Bureau of Indian Education 2022–23 Technical Report

Grades 5, 8, and 11 Science

Prepared by Cognia and the Bureau of Indian Education

cognia™

Table Of Contents

Appendix B   Performance Level Descriptors

Appendix C   Accommodation Frequencies

Appendix D   Scorer Qualification Rates

Appendix E   Classical Item Statistics

Appendix F   Item Response Theory Parameters

Appendix G   Test Characteristic Curves & Test Information Functions

Appendix H   Raw to Scaled Score Lookup Tables

Appendix I   Decision Accuracy and Consistency Results

Appendix J   BIE Science Reporting Business Requirements

Appendix K   Score Report Interpretation Quick Guide

Appendix L   Cumulative Scaled Score Distributions

Appendix M   Scaled Score Descriptive Statistics

List of Tables and Figures

Chapter 1. Introduction to the Assessment Program

1.1 Purposes and Uses of the Assessment Program

The Bureau of Indian Education (BIE—see Appendix A for a list of acronyms) science assessment is a summative assessment for science, administered to students in grades 5, 8, and 11. It is designed to provide evidence to determine a student's grade-level proficiency and progress toward college and/or career readiness, as defined by the Bureau, by showing BIE students have mastered the Next Generation Science Standards. The BIE Science Assessment is a key component of BIE's Every Student Succeeds Act (ESSA) plan to meet ESSA's general assessment requirements.

As the BIE Science Assessment is an end-of-grade/end-of-course single measure, interpretations and uses of its scores should be supplemented with additional measures, including information from classroom summative, interim, and formative assessments in science. In keeping with the practices set forth in *Standards for Educational and Psychological Testing*, each student's score should be used as part of a body of evidence regarding mastery and should not be used in isolation to make high-stakes decisions (AERA, APA, & NCME, 2014). Hence, aggregation of student scores on the BIE Science Assessment at the school or district levels, or for the entire BIE, is generally a more reliable indicator of program success, particularly when monitored over the course of several years.

The BIE Science Assessment is designed to provide point-in-time information about the academic achievement and progress of students. Student results are reported according to academic achievement descriptors utilizing scale scores for each of four performance levels: Novice, Nearing Proficiency, Proficient and Advanced. The results from these assessments provide educators and the public with information to guide the creation of future educational practices to meet the needs of students, while monitoring the continuous improvement efforts of schools, districts, and the BIE in achieving a world-class education system for all students.

1.2 Statements of Intended Score Interpretations and Uses (SIUs)

The phrase "intended score interpretations for uses" appears several times in *Standards for Educational and Psychological Testing* and is the core of the field's views on validity and validation. For the BIE Science Assessment and other assessment programs, the phrase refers broadly to test scores (e.g., total test scale scores, aggregations of test scores, the percentages of students at or above standard), and other test performance information elements, such as the definition of "at or above standard" in the performance level descriptors (PLDs—see Appendix B).

1.2.1 Primary Intended BIE Science Assessment Score Interpretations and Uses

Primary Score Interpretations:

- For Elementary and Middle Schools, performance on the BIE Science Assessment indicates student mastery of grade levels 3–5 and 6–8 expectations for integration of Science and Engineering Practices (SEP), Disciplinary Core Ideas (DCI), and Crosscutting Concepts (CCC) as presented in the standards, which is the progression for the next level of science curriculum and is a predictor of being on track for college and career readiness.
- For High School, the BIE Science Assessment is designed to measure whether students are on track to be ready for college or career, as defined by the BIE, by showing they have mastered the Next Generation Science Standards, which require the integration of Science and Engineering Practices, Disciplinary Core Ideas, and Crosscutting Concepts to explain phenomena and solve problems.

Primary Score Uses:

Educators, administrators, and other stakeholders at the school level can use the Science Assessment and its results to (a) monitor trends in student performance, (b) design professional development for teachers, and (c) drive accountability results.

- Teachers can use the BIE Science Assessment and its results to better integrate assessment with their instructional planning.
- Parents can use the BIE Science Assessment and its results to get information about (a) what their child knows and can do, and (b) their child's progress over time.

The intended score interpretation and uses stated here align with the original statements of intended score interpretations and uses in the National Center and State Collaborative 2015 Operational Assessment Technical Manual.

The BIE Science Assessment is designed, developed, and implemented to support three intended SIUs, according to the broad interpretation of the phrase above. These interpretations and uses are applicable to assessments in general and to specific applications with individual students and groups of students, as described below.

SIU 1: Intended Score Interpretation

The BIE Science Assessment provides reliable and valid information about important knowledge and skills in grade-level science attained by general education students.

- Claim 1.1: The content of the tests represents the content of the standards.
- Claim 1.2: The test items are construct-relevant.
- Claim 1.3: Test scores on the BIE Science Assessments provide reliable information about student performance and accurate classifications into performance levels.
- Claim 1.4: Item and test scoring are implemented accurately; approved scoring rules are implemented accurately.

SIU 2: Intended Score Use for Individual Students

Scale scores can be used to compare an individual student's performance to the performance of other students in BIE.

- Claim 2.1: Educators and other stakeholders at the school level can use results from the BIE Science Assessment to describe and monitor student achievement status with respect to mastery of the content standards.

SIU 3: Intended Score Use for Groups of Students

SIU statements for groups of students are applicable to aggregate reporting of student subgroups (e.g., English learners, students with disabilities, racial/ethnic subgroups) within those levels of aggregation.

- Claim 3.1: Educators can use results from the BIE Science Assessment to support instructional planning for groups of students.
- Claim 3.2: School and BIE stakeholders can use results from the BIE Science Assessment to make comparisons between organizations.

Claims, subclaims, and evidence that support the intended interpretations and uses of BIE Science Assessment scores are provided in Chapter 11.

1.3 Introduction to Validity Arguments for the Program: Rationales for the Approach

This report documents test development procedures and psychometric outcomes for the 2023 BIE Science Assessment. These technical aspects of the program contribute to the accumulation of validity evidence to support BIE Science Assessment score interpretations and uses. Because the interpretations of test scores, not the test itself, are evaluated for validity, this report presents documentation to substantiate intended interpretations (AERA et al., 2014). Subsequent chapters of this report discuss test development, test alignment, test administration, scoring, equating, item analyses, reliability, scaled scores, performance levels, and reporting. Each of these topics contributes important information toward establishing the validity of the assessment program. Note, however, that this report does not include certain aspects of a comprehensive validity argument that could be important to consider when drawing conclusions about validity. For example, additional sources of validity evidence might speak to the extent to which BIE Science Assessment scores converge with other measures of the same or similar constructs and diverge from measures of different constructs and consequences that arise from scores at the student, school, and district levels, as well as from scores for BIE as a whole.

The *Standards for Educational and Psychological Testing* (AERA et al., 2014) also gives a framework for describing sources of evidence that should be considered when constructing a validity argument. These sources include evidence based on the following five areas: test content, response processes, internal structure, relationship to other variables, and consequences of testing. These sources address different aspects of supporting evidence for validity arguments; they are not distinct types of validity. Instead, each contributes to a body of evidence about the comprehensive validity of score interpretations and uses and the intended interpretations and uses. Moreover, these sources represent only a partial list of sources of evidence from the BIE Science Assessment design, development, test administration, analysis, and reporting processes that are relevant to the overall validity arguments for intended interpretations and uses of BIE Science Assessment scores and other information.

Chapter 2. Overview of the Program

2.1 History of the Program

BIE's mission is to provide quality education opportunities from early childhood through life in accordance with a tribe's needs for cultural and economic well-being, in keeping with the wide diversity of Indian tribes as distinct cultural and governmental entities. To better support this mission, the BIE began the process in 2020 of administering a unique science assessment. BIE Science was scheduled to have its first administration in Spring 2021, and due to unforeseen constraints, used the New Mexico Assessment of Science Readiness assessments. Beginning in 2022, BIE administered a unique science assessment separate from any other state, the BIE Science Assessment. After its first administration in Spring 2022, BIE Science had its second administration in Spring 2023 and continues to work with Cognia to continue developing culturally relevant items and test forms. BIE Schools have the option to test both online and on paper, depending on their specific technology access and needs.

The window for the 2023 BIE Science administration with Cognia was April 3–May 12, 2023.

2.2 Stakeholder Involvement

Cognia and the BIE work together on all aspects of the implementation of the BIE Science Assessment program. The BIE also works with several stakeholder groups for input into the implementation of the program.

2.2.1 Educator Committees

Educator committees are periodically convened for the purpose of content development. The committees include those listed below, with the details of each committee found in chapter 4.

- National Item Review Committee: Cognia convenes a national item review committee to review the content of the items that are created. BIE educators comprise two seats per grade/content span for those committees.
- National Bias Review Committee: Cognia convenes a national bias committee to look for bias and sensitivity concerns in the content that is created. BIE educators comprise two seats on that committee.
- National Range Finding Review Committee: Cognia convenes a national range finding committee to review a range of student responses to open-ended field-tested items, assign score points in accordance with the item's rubric, and establish basic scoring rules and guidelines for each item. BIE educators comprise two seats on that committee.

2.3 Student Participation

BIE policy defines student participation on a BIE Science Assessment as attempting five or more items on the given assessment. Table 2-1 provides participation rates, as a function of assessment subject (science), and background/demographic variable. The number of students participating in the BIE Science Assessment in Spring 2023 per grade ranged from approximately 900 to 1,900.

Table 2-1 Participation Number by Grade, as a function of demographic variable

Type	Grade 5	Grade 8	Grade 11
All	1,937	1,662	870
Female	924	788	429
Male	1,011	874	441
Gender Undefined	2	0	0
American Indian or Alaska Native	1,920	1,642	848
Asian	4	6	2
Black	1	1	1
Multi-racial	8	7	1
Currently Receiving LEP Services	313	209	150
Not Receiving LEP Services	310	278	198
Special Ed	320	262	129
Non-Special Ed	1,613	1,396	721
Economically Disadvantaged Students	1,867	1,601	817
Non-Economically Disadvantaged Students	40	40	27
Gifted Students	104	129	59
Non-Gifted Students	89	87	77
Title1 Students	1,874	1,609	817
Non-Title1 Students	42	35	30

The BIE Science Assessments were administered in either computer-based or paper-based formats. Most students utilized computer-based administration. Table 2-2 contains the number of students utilizing computer-based or paper-based administrations. Additionally, the usage statistics on accommodation(s) and accessibility feature(s), as a function of grade are available in Appendix C. Only students who met the attemptedness rule (i.e., attempted 5 or more items) contributed to the frequencies in the aforementioned tables. Of the students that participated in the Spring 2023 administration, the table below indicates numbers of students who were assessed in each mode.

Table 2-2. Number of Participating Students, as a Function of Grade

Grade	Computer-Based	Paper-Based
5	1731	206
8	1562	100
11	788	82

The small numbers of students per grade in Spring 2023 (the total number of BIE students who participated in the 2023 science assessments were less than 2,000 per grade) impacted the analyses reported in this document in a variety of ways, such as increased levels of sampling, measurement, and/or estimation error in test reliabilities (overall and subgroup), decision accuracies and consistencies, classical item statistics, dimensionality effect sizes, scaled score distributions, and performance level distributions. .

Chapter 3. Test Content

3.1 Content Standards

Test content, including items and stimuli, for the BIE Science Assessment was developed according to the *Next Generation Science Standards*. These standards are the basis for the test designs developed for each grade and are used to inform the development of items. Each item is designed to measure a specific standard, or performance expectation, and align to multiple dimensions of the standard (Disciplinary Core Idea, Science and Engineering Practice, Crosscutting Concept).

The specific content standards were subsequently grouped into categories for the purpose of communicating with students, families, and educators. The content standards that are eligible to be included in the BIE Science Assessment are described in the following sections.

3.1.1 Eligible Standards

The BIE Science Assessment assesses the Next Generation Science Standards at grades 5, 8, and 11. All standards are eligible for assessment. However, because of the number of standards per grade, not all standards can be tested every year. The design of the BIE Science Assessment allows for all assessable standards to be included in the assessment at least once within a three-year period.

3.2 Assessment Design

Table 3.1 provides a summary of the number of items and points by item type, usage (i.e., operational items or field-test items), and estimated testing time for each grade level and content area of the BIE Science Assessment. The test is administered in three sessions. Test forms contain core operational items, matrix operational items, and matrix field-test items. There are two sets of operational items, set A and set B, differing in the standalone MS-2 items that are in the set (but still following the same content blueprint), to support sufficient assessment of all content standards over time. All operational items count toward student score, with the core operational items being common across both forms and the matrix operational items being administered across the two different operational forms. Matrix field-test items are items administered to subsets of students to "try out" performance (with different students receiving different field-test items), and therefore do not count toward student score. The types of items on the BIE Science Assessment are 1-point machine-scored items (MS–1), 2-point machine scored items (MS-2), and 4–point constructed-response items (OE-4). Additional item-type descriptions can be found in section 3.2.4.

Table 3-1. Student Testing Experience

Grades 5, 8	Cluster/Passage Items			Standalone Items			
	Stimulus/Passage	MS-1	MS-2	MS-2	OE	Total Items	Total Points
Core Operational Items	6	12	12	0	3	27	48
Matrix Operational Items	0	0	0	8	0	8	16
Matrix Field Test Items	2	4	4	4	1	13	24
Total Student Experience	**8**	**16**	**16**	**12**	**4**	**48**	**88**
					Estimated Testing Time (min)		**150 (50/50/50)**

Grade 11	Cluster/Passage Items			Standalone Items			
	Stimulus/Passage	MS-1	MS-2	MS-2	OE	Total Items	Total Points
Core Operational Items	6	12	12	0	3	27	48
Matrix Operational Items	0	0	0	10	0	10	20
Matrix Field Test Items	2	4	4	5	1	14	26
Total Student Experience	**8**	**16**	**16**	**15**	**4**	**51**	**94**
					Estimated Testing Time (min)		**165 (55/55/55)**

BIE Science Assessment Specifications

The reporting categories for the BIE Science Assessment are based on the science domains in the Next Generation Science Standards. Target percentages for the distribution of operational test points for each of the reporting categories reflect the distribution in the standards, so as not to over or underrepresent content. Because of their small percentage of standards overall, items aligned to standards in Engineering, Technology, and Applications of Science are reported under the reporting category domain that matches the context of the design problem presented. Percentages for the three reported domains are shown in the tables in the next section.

Specifications for the full test blueprints for the construction of the operational forms reflect the reporting category specifications. These constructs represent key aspects of the standards to which items are aligned; as such, the percentage of operational test points for each should be maintained from year to year. Note that some of the points for each reporting category come from clusters (a grouping of four items - 2 MS-1 and 2 MS-2 - all associated with a common stimulus) and some points come from standalone/discrete items.

Form 1 is used to create print and accommodated forms. On the print (PBT) form, any technology-enhanced item is replaced with an equivalent multiple-choice or multi-select item for administration. All other items on the PBT form remain the same as the online form.

Table 3-2. Grades 5, 8, 11 – BIE Science Assessment Operational Test Blueprint

Grade 5					
Reporting Category	Ideal Number of Clusters	Ideal Number of Standalone MS-2	Ideal Number of Standalone OE	Ideal Number of Core Points	Ideal Percent of Core Points (+/- 4%)
Practices and Crosscutting Concepts in Physical Sciences	2	4-6	1	24-28	40%
Practices and Crosscutting Concepts in Life Sciences	2	1-3	1	18-22	30%
Practices and Crosscutting Concepts in Earth and Space Sciences	2	1-3	1	18-22	30%
Grade 8					
Reporting Category	Ideal Number of Clusters	Ideal Number of Standalone MS-2	Ideal Number of Standalone OE	Ideal Number of Core Points	Ideal Percent of Core Points (+/- 4%)
Practices and Crosscutting Concepts in Physical Sciences	2	2-4	1	20-24	35%
Practices and Crosscutting Concepts in Life Sciences	2	2-4	1	20-24	35%
Practices and Crosscutting Concepts in Earth and Space Sciences	2	1-3	1	18-22	30%
Grade 11					
Reporting Category	Ideal Number of Clusters	Ideal Number of Standalone MS-2	Ideal Number of Standalone OE	Ideal Number of Core Points	Ideal Percent of Core Points (+/- 4%)
Practices and Crosscutting Concepts in Physical Sciences	2	3-5	1	22-26	35%
Practices and Crosscutting Concepts in Life Sciences	2	3-5	1	22-26	35%
Practices and Crosscutting Concepts in Earth and Space Sciences	2	1-3	1	18-22	30%

3.2.1 Content Coverage Blueprint

The distribution of emphasis for BIE content standards in Science for the Spring 2023 assessment is shown in Table 3-3. Assessable standards cover physical sciences, life sciences, earth and space sciences, and engineering, technology, and applications of science (ETS).

Table 3-3. Distribution of Emphasis Across Content Standards in Terms of Percentage of Total Test Points by Grade—Science Grades 5, 8, 11—Spring 2023*

Standards Category	Grade 5		Grade 8		Grade 11	
	Total Points	% of Total Core Points	Total Points	% of Total Core Points	Total Points	% of Total Core Points
Physical Sciences	24	37.50%	22	34.38%	22	32.35%
Life Sciences	20	31.25%	22	34.38%	24	35.29%
Earth and Space Sciences	20	31.25%	20	31.25%	22	32.35%
ETS	0	0%	0	0%	0	0%
Grand Total	64	100.00%	64	100.00%	68	100.00%

* 5 points are also coded to ETS in G5, 3 points in G11 and G8.

3.2.2 Operational Section

As mentioned at the start of section 3.2, there are two sets of operational items, set A and set B, differing in the standalone MS-2 items that are in the set (but still following the same content blueprint). Table 3-4 shows the reporting categories in the BIE Science Assessment test design, and the maximum possible number of raw-score points students could earn in each reporting category on the Spring 2023 assessment for both operational forms. Note: Because only operational items are counted toward students' scaled scores, only operational items are reflected in this table. The number of items and item types that are used to achieve these distributions are provided in the tables at the beginning of section 3.2. Any items aligned to standards in ETS are reported under the reporting category domain that matches the context of the design problem presented.

Table 3-4 Distribution of Raw-Score Points Across Reporting Categories by Grade—Science Grades 5, 8, 11—Spring 2023

Reporting Category	Grade 5		Grade 8		Grade 11	
	Total Points	% of Total Core Points	Total Points	% of Total Core Points	Total Points	% of Total Core Points
Practices and Crosscutting Concepts in Physical Sciences	24	37.50%	22	34.38%	22	32.35%
Practices and Crosscutting Concepts in Life Sciences	20	31.25%	22	34.38%	24	35.29%
Practices and Crosscutting Concepts in Earth and Space Sciences	20	31.25%	20	31.25%	22	32.35%
Grand Total	**64**	**100.00%**	**64**	**100.00%**	**68**	**100.00%**

3.2.3 Field-Test Sections

All items are appropriately field-tested prior to operational use. Items for the BIE Science Assessment came from a national item bank, and these items were field tested with BIE students and/or other students using the same item bank content. For the field testing done within the BIE Science Assessment forms, the test utilizes a matrix design that embeds the field-test items within each form. Matrix field-test items are items administered to subsets of students to "try out" performance (with different students receiving different field-test items), and therefore do not count toward student scores. In grades 5 and 8, the BIE Science Assessment contains a total of 13 field-test items per form: two clusters (with four items each), four MS-2 standalones, and one OE-4. The grade 11 test contains 14 field-test items per form: two clusters (with four items each), five MS-2 standalones, and one OE-4.

3.2.4 Item Types

Item types are chosen to best balance the desires for making efficient use of limited testing time and providing coverage of a broad range of knowledge and skills. The item types used on the BIE Science Assessment and the functions of each are listed below.

The item types on the BIE Science Assessment include machine-scored 1-point items (MS-1), machine-scored 2-point items (MS-2), and open-ended items (OE-4). Some of the MS-1 and MS-2 items are grouped together in clusters.

MS-1 items may be multiple choice, multiple select, or technology enhanced (e.g., drag-and-drop, hot spot, drop-down selections). MS-1 items are only found in clusters. They are all machine-scored as correct only; partial credit is not awarded.

MS-2 items have two parts (Part a and Part b) for students to answer. These items may combine multiple choice, multiple select, and/or technology-enhanced interactions across the two parts. MS-2 items are included in clusters and as standalone items. They are all machine scored, and students may earn 2, 1, or 0 points across Part a and Part b.

An item cluster is a set of items all associated with a common stimulus. Clusters contain four items, with two of the items being worth 1 point (MS-1) and two of the items being worth 2 points (MS-2). The clusters typically align to two PEs, and all clusters measure all three dimensions of the PEs being assessed.

Open-ended items (OE-4) are worth 4 points. These items require students to write an extended response to a prompt. The prompt may be a single prompt, or more typically, the items are written with multiple, scaffolded parts for students to respond to. These items are hand-scored, with scorers using a rubric and scoring notes to evaluate responses on a scale from 0–4.

Each type of item on the assessment is worth a specific number of points in the student's total Science score, as shown in Table 3-5.

Table 3-5. Science Item Types

Item Type	Maximum Number of Points Available
MS-1	1
MS-2	2
OE-4	4

3.2.5 Passage Types

On the BIE Science Assessment, all clusters are written with an extended, rich stimulus that may be one or more paragraphs in length. The stimulus must present a single, rich science phenomenon or engineering design problem aligned to the standards/performance expectations being assessed. The phenomenon or problem must launch and support a single storyline, or sequence of sense-making, which is carried out in the items.

The stimulus may present any variety of elements to provide the necessary information related to the phenomenon or problem and the storyline: text paragraphs, graphs, data tables, models, drawings, etc. All information in the stimulus should be necessary, but not conceptually sufficient, for students to respond (i.e., students must also use their own knowledge of the constructs in the standards to answer the items, rather than simply identify given information), and the stimulus must provide enough information to allow students to engage in the SEPs, DCIs, and CCCs of the targeted standards as they respond to items.

Chapter 4. Test Development

4.1 Overview of General Approach

This chapter provides an overview of the development of the BIE Science Assessment, including test and item specifications, item reviews, and test construction.

According to the Standards for Educational and Psychological Testing (AERA, APA, & NCME, 2014), "important validity evidence can be obtained from an analysis of the relationship between a test's content and the construct it is intended to measure" (p. 14). Accordingly, the descriptions of the test development procedures included in this chapter provide evidence that supports both the content and construct validity of the assessments.

4.2 Item Specifications

The test design for the BIE Science Assessment is based on the three content domains of Physical Sciences, Life Sciences, and Earth and Space Sciences. Items are expected to align to the multiple dimensions of the standards (Disciplinary Core Ideas, Science and Engineering Practices, Crosscutting Concepts) in each domain, such that every item is at least two-dimensional, if not three-dimensional. To emphasize this multi-dimensional nature of the items, the names of the reporting categories incorporate the three dimensions (Practices and Crosscutting Concepts in Physical Sciences, Practices and Crosscutting Concepts in Life Sciences, Practices and Crosscutting Concepts in Earth and Space Sciences). Students are expected to demonstrate sense-making by using core ideas, practices, and crosscutting concepts together to respond to items.

Items assessing Engineering, Technology, and Applications of Science are reported within the Physical, Life, or Earth and Space Sciences category, depending on the content match of the design problem presented in the item.

As content support, students taking the Grade 11 test are provided with a Periodic Table reference sheet. No items on the assessment require a calculator or other mathematical tools to answer.

Cognitive Complexity

In addition to being created according to subject-area content standards, each item on the BIE Science Assessment is assigned information about its cognitive complexity.

Because the items on the BIE Science Assessment are NGSS-aligned, the cognitive complexity of the items is evaluated with a different framework than Depth of Knowledge (DOK). This framework, Cognitive Complexity Framework for SSIB, is based on Achieve's *A Framework to Evaluate Cognitive Complexity in Science* (September 2019).

Under the Cognitive Complexity Framework for SSIB, four indicators are used to classify the cognitive complexity of each item: stimulus, science and engineering practice, disciplinary core idea, and crosscutting concept. For each indicator, the classification in terms of high, medium, or low complexity is based on how the students are using the indicator to respond to the item – specifically, to what degree does students' engagement with the indicator contribute to the level of sense-making required by the item.

The evaluation of cognitive complexity is done at the individual item level. For an operational BIE Science Assessment form, after summing the operational points that reflect cognitive complexity at each

complexity level across all four indicators, the target distribution is that at least 10% of the points should be high cognitive complexity and no more than 35% of the points should be low cognitive complexity.

