content for Simulation interactions only. Simulation interactions consist of an animation tool, a set of input tools, and an output table. Students select parameters from the input tools to influence the animation. After the animation runs, the simulation results appear in the output table. Students can run multiple trials with different parameters to insert additional rows into this table.



Appendix H

Shared Science Assessment Item Bank

Shared Science Assessment Item Bank

Table H-1. Spring 2022 Shared Science Assessment Operational and Field-Test Item Bank by Performance Expectation, Elementary School

Science Discipline	Disciplinary Core Idea	Performance Expectation	ICCR Items	West Virginia Items	MOU Items[a]	Total Bank Items
Earth and Space Sciences	ESS1	4-ESS1-1	2	1	10	13
		5-ESS1-1	2	0	14	16
		5-ESS1-2	8	0	13	21
	ESS2	3-ESS2-1	3	0	9	12
		3-ESS2-2	2	1	8	11
		4-ESS2-1	4	0	11	15
		4-ESS2-2	3	0	13	16
		5-ESS2-1	0	2	8	10
		5-ESS2-2	3	1	11	15
	ESS3	3-ESS3-1	3	0	8	11
		4-ESS3-1	7	0	5	12
		4-ESS3-2	7	1	12	20
		5-ESS3-1	3	1	11	15
Life Sciences	LS1	3-LS1-1	3	0	8	11
		4-LS1-1	9	1	10	20
		4-LS1-2	1	2	12	15
		5-LS1-1	3	3	12	18
	LS2	3-LS2-1	4	0	9	13
		5-LS2-1	2	0	13	15
	LS3	3-LS3-1	2	1	9	12
		3-LS3-2	3	0	6	9
	LS4	3-LS4-1	3	0	13	16
		3-LS4-2	7	0	5	12
		3-LS4-3	4	0	8	12
		3-LS4-4	5	0	7	12
Physical Sciences	PS1	5-PS1-1	5	0	10	15

Science Discipline	Disciplinary Core Idea	Performance Expectation	ICCR Items	West Virginia Items	MOU Items[a]	Total Bank Items
		5-PS1-2	2	0	11	13
		5-PS1-3	5	0	11	16
		5-PS1-4	2	1	8	11
	PS2	3-PS2-1	4	1	7	12
		3-PS2-2	4	0	7	11
		3-PS2-3	3	2	6	11
		3-PS2-4	2	0	7	9
		5-PS2-1	2	1	7	10
	PS3	4-PS3-1	5	0	13	18
		4-PS3-2	5	1	9	15
		4-PS3-3	2	0	11	13
		4-PS3-4	5	0	13	18
		5-PS3-1	3	0	10	13
	PS4	4-PS4-1	2	0	10	12
		4-PS4-2	1	0	11	12
		4-PS4-3	3	0	9	12
Total			**148**	**20**	**405**	**573**

Note. [a]Other MOU states include Connecticut, Hawaii, Idaho, Montana, MSSA (Rhode Island and Vermont), Oregon, Utah, and Wyoming.

Table H-2. Spring 2022 Shared Science Assessment Operational and Field-Test Item Bank by Performance Expectation, Middle School

Science Discipline	Disciplinary Core Idea	Performance Expectation	ICCR Items	West Virginia Items	MOU Items[a]	Total Bank Items[b]
Earth and Space Sciences	ESS1	MS-ESS1-1	6	0	7	13
		MS-ESS1-2	4	1	5	10
		MS-ESS1-3	2	0	9	11
		MS-ESS1-4	3	0	9	12
	ESS2	MS-ESS2-1	4	0	7	11
		MS-ESS2-2	3	1	9	13
		MS-ESS2-3	3	0	8	11
		MS-ESS2-4	2	0	7	9
		MS-ESS2-5	2	1	5	8
		MS-ESS2-6	2	0	4	6
	ESS3	MS-ESS3-1	2	0	9	11
		MS-ESS3-2	3	1	7	11
		MS-ESS3-3	2	0	10	12
		MS-ESS3-4	2	0	10	12
		MS-ESS3-5	5	0	7	12
Life Sciences	LS1	MS-LS1-1	1	0	7	8
		MS-LS1-2	2	0	10	12
		MS-LS1-3	1	1	4	6
		MS-LS1-4	4	0	5	9
		MS-LS1-5	3	0	8	11
		MS-LS1-6	3	2	5	10
		MS-LS1-7	3	0	6	9
		MS-LS1-8	5	0	6	11
	LS2	MS-LS2-1	5	1	10	16
		MS-LS2-2	3	0	6	9
		MS-LS2-3	3	0	10	13
		MS-LS2-4	9	0	8	17
		MS-LS2-5	4	0	8	12

Science Discipline	Disciplinary Core Idea	Performance Expectation	ICCR Items	West Virginia Items	MOU Items[a]	Total Bank Items[b]
	LS3	MS-LS3-1	2	1	9	12
		MS-LS3-2	3	1	8	12
	LS4	MS-LS4-1	5	0	8	13
		MS-LS4-2	1	0	7	8
		MS-LS4-3	2	0	8	10
		MS-LS4-4	4	0	7	11
		MS-LS4-5	4	0	7	11
		MS-LS4-6	2	0	5	7
Physical Sciences	PS1	MS-PS1-1	2	0	8	10
		MS-PS1-2	3	0	9	12
		MS-PS1-3	3	1	5	9
		MS-PS1-4	2	0	10	12
		MS-PS1-5	1	1	10	12
		MS-PS1-6	2	0	5	7
	PS2	MS-PS2-1	1	0	9	10
		MS-PS2-2	3	0	7	10
		MS-PS2-3	1	0	7	8
		MS-PS2-4	1	2	8	11
		MS-PS2-5	0	0	10	10
	PS3	MS-PS3-1	3	0	9	12
		MS-PS3-2	3	1	8	12
		MS-PS3-3	6	1	5	12
		MS-PS3-4	2	0	6	8
		MS-PS3-5	5	0	6	11
	PS4	MS-PS4-1	4	0	6	10
		MS-PS4-2	5	1	6	12
		MS-PS4-3	2	1	7	10
Total			**163**	**18**	**406**	**587**

Note. [a]Other MOU states include Connecticut, Hawaii, Idaho, Montana, MSSA (Rhode Island and Vermont), Oregon, Utah, and Wyoming. [b]Count excludes eight MOU items that do not align to the NGSS.

Appendix I

Adaptive Algorithm Design

TABLE OF CONTENTS

1. INTRODUCTION, BACKGROUND, AND DEFINITIONS 2
1.1 Blueprint 3
1.2 Content Value 4
1.2.1 Content Value for Single Items 5
1.2.2 Content Value for Sets of Items 6
1.3 Information Value 7
1.3.1 Individual Information Value 7
1.3.2 Binary Items 7
1.3.3 Polytomous Items 7
1.3.4 Item Group Information Value 10
2. ENTRY AND INITIALIZATION 11
2.1 Item Pool 11
2.2 Adjust Segment Length 11
2.3 Initialization of Starting Theta Estimates 11
2.4 Insertion Of Embedded Field-Test Items 12
3. ITEM SELECTION 13
3.1 Trimming The Custom Item Pool 13
3.2 Recycling Algorithm 14
3.3 ADAPTIVE ITEM SELECTION 14
3.4 Selection of The Initial Item 15
3.5 Exposure Control 15
4. TERMINATION 16
APPENDIX 1. DEFINITIONS OF USER-SETTABLE PARAMETERS 17
APPENDIX 2. SUPPORTING DATA STRUCTURES 19
ADDENDUM. ADJUSTMENTS TO THE USE OF ITEM CLUSTERS 20

Adaptive Item Selection Algorithm

1. Introduction, Background, and Definitions

This document describes the adaptive item selection algorithm. The item selection algorithm is designed to cover a standards-based blueprint, which may include content, cognitive complexity, and item type constraints. The item selection algorithm will also include:

- the ability to customize an item pool based on access constraints and screen items that have been previously viewed or may not be accessible for a given individual;
- a mechanism for inserting embedded field-test items; and
- a mechanism for delivering "segmented" tests in which separate parts of the test are administered in a fixed order.

This document describes the algorithm and the design for its implementation for the test delivery system (TDS). The implementation builds extensively on the algorithm implemented in the Cambium Assessment, Inc (CAI)'s TDS and incorporates substantial CAI intellectual property. CAI will release the algorithm and the implementation described here under the same open-source license under which the rest of the open-source system is released.

The general approach described here is based on a highly parameterized multiple-objective utility function. The objective function includes:

- a measure of content match to the blueprint;
- a measure of overall test information; and
- measures of test information for each reporting category on the test.

We define an objective function that measures an item's contribution to each of these objectives, weighting them to achieve the desired balance among them. Equation (1) sketches this objective function for a single item.

$$f_{ijt} = w_2 \frac{1}{\sum_{r=1}^{R} d_{rj}} \sum_{r=1}^{R} s_{rit} p_r d_{rj} + w_1 \sum_{k=1}^{K} q_k h_{1k}(v_{kijt}, V_{kit}, t_k) + w_0 h_0(u_{ijt}, U_{it}, t_0) \qquad (1)$$

where the term w represents user-supplied weights that assign relative importance to meeting each of the objectives d_{rj} indicates whether item j has the blueprint-specified feature r, and p_r is the user-supplied priority weight for feature r. The term s_{rit} is an adaptive control parameter that is described. In general, s_{rit} increases for features that have not met their designated minimum as the end of the test approaches.

The remainder of the terms represents an item's contribution to measurement precision:

- v_{kijt} is the value of item j toward reducing the measurement error for reporting category k for examinee i at selection t; and
- u_{ijt} is the value of item j in terms of reducing the overall measurement error for examinee i at selection t.

The terms U_{it} and V_{kit} represent the total information overall and on reporting category k, respectively.

The term q_k is a user-supplied priority weight associated with the precision of the score estimate for reporting category k. The terms t represent precision targets for the overall score (t_0) and each score reporting category score. The functions $h(.)$ are given by:

$$h_0(u_{ijt}, U_{it}, t_0) = \begin{cases} au_{ijt} \text{ if } U_{it} < t_0 \\ bu_{ijt} \text{ otherwise} \end{cases}$$

$$h_{1k}(v_{kijt}, V_{kit}, t_k) = \begin{cases} c_k v_{kijt} \text{ if } V_{kit} < t_k \\ d_k v_{kijt} \text{ otherwise} \end{cases}$$

Items can be selected to maximize the value of this function. This objective function can be manipulated to produce a pure, standards-free adaptive algorithm by setting w_2 to zero or a completely blueprint-driven test by setting $w_1 = w_0 = 0$. Adjusting the weights to optimize performance for a given item pool will enable users to maximize information subject to the constraint that the blueprint is virtually always met.

We note that the computations of the content values and information values generate values on very different scales, and that the scale of the content value varies as the test progresses. Therefore, we normalize both the information and content values before computing the value of Equation (1). This normalization is given by $x = \begin{cases} 1 \; if \; min = max \\ \frac{v-min}{max-min} \; otherwise \end{cases}$, where min and max represent the minimum and maximum, respectively, of the metric computed over the current set of items or item groups.

The remainder of this section describes the overall program flow, the form of the blueprint, and the various value calculations employed in the objective function. Subsequent sections describe the details of the selection algorithm.

1.1 BLUEPRINT

Each test will be described by a single blueprint for each segment of the test and will identify the order in which the segments appear. The blueprint will include:

- an indicator of whether the test is adaptive or fixed form;
- termination conditions for the segment, which are described in a subsequent section;
- a set of nested content constraints, each of which is expressed as:

- the minimum number of items to be administered within the content category;
- the maximum number of items to be administered within the content category;
- an indication of whether the maximum should be deterministically enforced (a "strict" maximum);
- a priority weight for the content category p_r;
- an explicit indicator as to whether this content category is a reporting category; and
- an explicit precision-priority weight (q_k) for each group identified as a reporting category.

- a set of non-nested content constraints, which are represented as:
  - a name for the collection of items meeting the constraint;
  - the minimum number of items to be administered from this group of items;
  - the maximum number of items to be administered from this group of items;
  - an indication of whether the maximum should be deterministically enforced (a "strict" maximum);
  - a priority weight for the group of items p_r;
  - an explicit indicator as to whether this named group will make up a reporting category; and
  - an explicit precision-priority weight (q_k) for each group identified as a reporting category.
  - The priority weights, p_r on the blueprint, can be used to express values in the blueprint match. Large weights on reporting categories paired with low (or zero) weights on the content categories below them may allow more flexibility to maximize information in a content category covering fewer fine-grained targets, while the reverse would mitigate toward more reliable coverage of finer-grained categories, with less content flexibility within reporting categories.