4.3 Item Review Committees and Processes

Items used on the BIE Science Assessment are developed to measure achievement on BIE's science standards, the Next Generation Science Standards. Cognia content specialists, in collaboration with BIE, ensure this alignment, and ongoing independent evaluations are held to verify alignment. In addition, independent reviews are scheduled to ensure that items and passages conform to bias and sensitivity guidelines.

4.3.1 Content and Item Reviews

The test developers at Cognia review newly developed items for:

- alignment to the intended content standard;
- item integrity, including content and structure, format, clarity, and possible ambiguity;
- desired correct responses;
- appropriateness and quality of graphics;
- appropriateness of scoring-guide descriptions and distinctions;
- completeness of associated item documentation (e.g., scoring guide, content codes, key, grade level, cognitive complexity); and
- appropriateness for the designated grade level.

Newly developed stimuli and items for the BIE Science Assessment also undergo review by nationally representative panels of content and assessment experts, including educators from across many states. The purpose of these reviews is to evaluate items and determine their suitability for assessment by answering the following four questions:

- Does the item align with the assigned content standard(s)?
- Is the content accurate?
- Are the content and context grade-level appropriate?
- Does the item provide maximum accessibility for all students?

4.3.2 Bias and Sensitivity Review

Bias and sensitivity review is an essential component of the passage- and item-review process. All Cognia content specialists receive training in bias and sensitivity issues. Controversial and biased topics are avoided in the test development process. Internal reviews include review of not only content but context, with a particular awareness of bias and sensitivity issues that are specific to BIE students.

Since no one person is well versed in the full spectrum of possible concerns, the bias and sensitivity review committee helps to ensure that all potential issues are identified. All stimuli and items undergo bias and sensitivity review prior to field-testing.

The bias and sensitivity review committee comprises a diverse group of people who represent a variety of national student subgroups, including panelists from diverse backgrounds. The people currently serving on this committee include business professionals, educators, a school administrator, an ESL tutor, graduate school students, and retired professionals. United States national racial and ethnic groups represented on this committee include African American, Asian American, Hispanic/Latino/Latina, Native

American, and White. These representatives have varied experiences with urban/suburban/rural environments and economically disadvantaged students.

Because of testing disruptions due to COVID-19, cycles of item development and field testing were also disrupted. Because of form reuse, the items to be field tested in Spring 2022 were not used in the 2022 test forms and were held for use in 2023. Thus, because the items had already undergone development and committee reviews prior to the 22-23 cycle, there were no educator review committees convened in this cycle.

Going forward, however, educator committees will again be periodically convened for the purpose of content development. Committees for both item content review (one committee per grade) and bias/sensitivity review (one committee for all grades) will be held, with BIE educators comprising two seats per committee.

4.4 Test Forms Construction

The Cognia content specialists and psychometricians work collaboratively to produce operational test forms using sequential and iterative procedures that support both the content and construct validity of the assessments.

4.4.1 Item and Stimulus Selection

After field-testing and item data review, Cognia test developers carefully select the items that will appear in the operational tests. In consultation with Cognia psychometricians, test developers consider the following criteria in selecting sets of items for the operational tests:

- **Content coverage/match to test design and blueprints.** The test designs and blueprints stipulate a specific number of items by item type and content distribution.
- **Item difficulty and complexity.** Item statistics are evaluated to ensure quality psychometric characteristics, as well as similar levels of difficulty and complexity from year to year.
- **"Cueing" items.** Items are reviewed for any information that might "cue" or provide information that would help to answer another item.

Test developers sort and lay out passages and items into test forms. During assembly of the test forms, the following criteria are considered:

- **Key patterns.** The sequence of keys (correct answers) is reviewed to ensure that their order appears random.
- **Option balance.** Selected-response (SR) items are balanced across forms so that key options are not markedly disproportionate.
- **Page fit.** Items always appear one per screen for online testing. Common science stimuli always appear to the left of the associated item.
- **Visual appeal.** Every effort is made to make each item as accessible as possible. Each item's presentation may differ slightly depending on the delivery method and size of the screen.

A reviewer designated by the BIE reviews the test form and, prior to approval, specifically considers the following criteria:

- **Construct validity.** The test content is evaluated to determine the degree to which the test measures what it claims, or purports, to be measuring and items/tasks are aligned to the appropriate indicator/standard/measurable outcome.

- **Key accuracy.** Item keys (and the number of designated keys) are reviewed to ensure accuracy.
- **Positive phrasing in item stems.** Items are checked to ensure that negative words such as "not" and/or "except" are rarely, if ever, used.
- **Specific determiners.** Words such as "always," "never," "totally," and "absolutely" are avoided whenever possible to prevent inadvertent cueing of correct or incorrect answers.
- **Clueing/clanging item associations.** The items on the test are reviewed to ensure that the answer to an item is not given away within another item on the same form (clueing) or that an item context is not too similar to another item on the same form (clanging).
- **Bias/sensitivity concerns.** The test is reviewed by all appropriate stakeholders within the BIE and assessment bureaus to ensure that the content is appropriate for students.
- **Errors or typos.** The test is reviewed to verify that the content and metadata are accurate and there appear to be no obvious human errors.

4.4.2 Selection Specifications to Meet Blueprint Requirements

All BIE Science Assessment items are appropriately field-tested prior to operational use. Once stimuli have been field-tested with a set of items, content specialists evaluate the statistics from the items associated with each stimulus. Often, items associated with a stimulus demonstrate a range of student performance, which is largely dependent upon factors inherent to each item. However, if a circumstance is encountered where many items associated with a stimulus are not performing as expected, this is evaluated carefully. While this scenario does not automatically mean the stimulus contains content that is not comprehensible or accessible, it does signal the need to thoroughly review the stimulus in relation to the item content and reevaluate the acceptability of the stimulus. Cognia assessment content specialists also review all the aspects of item content, and this is especially important when data indicate that further scrutiny is warranted.

The process for item data includes the following information for all field-tested items:

- classical item difficulty for all items (i.e., *p*-value)
- score distributions for polytomous items
- item option selection distribution for multiple-choice items
- 10 most frequent student responses for multi-select items
- item-total and option-total correlations
- Item Response Theory (IRT) statistics

Differential item functioning (DIF) using the standardization DIF procedure (Dorans & Kulick, 1986)[1] to produce classifications among student subgroups, such as gender, economically disadvantaged status, Limited English Proficiency (LEP) status. The flags listed in Tables 4-1 and 4-2 are used to identify those items that require an additional level of scrutiny.

Table 4-1. Criteria for Flagged Items Based on Classical Test Theory (CTT) Statistics

Item-Flagging Criteria	Concern
If *p*-value of keyed response < 0.10	Item too difficult
If *p*-value of keyed response > 0.90	Item too easy
If *p*-value of distractor* > *p*-value of keyed response	Possible mis-key
If *p*-value of distractor* > 0.35	Possible second correct option
If item-total correlation < 0.15	Poorly discriminating item
If item-total correlation < 0.00	Non-discriminating or negatively discriminating item
If DIF analysis is B or C	Possible bias in item (B, B-, C, C-)

Note: These analyses examine item score and item option selection distribution for polytomous and selected-response items, respectively.

Table 4-2. Criteria for Flagged Items Based on Item Response Theory (IRT) Statistics

Item-Flagging Criteria	Concern
If IRT *a*-parameter < 0.30	Poorly discriminating
If IRT *b*-parameter < -3.00	Easy item
If IRT *b*-parameter > 3.00	Hard item
If IRT *c*-parameter > 0.35	Low ability students answer correctly (i.e., guessing)
If IRT *b*-parameter standard error of estimation > 0.3	Uncertainty around the item difficulty

The item content of each flagged item is reviewed and discussed by Cognia content specialists, with additional feedback from a group of educators, to make a decision regarding acceptability of the item. At the end of the process, all field-tested items are designated with a status of "Accept," "Rework," or "Reject." Accepted items become eligible for operational testing. Rework items are eligible to be edited and field-tested again so new item data can be generated. Rejected items are removed from the pool of items eligible for operational testing.

Cognia understands that item-level data review must be conducted thoroughly and carefully because of the impacts on test construction, which need to be consistent from administration to administration. Being experts in their respective content areas, Cognia's content specialists also understand that some assessed standards are typically more challenging for students than others, and the specialists can simultaneously make good decisions about both content and data in accepting or rejecting items for operational use based on the item statistics. Finally, Cognia understands that items with DIF statistic flags need to be scrutinized for potential sources of bias. While a flag does not automatically mean the item contains biased content, it does signal the need to thoroughly review the item content and evaluate the ways in which the different focal groups would have access and ability to answer the item to ensure it is fair for all students.

---

[1] *DIF occurs when an item has difficulty measures that vary across contexts for similarly able subgroups of examinees. DIF procedures are designed to identify items on which the performances of certain subgroups of interest differ from each other after controlling for construct-relevant achievement. In order to ensure meaningful results, DIF statistics are not computed for populations containing less than 200 students in both subgroups. Analysis was conducted using field-test data to detect potential DIF at the item level. The standardization DIF procedure (Dorans & Kulick, 1986) was employed to evaluate subgroup differences. The computed DIF indices have a theoretical range of -1.0 to 1.0 for multiple-choice items. Critical values are defined as 0.05 and 0.10 and the values are flagged as statistically significant using alpha = 0.05. If the absolute value of standardized DIF is equal to or greater than 0.10, the item is classified "C" DIF; items with absolute values greater than or equal to 0.05 are classified as "B" DIF; otherwise, items are classified as "A" DIF.*

4.4.3 Cognitive Processes

Cognitive-process assesses whether students' cognitive skills and processes align with those specified in the construct domains defined by test developers for all students and subgroups.

Subject matter experts, including educators, pay close attention to cognitive processes through test review. A team of science content specialists developed and internally reviewed items for BIE Science (Cognia's Science Secure Item Bank). Subsequently, state/entity department of education content and/or assessment specialists, along with science educators, reviewed the items. Their input ensured that the tasks effectively measured the intended cognitive ability. Science educators participated in the review of all items developed and field tested, in addition to Cognia content expert reviews for alignment, content accuracy, and cognitive complexity levels. Committee review meetings included discussions and confirmation of target students' characteristics, such as various learning targets, use of different science dimensions, and the degree of cognitive processing elicited by the stimulus and application of each dimension. Educators and specialists with expertise in science content provided feedback in these committee meetings, indicating that the test content did not require extraneous cognitive processes for engagement.

Chapter 5. Test Administration

Orderly and secure test administrations are necessary to protect secure test content and ensure that test data are validity-interpretable to meet score reporting and accountability reporting requirements.

5.1 Roles and Responsibilities

The Test Coordinator's Manual emphasizes that School Test Coordinators (STCs) play a crucial role as the primary source of assessment information for school stakeholders. STCs are responsible for keeping their schools informed about assessment policies, changes, and providing resources to teachers.

Manuals, including the Test Coordinator's Manual and the Test Administrator's Manual, ensure consistent administration procedures across schools, emphasizing test security and ethical administration. These documents are accessible on the BIE Help and Support Website.

Additional staff, such as Test Administrators (TAs), are vital for a successful BIE Science Assessment. TAs must follow procedures outlined in the Test Administrator's Manual and attend training provided by the STC. If additional staff are needed, those who have received training and signed the Confidentiality Agreement may assist with one-to-one accommodations. STCs are appointed at the local level.

5.2 Test Administrator's Manual

For Spring 2023, the Test Administrator's Manual (TAM) outlined the steps to follow before, during, and after administration of the Spring 2023 BIE Science Assessment. Understanding of and compliance with each of these steps is vital for successful administration.

The TAM covers administration policies such as security guidelines and administration information, accessibility features and accommodations including requirements for computer-based tests (CBT) and paper-based tests (PBT), preparing for CBTs and PBTs, administering CBTs and PBTs, directions and scripts for use during CBT and PBT administrations, and what to do at the completion of CBTs and PBTs.

5.3 TA and Proctor Training Requirements and 2023 Test Administrations

All TAs and proctors involved in test administration, preparation, and security are required to attend training provided by the STC. Training should include information on test security policies and procedures, test administration procedures, documentation and provision of testing accommodations, and the importance of strictly following all directions in the manuals.

5.4 Testing Irregularity Reports

During the Spring 2023 BIE Science testing window, test administrators and coordinators were trained to report test administration irregularities. The BIE defines a testing irregularity as any incident in the handling or administration of a test that results in questioning the accuracy of the data or security of the test that may or may not result in an invalidation.

Three test irregularities were submitted to BIE for the following reasons:

- Test tickets were misplaced for an entire class. After investigation, the school determined the test tickets were inadvertently discarded. The test administrator reprinted the log-in tickets and resumed testing as planned. All student responses were submitted and scored.
- Grade 5 test materials were left unattended in a classroom with students. They remained in a storage bin and there was no evidence that any materials were viewed by students.
- A student was using an electronic device while completing their test. The cell phone was confiscated, but it was determined there was no use of the phone in relation to the test. The student completed the last question and submitted the test.

5.5 Test Security

The BIE Assessment Program requires that the BIE Science Assessment be treated with the highest level of test security and accountability. The security of BIE materials must be maintained before, during, and after the test administration. TAs, proctors, and school test coordinators are required to follow the guidelines in the TAM for distributing, collecting, and returning testing materials. All testing personnel are required to have access to a central, locked storage space for safekeeping of test materials until print materials are returned to Cognia.

To maintain the validity of the tests administered in the assessment program, keeping all test questions secure is absolutely necessary. If security is breached or compromised, the assessment results may not be valid. If one student or school has advantages not awarded to another, the test administration is no longer standardized and loses the important distinction of being appropriate for program accountability.

TAs must follow these security guidelines before, during, and after testing:

- Receive training on test security and administration by the STC.
- Complete the BIE Confidentiality Agreement and return it to the STC.
- Follow the testing schedule established by the school.
- Ensure TA is not assigned to a classroom in which a relative is being tested.
- Carry out standard examination procedures.
- Ensure secure test materials are secured in a central and locked area when not in use.
- Use the security checklist or a similar tracking tool daily, as provided by the STC, during test administration to check in and check out all test materials.
- Report any possible breaches of security to the STC immediately. Examples of security breaches include but are not limited to:

  improper handling of test materials, such as
  - someone reproducing any student responses,
  - allowing any unauthorized access to test materials before, during, or after testing, or
  - leaving test materials (including computers being used for CBTs) unsecured when the TA or a proctor is not in the classroom, and

  improper test administration procedures, such as
  - coaching students during testing,
  - altering student responses in any way, or
  - stray mark cleanup, including but not limited to erasing double-marks or lightly marked answers.

School staff are prohibited from studying or discussing online test questions in any manner, either among themselves or with students before, during, or after testing.

Chapter 6. Scoring: Scope of Work, Processes, and Procedures

6.1 Scope of Work

The operational items on the 2022-23 BIE Science Assessment included 4-point open-ended response items. Table 6-1 outlines the number and type of each item per grade.

Table 6-1. Overview of BIE Science Assessment Scope-of-Work 2022−2023

Grade 5	Grade 8	Grade 11
OP – 3 OE4	OP – 3 OE4	OP – 3 OE4

OP=Operational, OE#= 4-point open-ended response item

6.2 Operational Scoring: Processes and Procedures

6.2.1 Score Verification of Multiple-Choice Items

For both computer-based tests (CBTs) and paper-based tests (PBTs), responses to multiple-choice items were compared to scoring keys using item analysis software. This robust software compared each student response to multiple-choice items to the respective answer key and assigned a maximum score of 1 point for correct responses and 0 points for incorrect answers. In PBTs, if students filled in multiple bubbles in response to one item, the response was assigned 0 points. At the end of an administration, a second independent validation of all the student responses was conducted to compare and validate results to ensure accurate machine scoring.

6.2.2 Scoring of Open-Ended Response Items

6.2.2.1 Personnel Structure

Cognia's personnel structure for scoring responses consisted of four hierarchical levels:

Figure 6−1. Cognia Scoring Staff
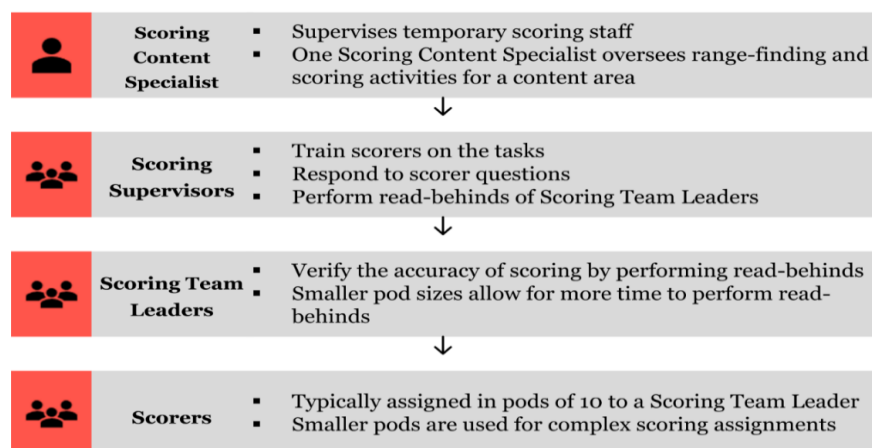
All responses were scored by fully vetted scorers who were supervised by Scoring Team Leaders (STLs). The Scoring Supervisors monitored the work of the STLs assigned to them. The Scoring Content Specialist monitored the work of the Scoring Supervisors, STL, and scorers. Scoring Content Specialists are full-time Cognia staff who report to the Scoring Content Group Manager, who in turn reports to the Director for Content and Quality in the Scoring Services department. This hierarchical structure whereby each level monitors the one below ensures reliable quality and consistency in scoring.

Scoring Content Specialist

The Scoring Content Specialist functioned as the primary lead for his or her designated content area and as a liaison between scoring activities and the Scoring Project Manager to ensure that established quality standards and production schedules were met.

During scoring, the Scoring Content Specialist was responsible for supervising all scoring staff working on the project, including Scoring Supervisors and STLs. The Scoring Content Specialist was also responsible for assuring the consistency and accuracy of scoring work performed by individual scorers and across groups of scorers.

Scoring Supervisors

Scoring Supervisors managed the scorer training and supervised the STLs and scorers working on a designated item and/or content. Scoring Supervisors worked closely with the STLs to ensure consistency and provide counsel and retraining to scorers, as necessary. In addition, Scoring Supervisors engaged in supervisory oversight and performed quality-control checks to ensure the consistency and accuracy of the STLs. Scoring Supervisors, who were responsible for monitoring training and conducting the retraining of scorers, were selected for their ability to instruct and for their level of expertise in their respective disciplines.

Scoring Team Leaders

The STLs were responsible for supervising and monitoring the group of scorers assigned to them. STLs worked closely with their scorers to maintain consistently accurate scoring. They provided quality checks, and they counseled scorers, as necessary. STLs were responsible for monitoring and maintaining accurate scoring of their assigned scorers. This included performing read-behinds on scorers and monitoring other quality-control measures. STLs were responsible for arbitrating responses scored by multiple scorers when the assigned scores varied by more than one score point. The arbitration process ensured that such responses received the necessary attention by providing an additional review before assigning a third and final resolution score. In addition to the essential quality control, the arbitration process provided continued opportunities for scorer training.

Because the read-behinds that the STLs performed moderated the scoring process and thus maintained the integrity of the scores, individuals chosen to fill STL positions were selected for their accuracy and content knowledge.

Scorers

Scorers are individuals who evaluate student responses and assign scores.

6.2.2.2 Scorer Recruitment

Cognia actively sought a diverse pool of scorers with a broad range of backgrounds: teachers, scientists, business professionals, graduate school students, retired educators, and the like.

The minimum requirement to assume a position as a scorer or Scoring Team Leader is 48 college credits, which include classes related to the content area being scored. Scoring Supervisors must hold a

bachelor's degree with classes related to the content area being scored. All potential scorers and leadership staff submitted documentation (e.g., résumés and/or transcripts) as evidence of meeting the education and experience requirements. Each scorer and leadership staff signed a binding non-disclosure/confidentiality agreement as well.

6.2.2.3 Scoring Platform

For the scoring of the 2022-23 BIE Science test administration, Cognia engaged OSCAR, an online scoring and reporting system, as the scoring platform due to its capabilities for facilitating online rangefinding. OSCAR ensures the security of student responses and test items. During scoring, no student names or schools/districts associated with viewed student work are visible to scorers, and all Scoring Services temporary associates are subject to the same non-disclosure requirements as full-time Cognia staff. Cognia maintained security by using a highly secure server-to-server interface, ensuring that access to all student response images was limited only to scorers, appropriate Cognia staff, and educators participating in the rangefinding activities.

Scorers evaluated most student responses from images rendered by the online testing platform and a small number of responses from scanned images of paper-based tests. Whether administered in an online or a paper/pencil environment, all responses were scored applying the same scoring criteria.

Prior to the beginning of scoring, databases were created for each grade to receive submitted student responses for each item to be scored. To provide maximum security for all testing and scoring materials, each scorer was required to log on to the scoring systems using a unique combination of an assigned username and password.

6.2.2.4 Leadership Training

Scoring Supervisors and select STLs were given a separate training session one day before scorer training. Scoring staff, including Scoring Supervisors and STLs, responsible for scoring student responses were required to achieve the same standard as scorers on item qualification sets: a minimum accuracy scoring rate of 70 percent exact, and 90 percent exact plus adjacent agreement (70/90).

6.2.2.5 Scorer Training

For the scoring of BIE science common operational items, all scorer training was conducted using pre-recorded, interactive training modules.

These modules allowed for self-paced, individual training. Modules were produced by experienced Scoring Supervisors who prepared all training materials for an image slide-show presentation which was overlaid with sound. The format of this training process replicated the traditional face-to-face group training led by a Scoring Supervisor. Each recording started with a discussion of the item and the rubric followed by a detailed discussion of each anchor paper and its rubric-based scoring rationale. After the conclusion of anchor paper training, scorers would gain access to a set of practice papers, to which they would apply the scoring standards as detailed in the rubric and as exemplified in the anchor papers to determine the correct score.

Scorers were given two opportunities to qualify. If scorers were unable to attain a score match of at least 70 percent exact and 90 percent adjacent agreement on the first qualifying set, they were retrained by discussing the responses contained in the first qualification set with respect to the score-point descriptions of the rubric and by comparing them to the responses of the anchor set. Following this retraining, scoring leadership would administer a second qualification set. If scorers achieved a scoring accuracy rate of at least 70 percent exact and 90 percent adjacent agreement on the second qualification set, then they were allowed to score student responses. Scorers who failed to pass the minimum

threshold were not allowed to score that item. They were either trained on another item or they were dismissed from the project. Appendix D captures the qualification rates for all grades.

6.2.2.6 Monitoring Scoring Quality

Scorers were required to demonstrate and maintain their ability to score student responses accurately and consistently throughout the scoring process. The OSCAR scoring system enabled scoring leadership to measure and monitor individual and group performance on each scored item in terms of accuracy and consistency, and in terms of read rate (scoring speed) and overall production rate on a constant, real-time basis. Scoring tools employed to measure scoring quality were as follows:

- Read-behind scoring
- Double-blind scoring
- Embedded validity responses
- Recalibration sets

Each scorer's performance on the above procedures was monitored and recorded by the scoring system and scoring leadership could review data related to the accuracy, consistency, and overall quality of scoring. Scoring leadership was always available to answer scorer questions. They also counseled and retrained scorers as needed to determine whether a scorer should continue scoring. If a scorer's performance did not meet the prescribed quality standards, scoring leadership initiated a process through which that scorer's work was invalidated and returned to the scoring queue of unscored responses to be re-scored by those scorers who demonstrated scoring accuracy at or above standard.

Read-Behind Scoring

Read-behind scoring allowed scoring leadership to monitor each scorer's scoring performance by way of an immediate real-time snapshot of the scorer's accuracy. The data generated by read-behind scoring presented leadership with opportunities to answer questions and to provide counsel to scorers who may have had trouble maintaining the scoring standards. If the scores assigned by the scorer and the STL were discrepant (more than one score point apart) or if there were a significant number of adjacent scores (one score point apart) between the scorer and the STL, scoring leadership then counseled and retrained the scorer. Scoring leadership determined when or whether these scorers were given access to resume operational scoring. Retrained scorers were subject to additional monitoring and read-behinds.

The number of read-behinds for each scorer varied depending on the accuracy of the scorer. BIE scoring specifications require a minimum of two read-behinds per item per hour per scorer. Consistently accurate scorers would only receive the minimum number of read-behinds, whereas scorers who exhibited difficulties in maintaining accuracy or consistency received additional read-behinds.

In addition to scorers, scoring leadership was also subject to quality assurance reviews, which were administered by the Scoring Content Specialist. They monitored scoring leadership's accuracy and consistency by reviewing the read-behind results.