An example of a blueprint specification appears in Appendix J-1.

1.2 CONTENT VALUE

Each item or item group will be characterized by its contribution to meeting the blueprint, given the items that have already been administered at any point. The contribution is based on the presence or absence of features specified in the blueprint and denoted by the term d in Equation (1). This section describes the computation of the content value.

1.2.1 Content Value for Single Items

For each constraint appearing in the blueprint (r), an item i either does or does not have the characteristic described by the constraint. For example, a constraint might require a minimum of four and a maximum of six algebra items. An item measuring algebra has the described characteristic, and an item measuring geometry, but algebra does not. To capture this constraint, we define the following:

- d_j is a feature vector in which the elements are d_{rj}, summarizing item j's contribution to meeting the blueprint. This feature vector includes content categories such as claims and targets as well as other features of the blueprint, such as Depth of Knowledge (DOK) and item type.
- S_{it} is a diagonal matrix, the diagonal elements of which are the adaptive control parameters s_{rit}.
- p is the vector containing the user-supplied priority weights p_r.

The scalar content value for an item is given by $C_{ijt} = d_j' S_{it} p$.

Letting z_{rit} represent the number of items with feature r administered to student i by iteration t, the value of the adaptive control parameters is:

$$s_{rit} = \begin{cases} m_{it}\left(2 - \dfrac{z_{rit}}{Min_r}\right) & if\ z_r < Min_r \\ 1 - \dfrac{z_{rit} - Min_r}{Max_r - Min_r} & if\ \ Min_r < z_{rit} < Max_r \\ \left(Max_r - z_{rit}\right) - 1 & if\ \ Max_r \le z_{rit} \end{cases}$$

The blueprint defines the minimum (Min_r) and maximum (Max_r) number of items to be administered with each characteristic (r).

The term $m_{it} = \frac{T}{T-t}$ where T is the total test length. This has the effect of increasing the algorithm's preference for items that have not yet met their minimums as the end of the test nears and the opportunities to meet the minimum diminish.

This increases the likelihood of selecting items for content that has not met its minimum as the opportunities to do so are used up. The value s is highest for items with content that has not met its minimum, declines for items representing content for which the minimum number of items has been reached but the maximum has not, and turns negative for items representing content that has met the maximum.

1.2.2 Content Value for Sets of Items

Calculation of the content value of sets of items is complicated by two factors:

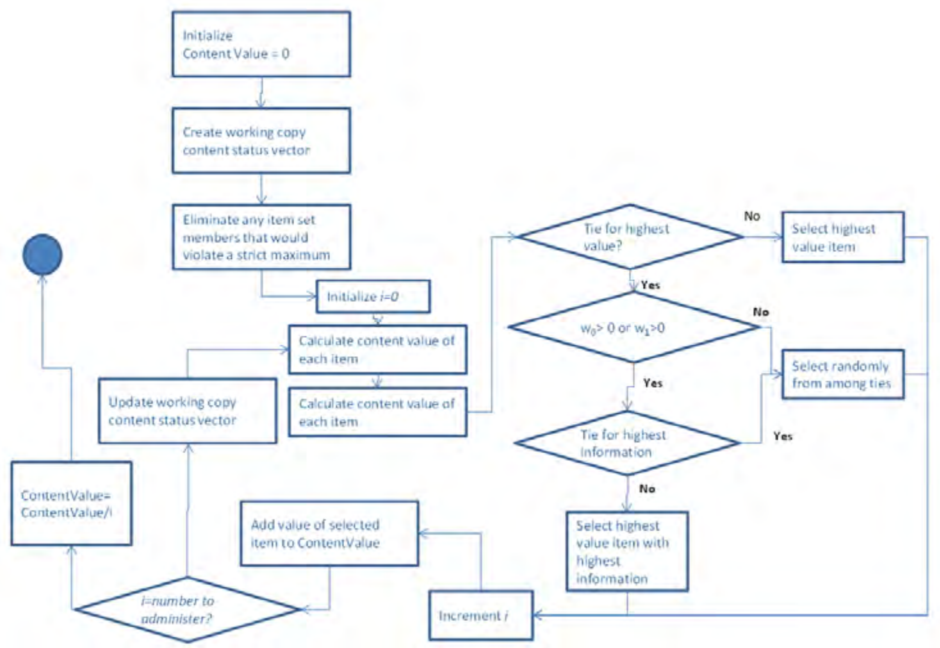
1. The desire to allow more items to be developed for each set and to have the most advantageous set of items administered.
2. The design objective of characterizing the information contribution of a set of items as the expected information over the working theta distribution for the examinee.

The former objective is believed to enhance the ability to satisfy highly constrained blueprints while still adapting to obtain good measurement for a broad range of students. The latter arises from the recognition that English Language Arts (ELA) tests will select one set of items at a time, without an opportunity to adapt once the passage has been selected.

The general approach involves successive selection of the highest content value item in the set until the indicated number of items in the set have been selected. Because the content value of an item changes with each selection, a temporary copy of the already-administered content vector for the examinee is updated with each selection such that subsequent selections reflect the items selected in previous iterations.

Exhibit A on the following page presents a flowchart for this calculation. Readers will note the check to determine whether $w_0 > 0$ or $w_1 > 0$. These weights, defined with Equation (1), identify the user-supplied importance of information optimization relative to blueprint optimization. In cases such as independent field tests, this weight may be set to zero, as it may not be desirable to make item administration dependent on the match to student performance. In more typical adaptive cases where item statistics will not be recalculated, favoring more informative items is generally better. The final measure of content value for the set of selected set of items is divided by the number of items selected to avoid a bias toward selection of sets with more items.

Exhibit A. Content Value Calculation for Item Sets



1.3 Information Value

Each item or item group also has value in terms of maximizing information, both overall and on reporting categories.