Double-Blind Scoring and Arbitration Resolution

Double-blind scoring refers to the process of two scorers independently scoring the same response. During this process, neither scorer has any knowledge of the other scorer's score. The double-blind process helps inform scoring leadership about the consistency of scoring among peer scorers who actively score an item. All responses in grades 5 and 8 had a minimum of 2 percent of responses double-blind scored by another human scorer. In grade 11, the AI engine that is integrated in the scoring platform was employed to support operational scoring. Using a sophisticated algorithm, the AI scoring engine is designed to learn from and mimic human scorers. The model was built using human-scored responses obtained during field-test scoring of these items. Model results were carefully evaluated to determine the

appropriate percent of human double-blind scoring. Human second reads were used to verify the reliability of the model in an operational setting. In grade 11, AI scoring was applied as the first score with a 10% human double-blind score.

During double-blind scoring, the scoring system distributes randomly selected responses assigned for double-blind scoring to different scorers without alerting either scorer. The scoring system then records each scorer's score and routes any scoring discrepancies of more than one point between the two scores to an arbitration response queue for resolution by the STL. The percentage of double-blind responses sent to arbitration by a scorer because of a difference in actual scores (i.e., not including blank or unreadable responses) should not have exceeded 10 percent. If a scorer's arbitration percentage exceeded this threshold, scoring leadership counseled, retrained, and/or dismissed the scorer.

Embedded Validity Responses

Validity responses are prescored responses that serve calibration purposes at the onset of scoring an item. Ten validity responses were embedded in the first one hundred live student responses and distributed to each scorer in randomized order. Scorers were not aware when they were scoring an embedded validity response as compared to a live student response. Scorers who demonstrated an accuracy rate of less than 70% exact on each composite score were counseled and the STL increased the number of read-behinds to ensure accuracy.

Recalibration Sets

A set of five calibration papers was administered starting with the second day of scoring an item. This set of five responses, selected by scoring leadership, was inserted into the scoring queue, and served as a refresher. It was used to gauge the scorers' ability to maintain accurate scoring of the item on days following their initial item training.

Interrater Reliability

Section 6.2.2.6 of this report describes in detail the processes that were implemented to monitor the quality of the hand-scoring of student responses for open-ended items. One of these processes was double-blind scoring: A minimum of 2 percent of student responses in grades 5 and 8 was randomly selected and routed to two different scorers to be scored independently. In grade 11, 10% of all AI-scored responses received a second human score. Results of the double-blind scoring in grades 5 and 8 were used during scoring to identify scorers who required retraining or other intervention and are presented here as evidence of the reliability of the BIE Science Assessment. In grade 11, the human double-blind scoring results were used to verify the accuracy of the AI scoring engine. Table 6-3 shows the number of included scores, and the percentage of exact and adjacent agreement as well as percent third reads for adjudication.

Table 6-2. Summary of Interrater Reliability Statistics for BIE Science across all OP by Grade

Grade	Total # of Responses Scored	Total # of Double-Blind Responses Scored	Total % Double-Blind Responses Scored	Score Categories	Score Point Ranges	% Exact	% Adjacent	% Third Reads
5	1,937	39	2.0	1	0-4	68	30	2
8	1,662	34	2.0	1	0-4	70	27	3
11	870	87	10.0	1	0-4	55	37	8

6.2.2.7 Score-of-Record Rules

Per scoring specifications, the final score-of-record was determined as follows:

- If there was an exact agreement between the scorer and the STL scores, no action was taken—the scorer's original score remained.
- If there was a difference between the scores, either adjacent or discrepant, the STL's score became the score-of-record. Adjacent scores differ by 1 point, while discrepant scores differ by more than 1 point.

Chapter 7. Classical Item and Test Analysis

As noted in the *Principles of Educational and Psychological Testing* (Brown, 1983), "A test is only as good as the items it contains." A complete evaluation of a test's quality must include an evaluation of each item. Both *Standards for Educational and Psychological Testing* (AERA et al., 2014) and *Code of Fair Testing Practices in Education* (Joint Committee on Testing Practices, 2004) include standards for identifying quality items. Items should assess only knowledge or skills that are identified as part of the domain being tested and should avoid assessing irrelevant factors. Items should also be unambiguous and free of grammatical errors, potentially insensitive content or language, and other confounding characteristics. In addition, items must not unfairly disadvantage students in particular racial, ethnic, or gender groups.

7.1 Classical Item Statistics

All operational items were evaluated in terms of classical item difficulty, which under classical test theory practices is defined as the average scored response on an item, divided by the maximum possible score for the item. Although this index is traditionally described as an estimate of item difficulty, it is properly interpreted as an easiness index. The greater in value a classical item difficulty is, the easier the item.

Items that are answered correctly by almost all students provide little information about differences in student abilities, but they do indicate knowledge or skills that have been mastered by most students. Similarly, items that are correctly answered by very few students provide little information about differences in student abilities, but they may indicate knowledge or skills that have not yet been mastered by most students. In general, to provide adequate measurement, classical difficulty indices should range from near-chance performance (e.g., 0.25 for four-option multiple-choice items) to 0.90, with a majority of items generally falling between around 0.4 to 0.7. However, on a standards-referenced assessment such as the BIE science, it is appropriate to include items with very low or very high item difficulty values to ensure sufficient content coverage.

A desirable characteristic of an item is for higher-ability students to perform better on the item than lower-ability students do. The correlation between student performance on a single item and total test score is a commonly used measure of this characteristic of the item. Within classical test theory, the item-total correlation is referred to as the item's classical discrimination because it indicates the extent to which successful performance on an item discriminates between high and low scores on the test. Each of the item-total correlations reported here is the Pearson correlation between scored responses on a given item and total raw scores. This Pearson correlation is commonly referred to as the point-biserial correlation (for a dichotomously scored item) and a point-polyserial correlation (for a polytomously scored item). The theoretical range of these correlations is −1.0 to +1.0, with a typical observed range from 0.2 to 0.6. Discrimination indices can be thought of as measures of how closely an item assesses the same knowledge and skills assessed by other items contributing to the criterion total score. That is, the discrimination index can be thought of as a measure of construct consistency.

A comparison of indices across grade levels is complicated because these indices are population-dependent. Direct comparisons would require that either the items or the students were common across groups. Since that is not the case, it cannot be determined whether differences in performance across grade levels are due to differences in student abilities, differences in item difficulties, or both.

Classical item difficulties and item-total correlations are provided in Appendix E and a summary table listed in Table 7-1. These statistics were calculated based on a national student sample, such as all the states and entities that used Cognia Secure Science Item Bank (SSIB.)

Table 7-1. Item Summary by Grade

Grade	Item Type	Item Mean
5	MC	0.44
5	OR	0.87
8	MC	0.36
8	OR	0.79
11	MC	0.42
11	OR	0.74

7.2 Total Test and Subscore Intercorrelations

When subscores are strongly related to each other, it implies a high internal consistency between subscores. The Pearson correlation matrices among the individual reporting categories (i.e., subscores) are shown in Tables 7-2 to 7-4. These correlations are based on BIE student performance, which includes their total test scale score and their three subscores.

Table 7-2. Pearson Correlations of Total Test and Subtest Raw Scores on BIE Science Grade 5, as a Function of Operational Form

Subtest	Number of Items	Total Test	Earth and Space Sciences	Physical Sciences	Life Sciences
Operational Form 1					
Total Test	35	1.00	--	--	--
Earth and Space Sciences	13	0.83	1.00	--	--
Physical Sciences	11	0.79	0.51	1.00	--
Life Sciences	11	0.85	0.58	0.48	1.00
Operational Form 2					
Total Test	35	1.000	--	--	--
Earth and Space Sciences	13	0.83	1.00	--	--
Physical Sciences	11	0.79	0.51	1.00	--
Life Sciences	11	0.85	0.58	0.48	1.00

Table 7-3. Pearson Correlations of Total Test and Subtest Raw Scores on BIE Science Grade 8, as a Function of Operational Form

Subtest	Number of Items	Total Test	Earth and Space Sciences	Physical Sciences	Life Sciences
Operational Form 1					
Total Test	35	1.000	--	--	--
Earth and Space Sciences	12	0.75	1.00	--	--
Physical Sciences	12	0.85	0.46	1.00	--
Life Sciences	11	0.81	0.43	0.55	1.00
Operational Form 2					
Total Test	35	1.000	--	--	--
Earth and Space Sciences	12	0.75	1.00	--	--
Physical Sciences	12	0.85	0.46	1.00	--
Life Sciences	11	0.81	0.43	0.55	1.00

Table 7-4. Pearson Correlations of Total Test and Subtest Raw Scores on BIE Science Grade 11, as a Function of Operational Form

Subtest	Number of Items	Total Test	Earth and Space Sciences	Physical Sciences	Life Sciences
Operational Form 1					
Total Test	37	1.000	--	--	--
Earth and Space Sciences	12	0.71	1.00	--	--
Physical Sciences	13	0.87	0.49	1.00	--
Life Sciences	12	0.76	0.26	0.49	1.00
Operational Form 2					
Total Test	37	1.000	--	--	--
Earth and Space Sciences	12	0.71	1.00	--	--
Physical Sciences	13	0.87	0.49	1.00	--
Life Sciences	12	0.76	0.26	0.49	1.00

Chapter 8. Psychometrics: Item Response Theory (IRT) Scaling and Equating

This chapter describes the procedures used to scale the BIE Science Assessment, where all test forms were pre-equated for the Spring 2023 administration.

8.1 IRT Models

All BIE items were previously calibrated using item response theory (IRT). IRT uses mathematical models to define a relationship between an unobserved measure of student proficiency, usually referred to as theta (θ), and the probability (p) of getting a dichotomous item correct or of getting a particular score on a polytomous item. In IRT, all items are assumed to be independent measures of the same construct (i.e., of the same θ). Another way to think of θ is as a mathematical representation of the latent trait of interest. Several common IRT models are used to specify the relationship between θ and p (Hambleton & van der Linden, 1997; Hambleton & Swaminathan, 1985). The process of determining the specific mathematical relationship between θ and p is called item calibration. After items are calibrated, they are defined by a set of parameters that specify a nonlinear, monotonically increasing relationship between θ and p. Once the item parameters are known, an estimate of θ for each student can be calculated. This estimate, $\hat{\theta}$, is considered to be an estimate of the student's performance. It has characteristics that may be preferable to those of raw scores for equating and scaling purposes.

For the BIE Science Assessments, the three-parameter logistic (3PL) model was used for dichotomous (selected-response) items and the Graded-Response Model (GRM) was used for polytomous (constructed-response) items. The 3PL model for dichotomous items can be defined as:

$$P_i(\theta_j) = P(U_i = 1|\theta_j) = c_i + (1 - c_i)\frac{\exp[Da_i(\theta_j - b_i)]}{1 + \exp[Da_i(\theta_j - b_i)]},$$

Where U indexes the scored response on an item,

i indexes items,

j indexes students,

a represents item discrimination,

b represents item difficulty,

c is the lower asymptote parameter, and

D is a normalizing constant equal to 1.701.

In the GRM for polytomous items, an item is scored in a $k + 1$ graded category that can be viewed as a set of k dichotomies. At each point of dichotomization (i.e., at each threshold), a two-parameter model

can be used. This implies that a polytomous item with a $k+1$ category can be characterized by k Item Category Threshold Curves (ICTCs) of the two-parameter logistic form:

$$P_{ik}^*(k|\theta_j) = P(U_i \geq k|\theta_j) = \frac{\exp[Da_i(\theta_j-b_i+d_{ik})]}{1+\exp[Da_i(\theta_j-b_i+d_{ik})]},$$

where   U indexes the scored response on an item,

       i indexes the items,

       j indexes students,

       k indexes threshold,

       a represents item discrimination,

       b represents item difficulty,

       d represents item category threshold, and

       D is a normalizing constant equal to 1.701.

After computing k ICTCs in the GRM, $k+1$ Item Category Characteristic Curves (ICCCs) are derived by subtracting adjacent ICTCs:

$$P_{ik}(\theta_j) = P(U_i = k|\theta_j) = P_{i(k-1)}^*(\theta_j) - P_{ik}^*(\theta_j),$$

where P_{ik} represents the probability that the score on item i falls in category k, and

       P_{ik}^* represents the probability that the score on item i falls above the threshold k

Note that $P_{i0}^* = 1$ and $P_{i(m+1)}^* = 0$.

The GRM is also commonly expressed as:

$$P_{ik}(k|\theta_j) = \frac{\exp[Da_i(\theta_j-b_i+d_k)]}{1+\exp[Da_i(\theta_j-b_i+d_k)]} - \frac{\exp[Da_i(\theta_j-b_i+d_{k+1})]}{1+\exp[Da_i(\theta_j-b_i+d_{k+1})]}.$$

The Item Characteristic Curve (ICC) for polytomous items is computed as a weighted sum of ICCCs, where each ICCC is weighted by a score assigned to a corresponding category:

$$E(U_i|\theta_j) = \sum_{k}^{m+1} w_{ik} P_{ik}(\theta_j)$$

See Lord and Novick (1968), Hambleton and Swaminathan (1985), and Baker and Kim (2004) for more information about item calibration and parameter estimation.

8.2 Dimensionality Analysis

Tests are constructed with multiple content-area subcategories and their associated knowledge and skills. Hence, the potential exists for dimensions being invoked beyond the common primary dimension. Generally, the content-area subcategories are highly correlated with each other, and the primary dimension they share typically explains an overwhelming majority of the variance in test scores. The

presence of just such a dominant primary dimension is the psychometric assumption that provides the foundation for the unidimensional item response theory (IRT) models that are used for scaling and equating of the BIE tests.

The purpose of dimensionality analysis is to investigate whether violation of the assumption of test unidimensionality is statistically detectable and, if so, (a) the degree to which unidimensionality is violated and (b) the nature of the multidimensionality. Findings from dimensionality analyses performed on the BIE operational items for science are reported below. (Note: Only operational items were analyzed since they are used for score reporting.)

The dimensionality analyses were conducted using the nonparametric IRT-based methods DIMTEST (Stout, 1987; Stout, Froelich, & Gao, 2001) and DETECT (Zhang & Stout, 1999). Both methods use the estimated average conditional covariances for item pairs as their basic statistical building block. A conditional covariance is the covariance between two items conditioned on expected total score for the rest of the test, and the average conditional covariance is obtained by averaging across every possible conditioning score. When a test is strictly unidimensional, all conditional covariances are expected to take on values within random noise of zero, indicating statistically independent item responses for examinees with equal expected total test scores. Nonzero conditional covariances are essentially violations of the principle of local independence, and local dependence implies multidimensionality. Thus, nonrandom patterns of positive and negative conditional covariances indicate multidimensionality.

DIMTEST is a hypothesis-testing procedure for detecting violations of local independence. The data are first divided into a training sample and a cross-validation sample. Then an exploratory analysis of the conditional covariances is conducted on the training sample data to find the cluster of items that display the greatest evidence of local dependence. The cross-validation sample is then used to test whether the conditional covariances of the selected cluster of items display local dependence, conditioned on total score on the non-clustered items. The DIMTEST statistic follows a standard normal distribution under the null hypothesis of unidimensionality.

The DETECT statistic is an effect-size measure of multidimensionality. As with DIMTEST, the data are first divided into a training sample and a cross-validation sample. The training sample is used to find a set of mutually exclusive and collectively exhaustive clusters of items that best fit a systematic pattern of positive conditional covariances for pairs of items from the same cluster and negative conditional covariances from different clusters. Next, the clusters from the training sample are used with the cross-validation sample data to average the conditional covariances. Within-cluster conditional covariances are summed; from this sum the between-cluster conditional covariances are subtracted. This difference is divided by the total number of item pairs, and this average is multiplied by 100 to yield an index of the average violation of local independence for an item pair. DETECT values less than 0.2 indicate very weak multidimensionality (or near unidimensionality); values of 0.2 to 0.4, weak multidimensionality; values of 0.4 to 1.0, moderate multidimensionality; and values greater than 1.0, strong multidimensionality (Roussos & Ozbek, 2006).

DIMTEST and DETECT were applied to the BIE science tests, Set A and Set B for each of grades 5, 8, and 11. First, each dataset was split into a training sample and a cross-validation sample. DIMTEST was then applied to each sample. As shown in Table 8-1, the DIMTEST null hypothesis of unidimensionality was not rejected for any of the datasets. Because the DIMTEST statistic did not find any evidence for rejecting the null hypothesis of unidimensionality, no DETECT analyses were needed or conducted for estimating the sizes of the violations of local independence,

In summary, the dimensionality analyses conducted for the BIE science assessments found no evidence of any violations of the assumption of unidimensionality and, thus, strongly support the use of the unidimensional IRT models whose implementation is detailed in this chapter.

Table 8-1. DIMTEST Hypothesis Testing Results, as a Function of Grade*

Content Area	Grade	P-value	Interpretation
	05	0.26	Does not detect violation of unidimensionality
Science (operational Set A)	08	0.72	Does not detect violation of unidimensionality
	11	0.76	Does not detect violation of unidimensionality
	05	0.08	Does not detect violation of unidimensionality
Science (operational Set B)	08	0.35	Does not detect violation of unidimensionality
	11	0.47	Does not detect violation of unidimensionality

8.3 Item Response Theory Results

The tables in Appendix F give the IRT item parameters of all common items on the Spring 2023 BIE Science Assessment by grade.

Test characteristic curves (TCCs) are based on the IRT item parameters and display the expected (average) raw score associated with each θ_j value between –4.0 and 4.0, or equivalently the expected (average) raw score associated with each observable scaled score (see Section 8.4 for details on scaled scores). Mathematically, the TCC is computed by summing the ICCs of all items that contribute to the raw score. Using the notation introduced in Section 7.1, the expected raw score at a given value of θ_j is

$$E\left(X|\theta_j\right) = \sum_{i=1}^{n} E(U_i|\theta_j),$$

where i indexes the items (and n is the number of items contributing to the raw score),

  j indexes students (here, θ_j runs from –4 to 4), and

  $E\left(X|\theta_j\right)$ is the expected raw score for a student of ability θ_j,

  U indexes the scored response on an item.

The expected raw score monotonically increases with θ_j, consistent with the notion that students of high ability tend to earn higher raw scores than do students of low ability. Most TCCs are "S-shaped"—flatter at the ends of the distribution and steeper in the middle.

Test information functions (TIFs) display the amount of statistical information the test provides at each value of θ_j, or equivalently display the amount of statistical information the test provides at each observable scaled score. TIFs depict test score precision across the entire latent trait continuum. There is an inverse relationship between the information of a test and its conditional standard error of measurement (CSEM). The CSEM at a given θ_j [$CSEM(\theta_j)$] is equal to the inverse of the square root of the statistical information at θ_j (e.g., Hambleton, Swaminathan, & Rogers, 1991). That is, the $CSEM(\theta_j)$ is equal to the inverse of the square root of the TIF at a given θ_j [$TIF(\theta_j)$], the expression for which can be written as follows:

$$CSEM\left(\theta_j\right) = \frac{1}{\sqrt{TIF(\theta_j)}}$$

Compared to the tails, TIFs are often higher near the middle of the θ distribution, where most students are located and where most items are sensitive by design.

Appendix G contains graphs of the TCC and TIF for the BIE science tests, Set A and Set B for each of grades 5, 8, and 11. Each TCC graph displays the expected raw score (on the vertical axis) for the entire theta scale (on the horizontal axis). Each TCC graph also has a set of vertical lines that indicate the values of the theta cut scores for the given grade. Each TIF graph displays test information value (on the vertical axis) at the entire theta scale (on the horizontal axis). Each TIF graph also has a set of vertical lines that indicate the values of the scaled score cut scores for the given subject and grade.

8.4 Equating

The purpose of equating is to ensure that scores obtained from different forms of a test are equivalent to each other. Equating may be used if multiple test forms are administered in the same year or to equate one year's forms to those given in the previous year.

The cut scores for 2023 BIE Science Assessments were established in 2022 by Cognia with a standard setting using previously calibrated items from the Cognia SSIB. As new SSIB items are field tested each year (for BIE and other SSIB contracts), the operational items on the forms serve as equating items to bring the field-test items onto the SSIB scale.

The pre-equating process uses item bank values of the IRT item parameters to place the pre-equated test form onto the established IRT scale. Equating ensures that students are not given an unfair advantage or disadvantage because the test form they took is easier or harder than those taken by other students.

8.5 Reported Total Test and Subtest Scale Scores

The θ scale used in IRT calibrations is not readily understood by most stakeholders. As such, reporting scales are used for BIE reporting. The reporting scales are linear transformations of the underlying θ scale. To obtain a student's scaled score on a given assessment, the student's raw score (i.e., total number of points earned) is translated into a value on the underlying θ scale using TCC mapping. The student's θ value is translated into a scaled score (*SS*) using the following linear equation:

$$SS = \beta_0 + \beta_1 \theta$$

where β_0 is an intercept constant and

β_1 is a slope constant,

The CSEM can also be translated into a scaled CSEM. Whereas values of the CSEM are on the θ scale, values of the scaled CSEM are on the reporting scale. The scaled CSEM is obtained via the following equation:

$$Scaled\ CSEM = \beta_1 \times CSEM(\theta)$$

Table 8-2 shows the slope and intercept terms used for the Spring 2023 BIE Science Assessment to calculate the scaled scores. See Appendix H for Raw to Scale Score Lookup Tables.

Table 8-2. Spring 2023 Scaled Score Slopes and Intercepts by Subject and Grade

Grade	Slope	Intercept
5	12.5	553.56675
8	10	855.1012
11	7.5	1159.7213

It is important to note that converting from raw scores to θ values to scaled scores does not change students' achievement-level classifications. Given the relative simplicity of raw scores, it is fair to question why scaled scores are reported instead of raw scores. Scaled scores make the reporting of grade-level results consistent across test forms and administrations. It is this uniformity across scaled scores that facilitates the understanding of student performance. The psychometric advantage of scaled scores over raw scores comes from there being linear transformations of θ. Since the θ scale is used for pre- or post-equating, scaled scores are comparable from one year to the next. Raw scores are not.

8.6 Performance Levels

The cut scores used for the Spring 2023 BIE Science Assessments are the cut scores that were established for the Cognia SSIB in 2022, on which the BIE Science summative assessments are based on. BIE conducted a standard validation in August 2022, however, due to low participation, BIE determined the cut scores set for the SSIB would be used. A standards validation was held in 2023 to verify these cut scores against the BIE Science Assessment. The cut scores on the theta scale and the reporting scale, used for the Spring 2023 BIE Science Assessments, are presented in Table 8-3.

Table 8-3. Spring 2023 Cutpoints on the Theta Metric and Reporting Scale by Grade

Theta			Scale Score				
Cut1	Cut2	Cut3	Min	Cut1	Cut2	Cut3	Max
-0.75048	0.51466	1.70117	500	544	560	574	590
-0.96101	0.48988	2.73095	800	845	860	882	890
-0.76114	0.03716	2.91134	1100	1154	1160	1181	1190

8.6.1 Percentages of Students in Each Performance Level

The empirical performance level distributions for the Spring 2023 administration of BIE Science Assessments are shown in Table 8-4.

Table 8-4. Performance Level Distribution as a Function of Grade*

Grade	Number of Students	Novice	Nearing Proficiency	Proficient	Advanced	% Novice	% Nearing Proficiency	% Proficient	% Advanced
5	1937	999	741	177	20	52%	38%	9%	1%
8	1662	479	928	254	1	29%	56%	15%	0%
11	870	497	267	105	1	57%	31%	12%	0%

Calculations based on those students attempting 5 or more items.

8.7 Differential Item Functioning

The *Code of Fair Testing Practices in Education* (Joint Committee on Testing Practices, 2004) explicitly states that subgroup differences in performance should be examined when sample sizes permit and that actions should be taken to ensure that differences in performance are due to construct-relevant, rather than irrelevant, factors. Chapter 3 of Standards for Educational and Psychological Testing (AERA et al., 2014) includes similar guidelines. As part of the effort to identify such problems, MSAA Science items were evaluated in terms of DIF statistics.

For items in the SSIB, the standardization DIF procedure (Dorans & Kulick, 1986) was employed to evaluate subgroup differences. The standardization DIF procedure is designed to identify items for which subgroups of interest perform differently, beyond the impact of differences in overall achievement. The DIF procedure calculates the difference in item performance for two groups of students (at a time) matched for achievement on the total test. Specifically, average item performance is calculated for students conditional on scale score. Then an overall average is calculated, weighting by the pooled scale score distribution so that it is the same for the two groups.