1.3.1 Individual Information Value

The information value associated with an item will be an approximation of information. The system will be designed to use generalized Item Response Theory (IRT) models; however, it will treat all items as though they offer equal measurement precision. This is the assumption made by the Rasch model, but in more general models, items known to offer better measurement are given preference by many algorithms. Subsequent algorithms are then required to control the exposure of the items that measure best. Ignoring the differences in slopes serves to eliminate this bias and help equalize exposure.

1.3.2 Binary Items

The approximate information value of a binary item will be characterized as $I_j(\theta) = p_j(\theta)(1 - p_j(\theta))$, where the slope parameters are artificially replaced with a constant.

1.3.3 Polytomous Items

In terms of information, the best polytomous item in the pool is the one that maximizes the expected information, $I_j(\theta)$. Formally, $I_j(\theta) > I_k(\theta)$ for all items $k \neq j$. The true value θ, however, remains unknown and is accessed only through an estimate, $\hat{\theta} \sim N(\bar{\theta}, \sigma_\theta)$. By definition of an expectation, the expected information $I_j(\theta) = \int I_j(t) f(t|\bar{\theta}, \sigma_\theta) dt$.

The intuition behind this result is illustrated in Exhibit B. In Exhibit B, each panel graphs the distribution of the estimate of θ for an examinee. The top panel assumes a polytomous item in which one step threshold (A1) matches the mean of the θ estimate distribution. In the bottom panel, neither step threshold matches the mean of the θ estimate distribution. The shaded area in each panel indicates the region in which the hypothetical item depicted in the panel provides more information. We see that approximately 2/3 of the probability density function is shaded in the lower panel, while the item depicted in the upper panel dominates in only about 1/3 of the cases. In this example, the item depicted in the lower panel has a much greater probability of maximizing the information from the item, despite the fact that the item in the upper panel has a threshold exactly matching the mean of the estimate distribution and the item in the lower panel does not.

Exhibit B. Two Example Items, with the Shaded Region Showing the Probability that the Item Maximizes Information for the Examinee Depicted

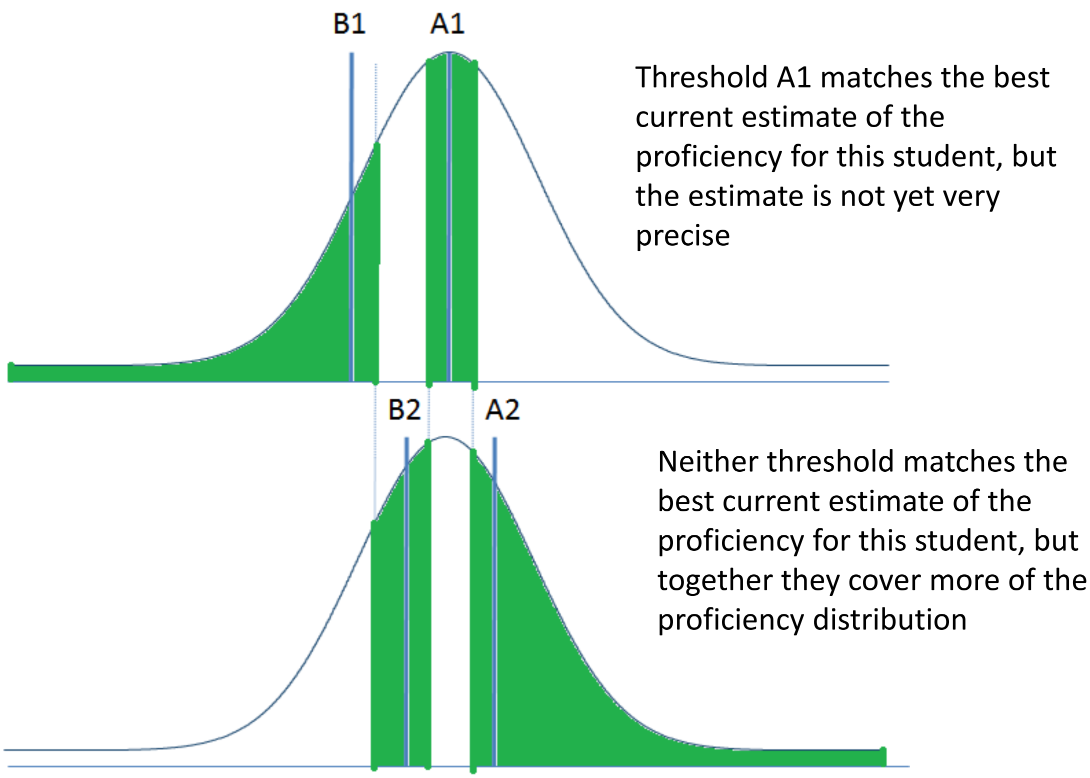
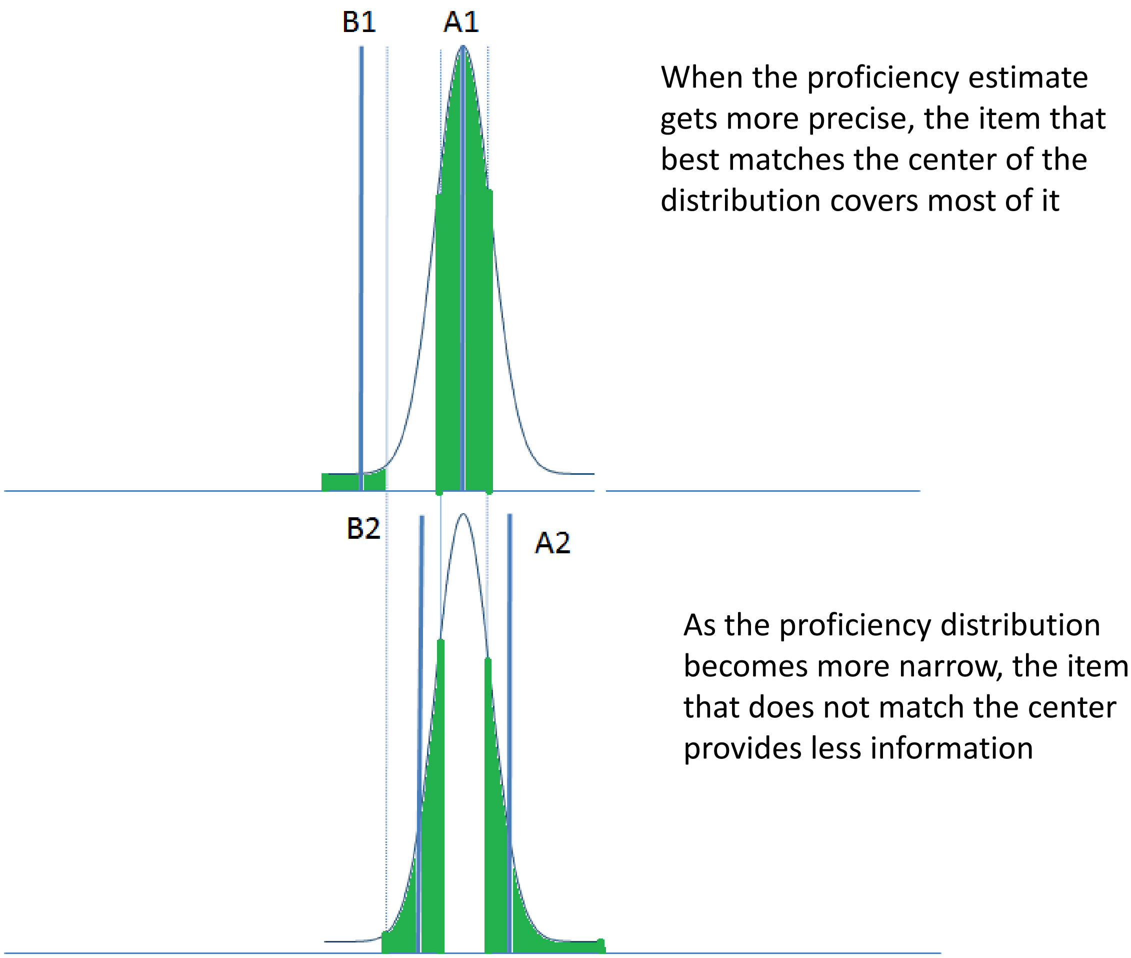


Exhibit C on the following page shows what happens to information as the estimate of this student's proficiency becomes more precise (later in the test). In this case, the item depicted in the top panel maximizes information about 65 to70 percent of the time, compared to about 30 to 35 percent for the item depicted in the lower panel. These are the same items depicted in the Exhibit B, but in this case, we are considering information for a student with a more precise current proficiency estimate.

Exhibit C. Two Example Items, with the Shaded Region Showing the Probability that the Item Maximizes Information for the Examinee Depicted



The approximate information value of polytomous items will be characterized as the expected information, specifically $E\left[I_j(\theta)|m_i, s_i\right] = \int \sum_{k=1}^{K} I_{jk}(t)\, p_j(k|t)\phi(t; m_i, s_i)dt$, where $I_{jk}(t)$ represents the information at t of response k to item j, $p_j(k|t)$ is the probability of response k to item j (artificially holding slope constant), given proficiency t, $\phi(.)$ represents the normal probability density function, and m_i and s_i represent the mean and standard deviation of examinee i's current estimated proficiency distribution.

We propose to use Gauss-Hermite quadrature with a small number of quadrature points (approximately five). Experiments show that we can complete this calculation for 1,000 items in fewer than 5 milliseconds, making it computationally reasonable.

As with the binary items, we propose to ignore the slope parameters to even exposure and avoid a bias toward the items with better measurement.

1.3.4 Item Group Information Value

Item groups differ from individual items in that a set of items will be selected for administration. Therefore, the goal is to maximize information across the working theta distribution. As with the polytomous items, we propose to use Gauss-Hermite quadrature to estimate the expected information of the item group.

In the case of multiple-item groups

$$E\left[I_g(\theta)|m_i, s_i\right] = \frac{1}{J_g}\int \sum_{j=1}^{J_g} I_{g(j)}(t)\,\phi(t; m_i, s_i)dt$$

Where $I_g(.)$ is the information from item group g, $I_{g(j)}$ is the information associated with item $j \in g$, for the J_g items in set g. In the case of polytomous items, we use the expected information, as described above.

2. Entry and Initialization

At startup, the system will

- create a custom item pool;
- initialize theta estimates for the overall score and each score point; and
- insert embedded field-test items.

2.1 Item Pool

At test startup, the system will generate a *custom item pool*, a string of item IDs for which the student is eligible. This item pool will include all items that

- are active in the system at test startup; and
- are not flagged as "access limited" for attributes associated with this student.

The list will be stored in ascending order of ID.

2.2 Adjust Segment Length

Custom item pools run the risk of being unable to meet segment blueprint minimums. To address this special case, the algorithm will adjust the blueprint to be consistent with the custom item pool. This capability becomes necessary when an accommodated item pool systematically excludes some content.

Let

$\boldsymbol{S}$ be the set of top-level content constraints in the hierarchical set of constraints, each consisting of the tuple (*name*, *min*, *max*, *n*);

$\boldsymbol{C}$ be the custom item pool, each element consisting of a set of content constraints $\boldsymbol{B}$;

$\boldsymbol{f}$, $\boldsymbol{p}$ integers represent item shortfall and pool count, respectively; and

$\boldsymbol{t}$ be the minimum required items on the segment.

For each s in $\boldsymbol{S}$, compute n as the sum of active operational items in $\boldsymbol{C}$ classified on the constraint.

$\boldsymbol{f}$ = summation over S $(min - n)$

$\boldsymbol{p}$ = summation over S (n)

if $t - f < p$, then $t = t - f$

2.3 Initialization of Starting Theta Estimates

The user will supply five pieces of information in the test configuration:

1. A default starting value if no other information is available

2. An indication whether prior scores on the same test should be used, if available
3. Optionally, the test ID of another test that can supply a starting value, along with
4. Slope and intercept parameters to adjust the scale of the value to transform it to the scale of the target test
5. A constant prior variance for use in calculation of working EAP scores

2.4 INSERTION OF EMBEDDED FIELD-TEST ITEMS

Each blueprint will specify

- the number of field-test items to be administered on each test;
- the first item position into which a field-test item may be inserted; and
- the last item position into which a field-test item may be inserted.

Upon startup, select randomly from among the field-test items or item sets until the system has selected the specified number of field-test items. If the items are in sets, the sets will be administered as a complete set, and this may lead to more than the specified number of items administered.