When differential performance between two groups occurs on an item (i.e., a DIF index in the "low" or "high" categories, explained below), it may or may not indicate item bias, e.g., caused by construct-irrelevant factors. Other construct-relevant reasons could also lead to DIF, such as course-taking patterns or differences in school curricula. On the other hand, if subgroup differences in performance can be traced to differential experience (such as geographical living conditions or access to technology), the inclusion of such items should be reconsidered.

For items in the SSIB, four subgroup comparisons were evaluated for DIF:

- Male compared with Female
- White compared with Black
- White compared with Hispanic
- White compared with Asian
- White compared with Multiracial
- Not economically disadvantaged status compared with economically disadvantaged

The DIF statistics were calculated based only on the members of the subgroup in question in the computations; values were calculated only for subgroups with 100 or more students. The tables in Appendix I present the number of items classified as either "low" or "high" DIF, overall and by group favored. Computed DIF indices have a theoretical range from -1.0 to 1.0 for selected-response items. Dorans and Holland (1993) suggested that index values between -0.05 and 0.05 should be considered negligible. The preponderance of the MSAA Science items fell within this range. Dorans and Holland further stated that items with values between -0.10 and -0.05 and those with values between 0.05 and 0.10 (i.e., "low" DIF) should be inspected to ensure that no possible effect is overlooked, and a list of low DIF items are listed in Appendix I, and that items with values outside the -0.10 to 0.10 range (i.e., "high" DIF) are more unusual and should be examined very carefully; thus, content experts conducted a review of items flagged for DIF.

No items used in BIE science assessments showed high DIF, a handful of items showed low DIF, as indicated in Table 8-5. These results indicate that the content bias reviews for science were conducted thoroughly.

Table 8-5. Number of Items with Low DIF

Subject	Grade	Student Subgroup	DIF Group Reference	DIF Group Focal	Number of Items favor_focal_low	Number of Items favor_reference_low
Science	05	Ethnic	White	Hispanic	--	2
		Gender	Female	Male	1	--
		Race	White	Asian	1	1
	08	Ethnic	White	Hispanic	--	2
		Gender	Female	Male	1	--
		Race	White	Multi-racial	--	1
	11	Gender	Female	Male	1	--
		Race	White	Multi-racial	1	--

Chapter 9. Score Reliability

9.1 IRT Marginal Reliabilities

IRT marginal reliability estimation is based on applying the standard classical test theory (CTT) formula, relating variances of true score, observed score, and measurement error, in the IRT setting. In CTT, the relationship between these variances is given by:

$$\sigma_X^2 = \sigma_T^2 + \sigma_E^2$$

where σ_X^2 is the observed-score variance,

       σ_T^2 is the true-score variance, and

       σ_E^2 is the error variance.

Starting from this equation, it can be shown that the formula for CTT reliability can be expressed by:

$$CTT\ Reliability = 1 - \frac{\sigma_E^2}{\sigma_X^2}.$$

IRT marginal reliability is based on extending the CTT model to an IRT framework (Samejima, 1994) and provides an IRT-based estimate of the overall test reliability. Error variance is estimated as the mean squared conditional standard error of measurement (CSEM) of the theta estimates across students within a grade. Observed-score variance is estimated as the variance of the theta estimates across students within a grade. Equivalently, the mean squared CSEM of the scale scores and the variance of the scale scores can be used in place of the CSEM of the theta estimates and the variance of the theta estimates, respectively. IRT marginal reliability is then given by the following formula:

$$IRT\ Marginal\ Reliability = 1 - \frac{\overline{CSEM(\theta)^2}}{Var(\hat{\theta})} = 1 - \frac{\overline{CSEM(SS)^2}}{Var(SS)},$$

where $\overline{CSEM(\theta)^2}$ is the mean squared CSEM,

       $\overline{CSEM(SS)^2}$ is the mean squared scaled CSEM,

       $Var(\hat{\theta})$ is the variance of theta estimates, and

       $Var(SS)$ is the variance of scaled scores.

Using this formula, IRT marginal reliability estimates were calculated for each grade. The reliability of a test can also be evaluated by simply examining the CSEMs themselves. CSEMs facilitate the interpretation of individual scale scores. With any given scale score estimate for a student, the reasonable limits of the true scale score for the student can be calculated by using the CSEM for the scale score.

Table 9-1 presents descriptive scale score statistics, IRT-based reliability, and mean scale score CSEMs by grade based on BIE student population. As shown in the table, most of the values reached levels associated with adequate reliability (0.70 or higher).

Table 9-1. IRT Marginal Reliability by Grade and Form

Grade	Scale Form	Min	Max	Mean	SD	IRT Marginal Reliability	Mean Scaled CSEM
5	1	506	588	544.73	10.90	0.84	4.29
	2	502	590	544.63	10.88	0.84	4.25
8	1	802	875	850.07	8.75	0.81	3.72
	2	800	886	851.35	9.57	0.84	3.62
11	1	1116	1172	1153.09	5.55	0.71	2.86
	2	1129	1182	1153.81	7.10	0.81	2.98

At the total test level and per grade, IRT marginal reliability ranged from 0.71 to 0.84. Note that IRT marginal reliability is partially dependent upon the variance in scaled scores. When present, range restriction in smaller samples can reduce the variance in scaled scores and therefore reduce the resulting value of IRT marginal reliability. Thus, while not necessarily ideal, the observed values of IRT marginal reliability may have been impacted due to the somewhat small sample sizes from the BIE student population taking the spring 2023 administration.

While subgroup reliability results are included in Table 9-2 for subgroups with at least 100 students, many of the subgroups have fewer than 100 students per subject and grade. Because the subgroup reliabilities are based on very small samples, no interpretations ought to be made about the adequacy of these subgroup reliabilities.

Given that, the results in Table 9-2 should be interpreted with appropriate levels of caution. Reliabilities are dependent not only on the measurement properties of a test, but also on the statistical distribution of the studied subgroup. Additionally, reliability estimates can be artificially depressed for subgroups with little variability in test scores (Draper & Smith, 1998).

Table 9-2. IRT Marginal Reliability by Grade and Subgroup

Grade	Type	Number of Student	Min	Max	Mean	SD	IRT Marginal Reliability	Mean Scaled CSEM
5	All	2,045	501	589	544.92	10.80	0.83	4.34
	Female	973	501	588	545.01	10.39	0.82	4.32
	Male	1,068	503	589	544.84	11.18	0.84	4.37
	Currently receiving LEP services	329	514	576	543.85	9.30	0.78	4.33
	Not receiving LEP services	341	503	577	546.11	11.10	0.85	4.29
	Economically Disadvantaged Students	1,964	501	589	544.84	10.82	0.83	4.35
	Gifted Students	110	519	588	550.45	12.61	0.89	4.18
	Non-gifted Students	101	514	579	544.78	10.16	0.81	4.31
	Title1 Students	1,972	501	589	544.91	10.73	0.83	4.34
8	All	1,764	800	885	848.95	9.93	0.83	3.95
	Female	830	803	874	848.88	9.55	0.82	3.94
	Male	934	800	885	849.00	10.26	0.84	3.96
	Currently receiving LEP services	215	829	871	848.72	8.16	0.77	3.86
	Not receiving LEP services	299	812	872	849.87	9.55	0.83	3.84
	Economically Disadvantaged Students	1,697	800	885	848.97	9.98	0.83	3.95
	Gifted Students	139	803	885	854.35	11.30	0.88	3.67
	Title1 Students	1,706	800	885	848.92	9.99	0.83	3.95
11	All	966	1111	1180	1153.50	6.16	0.76	2.90
	Female	482	1111	1174	1153.53	6.03	0.74	2.90
	Male	484	1130	1180	1153.47	6.29	0.77	2.91
	Currently receiving LEP services	167	1139	1172	1154.99	5.45	0.74	2.74
	Not receiving LEP services	223	1141	1180	1154.36	5.76	0.76	2.79
	Economically Disadvantaged Students	902	1111	1180	1153.50	6.18	0.76	2.91
	Title1 Students	906	1111	1180	1153.55	6.23	0.76	2.91

9.2 Decision Accuracy and Consistency

While related to reliability, the accuracy and consistency of classifying students into achievement categories are even more important statistics in a standards-based reporting framework (Livingston & Lewis, 1995). After the achievement levels were specified and students were classified into those levels, empirical analyses were conducted to estimate the statistical accuracy and consistency of the classifications.

Accuracy refers to the extent to which decisions based on test scores match decisions that would have been made if the scores did not contain any measurement error. Evaluation of decision accuracy is essential, considering all test scores contain measurement error. Consistency measures the extent to which classification decisions based on test scores match the decisions based on scores from a second, parallel form of the same test. Consistency can be evaluated directly from actual responses to test items if two complete and parallel forms of the test are given to the same group of students. In operational test programs, however, such a design is usually impractical.

However, techniques have been developed to estimate both the accuracy and the consistency of classification decisions based on a single administration of a test. The Rudner (2001, 2005) technique was used for the BIE Science assessments because it can be easily applied to data that are scored in the IRT theta metric or any linear transformation of this metric, such as the scale scores. The applicability of the Rudner technique to IRT-based metrics distinguishes this method from methods based on observed scores, such as the Lewis and Livingston (1995) method.

For details of the Rudner method, refer to Rudner (2001, 2005); given here is a brief review of the basic idea behind the method. Using an examinee's estimated scale score and standard error, assuming a normal probability distribution, the method first calculates for all examinees at a fixed value of true scale score, the expected proportion whose observed scale score is in an interval [a,b]. Then, by summing over all examinees whose true scale scores are in an interval [c,d], the method yields the expected proportion of all examinees whose true scale score is in [c,d] and whose observed scale score is in [a,b]. By setting [a,b] and [c,d] to correspond to the true score intervals defined by the cut scores yields the elements of a classification table that shows the expected proportion of all examinees with observed and true scale scores in each cell. These proportions can then be used to calculate both classification accuracy and classification consistency estimates.

For the classification accuracy tables, cell [*i*, *j*] represents the estimated proportion of students whose true scale score fell into classification *i* (where *i* = 1 to 4, for the four achievement levels) and whose observed scale score fell into classification *j* (where *j* = 1 to 4). The sum of the diagonal entries (i.e., the proportion of students who's true and observed classifications matched) signified overall accuracy.

Another way to measure consistency is to use κ (kappa; Cohen, 1960), which indicates the proportion of consistent classifications after removing the proportion of consistent classifications that would be expected by chance. It is calculated using the following formula:

$$\kappa = \frac{(Observed\ agreement) - (Chance\ agreement)}{1 - (Chance\ agreement)} = \frac{\sum_i C_{ii} - \sum_i C_{i.} C_{.i}}{1 - \sum_i C_{i.} C_{.i}},$$

where $C_{i.}$ is the proportion of students whose observed achievement level would be Level i (where i = 1–3) on the first hypothetical parallel form of the test;

$C_{.i}$ is the proportion of students whose observed achievement level would be Level i (where i = 1–3) on the second hypothetical parallel form of the test; and

C_{ii} is the proportion of students whose observed achievement level would be Level i (where i = 1–3) on both hypothetical parallel forms of the test.

Because κ is corrected for chance, its values are lower than are other consistency estimates.

Figure 9-1 shows the overall decision accuracy for science by grade level. More details on decision accuracy and consistency (DAC) are provided in Appendix I. Table I-1 in Appendix I includes overall accuracy and consistency indices, along with kappa. Accuracy and consistency values conditional on performance level are also provided in Table I-1. For these calculations, the denominator is the proportion of students associated with a given performance level. Following is an example from Table I-1, looking at Level 1 for grade 5 for scale form 1.

- The conditional accuracy value was 0.84. This indicates that among the students whose true scale scores placed them in Level 1, 84% would be expected to be in this same level again when categorized according to their observed scale scores.
- The *consistency* value was 0.76. This indicates that among the students whose *observed scale scores* placed them in Level 1, 76% would be expected to be in this same level again if a second parallel test form were used.

For some testing situations, the greatest concern may be decisions regarding level thresholds. For example, in testing done for the Every Student Succeeds Act (ESSA) accountability purpose, the primary concern is distinguishing between students who are proficient and those and those who are not yet proficient. Table I-2 in Appendix I provides accuracy and consistency estimates at each cutpoint, as well as false positive and false negative decision rates. A false positive rate is the proportion of students whose observed scores were above the cut and whose true scores were below the cut. A false negative rate is the proportion of students whose observed scores were below the cut and whose true scores were above the cut.

Figure 9-1a. Overall Decision Accuracy for Science by Grade for Form 1 (Set A)

Overall Decision Accuracy for BIE Science Set A

Overall Decision Accuracy for BIE Science Set B
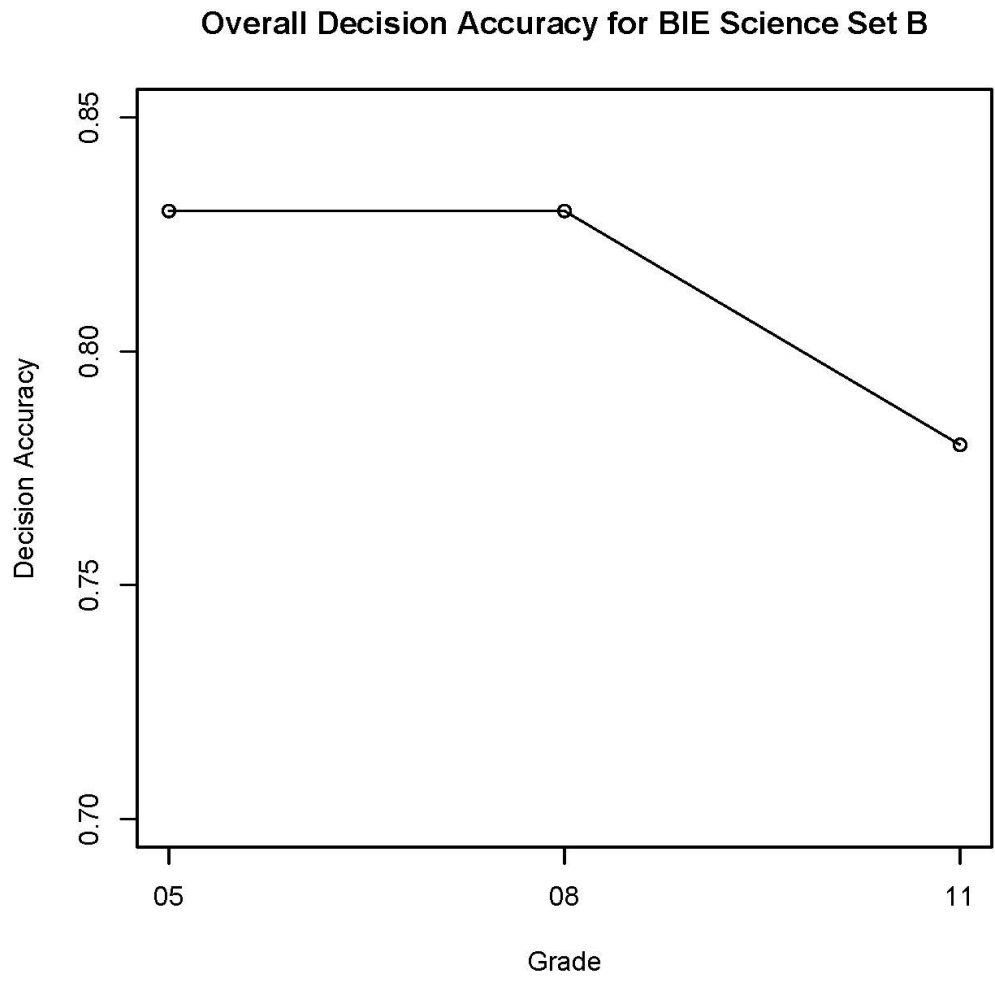
Chapter 10. Score Reporting

10.1 Individual Student Reports

The individual student report (ISR) is distributed to schools and parents. The report is produced in color. One copy of the report is sent home to parents by the school and one copy of the report is retained by the school. The report is designed to print on one 8 ½- x 11-inch sheet of paper printed duplex. The front of the report contains demographic identifying information about the student. It also contains a letter to Parents and Guardians. A list of resources is presented on the front page as well. A graphical display of the scale and achievement levels is shown, as well as the achievement level descriptors. It should be noted that Cognia carried forward the agreed upon report design from the previous year's administration.

The back page of the report contains all the students' performance information. Each student's performance on the BIE Science Assessment is described in the ISR. Based on the student's earned score the student is assigned to one of the following achievement levels: Nearing Proficiency, Proficient, and Advanced. The report contains scale scores, achievement levels, standard error, and reporting category performance indicators. The report also contains the percentage of students in each performance indicator for each reporting category. The student's scale score is compared to the average scale score at the school, district, and overall BIE levels. The report provides a descriptor for the earned achievement level and the level above if the earned achievement level is not Advanced. Examples are provided along with the descriptors. For additional information concerning the individual student report, see Appendix J–BIE Science Reporting Business Requirements and Appendix K (Score Report Interpretation Quick Guide) for the sample reports.

10.2 Scale Score

A scale score is a numerical value that summarizes student performance. Not all students respond to the same set of test items, so each student's scaled score accounts for the slight differences in difficulty among the various forms and administrations of the test. The resulting scale score allows for an appropriate comparison across test forms and administration years within a grade or course and content area. BIE reports provide overall scale scores for science, which determine a student's achievement level for the content area. The scale score range is XX00-XX90 where XX=the student's tested grade. For example, in Grade 5 the range is 500-590.

For example, a student who earns an overall scale score of 800 on one form of the grade 8 assessment would be expected to earn an overall scale score of 800 on any other form of the grade 8 assessment. Furthermore, the student's overall scale score and level of mastery of concepts and skills would be comparable to a student who took the same assessment the previous year or the following year. For cumulative scale-score distributions see Appendix L; for scaled score descriptive statistics, see Appendix M.

10.3 Achievement Level

Each BIE achievement level is a broad category that is defined by a student's overall scale score and is used to report overall student performance by describing how well students met the expectations for their grade levels. There are four achievement levels for the Spring 2023 BIE Science Assessments:

Novice. Students demonstrate evidence of **emerging** understanding and use of college and career readiness knowledge, skills, and abilities.

Nearing Proficiency. Students demonstrate evidence of **partial** understanding and use of college and career readiness knowledge, skills, and abilities.

Proficient. Students demonstrate evidence of **satisfactory** understanding and use of college- and career-readiness knowledge, skills, and abilities.

Advanced. Students demonstrate evidence of **thorough** understanding and use of college and career readiness knowledge, skills, and abilities.

These PLDs are referred to as Policy Definitions for reporting BIE performance in science. Range PLDs describe the knowledge and skills that students throughout the range of each proficiency level are expected to be able to demonstrate in each grade. For example, in line with the nature of the science standards, the science range PLDs combine science and engineering practices, disciplinary core ideas, and crosscutting concepts that students in grades 5, 8, and 11 are expected to integrate and demonstrate. The range PLDs appear in Appendix B.

10.4 Standards Performance Indicators

Standards performance for BIE assessments is reported using words and colors that indicate whether the student performed above standard, at/near standard, or below standard in each standard. Additional information about standard performance indicators is in the Score Report Interpretation Quick Guide, Appendix K in this document. Performance indicators for each standard are summarized at the overall level and presented as whole number percentages in a horizontal stacked bar that is color coded to represent the performance indicators.

10.5 Comparison of Student Performance at the School, District, and Overall Level

The last section of the back page is a table containing the Achievement levels, Achievement Level Descriptors, scale score range for each achievement level and the student's overall science scale score on a bar graph compared to the scale score average at the school, district and overall (BIE) level. The student's scale score bar is colored orange. The other bars are colored in the same shade of blue.

10.6 Student Roster

A student roster is produced for each participating school. The roster lists all students in the reporting dataset. The student is either reported with a tested status or a not-tested reason. See the Reporting Business Requirements document for more information on participation statuses. The Roster is printed in color on 8 ½- x 11-inch paper duplex. The report is oriented landscape. The report lists the following information about each student:

Student Name

Student's NASIS ID

Student's Grade

Student's Gender

Student's Testing Status

Student's Scale Score

Student's Achievement Level

The report is marked as confidential.

The roster is printed and shipped to the schools.

Chapter 11. Validity Arguments to Support Intended Score Interpretation and Uses

This chapter presents the primary intended score interpretation and three primary intended score uses. This chapter also presents the claims and subclaims that underlie these four score interpretations and uses (SIUs) and the evidence that supports the claims and subclaims. The BIE validity argument model is introduced and applied to develop validity arguments to support the four SIUs.

The *Standards for Educational and Psychological Testing* (2014) defines validity as "the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests" (p. 11). Elaborating on that definition, the *Standards* asserts that "it is the interpretations of test scores for proposed uses that are evaluated, not the test itself" (p. 11) and that "validation logically begins with an explicit statement of the proposed interpretation of test scores, along with a rationale for the relevance of the interpretation to the proposed use" (p. 11). This definition applies specifically to intended interpretations and uses of test scores, rather than to the broader program of curriculum and instruction in which a testing program is embedded or to the surrounding education and school improvement policies and aspirations for student learning.

Further, the *Standards* states that "a sound validity argument integrates various strands of evidence into a coherent account of the degree to which existing evidence and theory support the intended interpretations of test scores for specific uses" (p. 21). We use these views in the *Standards,* that evidence must be used to support score interpretation and use claims, as the basis for the BIE validity argument model, which we describe next.

Emerging practice in state assessment programs is to construct validity arguments based on Toulmin's model of argumentation (Toulmin. 1958), Chapelle's proposed practice-oriented adaptation (2021), and Kane's formulation of validity arguments (2013). A model for BIE validity arguments, derived from these three conceptualizations, is shown in Figure 11-1. The first panel shows the BIE model; the second panel is an illustration for a validity argument for a score interpretation and use statement.

Figure 11-1. BIE Validity Argument Model
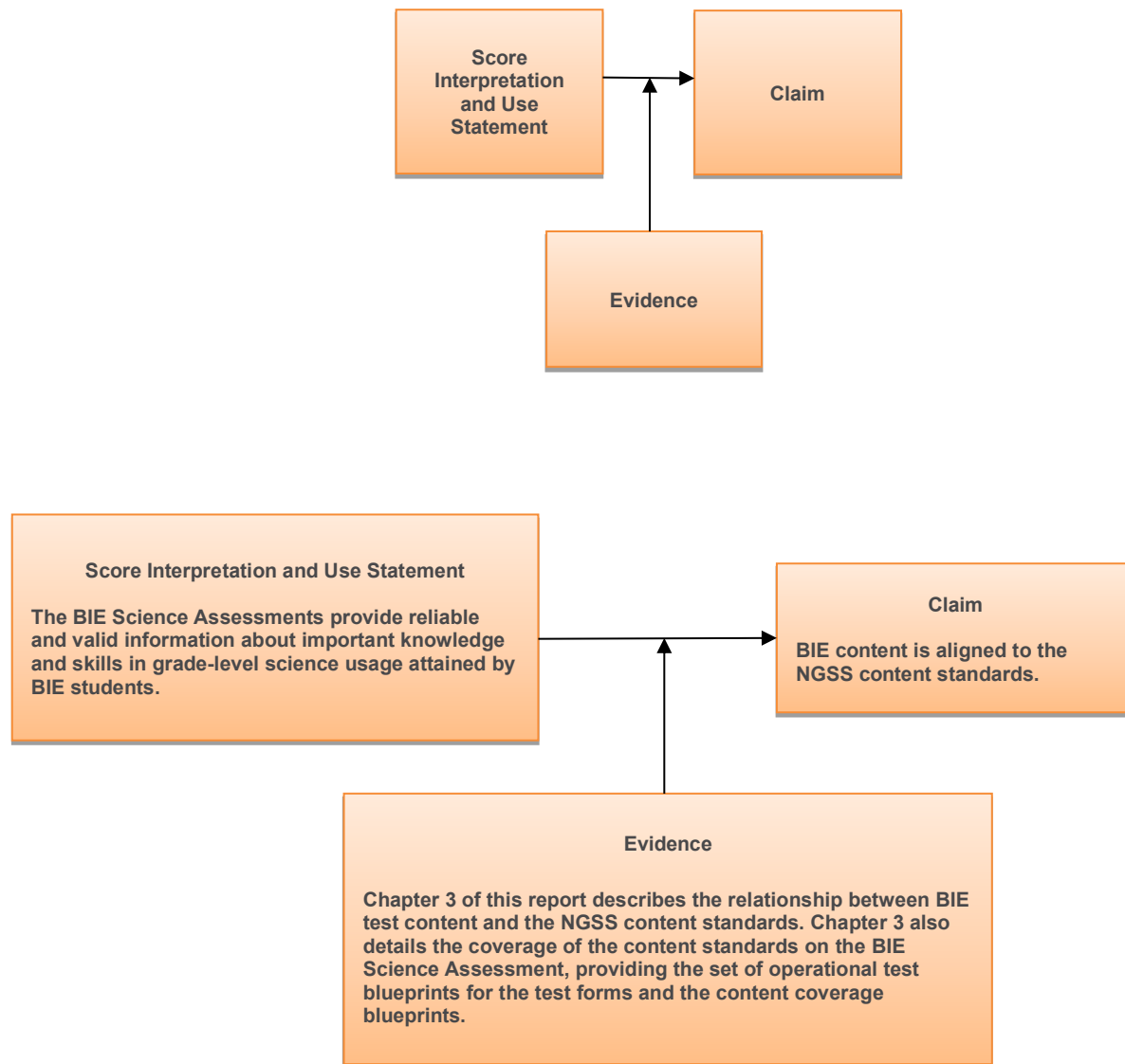
Adapted from Chapelle (2021) Figures 2.1-2.3, Kane (2013) Figure 1, and Toulmin (1958).