The probability of selection will be given by $p_j = \frac{\sum_{j=1}^{K} K_j}{\sum_{j=1}^{K} a_j K_j} a_j K_j \frac{m}{N_J}$, where

p_j represents the probability of selecting the item;

m is the targeted number of field-test items;

N_j is the total number of active items in the field-test pool;

K_j is the number of items in item set j; and

a_j is a user-supplied weight associated with each item (or item set) to adjust the relative probability of selection.

The a_j variables are included to allow for operational cases in which some items must complete field testing sooner or enter field testing later. While using this parameter presents some statistical risk, not doing so poses operational risks.

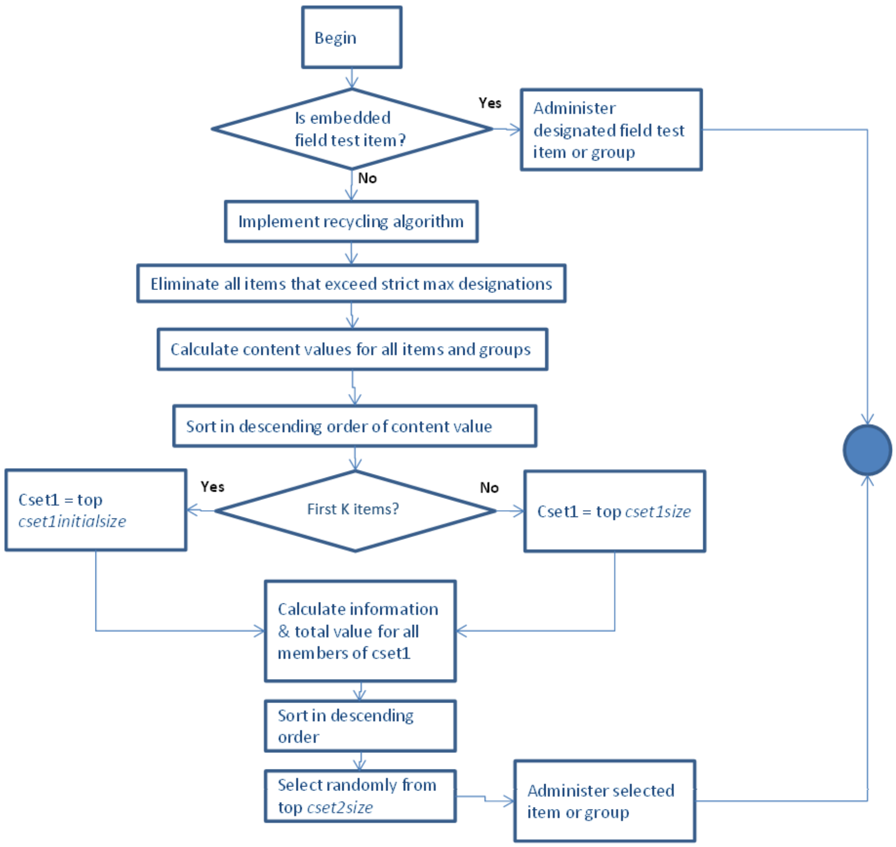
For each item set, generate a uniform random number r_j on the interval $\{0,1\}$. Sort the items in ascending order by $\frac{r_j}{p_j}$. Sequentially select items, summing the number of items in the set. Stop the selection of field-test items once $FTNMin \leq m \leq FTNMax = \sum_{j=0} K_j$.

Next, each item is assigned to a position on the test. To do so, select a starting position within $f - FTMax - FTMin$ positions from $FTMin$, where $FTMax$ is the maximum allowable position for field-test items and $FTMin$ is the minimum allowable position for field-test items. $FTNMin$ and $FTNMax$ refer to the minimum and maximum number of field-test items, respectively. Distribute the items evenly within these positions.

3. Item Selection

Exhibit D summarizes the item selection process. If the item position has been designated for a field-test item, administer that item. Otherwise, the adaptive algorithm kicks in.

Exhibit D. Summary of Item Selection Process



This approach is a "content first" approach designed to optimize match to blueprint. An alternative, "information first" approach, is possible. Under an information first approach, all items within a specified information range would be selected as the first set of candidates, and subsequent selection within that set would be based, in part, on content considerations. The engine is being designed so that future development could build such an algorithm using many of the calculations already available.

3.1 Trimming The Custom Item Pool

At each item selection, the active item pool is modified in four steps:

1. The custom item pool is intersected with the active item pool, resulting in a custom active item pool.
2. Items already administered on this test are removed from the custom active item pool.

3. Items that have been administered on prior tests are tentatively removed (see Section 3.2, Recycling Algorithm).

4. Items that measure content that has already exceeded a strict maximum are tentatively removed from the pool, removing entire sets containing items that meet this criterion.

3.2 Recycling Algorithm

When students are offered multiple opportunities to test, or when prior tests have been started and invalidated, students will have seen some of the items in the pool. The trimming of the item pool eliminates these items from the pool. It is possible that in such situations, the pool may no longer contain enough items to meet the blueprint.

Hence, items that have been seen on previous administrations may be returned to the pool. If there are not enough items remaining in the pool, the algorithm will recycle items (or item groups) with the required characteristic that is found in insufficient numbers. Working from the least recently administered group, items (or item groups) are reintroduced into the pool until the number of items with the required characteristics meets the minimum requirement. When item groups are recycled, the entire group is recycled rather than an individual item. Items administered on the current test are never recycled.

3.3 ADAPTIVE ITEM SELECTION

Selection of items will follow a common logic, whether the selection is for a single item or an item group. Item selection will proceed in the following three steps:

1. Select Candidate Set 1 (*cset1*).
    a. Calculate the content value of each item or item group.
    b. Sort the item groups in descending order of content value.
    c. Select the top *cset1size*, a user-supplied value that may vary by test.
2. Select Candidate Set 2 (*cset2*).
    a. Calculate the information values for each item group in *cset1*.
    b. Calculate the overall value of each item group in *cset1* as defined in Equation (1).
    c. Sort *cset2* in descending order of value.
    d. Select the top *cset2size* item groups, where *cset2size* is a user-supplied value that may vary by test.
3. Select the item or item group to be administered.
    a. Select randomly from *cset2* with uniform probability.