The first panel in Figure 11-1, displays a generic representation of a BIE SIU statement and how it connects directly to a corresponding validity argument and claims and subclaims, shown at the right. The evidence that supports the BIE SIUs, claims, and validity arguments connects directly to the claims-validity argument pairings. The second panel displays a specific example of this relationship.

In the BIE validity argument model, the overall validity argument is that the existing design, procedural, and psychometric evidence supports the intended score interpretations and uses. Each of the interpretations and uses represents a set of claims, subclaims, and validity arguments that require supporting evidence to connect the evidence to the score interpretation and use. This line of reasoning and argumentation creates supported validity arguments. The remaining sections of this chapter describe the a) intended SIUs; b) claims, subclaims, and validity arguments, which connect the BIE design, procedural, psychometric, and other program information to the SIUs; and c) evidence that supports each SIU and validity argument (for which detail is provided in the previous chapters).

The relationships among the score interpretations and uses, claims, and validity arguments appear in Table 11-1. Each entry in the table is presented following the table, with descriptions and summaries of the supporting evidence.

Table 11-1. Relationships Among Score Interpretations and Uses, Claims, and Sub-Claims, and Supporting Evidence

Claims	Claims and Subclaims that Support Score Interpretations and Uses
	SIU 1: Primary Intended Score Interpretation
	The BIE Science Assessment provides reliable and valid information about important knowledge and skills in grade-level science attained by BIE general education students.

Claim 1.1: The content of the tests represents the content of the standards.

    1.1.1   BIE content is aligned to Next Generation Science Standards.

    1.1.2   BIE items are aligned to Next Generation Science Standards.

Claim 1.2: The test items are construct-relevant.

    1.2.1.   Items require application of the knowledge, skills, and abilities (KSAs) of the targeted construct.

    1.2.2.   Items are free of bias and sensitivity issues.

    1.2.3   Students' cognitive skills and processes match those identified in the construct domains for all students and for each subgroup.

Claim 1.3: Test scores on the BIE Science Assessments provide reliable information about student performance and accurate classifications into performance levels.

    1.3.1.   BIE scores and performance level categorizations are adequately reliable for their intended purpose.

    1.3.2.   Item characteristics support intended interpretations about all students who take the BIE Science Assessment.

    1.3.3.   Test characteristics support intended interpretations about all students who take the BIE Science Assessment.

Claim 1.4: Item and test scoring are implemented accurately.

    1.4.1.   Machine-scored items were scored accurately.

    1.4.2.   Constructed-response item scoring training and monitoring procedures met industry standards.

	SIU 2: Intended Score Use for Individual Students

Claim 2.1: Educators, schools, and districts can use results from the BIE Science Assessment to describe student achievement status with respect to mastery of the content standards.

    2.1.1.   BIE test scores and performance level categorizations of individual students are adequately reliable and valid measures of student achievement status with respect to mastery of the content standards.

	SIU 3: Intended Score Use for Groups of Students

Claim 3.1: Educators can use results from the BIE Science Assessment to support instructional planning for groups of students.

    3.1.1.   Teachers find the performance level descriptors and their students' performance levels useful for planning instruction, especially for students whose test scores fall within performance levels 1 and 2.

    3.1.2.   Teachers find their students' scale score information useful for planning instruction, especially for students whose test scores fall within performance levels 1 and 2.

Claim 3.2: Schools and districts can use results from the BIE Science Assessment to make comparisons between organizations (e.g., schools, districts).

    3.2.1.   BIE scores and performance levels for groups of students are adequately reliable and valid to enable school, district, and state leaders to monitor changes in means, standard deviations, and performance level percentages for classroom, school, district, and state groups.

    3.2.2.   BIE scores and proficiency level categorizations of groups of students are adequately reliable and valid to enable monitoring of grade-level performance and student-cohort performance.

Evidence that supports SIUs and claims in BIE validity arguments is summarized below, using a relevance-rating scale, with rating levels defined in Table 11-2.

Table 11-2. Relevance and Completeness or Completeness of Evidence in Support of SIUs and Claims Underlying Validity Arguments for BIE Score Interpretations and Uses

Complete Evidence	When all required pieces of relevant evidence are provided to support a validity argument
Moderate to Substantial Evidence	When several pieces of relevant evidence are provided, but not all required pieces of evidence are provided
Limited Evidence	When only one or two pieces of evidence are provided, where the evidence may be only marginally relevant, or where more than one or two pieces of evidence are required
No Evidence	When no relevant evidence exists

11.1 Primary Intended Score Interpretation

The primary intended score interpretation for the BIE Science Assessment (SIU 1) states that the BIE Assessments provide reliable and valid information about important knowledge and skills in grade-level science usage attained by general education students.

Claim 1.1. The content of the tests represents the content of the standards.

Items used on the BIE Science Assessment are developed to measure achievement on the Next Generation Science Standards in the assessed content areas. Cognia content specialists, in collaboration with the BIE, ensure this alignment, and ongoing independent evaluations are held to verify alignment. In addition, independent reviews are scheduled to ensure that items and passages conform to bias and sensitivity guidelines.

Subclaim1.1.1. BIE Science Assessment content is aligned to Next Generation Science Standards.

> **Evidence**: Chapter 3 of this report describes the relationship between BIE test content and the Next Generation Science Standards. Chapter 3 also details the coverage of the content standards on the BIE Science Assessment, providing the set of operational test blueprints for test forms and the content coverage blueprints.

> BIE Science Assessments uses Cognia Secure Science Item Bank (SSIB) and has the exact same test design as well as uses many of the same items on the test as the other state that had the alignment study. Theoretically, the alignment study results on the other state should still hold for BIE Science Assessments and no additional alignment study is needed.

> **Summary of evidence**: Complete evidence.

Subclaim 1.1.2. BIE Science Assessment items are aligned to the Next Generation Science Standards.

> **Evidence**: Chapter 4 describes the item specifications and standardized item writer training in support of new item development. Chapter 4 also details the item review process performed by item review committees, to ensure item content alignment to the intended content standard. Content alignment for BIE science item bank and test forms have been verified by an independent alignment study in other states. Therefore, an additional alignment study that is specific to BIE assessments is not needed.

Summary of evidence: Complete evidence.

Subclaim 1.2.1. Items require the application of the KSAs of the targeted construct.

Evidence: The 2023 operational items are aligned with the Next Generation Science Standards. The evidence for element 1.2.1 is directly linked to the subclaims 1.1.1 and 1.1.2 above.

Summary of evidence: Complete evidence.

Subclaim 1.2.2. Items are free of bias and sensitivity issues.

Evidence*: During the item development process, the items followed a rigorous development cycle that includes reviews by subject matter experts and by Item Content and Bias and Sensitivity panelists. The item development process also includes data reviews, during which item-level statistics—including differential item functioning (DIF) statistics—are reviewed. See Chapter 8 for a detailed description of the item review process.

Summary of evidence*: Complete evidence, based on current Cognia procedures for the Spring 2023 testing season.

Subclaim 1.2.3. Students' cognitive skills and processes match those identified in the construct domains for all students and for each subgroup.

Evidence*: Cognitive-process evidence examines the extent to which students' cognitive skills and processes match those identified in the construct domains defined by test developers for all students and for each subgroup.

Evidence of cognitive processes can be obtained through subject matter experts (including educators) review or a study. The items for BIE Science (Cognia's Science Secure Item Bank) were developed and internally reviewed by a team of science content specialists, and then the items were reviewed by state/entity department of education content and/or assessment specialists and science educators from the field. Their input helped ensure the tasks would measure the intended cognitive ability. In addition to Cognia content expert reviews of the items for alignment, content accuracy, and cognitive complexity levels, science educators were involved in the review of all items developed and field tested. When reviewing items in committee review meetings, target students' characteristics (such as various learning targets, use of the different science dimensions, and the degree of cognitive processing elicited by the stimulus and application of each dimension) were all discussed and confirmed. The feedback from educators/specialists with expertise with science content in these committee meetings indicated that the test content did not require extraneous cognitive processes for engagement.

Summary of evidence*: Complete evidence, based on current Cognia procedures for the Spring 2023 testing season.

Claim 1.3: Test scores on the BIE Science Assessment provide reliable information about student performance and accurate classifications into performance levels.

Subclaim 1.3.1. BIE scores and performance level categorizations are adequately reliable for their intended purpose.

Evidence:
Score Reliability: Chapter 9 provides a description of IRT reliability theory and interpretation and a review of the relevant equations. Table 9-1 contains the reliability

results by grade. Table 9-2 also contains reliability results disaggregated by student subgroups. These reliability estimates are consistent with industry standards, which can be observed in technical reports posted online by other state assessment programs.

Scale Score Standard Errors: Chapter 8 provides a description of calculation and interpretation of the scaled scores and Chapter 9 provides a description of the calculation of the standard error for a scaled score. The average standard error for reported scaled scores is reported in Table 9-1. The scale score standard error can be compared to the scale score range and the scale score standard deviation to provide some context for interpretation. These standard error estimates are consistent with industry standards, which can be observed in technical reports posted online by other state assessment programs.

Decision Consistency and Accuracy Estimates: Decision accuracy is an estimate of the probability that the observed classification is the true classification. Decision consistency is an estimate of the probability that students would receive the same classification if they tested twice on parallel forms. Chapter 9 describes the theory and equations underlying the estimation of classification accuracy and consistency. Decision accuracy and consistency results are provided in Appendix I. These decision consistency and accuracy estimates are consistent with industry standards, which can be observed in technical reports posted online by other state assessment programs.

Summary of evidence: Complete evidence.

Subclaim 1.3.2. Item characteristics support intended interpretations about all students who take the BIE Science Assessment.

Evidence: The psychometric characteristics most pertinent to evaluating the adequacy of individual items are the estimated item parameters. The item parameter estimates are provided in Appendix F. For dichotomously scored items, the item parameters include the discrimination, difficulty, and lower asymptote parameters. For polytomously scored items, the item parameter estimates include the discrimination, location, and item-category parameters. All items undergo statistical analyses at the time of field-testing, including classical, DIF, and IRT analyses. As stated in Chapter 4, the results of these analyses are reviewed in Data Review meetings with national subject matter experts. After field-testing and prior to operational administration, items from the previous operational administration are reviewed for their item information function (IIF) contributions at the performance level cuts to evaluate and rate the quality of each item. After each operational administration, dimensionality analyses are also conducted to determine the adequacy of the unidimensional IRT model used for scaling, equating, and scoring the BIE students.

Summary of evidence: Complete evidence.

Subclaim 1.3.3. Test characteristics support intended interpretations about all students who take the BIE Science Assessment.

Evidence:

High correlations (e.g., greater than or equal to 0.7) among content area subdomain indicators and the relatively low reliability of these indicators demonstrate that such indicators must be interpreted and used cautiously, and in conjunction with other information about student achievement and learning needs in these areas.

Dimensionality: Dimensionality analysis was conducted on each grade-level test. Chapter 7, section 7.2, provides a detailed description of the dimensionality hypothesis testing and effect-size estimation methods and provides dimensionality results. No statistically significant violations of local independence were noted.

Conditional Standard Errors of Measurement: Chapter 9 provides a detailed description of the psychometric model that was fitted to the data, the test information function (TIF), and the inverse transformation of the TIF into the Conditional Standard Error of Measurement (CSEM). The TIF and CSEM are inverse transformations of each other. Whereas the TIF indicates test score precision, the CSEM indicates the converse, i.e., test score imprecision or measurement error. The TIF and its analogue, the CSEM, are the most pertinent products of the psychometric model in evaluating the adequacy of a test (form). By examining the value of CSEM at each of the performance level cut scores, the psychometric appropriateness and accuracy of each test can be evaluated.

> **Content Coverage**: Subclaims 1.1.1, 1.1.2, and 1.2.1 above detail the evidence in support of the content coverage and the alignment of the content to the BIE standards.
>
> **Scoring**: Subclaims 1.4.1 and 1.4.2 detail the evidence in support of accurate item and test scores.

Summary of evidence*:* Complete evidence.

Claim 1.4: Item and test scoring are implemented accurately.

Subclaim 1.4.1. Machine-scored items were scored accurately.

Evidence*:* As described in Section 6.2.1 of Chapter 6 and in Chapter 7, a classical item analysis on the set of machine-scored items is performed prior to scaling and equating. This ensures that for each machine-scored item, the response designated as the correct response was indeed the correct response.

Summary of evidence*:* Complete evidence.

Subclaim 1.4.2. Constructed-response item scoring training and monitoring procedures met industry standards.

Evidence*:* As detailed in Chapter 6, scorer recruitment, training, and qualification and scoring-monitoring procedures follow industry standards. Section 6.2.2, Scoring of Open-Ended Response Items, describes all the procedures that are used to ensure the accuracy of the scoring for the open-ended (constructed) response items, including administrator training and monitoring, benchmarking and identification of scoring materials, scorer recruitment and qualifications, scoring leadership, qualification, specific scoring rules to ensure accuracy, monitoring of quality control, quality reports, and interrater reliability.

Summary of evidence: Complete evidence.

11.2 Primary Intended Score Uses

11.2.1 Intended Score Use for Individual Students

Claim 2.1: Educators and school and district administrators can use results from the BIE Science Assessment to describe and monitor student achievement status with respect to mastery of the content standards.

Subclaim 2.1.1. BIE test scores and performance level categorizations of individual students are adequately reliable and valid measures of student achievement status with respect to mastery of the content standards.

> **Evidence**:
>> Scaled Score Standard Errors: Chapter 8 provides a description of calculation and interpretation of the scaled scores and Chapter 9 provides a description of the calculation of the standard error for a scaled score. The average standard error for reported scaled scores is reported in Appendix H. The scaled score standard error can be compared to the scaled score range and the scaled score standard deviation to provide some context for interpretation.
>>
>> Decision Consistency and Accuracy Estimates: Decision accuracy is an estimate of the probability that the observed classification is the true classification. Decision consistency is an estimate of the probability that students would receive the same classification if they tested twice on parallel forms. Chapter 9 describes the theory and equations underlying the estimation of classification accuracy and consistency. Decision accuracy and consistency results are provided in Appendix I.
>>
>> **Content Coverage**: Subclaims 1.1.1, 1.1.2, and 1.2.1 above detail the evidence in support of the content coverage and the alignment of the content to the BIE standards.
>>
>> Scoring: Subclaims 1.4.1 and 1.4.2 detail the evidence in support of accurate item and test scores.
>
> **Summary of evidence**: Complete evidence.

11.2.2 Intended Score Use for Groups of Students

Claim 3.1: Educators can use results from the BIE Science Assessment to support instructional planning for groups of students.

Subclaim 3.1.1. Teachers find the performance level descriptors and their students' performance levels useful for planning instruction, especially for students whose test scores fall within performance levels 1 and 2.

> **Evidence**: None.
>
> **Summary of evidence**: No evidence. An example of a source of evidence could be a survey of teachers to begin to understand the degree to which teachers find BIE performance level descriptors and their students' performance levels useful for planning instruction.

Subclaim 3.1.2. Teachers find their students' scale score information useful for planning instruction, especially for students whose test scores fall within performance levels 1 and 2.

> **Evidence**: None.

Summary of evidence: No evidence. An example of a source of evidence could be a survey of teachers to begin to understand the degree to which teachers find BIE scores useful for planning instruction.

Claim 3.2: Schools, districts, and state-level stakeholders can use results from the BIE Science Assessment to make comparisons between organizations (e.g., schools, districts).

Subclaim 3.2.1. BIE scores and performance levels for groups of students are adequately reliable and valid to enable school, district, and state leaders to monitor changes in means, standard deviations, and performance level percentages for classroom, school, district, and state groups.

Evidence: Evidence for the reliability and validity of the scores and the corresponding scoring processes is presented above under Claim 1.3, which cites Chapter 6 on scoring, Chapter 8 on IRT scaling and equating, and Chapter 9 on IRT reliability and decision accuracy and consistency. The reliability of aggregated scores (e.g., means) is typically as high as or higher than individual score reliabilities (e.g., Brennan, 1995). Appendix I contains the decision accuracy and consistency results for the overall test as well as by performance level and by cut score. Subclaims 1.1.1, 1.1.2, and 1.2.1 above detail the evidence in support of the content coverage and the alignment of the content to the BIE standards. Subclaims 1.4.1 and 1.4.2 detail the evidence in support of accurate item and test scores.

Summary of evidence: Moderate to substantial evidence.

Subclaim 3.2.2. BIE scores and proficiency level categorizations of groups of students are adequately reliable and valid to enable monitoring of grade-level performance and student-cohort performance.

Evidence: Evidence for the reliability and validity of the scores and the corresponding scoring processes is presented above under Claim 1.3, which cites Chapter 6 on scoring, Chapter 8 on IRT scaling and equating, and Chapter 9 on IRT reliability and decision accuracy and consistency. The reliability of aggregated scores (e.g., means) is typically as high as or higher than individual score reliabilities (e.g., Brennan, 1995). Appendix I contains the decision accuracy and consistency results for the overall test as well as by performance level and by cut score. Subclaims 1.1.1, 1.1.2, and 1.2.1 above detail the evidence in support of the content coverage and the alignment of the content to the BIE standards. Subclaims 1.4.1 and 1.4.2 detail the evidence in support of accurate item and test scores.

Summary of evidence: Moderate to substantial evidence.

11.3 Conclusions and Next Steps

The majority of the claims and subclaims that support the four claims—that is, the primary intended score interpretations and three intended score uses—are supported by solid evidence. These claims and subclaims and their supporting evidence comprise the validity arguments for BIE scores. Table 11-3 summarizes the relevance ratings for each claim and subclaim.

Table 11-3 indicates the following:

Primary Score Intended Score Interpretation

Of the three claims and nine subclaims that support the intended score interpretation, eight sets of evidence are complete, and one set of evidence is moderate to substantial.

Intended Score Use for Individual Students

The one claim has one supporting subclaim that is moderate to substantial.

Intended Score Use for Groups of Students

Of the two claims and four supporting subclaim evidence, two sets of evidence are moderate to substantial and two subclaims do not currently have evidence.

Table 11-3. Status of Evidence for All SIUs, Claims, and Subclaims

SIUs, Claims, and Subclaims	Relevance of the Evidence to the Validity Argument			
	No Evidence Exists Currently	Limited	Moderate to Substantial	Complete
SIU 1: Primary Intended Score Interpretation				
The BIE Science Assessment provides reliable and valid information about important knowledge and skills in grade-level science usage attained by general education students.				
1.1.1. BIE content is aligned to Next Generation Science Standards.				X
1.1.2. BIE items are aligned to Next Generation Science Standards.				X
1.2.1. Items require application of the KSAs of the targeted construct.				X
1.2.2. Items are free of bias and sensitivity issues.				X
1.2.3 Students' cognitive skills and processes match those identified in the construct domains for all students and for each subgroup.				X
1.3.1. BIE scores and performance level categorizations are adequately reliable for their intended purpose.				X
1.3.2. Item characteristics support intended interpretations about all students who take the BIE Science Assessment.				X
1.3.3. Test characteristics support intended interpretations about all students who take the BIE Science Assessment.	X			

SIUs, Claims, and Subclaims	Relevance of the Evidence to the Validity Argument			
	No Evidence Exists Currently	Limited	Moderate to Substantial	Complete
1.4.1. Machine-scored items were scored accurately.				X
1.4.2. Constructed-response item scoring training and monitoring procedures met industry standards.				X
SIU 2: Intended Score Use for Individual Students Scale scores can be used to compare an individual student's performance to the performance of other students in BIE.				
2.1.1. BIE test scores and performance level categorizations of individual students are adequately reliable and valid measures of student achievement status with respect to mastery of the content standards.			X	
SIU 3: Intended Score Use for Groups of Students SIU statements for groups of students are applicable to aggregate reporting of student subgroups (e.g., English learners, students with disabilities, racial/ethnic subgroups) within those levels of aggregation.				
3.1.1. Teachers find the performance level descriptors and their students' performance levels useful for planning instruction, especially for students whose test scores fall within performance levels 1 and 2.	X			
3.1.2. Teachers find their students' scale score information useful for planning instruction, especially for students whose test scores fall within performance levels 1 and 2.	X			
3.2.1. BIE scores and performance levels for groups of students are adequately reliable and valid to enable school, district, and state leaders to monitor changes in means, standard deviations, and performance level percentages for classroom, school, district, and state groups.			X	
3.2.2. BIE scores and proficiency level categorizations of groups of students are adequately reliable and valid to enable monitoring of grade-level performance and student-cohort performance.			X	

11.3.1 Research Agenda

The Score Card ratings provide a road map for a research agenda for the BIE science program. Specifically, the BIE and Cognia can work together to identify the highest priority claims and subclaims for which *No Evidence Exists Currently* and the evidence is *Limited* and plan studies to gather relevant evidence and strengthen validity arguments. This will be a topic of discussion and planning for more immediate and longer-term efforts during the 2023–2024 school year.

References

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

Baker, F. B., & Kim, S.H. (2004). *Item response theory: Parameter estimation techniques* (2nd ed.). New York: Marcel Dekker, Inc.

Brennan, R. L. (1995). The conventional wisdom about group mean scores. *Journal of Educational Measurement*, *32*(4), 385 –396.

Brown, F. G. (1983). *Principles of educational and psychological testing* (3rd ed.). Fort Worth: Holt, Rinehart, and Winston.

Chapelle, C. A. (2021). *Argument-based validation in testing and assessment*. Sage.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*, 37–46.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297–334.

Dorans, N. J., & Kulick, E. (1986). Demonstrating the utility of the standardization approach to assessing unexpected differential item performance on the Scholastic Aptitude Test. *Journal of Educational Measurement*, *23*, 355–368.

Draper, N. R., & Smith, H. (1998). *Applied regression analysis* (3rd ed.). New York, NY: John Wiley and Sons.

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston, MA: Kluwer Academic Publishers.

Hambleton, R. K., & van der Linden, W. J. (1997). *Handbook of modern item response theory*. New York, NY: Springer-Verlag.

Joint Committee on Testing Practices. (2004). *Code of fair testing practices in education*. Washington, D.C.: National Council on Measurement in Education.

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, *50*(1), 1–73.

Livingston, S. A., & Lewis, C. (1995). Estimating the consistency and accuracy of classifications based on test scores. *Journal of Educational Measurement*, *32*, 179–197.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of meta test scores*. Reading, MA: Addison-Wesley.

Persuasiveness. (2019). In *Vocabulary.com* Retrieved from https://www.lexico.com/en/definition/persuasive

Plausibility. (2019). In *Lexico.com* Retrieved from https://www.lexico.com/en/definition/plausibility

Relevance. (2019). In *Lexico.com* Retrieved from https://www.lexico.com/en/definition/relevance

Roussos, L. A., & Ozbek, O. Y. (2006). Formulation of the DETECT population parameter and evaluation of DETECT estimator bias. *Journal of Educational Measurement*, *43*, 215–243.

Samejima, F. (1994). Estimation of reliability coefficients using the test information function and its modifications. *Applied Psychological Measurement, 18,* 229–244.

Stout, W. F. (1987). A nonparametric approach for assessing latent trait dimensionality. *Psychometrika*, *52*, 589–617.

Stout, W. F., Froelich, A. G., & Gao, F. (2001). Using resampling methods to produce an improved DIMTEST procedure. In A. Boomsma, M. A. J. van Duign, & T. A. B. Snijders (Eds.), *Essays on item response theory* (pp. 357–375). New York, NY: Springer-Verlag.

Toulmin, S. (1958). *The uses of argument*. Cambridge, UK: Cambridge University Press.

Zhang, J., & Stout, W. F. (1999). The theoretical DETECT index of dimensionality and its application to approximate simple structure. *Psychometrika*, *64*, 213–249.

Appendices

APPENDIX A
LIST OF ACRONYMS

Common Acronyms Used in Assessment Reports	
3PL	three-parameter logistic
AERA	American Educational Research Association
APA	American Psychological Association
BIE	Bureau of Indian Education
CBT	computer-based test
CR	constructed response items
CSEM	conditional standard error of measurement
CTT	classical test theory
DAC	decision accuracy and consistency
DETECT	Dimensionality Evaluation to Enumerate Contributing Traits
DIF	differential item functioning
DIMTEST	computer program used by Cognia
DOK	depth of knowledge
ESSA	Every Student Succeeds Act
ETS	Engineering, Technology, and the application of science
GRM	Graded-Response Model
ICC	item characteristic curve
ICCC	Item Category Characteristic Curve
ICTC	Item Category Threshold Curve
IIF	item information function
IRT	item response theory
ISR	individual student report
KSA	knowledge, skills, and abilities
LEP	limited English proficiency
MS	machine scored items
NCME	National Council on Measurement in Education
OE	open-ended items
PADDI	Principled Assessment Design, Development, and Implementation
PBT	paper-based test
PE	performance expectation
PLD	performance level descriptor
SEM	standard error of measurement
SIU	score interpretation and use
SR	selected response items
SSIB	Cognia's Summative Science Item Bank
STC	School Test Coordinator
STL	Scoring Team Leader
TA	Test Administrator
TAM	Test Administrator's Manual
TCM	Test Coordinator's Manual
TCC	test characteristic curve
TIF	test information function

APPENDIX B

PERFORMANCE LEVEL DESCRIPTORS

BIE Science Proficiency Level Descriptors

Grade 5

Policy PLDs

Policy PLDs define the knowledge and skill level expectations for all grades and content areas for the BIE Science Assessment.

Level 4. Advanced

Students demonstrate evidence of **thorough** understanding and use of college and career readiness knowledge, skills, and abilities.

Level 3. Proficient

Students demonstrate evidence of **satisfactory** understanding and use of college and career readiness knowledge, skills, and abilities.

Level 2. Nearing Proficiency

Students demonstrate evidence of **partial** understanding and use of college and career readiness knowledge, skills, and abilities.

Level 1. Novice
Students demonstrate evidence of **emerging** understanding and use of college and career readiness knowledge, skills, and abilities.

Range PLDs

Range PLDs describe the knowledge and skills that students throughout the range of each proficiency level are expected to be able to demonstrate. In line with the nature of the science standards, the statements combine science and engineering practices, disciplinary core ideas, and crosscutting concepts that students are expected to integrate and demonstrate.

Advanced

Students at the **Advanced** level demonstrate evidence of thorough understanding and use of all three dimensions (science and engineering practices, crosscutting concepts, and disciplinary core ideas) to make sense of phenomena and/or to design solutions to problems in the physical, life, and Earth and space sciences. In addition to demonstrating the skills and understandings at the Proficient level, students

performing at the Advanced level can be expected to be able to demonstrate knowledge and skills like in the following examples, as evidence of thorough understanding and use of the NGSS Standards:

- Develop, use, and analyze a model to describe and explain phenomena using an understanding of matter as tiny particles, describe quantities that should be measured to explain phenomena using an understanding of conservation of matter during physical and chemical changes, describe observations and measurements that can be used to identify materials based on their properties, and plan and conduct an investigation to determine whether a new substance with different properties is formed when two substances are mixed. (PS1)

- Plan and conduct an investigation to provide multiple pieces of evidence about phenomena using an understanding of the effects of balanced and unbalanced forces on the motion of an object, predict the future motion of an object based on complex patterns in observations and measurements, ask detailed questions to describe phenomena using an understanding of cause and effect of electric and magnetic interactions between objects not in contact with each other, thoroughly define a simple design problem that can be solved using magnets, and support an argument with multiple pieces of evidence about phenomena using an understanding that the gravitational force of Earth on objects is directed down. (PS2)

- Construct an explanation supported by multiple pieces of evidence about phenomena using an understanding of the relationship between the speed and energy of an object; provide and analyze evidence that energy can be transferred from place to place; predict and explain outcomes for the changes in energy that occur when objects collide; thoroughly design, test, and refine a device that converts energy from one form to another; and use models to explain phenomena using an understanding that food energy was once energy from the Sun. (PS3)

- Develop models to explain phenomena using an understanding that waves can cause objects to move, and that light allows objects to be seen; and compare and explain multiple solutions that use patterns to transfer information. (PS4)

- Develop models to explain phenomena using an understanding of the diversity and commonalities of the life cycles of organisms, construct an argument supported by multiple pieces of evidence about phenomena using an understanding that plants and animals have internal and external structures that support life functions, explain phenomena using an understanding that animals receive, process, and respond to information from their senses, and support an argument with multiple pieces of evidence about phenomena using an understanding that plants get the materials they need for growth chiefly from air and water. (LS1)

- Construct an argument supported by multiple pieces of evidence about phenomena using an understanding that some animals form groups that help members survive; and develop a model to explain phenomena using an understanding of the movement of matter among plants, animals, decomposers, and the environment. (LS2)

- Analyze and interpret data to provide multiple pieces of evidence about phenomena using an understanding that plants and animals have inherited traits and that variation of these traits exists in groups of similar organisms; and support an explanation with multiple pieces of evidence about phenomena using an understanding that traits can be influenced by the environment. (LS3)

- Analyze and interpret fossil data to provide multiple pieces of evidence of organisms and the environments in which they lived; construct an explanation supported by multiple pieces of evidence of phenomena using an understanding that variation among individuals of the same species is a survival advantage and that in a particular habitat some organisms survive well, some survive less well, and some cannot survive at all; and make a claim supported by multiple pieces of evidence about the merit of a solution to a problem caused by changes to the environment and the types of plants and animals that live there. (LS4)

- Explain phenomena using an understanding that patterns in rock formations and fossils in rock layers provide evidence to support an explanation for changes in a landscape over time, support an argument about phenomena using multiple pieces of evidence and an understanding that differences in the apparent brightness of the Sun compared to other stars is due to their relative distances from Earth, and represent and explain data in graphical displays to reveal patterns of daily changes in shadows, day and night, and the seasonal appearance of stars. (ESS1)
- Represent data to explain typical seasonal weather conditions, combine and synthesize information to describe climates in different regions of the world, provide multiple pieces of evidence for phenomena using an understanding of the effects of weathering and the rate of erosion, analyze and interpret data from maps to explain patterns of Earth's features, develop a model to explain phenomena using an understanding of how multiple systems on Earth interact, and describe and graph the percentages of water to provide multiple pieces of evidence about the distribution of water on Earth. (ESS2)
- Make a claim supported by multiple pieces of evidence about the merit of a design solution that reduces the impacts of a weather-related hazard, combine and synthesize information to explain that energy and fuels are derived from natural resources, how their uses affect the environment, and ways individual communities protect Earth's resources and the environment, and generate and compare multiple solutions to reduce the impacts of several natural Earth processes on humans. (ESS3)
- Define a simple design problem including detailed criteria for success and constraints; generate and compare multiple detailed solutions to a problem; and plan and carry out fair tests to identify more than one way to improve a model or prototype. (ETS1)

Proficient

Students at the **Proficient** level demonstrate evidence of satisfactory understanding and use of all three dimensions (science and engineering practices, crosscutting concepts, and disciplinary core ideas) to make sense of phenomena and/or to design solutions to problems in the physical, life, and Earth and space sciences. In addition to demonstrating the skills and understandings at the Nearing Proficiency level, students performing at the Proficient level can be expected to be able to demonstrate knowledge and skills like in the following examples, as evidence of satisfactory understanding and use of the NGSS Standards:

- Develop and use a model to describe phenomena using an understanding of matter as tiny particles, measure and graph quantities to describe phenomena using an understanding of conservation of matter during physical or chemical changes, make observations and measurements to identify materials based on their properties, and conduct an investigation to determine whether a new substance with different properties is formed when two substances are mixed. (PS1)
- Plan and conduct an investigation to provide one piece of evidence about phenomena using an understanding of the effects of balanced and unbalanced forces on the motion of an object, predict the future motion of an object based on patterns in observations and measurements, ask questions to describe phenomena using an understanding of cause and effect of electric or magnetic interactions between objects not in contact with each other, define a simple design problem that can be solved using magnets, and support an argument with one piece of evidence about phenomena using an understanding that the gravitational force of Earth on objects is directed down. (PS2)
- Construct an explanation supported by one piece of evidence about phenomena using an understanding of the relationship between the speed and energy of an object; provide evidence that energy can be transferred from place to place; predict outcomes for the changes in energy that occur when objects collide; design, test, and refine a device that converts energy from one form to another; and use models to describe phenomena using an understanding that food energy was once energy from the Sun. (PS3)

- Develop models to describe phenomena using an understanding that waves can cause objects to move, and that light allows objects to be seen; and compare multiple solutions that use patterns to transfer information. (PS4)
- Develop models to describe phenomena using an understanding of the diversity and commonalities of the life cycles of organisms, construct an argument supported by one piece of evidence about phenomena using an understanding that plants and animals have internal and external structures that support life functions, describe phenomena using an understanding that animals receive, process, and respond to information from their senses, and support an argument with one piece of evidence about phenomena using an understanding that plants get the materials they need for growth chiefly from air and water. (LS1)
- Construct an argument supported by one piece of evidence about phenomena using an understanding that some animals form groups that help members survive; and develop a model to describe phenomena using an understanding of the movement of matter among plants, animals, decomposers, and the environment. (LS2)
- Analyze and interpret data to provide one piece of evidence about phenomena using an understanding that plants and animals have inherited traits and that variation of these traits exists in groups of similar organisms; and support an explanation with one piece of evidence about phenomena using an understanding that traits can be influenced by the environment. (LS3)
- Analyze and interpret fossil data to provide one piece of evidence of organisms and the environments in which they lived; construct an explanation supported by one piece of evidence of phenomena using an understanding that variation among individuals of the same species is a survival advantage and that in a particular habitat some organisms survive well, some survive less well, and some cannot survive at all; and make a claim supported by one piece of evidence about the merit of a solution to a problem caused by changes to the environment and the types of plants and animals that live there. (LS4)
- Describe phenomena using an understanding that patterns in rock formations and fossils in rock layers provide evidence to support an explanation for changes in a landscape over time, support an argument about phenomena using one piece of evidence and an understanding that differences in the apparent brightness of the Sun compared to other stars is due to their relative distances from Earth, and represent data in graphical displays to reveal patterns of daily changes in shadows, day and night, and the seasonal appearance of stars. (ESS1)
- Represent data to describe typical seasonal weather conditions, combine information to describe climates in different regions of the world, provide one piece of evidence for phenomena using an understanding of the effects of weathering or the rate of erosion, analyze and interpret data from maps to describe patterns of Earth's features, develop a model to describe phenomena using an understanding of how Earth's systems interact, and describe and graph the percentages of water to provide one piece of evidence about the distribution of water on Earth. (ESS2)
- Make a claim supported by one piece of evidence about the merit of a design solution that reduces the impacts of a weather-related hazard, combine information to describe how energy and fuels are derived from natural resources, how their uses affect the environment, and ways individual communities protect Earth's resources and the environment, and generate and compare multiple solutions to reduce the impacts of natural Earth processes on humans. (ESS3)
- Define a simple design problem including criteria for success and constraints; generate and compare multiple solutions to a problem; and plan and carry out fair tests to identify one way to improve a model or prototype. (ETS1)

Nearing Proficiency

Students at the **Nearing Proficiency** level demonstrate evidence of partial understanding and use of all three dimensions (science and engineering practices, crosscutting concepts, and disciplinary core ideas) to make sense of phenomena and/or to design solutions to problems in the physical, life, and Earth and space sciences. Students performing at the Nearing Proficiency level can be expected to be able to demonstrate knowledge and skills like in the following examples, as evidence of partial understanding and use of the NGSS Standards:

- Use a model to describe phenomena using an understanding of matter as tiny particles, graph quantities to describe phenomena using an understanding of conservation of matter during physical or chemical changes, make observations to identify materials based on their properties, and use data to determine whether a new substance with different properties is formed when two substances are mixed. (PS1)
- Conduct an investigation to provide one piece of evidence about phenomena using an understanding of the effects of balanced or unbalanced forces on the motion of an object, predict the future motion of an object based on simple patterns in observations and measurements, ask questions using an understanding of cause and effect of electric or magnetic interactions between objects not in contact with each other, partially define a simple design problem that can be solved using magnets, and make a claim about phenomena using an understanding that the gravitational force of Earth on objects is directed down. (PS2)
- Describe phenomena using an understanding of the relationship between the speed and energy of an object, explain that energy can be transferred from place to place, describe the changes in energy that occur when objects collide, describe elements of a device that converts energy from one form to another, and describe phenomena using an understanding that food energy was once energy from the Sun. (PS3)
- Use models to describe phenomena using an understanding that waves can cause objects to move or that light allows objects to be seen; and describe a solution that uses patterns to transfer information. (PS4)
- Use models to describe phenomena using an understanding of the diversity and commonalities of the life cycles of organisms, make a claim about phenomena using an understanding that plants and animals have internal and external structures that support life functions, describe phenomena using an understanding that animals receive or process or respond to information from their senses, and make a claim about phenomena using an understanding that plants get the materials they need for growth chiefly from air and water. (LS1)
- Make a claim about phenomena using an understanding that some animals form groups that help members survive; and use a model to describe phenomena using an understanding of the movement of matter among plants, animals, decomposers, and the environment. (LS2)
- Use data to describe phenomena using an understanding that plants and animals have inherited traits OR that variation of these traits exists in groups of similar organisms; and make a claim about phenomena using an understanding that traits can be influenced by the environment. (LS3)
- Use fossil data to describe organisms and the environments in which they lived; describe phenomena using an understanding that variation among individuals of the same species is a survival advantage or that in a particular habitat some organisms survive well, some survive less well, and some cannot survive at all; and describe a solution to a problem caused by changes to the environment and the types of plants and animals that live there. (LS4)
- Describe phenomena using an understanding that patterns in rock formations or fossils in rock layers provide evidence to support changes in a landscape over time, describe phenomena using

an understanding that differences in the apparent brightness of the Sun compared to other stars is due to their relative distances from Earth, and use data in graphical displays to reveal patterns of daily changes in shadows or day and night or the seasonal appearance of stars. (ESS1)

- Use data to describe typical seasonal weather conditions, describe climates in different regions of the world, describe phenomena using an understanding of the effects of weathering or the rate of erosion, use maps to describe patterns of Earth's features, use a model to describe phenomena using an understanding of how Earth's systems interact, and provide a description of the distribution of water on Earth. (ESS2)
- Describe a design solution or component of a design solution that reduces the impacts of a weather-related hazard, use information to determine that energy and fuels are derived from natural resources or how their uses affect the environment or ways individual communities protect Earth's resources and the environment, and describe a solution or component of a solution to reduce the impacts of natural Earth processes on humans. (ESS3)
- Define a simple design problem including at least one criterion for success or one constraint; generate a solution to a problem or compare two solutions to a problem; and use results of a fair test to identify one way to improve a model or prototype. (ETS1)

Novice

Students at the **Novice** level demonstrate evidence of emerging understanding and use of all three dimensions (science and engineering practices, crosscutting concepts, and disciplinary core ideas) to make sense of phenomena and/or to design solutions to problems in the physical, life, and Earth and space sciences.

BIE Science Proficiency Level Descriptors

Grade 8

Policy PLDs

Policy PLDs define the knowledge and skill level expectations for all grades and content areas for the BIE Science Assessment.

Level 4. Advanced

Students demonstrate evidence of **thorough** understanding and use of college and career readiness knowledge, skills, and abilities.

Level 3. Proficient

Students demonstrate evidence of **satisfactory** understanding and use of college and career readiness knowledge, skills, and abilities.

Level 2. Nearing Proficiency

Students demonstrate evidence of **partial** understanding and use of college and career readiness knowledge, skills, and abilities.

Level 1. Novice

Students demonstrate evidence of **emerging** understanding and use of college and career readiness knowledge, skills, and abilities.

Range PLDs
Range PLDs describe the knowledge and skills that students throughout the range of each proficiency level are expected to be able to demonstrate. In line with the nature of the science standards, the statements combine science and engineering practices, disciplinary core ideas, and crosscutting concepts that students are expected to integrate and demonstrate.

Advanced

Students at the **Advanced** level demonstrate evidence of thorough understanding and use of all three dimensions (science and engineering practices, crosscutting concepts, and disciplinary core ideas) to make sense of phenomena and/or to design solutions to problems in the physical, life, and Earth and space sciences. In addition to demonstrating the skills and understanding at the Proficient level, students performing at the Advanced level can be expected to be able to demonstrate knowledge and skills like in the following examples, as evidence of thorough understanding and use of the NGSS Standards:

- Develop, use, and analyze models to describe or explain phenomena using an understanding of the structure of matter; to predict, describe, and explain phenomena using an understanding of changes in particle motion and state; and to provide evidence for and describe phenomena using an understanding of conservation of mass during multiple physical and chemical changes. (PS1)

- Plan, carry out, and refine investigations to provide evidence to explain phenomena using an understanding of the effects of forces, interactions, and mass on the motion of objects, as well as analyze various data to evaluate claims about phenomena using an understanding of gravitational interactions in systems, and to design and/or compare multiple solutions to a problem using an understanding of systems of colliding objects. (PS2)

- Plan, carry out, and refine investigations, use and analyze data, and develop, use, and analyze models to explain phenomena using an understanding of various relationships involving kinetic and potential energy in systems, as well as apply such understanding to design, test, and evaluate devices to solve problems related to energy transfer and to support and/or evaluate claims about phenomena related to energy transfer. (PS3)

- Develop, use, and apply mathematical representations, patterns, and models to explain phenomena using an understanding of wave properties and relationships and wave interactions with various materials, and synthesize multiple sources of information using an understanding of signal types to evaluate claims about phenomena related to reliability of digital and analog signals. (PS4)

- Use multiple pieces of evidence from investigations of phenomena to explain that living things are made of cells; develop and use models of phenomena to describe the function of a cell and its parts and to describe how food is rearranged in organisms through chemical reactions to support growth and/or release energy; support arguments about phenomena using multiple pieces of evidence and an understanding that the body is a system of interacting subsystems composed of cells; support an explanation for how several animal behaviors and several specialized plant structures affect the probability of reproductive success; use multiple pieces of evidence and an understanding of environmental and genetic factors to explain phenomena about how those factors influence the growth of organisms; construct an explanation using multiple pieces of evidence to explain the role of photosynthesis in the cycling of matter and flow of energy into and out of organisms; synthesize multiple sources of information about phenomena and an understanding of behaviors of organisms to determine that sensory receptors respond to stimuli by sending messages to the brain for immediate behavior or for storage as memories. (LS1)

- Analyze and interpret data about phenomena to provide evidence to explain multiple ways that resource availability affects populations, construct an argument supported by multiple pieces of evidence that populations are affected by changing physical and biological components of an ecosystem, develop and revise models to explain phenomena using an understanding of the cycling of matter and energy in an ecosystem, describe phenomena using an understanding of interactions among organisms and predict multiple patterns of interactions among organisms across multiple ecosystems, and evaluate multiple competing design solutions for phenomena that involve maintaining biodiversity in ecosystems. (LS2)

- Develop and use multiple models to explain phenomena using an understanding of how genetic mutations affect proteins resulting in harmful, beneficial, or neutral effects on an organism, and to explain phenomena using an understanding of how asexual reproduction results in offspring with identical genetic information and to explain how sexual reproduction results in offspring with genetic variation. (LS3)

- Analyze and interpret multiple pieces of data about phenomena using an understanding of the fossil record and modern organisms to show patterns in the change of life forms over time, apply multiple scientific ideas about phenomena to construct an explanation for the anatomical similarities and differences among modern and fossil organisms to infer evolutionary relationships, analyze pictorial data to compare similarities in embryological development across multiple familiar and unfamiliar species to identify evolutionary relationships, use multiple pieces of evidence and mathematical representations to explain phenomena using an understanding of how variation in genetic traits provides advantages to some individuals within a population and to support explanations of increases and decreases in specific traits over time, and explain phenomena by synthesizing multiple pieces of information about ways technologies have changed the way humans influence the inheritance of desired traits in organisms. (LS4)

- Develop, use, and revise a model of the Earth-Sun-Moon system to describe phenomena using an understanding of the cyclic pattern of the seasons and to describe phenomena using an understanding of the role of gravity in the motions within the solar system and galaxies, analyze and interpret data on multiple phenomena related to the scale properties of objects in the solar system, and use multiple pieces of evidence and an understanding of rock strata to explain phenomena about how the geologic time scale is used to organize Earth's history. (ESS1)

- Develop models to describe phenomena using an understanding of the flow of energy that drives the cycling of Earth's materials, use multiple pieces of evidence to explain phenomena using an understanding of how geoscience processes have changed Earth's surface at varying time and spatial scales, analyze and interpret multiple pieces of data to explain phenomena using an understanding of the evidence that supports past plate motions on Earth, develop models to describe phenomena using an understanding of the water cycle including energy and gravity, explain weather phenomena synthesizing and using evidence and an understanding of the interactions of air masses, and develop and use models to describe phenomena using an understanding of how unequal heating and Earth's rotation result in climate, atmospheric, and ocean circulation patterns. (ESS2)

- Use evidence to explain multiple phenomena using an understanding of how geoscience processes have resulted in uneven distribution of Earth's natural resources, analyze and interpret multiple pieces of data on natural hazard phenomena to forecast future catastrophic events and to inform the development of technologies to mitigate their effects, apply scientific principles to design a successful solution for monitoring and minimizing the human impacts on the environment, use multiple pieces of evidence to support an argument about phenomena using an understanding of how increases in human population impact Earth's systems, and ask multiple questions about phenomena to clarify evidence of multiple factors that have caused the rise in global temperatures. (ESS3)

- Define the criteria and constraints of a design problem with sufficient precision to ensure an optimal solution and using an understanding of scientific principles and potential impacts on people and the environment and an understanding of how those impacts may limit possible solutions, use a systematic process to evaluate how well multiple competing design solutions meet required criteria and constraints, analyze data from tests of multiple different design solutions to identify the best characteristics of each solution that can be combined into a new solution that will better meet

criteria for success, and develop a realistic model of a proposed object, tool, or process that generates data while it is repeatedly tested and modified until an optimal design is achieved. (ETS1)

Proficient

Students at the **Proficient** level demonstrate evidence of satisfactory understanding and use of all three dimensions (science and engineering practices, crosscutting concepts, and disciplinary core ideas) to make sense of phenomena and/or to design solutions to problems in the physical, life, and Earth and space sciences. In addition to demonstrating the skills and understanding at the Nearing Proficiency level, students performing at the Proficient level can be expected to be able to demonstrate knowledge and skills like in the following examples, as evidence of satisfactory understanding and use of the NGSS Standards:

- Develop and use models to describe phenomena using an understanding of the structure of matter; to predict and describe phenomena using an understanding of changes in particle motion and state; and to describe phenomena using an understanding of conservation of mass during one or two physical and chemical changes. (PS1)
- Plan and carry out investigations to produce data and/or provide evidence about phenomena using an understanding of the effects of forces, interactions, and mass on the motion of objects, as well as use direct data to support claims about phenomena using an understanding of gravitational interactions in systems and to design a solution to a problem using an understanding of systems of colliding objects. (PS2)
- Plan investigations, use data, and develop or use models to describe phenomena using an understanding of relationships involving kinetic and potential energy in systems, as well as apply such understanding to design and test devices to solve problems related to energy transfer and to support claims about phenomena related to energy transfer. (PS3)
- Use mathematical representations and patterns, and develop and use models, to describe phenomena using an understanding of wave properties and relationships and wave interactions with various materials and use multiple sources of information and an understanding of signal types to support claims about phenomena related to reliability of digital and analog signals. (PS4)
- Use one or two pieces of evidence from investigations of phenomena to explain that living things are made of cells, develop and use models of phenomena to describe the function of a cell and its parts and to describe how food is rearranged in organisms through chemical reactions to support growth and/or release energy, support arguments about phenomena using one or two pieces of evidence and an understanding that the body is a system of interacting subsystems composed of cells, support an explanation for how some animal behaviors and some specialized plant structures affect the probability of reproductive success, use one or two pieces of evidence and an understanding of environmental and genetic factors to explain phenomena about how those factors influence the growth of organisms, construct an explanation based on one or two pieces of evidence to explain the role of photosynthesis in the cycling of matter and flow of energy into and out of organisms, use information about phenomena and an understanding of behaviors of organisms to determine that sensory receptors respond to stimuli by sending messages to the brain for immediate behavior or for storage as memories. (LS1)
- Analyze and interpret data about phenomena to provide evidence to explain one or two ways that resource availability affects populations, construct an argument supported by one or two pieces of evidence that populations are affected by changing physical or biological components of an ecosystem, develop models to describe phenomena using an understanding of the cycling of matter and energy in an ecosystem, describe phenomena using an understanding of interactions among organisms and predict one or two patterns of interactions among organisms across multiple

ecosystems, and evaluate two competing design solutions for phenomena that involve maintaining biodiversity in ecosystems. (LS2)