Note that a “pure adaptive” test, without regard to content constraints, can be achieved by setting *cset1size* to the size of the item pool and w_2, the weight associated with meeting content constraints in Equation (1), to zero. Similarly, linear on-the-fly tests can be constructed by setting w_0 and w_1 to zero.

3.4 Selection of The Initial Item

Selection of the initial item can affect item exposure. At the start of the test, all tests have no content already administered, so the items and item groups have the same content value for all examinees. In general, it is a good idea to spread the initial item selection over a wider range of content values. Therefore, we define an additional user-settable value, *cset1initialsize*, which is the size of Candidate Set 1 on the first K items only, where K is the number of reporting categories. Similarly, we define *cset2initialisize*.

3.5 Exposure Control

This algorithm uses randomization to control exposure and offers several parameters that can be adjusted to control the tradeoff between optimal item allocation and exposure control. The primary mechanism for controlling exposure is the random selection from $CSET2$, the set of items or item groups that best meet the content and information criteria. These represent the "top k" items, where k can be set. Larger values of k provide more exposure control at the expense of optional selection.

In addition to this mechanism, we avoid a bias toward items with higher measurement precision by treating all items as though they measured with equal precision by ignoring variation in the slope parameter. This has the effect of randomizing over items with differing slope parameters. Without this step, it would be necessary to have other *post hoc* explicit controls to avoid the overexposure of items with higher slope parameters, an approach that could lead to different test characteristics over the course of the testing window.

4. TERMINATION

The algorithm will have configurable termination conditions. These may include

- administering a minimum number of items in each reporting category and overall;
- achieving a target level of precision on the overall test score;
- achieving a target level of precision on all reporting categories; and
- achieving a score insufficiently distant from a specified score with sufficient precision (e.g., less than two standard errors below proficient). Cambium Assessment, Inc (CAI) envisions this being used in conjunction with other termination conditions to allow very high or very low achieving students to continue on to a segment that contains items from adjacent grades but barring other students from those segments.

We will define four user-defined flags indicating whether each of these is to be considered in the termination conditions (*TermCount*, *TermOverall*, *TermReporting*, *TermTooClose*). A fifth user-supplied value will indicate whether these are taken in conjunction or if satisfaction of any one of them will suffice (*TermAnd*). Reaching the minimum number of items is always a necessary condition for termination.

In addition, two conditions will each individually and independently cause termination of the test:

1. Administering the maximum number of items specified in the blueprint
2. Having no items in the pool left to administer

APPENDIX 1. DEFINITIONS OF USER-SETTABLE PARAMETERS

This appendix summarizes the user-settable parameters in the adaptive algorithm.

Parameter Name	Description	Entity Referred to by Subscript Index
w_0	Priority weight associated with overall information	N/A
w_1	Priority weight associated with reporting category information	N/A
w_2	Priority weight associated with match to blueprint	N/A
q_k	Priority weight associated with a specific reporting category	reporting categories
p_r	Priority weight associated with a feature specified in the blueprint (These inputs appear as a component of the blueprint.)	features specified in the blueprint
a	Parameter of the function $h(.)$ that controls the overall information weight when the information target has not yet been hit	N/A
b	Parameter of the function $h(.)$ that controls the overall information weight after the information target has been hit	N/A
c_k	Parameter of the function $h(.)$ that controls the information weight when the information target has not yet been hit for reporting category k	reporting categories
d_k	Parameter of the function $h(.)$ that controls the information weight after the information target has been hit for reporting category k	reporting categories
cset1size	Size of candidate pool based on contribution to blueprint match	N/A
cset1initialsize	Size of candidate pool based on contribution to blueprint match for the first K items or item sets selected	N/A
cset2size	Size of final candidate pool from which to select randomly	N/A
cset2initialsize	Size of candidate pool based on contribution to blueprint match and information for the first item or item set selected	
t_0	Target information for the overall test	N/A
t_k	Target information for reporting categories	reporting categories
startTheta	A default starting value if no other information is available	N/A
startPrevious	An indication of whether previous scores on the same test should be used, if available	N/A
startOther	The test ID of another test that can supply a starting value, along with startOtherSlope	N/A
startOtherSlope	Slope parameter to adjust the scale of the value to transform it to the scale of the target test	N/A

Parameter Name	Description	Entity Referred to by Subscript Index
startOtherInt	Intercept parameter to adjust the scale of the value to transform it to the scale of the target test	N/A
$FTMin$	Minimum position in which field-test items are allowed	N/A
$FTMax$	Maximum position in which field-test items are allowed	N/A
$FTNMin$	Target minimum number of field-test items	N/A
$FTNMax$	Target maximum number of field-test items	N/A
a_j	Weight adjustment for individual embedded field-test items used to increase or decrease their probability of selection	field-test items
AdaptiveCut	The overall score cutscore, usually proficiency, used in consideration of *TermTooClose*	
TooCloseSEs	The number of standard errors below which the difference is considered “too close” to the adaptive cut to proceed. In general, this will signal proceeding to a final segment that contains off-grade items.	
TermOverall	Flag indicating whether to use the overall information target as a termination criterion	N/A
TermReporting	Flag to indicate whether to use reporting category information target as a termination criterion	N/A
TermCount	Flag to indicate whether to use minimum test size as a termination condition	N/A
TermTooClose	Terminate if you are not sufficiently distant from the specified adaptive cut	
TermAnd	Flag to indicate whether the other termination conditions are to be taken separately or conjunctively	N/A

APPENDIX 2. SUPPORTING DATA STRUCTURES

Cambium Assessment, Inc (CAI) Cautions and Caveats

- Use of standard error termination conditions will likely cause inconsistencies between the blueprint content specifications, and the information criteria will cause unpredictable results, likely leading to failures to meet blueprint requirements.
- The field-test positioning algorithm outlined here is very simple and will lead to deterministic placement of field-test items.

ADDENDUM. ADJUSTMENTS TO THE USE OF ITEM CLUSTERS

Cambium Assessment, Inc (CAI) adjusted the adaptive algorithm to the use of item clusters as follows:

- Using marginal maximum likelihood estimator (MMLE) to update proficiency estimates, marginalizing out cluster effects.
- Normalizing the information by the number of assertions within an item, to avoid over-selection of item clusters and stand-alone items with more assertions.