- Develop and use one or two models to explain phenomena using an understanding of how genetic mutations affect proteins resulting in harmful, beneficial, or neutral effects on an organism and to explain phenomena using an understanding of how asexual reproduction results in offspring with identical genetic information and how sexual reproduction results in offspring with genetic variation. (LS3)

- Analyze and interpret one or two pieces of data about phenomena using an understanding of the fossil record and modern organisms to show patterns in the change of life forms over time, apply one or two scientific ideas about phenomena to construct an explanation for the anatomical similarities and differences among modern and fossil organisms to infer evolutionary relationships, analyze pictorial data to compare similarities in embryological development across multiple familiar species to identify evolutionary relationships, use one or two pieces of evidence or mathematical representations to explain phenomena using an understanding of how variation in genetic traits provides advantages to some individuals within a population and to support explanations of increases and decreases in specific traits over time, and explain phenomena by synthesizing one or two pieces of information about technologies that have changed the way humans influence the inheritance of desired traits in organisms. (LS4)

- Develop and use a model of the Earth-Sun-Moon system to describe phenomena using an understanding of the cyclic pattern of the seasons and to describe phenomena using an understanding of the role of gravity in the motions within the solar system and galaxies, analyze and interpret data on one or two phenomena related to the scale properties of objects in the solar system, and use one or two pieces of evidence and an understanding of rock strata to explain phenomena about how the geologic time scale is used to organize Earth's history. (ESS1)

- Develop a model to describe phenomena using an understanding of the flow of energy that drives cycling of Earth's materials, use one or two pieces of evidence to explain phenomena using an understanding of how geoscience processes have changed Earth's surface at varying time and spatial scales, analyze and interpret one or two pieces of data to explain phenomena using an understanding of the evidence that supports past plate motions on Earth, develop a model to describe phenomena using an understanding of the water cycle including energy and gravity, explain weather phenomena using evidence and an understanding of the interactions of air masses, and develop and use a model to describe phenomena using an understanding of how unequal heating and Earth's rotation result in climate, atmospheric, or ocean circulation patterns. (ESS2)

- Use evidence to explain one or two phenomena using an understanding of how geoscience processes have resulted in uneven distribution of Earth's natural resources, analyze and interpret one or two pieces of data on natural hazard phenomena to forecast future catastrophic events and to inform the development of technologies to mitigate their effects, apply scientific principles to design a solution for monitoring and minimizing human impacts on the environment, use one or two pieces of evidence to support an argument about phenomena using an understanding of how increases in human population impact Earth's systems, and ask one or two questions about phenomena to clarify evidence of one or two factors that have caused the rise in global temperatures. (ESS3)

- Define the criteria and constraints of a design problem with sufficient precision to ensure a successful solution and using an understanding of scientific principles and potential impacts on people and the environment and an understanding of how those impacts may limit possible solutions, use a systematic process to evaluate how well two competing design solutions meet required criteria and constraints, analyze data from tests of two different design solutions to identify

the best characteristics of each solution that can be combined into a new solution that will better meet criteria for success, and develop a model of a proposed object, tool, or process that generates data while it is repeatedly tested and modified until an optimal design is achieved. (ETS1)

Nearing Proficiency

Students at the **Nearing Proficiency** level demonstrate evidence of partial understanding and use of all three dimensions (science and engineering practices, crosscutting concepts, and disciplinary core ideas) to make sense of phenomena and/or to design solutions to problems in the physical, life, and Earth and space sciences. Students performing at the Nearing Proficiency level can be expected to be able to demonstrate knowledge and skills like in the following examples, as evidence of partial understanding and use of the NGSS Standards:

- Use models to identify the structure of matter as it relates to phenomena; to describe basic phenomena using an understanding of changes in particle motion and state; and to identify that mass is conserved during physical or chemical changes that take place in various phenomena. (PS1)
- Identify or describe parts of investigations about phenomena using an understanding of the effects of forces, interactions, and mass on the motion of objects, describe aspects of phenomena using one or two pieces of data and an understanding of one gravitational interaction in a system, and identify a design or elements of a solution to a problem related to colliding objects. (PS2)
- Describe parts of investigations, use data, and use models to describe aspects of phenomena using an understanding of some relationships involving kinetic energy in systems; design and test a device for problems related to energy transfer; and identify principles that support claims about energy transfer in phenomena. (PS3)
- Use mathematical representations, patterns, and models to identify wave properties and wave interactions with various materials as they relate to phenomena and use one or two sources of information to identify that digital signals are more reliable than analog signals as demonstrated in various phenomena. (PS4)
- Use evidence from an investigation of a phenomenon to explain that living things are made of cells, use models of phenomena to describe the function of a cell and some of its parts and to describe that food is rearranged in organisms into new substances to support growth or to release energy, make claims about phenomena using evidence and a partial understanding that the body is a system of interacting subsystems composed of cells, describe some animal behaviors or specialized plant structures that may affect reproductive success, use evidence and a partial understanding of environmental or genetic factors to describe phenomena about how those factors influence the growth of organisms, construct an explanation based on one piece of evidence to describe the role of photosynthesis in the cycling of matter or flow of energy into and out of organisms, and use information about phenomena to describe that organisms use their senses to respond to stimuli immediately or to store information as memories. (LS1)
- Use data about phenomena to describe one way resource availability affects populations, make a claim supported by evidence that populations are affected by changing components of an ecosystem, use models to describe phenomena about the cycling of matter or energy in an ecosystem, use a partial understanding of interactions among organisms to describe one interaction between organisms, and describe a design solution or components of a solution for phenomena that involve maintaining biodiversity in ecosystems. (LS2)
- Use models to partially explain phenomena using an understanding of one way that genetic mutations affect organisms and to describe phenomena using an understanding of asexual

reproduction resulting in offspring with identical genetic information, or about sexual reproduction resulting in offspring with genetic variation. (LS3)

- Use data about phenomena using a partial understanding of the fossil record and modern organisms to show patterns in the change of life forms over time, apply a scientific idea about phenomena to describe an anatomical similarity or difference between modern and fossil organisms, use pictorial data to describe similarities in embryological development across multiple species, use evidence or one mathematical representation to describe phenomena using a partial understanding of variation in genetic traits among individuals within a population or to describe increases and decreases in specific traits over time, and describe phenomena using a partial understanding about technologies that have changed the way humans influence the inheritance of desired traits in organisms. (LS4)

- Use a model of the Earth-Sun-Moon system to describe phenomena using a partial understanding of the cyclic pattern of the seasons and to describe phenomena using a partial understanding of the role of gravity in the motions within the solar system or galaxies, use data on phenomena related to the scale properties of objects in the solar system, and use evidence and a partial understanding of rock strata to describe some aspects about how the geologic time scale is related to Earth's history. (ESS1)

- Use a model to describe phenomena using a partial understanding of the flow of energy that drives the cycling of Earth's materials, use evidence to explain phenomena using a partial understanding of how geoscience processes have changed Earth's surface, use data to explain phenomena using a partial understanding of the evidence that supports past plate motions on Earth, use a model to describe phenomena using an understanding of the water cycle, describe weather phenomena using evidence and a partial understanding of the interactions of air masses, and use a model to describe phenomena using an understanding of how unequal heating and Earth's rotation result in some climate, atmospheric, or ocean circulation patterns. (ESS2)

- Use evidence to explain a phenomenon using a partial understanding of how geoscience processes have resulted in uneven distribution of some of Earth's natural resources, use data on natural hazard phenomena to forecast future catastrophic events or to inform the development of one technology that could be used to mitigate their effects, identify human impacts on the environment or design parts of a solution for monitoring or minimizing the human impacts, use evidence to make a claim about phenomena using a partial understanding of how increases in human population impact Earth's systems, and ask one question about a phenomenon to clarify evidence of one factor that has caused the rise in global temperatures. (ESS3)

- Define one criterion or constraint of a design problem using an understanding of scientific principles and/or potential impacts on people and the environment and identify one way those impacts may limit possible solutions, use a systematic process to evaluate how well a design solution meets required criteria or constraints, analyze data from tests of a design solution to identify a characteristic of the solution that is necessary to meet the criteria for success, and develop a partial model of a proposed object, tool, or process that can be tested and modified. (ETS1)

Novice

Students at the **Novice** level demonstrate evidence of emerging understanding and use of all three dimensions (science and engineering practices, crosscutting concepts, and disciplinary core ideas) to make sense of phenomena and/or to design solutions to problems in the physical, life, and Earth and space sciences.

BIE Science Proficiency Level Descriptors

Grade 11

Policy PLDs

Policy PLDs define the knowledge and skill level expectations for all grades and content areas for the BIE Science Assessment.

Level 4. Advanced

Students demonstrate evidence of **thorough** understanding and use of college and career readiness knowledge, skills, and abilities.

Level 3. Proficient

Students demonstrate evidence of **satisfactory** understanding and use of college and career readiness knowledge, skills, and abilities.

Level 2. Nearing Proficiency

Students demonstrate evidence of **partial** understanding and use of college and career readiness knowledge, skills, and abilities.

Level 1. Novice

Students demonstrate evidence of **emerging** understanding and use of college and career readiness knowledge, skills, and abilities.

Range PLDs

Range PLDs describe the knowledge and skills that students throughout the range of each proficiency level are expected to be able to demonstrate. In line with the nature of the science standards, the statements combine science and engineering practices, disciplinary core ideas, and crosscutting concepts that students are expected to integrate and demonstrate.

Advanced
Students at the **Advanced** level demonstrate evidence of thorough understanding and use of all three dimensions (science and engineering practices, crosscutting concepts, and disciplinary core ideas) to make

sense of phenomena and/or to design solutions to problems in the physical, life, and Earth and space sciences. In addition to demonstrating the skills and understanding at the Proficient level, students performing at the Advanced level can be expected to be able to demonstrate knowledge and skills like in the following examples, as evidence of thorough understanding and use of the NGSS Standards:

- Use an understanding of the periodic table to predict multiple relative properties of elements, plan and conduct an investigation of phenomena related to multiple bulk scale properties of substances and explain how these relate to the strength of electrical forces between particles, explain phenomena about the release or absorption of energy by a chemical system by developing models to show changes in total bond energy, use multiple pieces of evidence to explain phenomena using an understanding of how changes in temperature and concentration affect reaction rate, explain phenomena about chemical systems at equilibrium using an understanding of how multiple conditions on the system could be changed to produce more or fewer products or more or fewer reactants, use multiple mathematical representations to support claims that mass is conserved during a chemical reaction, and develop multiple models to describe phenomena using an understanding of the changes in the nucleus of an atom and the energy released during fission, fusion, and radioactive decay. (PS1)

- Analyze and use multiple pieces of data from phenomena to show that f = ma; use multiple mathematical representations of phenomena and an understanding of momentum to support the claim that the total momentum of a system is conserved when there is no net force on the system; apply multiple scientific and engineering ideas to design, evaluate, and refine multiple devices that minimize force on an object during a collision; use mathematical representations of Newton's law of gravitation and Coulomb's law to describe and make predictions about familiar and unfamiliar phenomena using an understanding of gravitational and electrostatic forces between objects; plan and conduct an investigation of phenomena to produce multiple pieces of evidence that prove an electric current produces a magnetic field and that a changing magnetic field produces electric current; and communicate multiple pieces of information about phenomena using an understanding of how the molecular structure of a material relates to its macroscopic properties and makes the material well suited for particular uses. (PS2)

- Create multiple computational models of phenomena to calculate changes in energy of a system when energy flows into and out of the system is known; develop and use multiple models to explain phenomena using an understanding of how energy at the macroscopic scale can be accounted for at the microscopic scale in energy associated with particle motion and relative position; design, build, and refine devices that convert one form of energy into another; plan and conduct an investigation of phenomena to provide multiple pieces of evidence using an understanding that when two components of different temperatures are combined within a closed system, both components eventually have the same temperature; and develop and use a model to explain phenomena related to the forces and the changes in energy between two objects interacting through electric fields and magnetic fields. (PS3)

- Explain phenomena using an understanding of multiple mathematical representations regarding relationships among frequency, wavelength, and speed of waves in various media; evaluate multiple questions about phenomena using an understanding of the advantages of using digital transmission and storage of information; use multiple phenomena to evaluate claims that electromagnetic radiation can be described using a wave or particle model; in the context of phenomena, evaluate multiple claims about the effects that different frequencies of electromagnetic radiation have on matter; and communicate technical information about phenomena using an understanding of how multiple specific technological devices use the principles of wave behavior and wave interactions to transmit and capture information and energy. (PS4)

- Use multiple pieces of evidence to explain phenomena using an understanding of how the structure of DNA determines the structure of proteins and how proteins carry out the functions of life through specialized cells; develop and use a complex model to describe phenomena using an understanding of the organization of interacting systems within multicellular organisms; plan and conduct an investigation to provide multiple pieces of evidence about phenomena that show that feedback mechanisms maintain homeostasis; use a complex model to describe phenomena using an understanding of how cell division and differentiation help produce and maintain complex organisms; use a complex model to describe phenomena using an understanding of how photosynthesis transforms light energy into stored chemical energy; use multiple pieces of evidence to construct and revise an explanation about phenomena using an understanding of how carbon, hydrogen, and oxygen from sugar molecules combine with other elements to form amino acids and other large carbon-based molecules; and use a complex model to describe phenomena using an understanding that cellular respiration is a chemical process that breaks the bonds in food and oxygen molecules and forms bonds in new compounds, which results in a net transfer of energy. (LS1)

- Use mathematical and computational representations to support explanations of phenomena using an understanding of factors that affect carrying capacity of ecosystems at different scales; use multiple pieces of evidence and mathematical representations to support and revise explanations of phenomena using an understanding of factors affecting biodiversity and populations in ecosystems of different scales and to support claims for the cycling of matter and flow of energy among organisms in an ecosystem; use evidence to construct and revise an explanation of phenomena using an understanding of the cycling of matter and flow of energy in aerobic and anaerobic conditions; develop models to describe phenomena using an understanding of the role of photosynthesis and cellular respiration in the cycling of carbon among Earth's spheres; evaluate multiple claims, pieces of evidence, and reasoning about phenomena involving complex interactions in ecosystems using an understanding that these interactions maintain relatively consistent numbers and types of organisms under stable conditions, but changing conditions may result in a new ecosystem; design, evaluate, and refine solutions for reducing impacts of human activities on the environment or biodiversity; and evaluate multiple pieces of evidence about phenomena using an understanding of the role of group behavior on individual and species' chances to survive and reproduce. (LS2)

- Ask multiple questions about phenomena to clarify relationships surrounding the role of DNA in chromosomes in coding the instructions for traits passed from parents to offspring; use multiple pieces of evidence to make and defend a claim about phenomena using an understanding that inheritable genetic variations may result from new genetic combinations through meiosis, viable errors during replication, and/or mutations caused by environmental factors; and apply multiple concepts of statistics and probability to explain phenomena using an understanding of the variation and distribution of expressed traits in a population. (LS3)

- Communicate multiple pieces of scientific information about phenomena using an understanding that common ancestry and biological evolution are supported by multiple lines of empirical evidence; use multiple pieces of evidence to construct an explanation about phenomena using an understanding that the process of evolution primarily results from four factors: the potential for a species to increase in number, the heritable genetic variation of individuals in a species due to mutation and sexual reproduction, competition for limited resources, and the proliferation of those organisms that are better able to survive and reproduce in the environment; apply multiple concepts of statistics and probability to support explanations of phenomena using an understanding that organisms with an advantageous heritable trait tend to increase in proportion to organisms lacking the trait; use multiple pieces of evidence to construct an explanation of phenomena using an

understanding of how natural selection leads to adaptation of populations; evaluate multiple pieces of evidence supporting claims about phenomena using an understanding that changes in environmental conditions may result in: increases in numbers of individuals of some species, the emergence of new species over time, and the extinction of other species; and create and revise a simulation to test a solution to mitigate adverse impacts of human activity on biodiversity. (LS4)

- Use multiple pieces of evidence to develop a model to describe phenomena using an understanding of the life span of the Sun and the role of nuclear fusion in the Sun's core to release energy that reaches Earth in the form of radiation; use multiple pieces of evidence and an understanding of the phenomena of light spectra, motion of distant galaxies, and composition of matter in the universe to construct an explanation of the big bang theory; communicate multiple scientific ideas about phenomena using an understanding of the way stars produce different elements over their life cycles; use mathematical and computational representations of phenomena to predict the motion of orbiting objects in the solar system; evaluate multiple pieces of evidence of phenomena using an understanding of past and current movements of continental and oceanic crust and the theory of plate tectonics to explain the ages of crustal rocks; and apply scientific reasoning and an understanding of multiple pieces of evidence from ancient Earth materials, meteorites, and other planetary surfaces to describe phenomena about Earth's formation and early history. (ESS1)

- Develop models to describe phenomena using an understanding of how Earth's internal and surface processes operate at different spatial and temporal scales to form continental and ocean floor features; analyze multiple types of geoscience data about phenomena to make a claim that one change to Earth's surface can create feedback that causes changes to other Earth systems; use multiple pieces of evidence to develop a model of Earth's interior to describe phenomena using an understanding of the cycling of matter by thermal convection; use models to describe phenomena using an understanding of how variations in the flow of energy into and out of Earth's systems result in changes in climate; plan and conduct investigations of phenomena related to the properties of water using an understanding of water's effects on Earth materials and surface processes; develop quantitative models to describe phenomena using an understanding of the cycling of carbon among the hydrosphere, atmosphere, geosphere, and biosphere; and use multiple pieces of evidence to construct an argument about phenomena using an understanding of the simultaneous coevolution of Earth's systems and life on Earth. (ESS2)

- Use multiple pieces of evidence to construct an explanation about phenomena using an understanding of how the availability of natural resources, occurrence of natural hazards, and changes in climate have influenced human activity; using an understanding of cost-benefit ratios, evaluate multiple competing design solutions for developing, managing, and utilizing energy and mineral resources; create computational simulations of phenomena to show the relationships among management of natural resources, the sustainability of human populations, and biodiversity; using an understanding of human impacts on natural systems, evaluate and refine a technological solution that reduces these impacts; analyze multiple pieces of geoscience data and multiple global climate models of phenomena to make a forecast of the current rate of climate change and associated future impacts to Earth systems; and use computational representations to describe phenomena using an understanding of the relationships among Earth systems and how those relationships are modified due to human activity. (ESS3)

- Analyze a major global challenge to specify multiple qualitative and quantitative criteria and constraints for solutions that account for multiple societal needs and wants; design an engineering solution to multiple complex real-world problems by breaking them down into smaller, more manageable problems; evaluate and refine a solution to a complex real-world problem based on prioritized criteria and trade-offs that account for a range of constraints, including cost, safety, reliability, and aesthetics, as well as social, cultural, and environmental impacts; and use a

computer simulation to model the impact of multiple proposed solutions to a complex real-world problem with numerous criteria and constraints on interactions within and between systems relevant to the problem. (ETS1)

Proficient

Students at the **Proficient** level demonstrate evidence of satisfactory understanding and use of all three dimensions (science and engineering practices, crosscutting concepts, and disciplinary core ideas) to make sense of phenomena and/or to design solutions to problems in the physical, life, and Earth and space sciences. In addition to demonstrating the skills and understanding at the Nearing Proficiency level, students performing at the Proficient level can be expected to be able to demonstrate knowledge and skills like in the following examples, as evidence of satisfactory understanding and use of the NGSS Standards:

- Use an understanding of the periodic table to predict one or two relative properties of elements, plan and conduct an investigation of phenomena related to one or two bulk scale properties of substances and explain how these relate to the strength of electrical forces between particles, explain phenomena about the release or absorption of energy by a chemical system by developing a model to show changes in total bond energy, use one or two pieces of evidence to explain phenomena using an understanding of how changes in temperature or concentration affect reaction rate, explain phenomena about chemical systems at equilibrium using an understanding of how one or two of the conditions on the system could be changed to produce more products, use one or two mathematical representations to support claims that mass is conserved during a chemical reaction, and develop one or two models to describe phenomena using an understanding of the changes in the nucleus of an atom and the energy released during fission, fusion, and radioactive decay. (PS1)

- Analyze and use one or two pieces of data from phenomena to show that $f = ma$; use one or two mathematical representations of phenomena and an understanding of momentum to support the claim that the total momentum of a system is conserved when there is no net force on the system; apply one or two scientific and engineering ideas to design, evaluate, and refine a device that minimizes force on an object during a collision; use mathematical representations of Newton's law of gravitation and Coulomb's law to describe and make predictions about familiar phenomena using an understanding of gravitational and electrostatic forces between objects; plan and conduct an investigation of phenomena to produce one or two pieces of evidence that prove an electric current produces a magnetic field and that a changing magnetic field produces electric current; and communicate one or two pieces of information about phenomena using an understanding of how the molecular structure of a material relates to its macroscopic properties and makes the material well suited for particular uses. (PS2)

- Create a computational model of phenomena to calculate changes in energy of a system when energy flows into and out of the system is known; develop and use one or two models to explain phenomena using an understanding of how energy at the macroscopic scale can be accounted for at the microscopic scale in energy associated with particle motion and relative position; design, build, and refine a device that converts one form of energy into another; plan and conduct an investigation of phenomena to provide one or two pieces of evidence using an understanding that when two components of different temperatures are combined within a closed system, both components eventually have the same temperature; and develop and use a model to explain phenomena related to the forces and the changes in energy between two objects interacting through electric or magnetic fields. (PS3)

- Explain phenomena using an understanding of one or two mathematical representations regarding relationships among frequency, wavelength, and speed of waves in various media; evaluate one or two questions about phenomena using an understanding of the advantages of using digital transmission and storage of information; use one or two phenomena to evaluate claims that

electromagnetic radiation can be described using a wave or particle model; in the context of phenomena, evaluate one or two claims about the effects that different frequencies of electromagnetic radiation have on matter; and communicate technical information about phenomena using an understanding of how one or two specific technological devices use the principles of wave behavior and wave interactions to transmit and capture information and energy. (PS4)

- Use evidence to explain phenomena using an understanding of how the structure of DNA determines the structure of proteins and how proteins carry out the functions of life through specialized cells; develop and use a model to describe phenomena using an understanding of the organization of interacting systems within multicellular organisms; plan and conduct an investigation to provide evidence about phenomena that show that feedback mechanisms maintain homeostasis; use a model to describe phenomena using an understanding of how cell division and differentiation help produce and maintain complex organisms; use a model to describe phenomena using an understanding of how photosynthesis transforms light energy into stored chemical energy; use evidence to construct and revise an explanation about phenomena using an understanding of how carbon, hydrogen, and oxygen from sugar molecules combine with other elements to form amino acids and/or other large carbon-based molecules; and use a model to describe phenomena using an understanding that cellular respiration is a chemical process that breaks the bonds in food and oxygen molecules and forms bonds in new compounds, which results in a net transfer of energy. (LS1)

- Use mathematical or computational representations to support explanations of phenomena using an understanding of factors that affect carrying capacity of ecosystems at different scales; use one or two pieces of evidence and one or two mathematical representations to support and revise explanations of phenomena using an understanding of factors affecting biodiversity and populations in ecosystems of different scales and to support claims for the cycling of matter and flow of energy among organisms in an ecosystem; use evidence to construct or revise an explanation of phenomena using an understanding of the cycling of matter and flow of energy in aerobic and anaerobic conditions; develop a model to describe phenomena using an understanding of the role of photosynthesis and cellular respiration in the cycling of carbon among Earth's spheres; evaluate one or two claims, pieces of evidence, and reasoning about phenomena involving complex interactions in ecosystems using an understanding that these interactions maintain relatively consistent numbers and types of organisms under stable conditions, but changing conditions may result in a new ecosystem; design, evaluate, and refine a solution for reducing impacts of human activities on the environment or biodiversity; and evaluate one or two pieces of evidence about phenomena using an understanding of the role of group behavior on individual and species' chances to survive and reproduce. (LS2)