Appendix J

WVGSA Science Assessment Item Pool

WVGSA Science Assessment Item Pool

Table J-1. Spring 2022 WVGSA Science Assessment Operational and Field-Test Item Pool by Performance Expectation, Grade 5

Science Discipline	Disciplinary Core Idea	Performance Expectation	ICCR Items	West Virginia Items	MOU Items	Total Pool Items
Earth and Space Sciences	ESS1	4-ESS1-1	2	1	5	8
		5-ESS1-1	1	0	6	7
		5-ESS1-2	5	0	3	8
	ESS2	3-ESS2-1	3	0	3	6
		3-ESS2-2	2	1	4	7
		4-ESS2-1	2	0	3	5
		4-ESS2-2	1	0	5	6
		5-ESS2-1	0	2	2	4
		5-ESS2-2	3	1	2	6
	ESS3	3-ESS3-1	3	0	2	5
		4-ESS3-1	3	0	1	4
		4-ESS3-2	1	1	6	8
		5-ESS3-1	3	1	3	7
Life Sciences	LS1	3-LS1-1	1	0	4	5
		4-LS1-1	8	1	5	14
		4-LS1-2	1	2	3	6
		5-LS1-1	2	2	3	7
	LS2	3-LS2-1	4	0	6	10
		5-LS2-1	1	0	3	4
	LS3	3-LS3-1	2	1	4	7
		3-LS3-2	1	0	2	3
	LS4	3-LS4-1	2	0	8	10
		3-LS4-2	7	0	3	10
		3-LS4-3	4	0	2	6
		3-LS4-4	2	0	2	4
Physical Sciences	PS1	5-PS1-1	2	0	5	7

Science Discipline	Disciplinary Core Idea	Performance Expectation	ICCR Items	West Virginia Items	MOU Items	Total Pool Items
		5-PS1-2	2	0	4	6
		5-PS1-3	4	0	3	7
		5-PS1-4	1	1	2	4
	PS2	3-PS2-1	2	1	6	9
		3-PS2-2	3	0	2	5
		3-PS2-3	1	2	2	5
		3-PS2-4	0	0	2	2
		5-PS2-1	1	1	4	6
	PS3	4-PS3-1	4	0	3	7
		4-PS3-2	4	1	3	8
		4-PS3-3	2	0	3	5
		4-PS3-4	1	0	2	3
		5-PS3-1	2	0	3	5
	PS4	4-PS4-1	1	0	3	4
		4-PS4-2	1	0	6	7
		4-PS4-3	1	0	0	1
Total			**96**	**19**	**143**	**258**

Note. [a]Other MOU states include Connecticut, Hawaii, Idaho, MSSA (Rhode Island and Vermont), Oregon, Utah, and Wyoming.

Table J-2. Spring 2022 WVGSA Science Assessment Operational and Field-Test Item Pool by Performance Expectation, Grade 8

Science Discipline	Disciplinary Core Idea	Performance Expectation	ICCR Items	West Virginia Items	MOU Items	Total Pool Items
Earth and Space Sciences	ESS1	6-MS-ESS1-1	5	0	2	7
		6-MS-ESS1-2	3	1	1	5
		6-MS-ESS1-3	2	0	3	5
		7-MS-ESS1-4	2	0	4	6
	ESS2	6-MS-ESS2-5	1	1	3	5
		6-MS-ESS2-6	2	0	1	3
		7-MS-ESS2-1	2	0	5	7
		7-MS-ESS2-2	2	1	4	7
		7-MS-ESS2-3	3	0	6	9
		7-MS-ESS2-4	1	0	4	5
	ESS3	6-MS-ESS3-2	2	1	4	7
		6-MS-ESS3-5	2	0	3	5
		7-MS-ESS3-1	1	0	1	2
		7-MS-ESS3-3	1	0	2	3
		8-MS-ESS3-4	2	0	4	6
Life Sciences	LS1	6-MS-LS1-6	2	2	2	6
		6-MS-LS1-7	2	0	2	4
		7-MS-LS1-1	0	0	3	3
		7-MS-LS1-2	2	0	5	7
		7-MS-LS1-3	0	1	3	4
		7-MS-LS1-8	2	0	3	5
		8-MS-LS1-4	2	0	2	4
		8-MS-LS1-5	2	0	3	5
	LS2	6-MS-LS2-1	4	1	5	10
		6-MS-LS2-2	2	0	3	5
		6-MS-LS2-3	1	0	4	5
		6-MS-LS2-4	7	0	5	12
		6-MS-LS2-5	1	0	5	6

Science Discipline	Disciplinary Core Idea	Performance Expectation	ICCR Items	West Virginia Items	MOU Items	Total Pool Items
	LS3	8-MS-LS3-1	1	1	3	5
		8-MS-LS3-2	2	1	5	8
	LS4	8-MS-LS4-1	5	0	3	8
		8-MS-LS4-2	0	0	5	5
		8-MS-LS4-3	1	0	3	4
		8-MS-LS4-4	3	0	2	5
		8-MS-LS4-5	1	0	1	2
		8-MS-LS4-6	2	0	1	3
Physical Sciences	PS1	8-MS-PS1-1	1	0	3	4
		8-MS-PS1-2	3	0	5	8
		8-MS-PS1-3	0	1	2	3
		8-MS-PS1-4	1	0	7	8
		8-MS-PS1-5	1	1	4	6
		8-MS-PS1-6	1	0	2	3
	PS2	7-MS-PS2-1	1	0	3	4
		7-MS-PS2-2	3	0	4	7
		7-MS-PS2-3	1	0	1	2
		7-MS-PS2-4	0	2	2	4
		7-MS-PS2-5	0	0	5	5
	PS3	7-MS-PS3-1	1	0	2	3
		7-MS-PS3-2	3	1	3	7
		7-MS-PS3-3	3	1	2	6
		7-MS-PS3-4	1	0	1	2
		7-MS-PS3-5	4	0	2	6
	PS4	6-MS-PS4-1	4	0	4	8
		6-MS-PS4-2	5	1	3	9
		6-MS-PS4-3	0	1	1	2
Total			**106**	**18**	**171**	**295**

Note. [a]Other MOU states include Connecticut, Hawaii, Idaho, MSSA (Rhode Island and Vermont), Oregon, Utah, and Wyoming.