- Ask one or two questions about phenomena to clarify relationships about the role of DNA in chromosomes in coding the instructions for traits passed from parents to offspring; use one or two pieces of evidence to make and defend a claim about phenomena using an understanding that inheritable genetic variations may result from new genetic combinations through meiosis, viable errors during replication, and/or mutations caused by environmental factors; and apply one or two concepts of statistics and probability to explain phenomena using an understanding of the variation and distribution of expressed traits in a population. (LS3)

- Communicate one or two pieces of scientific information about phenomena using an understanding that common ancestry and biological evolution are supported by multiple lines of empirical evidence; use one or two pieces of evidence to construct an explanation about phenomena using an understanding that the process of evolution primarily results from four factors: the potential for a species to increase in number, the heritable genetic variation of individuals in a species due to

mutation and sexual reproduction, competition for limited resources, and the proliferation of those organisms that are better able to survive and reproduce in the environment; apply one or two concepts of statistics and probability to support explanations of phenomena using an understanding that organisms with an advantageous heritable trait tend to increase in proportion to organisms lacking the trait; use one or two pieces of evidence to construct an explanation of phenomena using an understanding of how natural selection leads to adaptation of populations; evaluate one or two pieces of evidence supporting claims about phenomena using an understanding that changes in environmental conditions may result in: increases in numbers of individuals of some species, the emergence of new species over time, and the extinction of other species; and create or revise a simulation to test a solution to mitigate adverse impacts of human activity on biodiversity. (LS4)

- Use one or two pieces of evidence to develop a model to describe phenomena using an understanding of the life span of the Sun and the role of nuclear fusion in the Sun's core to release energy that reaches Earth in the form of radiation; use one or two pieces of evidence and an understanding of the phenomena of light spectra, motion of distant galaxies, and composition of matter in the universe to construct an explanation of the big bang theory; communicate one or two scientific ideas about phenomena using an understanding of the way stars produce different elements over their life cycles; use mathematical or computational representations of phenomena to predict the motion of orbiting objects in the solar system; evaluate one or two pieces of evidence of phenomena using an understanding of past and current movements of continental and oceanic crust and the theory of plate tectonics to explain the ages of crustal rocks; and apply scientific reasoning and an understanding of one or two pieces of evidence from ancient Earth materials, meteorites, and other planetary surfaces to describe phenomena about Earth's formation and early history. (ESS1)

- Develop a model to describe phenomena using an understanding of how Earth's internal and surface processes operate at different spatial and temporal scales to form continental and ocean floor features; analyze one type of geoscience data about phenomena to make a claim that one change to Earth's surface can create feedback that causes changes to other Earth systems; use one or two pieces of evidence to develop a model of Earth's interior to describe phenomena using an understanding of the cycling of matter by thermal convection; use a model to describe phenomena using an understanding of how variations in the flow of energy into and out of Earth's systems result in changes in climate; plan and conduct an investigation of phenomena related to the properties of water using an understanding of water's effects on Earth materials and surface processes; develop a quantitative model to describe phenomena using an understanding of the cycling of carbon among the hydrosphere, atmosphere, geosphere, and biosphere; and use one or two pieces of evidence to construct an argument about phenomena using an understanding of the simultaneous coevolution of Earth's systems and life on Earth. (ESS2)

- Use one or two pieces of evidence to construct an explanation about phenomena using an understanding of how the availability of natural resources, occurrence of natural hazards, and changes in climate have influenced human activity; using an understanding of cost-benefit ratios, evaluate two competing design solutions for developing, managing, and utilizing energy and mineral resources; create a computational simulation of phenomena to show the relationships among management of natural resources, the sustainability of human populations, and biodiversity; using an understanding of human impacts on natural systems, evaluate or refine a technological solution that reduces these impacts; analyze one or two pieces of geoscience data and one or two global climate models of phenomena to make a forecast of the current rate of climate change and associated future impacts to Earth systems; and use a computational representation to describe phenomena using an understanding of the relationships among Earth systems and how those relationships are modified due to human activity. (ESS3)

- Analyze a major global challenge to specify one or two qualitative and quantitative criteria and constraints for solutions that account for one or two societal needs and wants; design an engineering solution to a complex real-world problem by breaking it down into smaller, more manageable problems; evaluate a solution to a complex real-world problem based on prioritized criteria and trade-offs that account for a range of constraints, including cost, safety, reliability, and aesthetics, as well as social, cultural, and environmental impacts; and use a computer simulation to model the impact of two proposed solutions to a complex real-world problem with numerous criteria and constraints on interactions within and between systems relevant to the problem. (ETS1)

Nearing Proficiency

Students at the **Nearing Proficiency** level demonstrate evidence of partial understanding and use of all three dimensions (science and engineering practices, crosscutting concepts, and disciplinary core ideas) to make sense of phenomena and/or to design solutions to problems in the physical, life, and Earth and space sciences. Students performing at the Nearing Proficiency level can be expected to be able to demonstrate knowledge and skills like in the following examples, as evidence of partial understanding and use of the NGSS Standards:

- Use a partial understanding of the periodic table to predict one relative property of elements, conduct an investigation of phenomena related to bulk scale properties of substances or construct a partial explanation of how these relate to the strength of electrical forces between particles, describe phenomena about the release or absorption of energy by a chemical system, explain phenomena using a partial understanding of how changes in temperature or concentration affect reaction rate, explain phenomena about chemical systems at equilibrium using a partial understanding of how one condition on the system could be changed to produce more products, use a mathematical representation to support a claim that mass is conserved during a chemical reaction, and use models to describe phenomena using a partial understanding of the changes in the nucleus of an atom or the energy released during fission, fusion, or radioactive decay. (PS1)

- Use data from phenomena to show that $f = ma$, use a mathematical representation of a phenomenon and a partial understanding of momentum to support the claim that the total momentum of a system is conserved, apply a scientific or an engineering idea to design a device that minimizes force on an object during a collision, use mathematical representations of Newton's law of gravitation or Coulomb's law to describe phenomena using a partial understanding of gravitational and electrostatic forces between objects, conduct an investigation of phenomena to produce evidence that an electric current produces a magnetic field or that a changing magnetic field produces electric current, and communicate information about phenomena using a partial understanding of how the molecular structure of a material relates to its macroscopic properties or makes the material well suited for particular uses. (PS2)

- Use a computational model of phenomena to calculate changes in energy of a system when energy flows into and out of the system is known; use models to explain phenomena using a partial understanding of how energy at the macroscopic scale can be accounted for at the microscopic scale in energy associated with particle motion or relative position; design a device that converts one form of energy into another; conduct an investigation of phenomena that provides evidence using a partial understanding that when two components of different temperatures are combined within a closed system, both components eventually have the same temperature; and use a model to explain phenomena related to the forces or the changes in energy between two objects interacting through electric or magnetic fields. (PS3)

- Explain phenomena using an understanding of a mathematical representation regarding one relationship among frequency, wavelength, and speed of waves in various media; ask questions

about phenomena using a partial understanding of the advantages of using digital transmission and storage of information; use one phenomenon to support a claim that electromagnetic radiation may be described using a wave or particle model; make a claim about phenomena related to the effects that different frequencies of electromagnetic radiation have on matter; and communicate information about phenomena using a partial understanding of how specific technological devices use the principles of wave behavior or wave interactions to transmit or capture information and energy. (PS4)

- Use evidence to explain phenomena using a partial understanding of how the structure of DNA determines the structure of proteins; use a model to describe phenomena using a partial understanding of the organization of interacting systems within multicellular organisms; conduct an investigation to provide evidence about phenomena that show that some feedback mechanisms maintain homeostasis; use a model to describe phenomena using a partial understanding of how cell division or differentiation helps produce or maintain a complex organism; use a model to describe phenomena using a partial understanding of how photosynthesis transforms light energy into stored chemical energy; describe that carbon, hydrogen, and oxygen from sugar molecules combine with other elements to form amino acids or other large carbon-based molecules; and use a model to describe phenomena using a partial understanding that cellular respiration is a chemical process that breaks the bonds in food or oxygen molecules, forms bonds in new compounds, or results in a net transfer of energy. (LS1)

- Use mathematical or computational representations to support explanations of phenomena using a partial understanding of factors that affect carrying capacity of ecosystems; use evidence or mathematical representations to describe factors affecting biodiversity or populations in ecosystems and to support a claim for the cycling of matter or flow of energy among organisms in an ecosystem; use evidence to construct an explanation of phenomena using a partial understanding of the cycling of matter and flow of energy in aerobic or anaerobic conditions; use a model to describe phenomena using an understanding of the role of photosynthesis or cellular respiration in the cycling of carbon; evaluate a claim about a phenomenon involving interactions in ecosystems using a partial understanding that these interactions maintain relatively consistent numbers and types of organisms under stable conditions, but changing conditions may result in a new ecosystem; identify a solution for reducing impacts of human activities on the environment or biodiversity; and use evidence to describe the role of group behavior on individual and species' chances to survive and reproduce. (LS2)

- Ask a question about a phenomenon to clarify relationships about the role of DNA in chromosomes in coding the instructions for traits passed from parents to offspring; make a claim about phenomena using a partial understanding that inheritable genetic variations may result from new genetic combinations through meiosis, viable errors during replication, or mutations caused by environmental factors; and describe phenomena using an understanding of the variation and distribution of expressed traits in a population. (LS3)

- Communicate scientific information about phenomena using a partial understanding that common ancestry and biological evolution are supported by empirical evidence; use evidence to construct an explanation about phenomena using a partial understanding that the process of evolution primarily results from one or two of the following factors: the potential for a species to increase in number, the heritable genetic variation of individuals in a species due to mutation and sexual reproduction, competition for limited resources, and the proliferation of those organisms that are better able to survive and reproduce in the environment; apply concepts of statistics or probability to explain phenomena using a partial understanding that organisms with an advantageous heritable trait tend to increase in proportion to organisms lacking the trait; explain phenomena using a partial understanding of how natural selection leads to adaptation of populations; use evidence to support

claims about phenomena using a partial understanding that changes in environmental conditions may result in one or two of the following: increases in numbers of individuals of some species, the emergence of new species over time, or the extinction of other species; and use a simulation to test a solution to mitigate adverse impacts of human activity on biodiversity. (LS4)

- Use evidence to develop a model to describe phenomena using a partial understanding of the life span of the Sun or the role of nuclear fusion in the Sun's core to release energy that reaches Earth in the form of radiation; use evidence and a partial understanding of the phenomena of light spectra, motion of distant galaxies, or composition of matter in the universe to construct an explanation of the big bang theory; communicate a scientific idea about a phenomenon using a partial understanding of the way stars produce different elements over their life cycles; use a mathematical or computational representation of phenomena to predict the motion of orbiting objects in the solar system; use evidence of phenomena using a partial understanding of past and current movements of continental and oceanic crust or the theory of plate tectonics to explain the ages of crustal rocks; and apply scientific reasoning or a partial understanding of evidence from ancient Earth materials, meteorites, and other planetary surfaces to describe phenomena about Earth's formation and early history. (ESS1)

- Use a model to describe phenomena using a partial understanding of how Earth's internal and surface processes operate to form continental and ocean floor features; use geoscience data about phenomena to make a claim that one change to Earth's surface can create feedback that causes changes to other Earth systems; develop a model of Earth's interior to describe phenomena using an understanding of the cycling of matter by thermal convection; use a model to describe phenomena using a partial understanding of how variations in the flow of energy into and out of Earth's systems result in changes in climate; conduct an investigation of phenomena related to the properties of water using a partial understanding of water's effects on Earth materials or surface processes; use a quantitative model to describe phenomena using a partial understanding of the cycling of carbon among the hydrosphere, atmosphere, geosphere, and biosphere; and use evidence to make a claim about phenomena using a partial understanding of the simultaneous coevolution of Earth's systems and life on Earth. (ESS2)

- Use evidence to construct an explanation about phenomena using a partial understanding of how the availability of natural resources, occurrence of natural hazards, or changes in climate have influenced human activity; using a partial understanding of cost-benefit ratios, evaluate a design solution for developing, managing, or utilizing energy or mineral resources; use a computational simulation of phenomena to show some of the relationships among management of natural resources, the sustainability of human populations, and biodiversity; using a partial understanding of human impacts on natural systems, describe a technological solution that reduces these impacts; use geoscience data or global climate models of phenomena to make a forecast of the current rate of climate change or associated future impacts to Earth systems; and use a computational representation to describe phenomena using a partial understanding of the relationships among Earth systems or how those relationships are modified due to human activity. (ESS3)

- Analyze a major global challenge to specify a qualitative or quantitative criterion or constraint for a solution that accounts for a societal need or want; describe one or two ways a complex real-world problem could be broken down into smaller, more manageable problems that could be solved through engineering; explain how a solution to a complex real-world problem meets required criteria or explain one or two trade-offs of the solution; and use a computer simulation to model the impact of a proposed solution to a complex real-world problem with two or three criteria and constraints on interactions within or between systems relevant to the problem. (ETS1)

Novice

Students at the **Novice** level demonstrate evidence of emerging understanding and use of all three dimensions (science and engineering practices, crosscutting concepts, and disciplinary core ideas) to make sense of phenomena and/or to design solutions to problems in the physical, life, and Earth and space sciences.

BIE Science Proficiency Level Descriptors

Grade 5

Policy PLDs

Policy PLDs define the knowledge and skill level expectations for all grades and content areas for the BIE Science Assessment.

Level 4. Advanced

Students demonstrate evidence of **thorough** understanding and use of college and career readiness knowledge, skills, and abilities.

Level 3. Proficient

Students demonstrate evidence of **satisfactory** understanding and use of college and career readiness knowledge, skills, and abilities.

Level 2. Nearing Proficiency

Students demonstrate evidence of **partial** understanding and use of college and career readiness knowledge, skills, and abilities.

Level 1. Novice

Students demonstrate evidence of **emerging** understanding and use of college and career readiness knowledge, skills, and abilities.

Borderline PLDs

Borderline PLDs describe the knowledge and skills that students, just barely within each proficiency level, are expected to be able to demonstrate. In line with the nature of the science standards, the statements combine science and engineering practices, disciplinary core ideas, and crosscutting concepts that students are expected to integrate and demonstrate.

Advanced

Students at the borderline of the **Advanced** level in most situations[1] can demonstrate evidence of thorough understanding and use of all three dimensions (science and engineering practices, crosscutting concepts, and disciplinary core ideas) to make sense of phenomena and/or to design solutions to problems in the physical, life, and Earth and space sciences. They also sometimes demonstrate the skills and understandings at the Proficient level rather than the skills and understandings of the Advanced level. Students performing at the borderline of the Advanced level can be expected in most situations[1] to be able to demonstrate knowledge and skills such as in the following examples, as evidence of just barely thorough understanding and use of the NGSS Standards:

- Develop, use, and analyze a model to describe and explain phenomena using an understanding of matter as tiny particles, describe quantities that should be measured to explain phenomena using an understanding of conservation of matter during physical and chemical changes, describe observations and measurements that can be used to identify materials based on their properties, and plan and conduct an investigation to determine whether a new substance with different properties is formed when two substances are mixed. (PS1)
- Plan and conduct an investigation to provide multiple pieces of evidence about phenomena using an understanding of the effects of balanced and unbalanced forces on the motion of an object, predict the future motion of an object based on complex patterns in observations and measurements, ask detailed questions to describe phenomena using an understanding of cause and effect of electric and magnetic interactions between objects not in contact with each other, thoroughly define a simple design problem that can be solved using magnets, and support an argument with multiple pieces of evidence about phenomena using an understanding that the gravitational force of Earth on objects is directed down. (PS2)
- Construct an explanation supported by multiple pieces of evidence about phenomena using an understanding of the relationship between the speed and energy of an object; provide and analyze evidence that energy can be transferred from place to place; predict and explain outcomes for the changes in energy that occur when objects collide; thoroughly design, test, and refine a device that converts energy from one form to another; and use models to explain phenomena using an understanding that food energy was once energy from the Sun. (PS3)
- Develop models to explain phenomena using an understanding that waves can cause objects to move and that light allows objects to be seen; and compare and explain multiple solutions that use patterns to transfer information. (PS4)

- Develop models to explain phenomena using an understanding of the diversity and commonalities of the life cycles of organisms, construct an argument supported by multiple pieces of evidence about phenomena using an understanding that plants and animals have internal and external structures that support life functions, explain phenomena using an understanding that animals receive, process, and respond to information from their senses, and support an argument with multiple pieces of evidence about phenomena using an understanding that plants get the materials they need for growth chiefly from air and water. (LS1)
- Construct an argument supported by multiple pieces of evidence about phenomena using an understanding that some animals form groups that help members survive; and develop a model to explain phenomena using an understanding of the movement of matter among plants, animals, decomposers, and the environment. (LS2)
- Analyze and interpret data to provide multiple pieces of evidence about phenomena using an understanding that plants and animals have inherited traits and that variation of these traits exists in groups of similar organisms; and support an explanation with multiple pieces of evidence about phenomena using an understanding that traits can be influenced by the environment. (LS3)
- Analyze and interpret fossil data to provide multiple pieces of evidence of organisms and the environments in which they lived; construct an explanation supported by multiple pieces of evidence of phenomena using an understanding that variation among individuals of the same species is a survival advantage and that in a particular habitat some organisms survive well, some survive less well, and some cannot survive at all; and make a claim supported by multiple pieces of evidence about the merit of a solution to a problem caused by changes to the environment and the types of plants and animals that live there. (LS4)
- Explain phenomena using an understanding that patterns in rock formations and fossils in rock layers provide evidence to support an explanation for changes in a landscape over time, support an argument about phenomena using multiple pieces of evidence and an understanding that differences in the apparent brightness of the Sun compared to other stars is due to their relative distances from Earth, and represent and explain data in graphical displays to reveal patterns of daily changes in shadows, day and night, and the seasonal appearance of stars. (ESS1)
- Represent data to explain typical seasonal weather conditions, combine and synthesize information to describe climates in different regions of the world, provide multiple pieces of evidence for phenomena using an understanding of the effects of weathering and the rate of erosion, analyze and interpret data from maps to explain patterns of Earth's features, develop a model to explain phenomena using an understanding of how multiple systems on Earth interact, and describe and graph the percentages of water to provide multiple pieces of evidence about the distribution of water on Earth. (ESS2)
- Make a claim supported by multiple pieces of evidence about the merit of a design solution that reduces the impacts of a weather-related hazard, combine and synthesize information to explain that energy and fuels are derived from natural resources, how their uses affect the environment, and ways individual communities protect Earth's

resources and the environment, and generate and compare multiple solutions to reduce the impacts of several natural Earth processes on humans. (ESS3)

- Define a simple design problem including detailed criteria for success and constraints; generate and compare multiple detailed solutions to a problem; and plan and carry out fair tests to identify more than one way to improve a model or prototype. (ETS1)

Proficient

Students at the borderline of the **Proficient** level in most situations[1] can demonstrate evidence of satisfactory understanding and use of all three dimensions (science and engineering practices, crosscutting concepts, and disciplinary core ideas) to make sense of phenomena and/or to design solutions to problems in the physical, life, and Earth and space sciences. They also sometimes demonstrate the skills and understandings at the Nearing Proficiency level rather than the skills and understandings of the Proficient level. Students performing at the borderline of the Proficient level can be expected in most situations[1] to be able to demonstrate knowledge and skills such as in the following examples, as evidence of just barely satisfactory understanding and use of the NGSS Standards:

- Develop and use a model to describe phenomena using an understanding of matter as tiny particles, measure and graph quantities to describe phenomena using an understanding of conservation of matter during physical or chemical changes, make observations and measurements to identify materials based on their properties, and conduct an investigation to determine whether a new substance with different properties is formed when two substances are mixed. (PS1)
- Plan and conduct an investigation to provide one piece of evidence about phenomena using an understanding of the effects of balanced and unbalanced forces on the motion of an object, predict the future motion of an object based on patterns in observations and measurements, ask questions to describe phenomena using an understanding of cause and effect of electric or magnetic interactions between objects not in contact with each other, define a simple design problem that can be solved using magnets, and support an argument with one piece of evidence about phenomena using an understanding that the gravitational force of Earth on objects is directed down. (PS2)
- Construct an explanation supported by one piece of evidence about phenomena using an understanding of the relationship between the speed and energy of an object; provide evidence that energy can be transferred from place to place; predict outcomes for the changes in energy that occur when objects collide; design, test, and refine a device that converts energy from one form to another; and use models to describe phenomena using an understanding that food energy was once energy from the Sun. (PS3)
- Develop models to describe phenomena using an understanding that waves can cause objects to move and that light allows objects to be seen; and compare multiple solutions that use patterns to transfer information. (PS4)
- Develop models to describe phenomena using an understanding of the diversity and commonalities of the life cycles of organisms, construct an argument supported by one piece of evidence about phenomena using an understanding that plants and animals have internal and external structures that support life functions, describe phenomena

using an understanding that animals receive, process, and respond to information from their senses, and support an argument with one piece of evidence about phenomena using an understanding that plants get the materials they need for growth chiefly from air and water. (LS1)

- Construct an argument supported by one piece of evidence about phenomena using an understanding that some animals form groups that help members survive; and develop a model to describe phenomena using an understanding of the movement of matter among plants, animals, decomposers, and the environment. (LS2)
- Analyze and interpret data to provide one piece of evidence about phenomena using an understanding that plants and animals have inherited traits and that variation of these traits exists in groups of similar organisms; and support an explanation with one piece of evidence about phenomena using an understanding that traits can be influenced by the environment. (LS3)
- Analyze and interpret fossil data to provide one piece of evidence of organisms and the environments in which they lived; construct an explanation supported by one piece of evidence of phenomena using an understanding that variation among individuals of the same species is a survival advantage and that in a particular habitat some organisms survive well, some survive less well, and some cannot survive at all; and make a claim supported by one piece of evidence about the merit of a solution to a problem caused by changes to the environment and the types of plants and animals that live there. (LS4)
- Describe phenomena using an understanding that patterns in rock formations and fossils in rock layers provide evidence to support an explanation for changes in a landscape over time, support an argument about phenomena using one piece of evidence and an understanding that differences in the apparent brightness of the Sun compared to other stars is due to their relative distances from Earth, and represent data in graphical displays to reveal patterns of daily changes in shadows, day and night, and the seasonal appearance of stars. (ESS1)
- Represent data to describe typical seasonal weather conditions, combine information to describe climates in different regions of the world, provide one piece of evidence for phenomena using an understanding of the effects of weathering or the rate of erosion, analyze and interpret data from maps to describe patterns of Earth's features, develop a model to describe phenomena using an understanding of how Earth's systems interact, and describe and graph the percentages of water to provide one piece of evidence about the distribution of water on Earth. (ESS2)
- Make a claim supported by one piece of evidence about the merit of a design solution that reduces the impacts of a weather-related hazard, combine information to describe how energy and fuels are derived from natural resources, how their uses affect the environment, and ways individual communities protect Earth's resources and the environment, and generate and compare multiple solutions to reduce the impacts of natural Earth processes on humans. (ESS3)
- Define a simple design problem including criteria for success and constraints; generate and compare multiple solutions to a problem; and plan and carry out fair tests to identify one way to improve a model or prototype. (ETS1)

Nearing Proficiency

Students at the borderline of the **Nearing Proficiency** level in most situations[1] can demonstrate evidence of partial understanding and use of all three dimensions (science and engineering practices, crosscutting concepts, and disciplinary core ideas) to make sense of phenomena and/or to design solutions to problems in the physical, life, and Earth and space sciences. They also sometimes demonstrate the skills and understandings at the Novice level rather than the skills and understandings of the Nearing Proficient level. Students performing at the borderline of the Nearing Proficiency level can be expected in most situations[1] to be able to demonstrate knowledge and skills such as in the following examples, as evidence of just barely partial understanding and use of the NGSS Standards:

- Use a model to describe phenomena using an understanding of matter as tiny particles, graph quantities to describe phenomena using an understanding of conservation of matter during physical or chemical changes, make observations to identify materials based on their properties, and use data to determine whether a new substance with different properties is formed when two substances are mixed. (PS1)
- Conduct an investigation to provide one piece of evidence about phenomena using an understanding of the effects of balanced or unbalanced forces on the motion of an object, predict the future motion of an object based on simple patterns in observations and measurements, ask questions using an understanding of cause and effect of electric or magnetic interactions between objects not in contact with each other, partially define a simple design problem that can be solved using magnets, and make a claim about phenomena using an understanding that the gravitational force of Earth on objects is directed down. (PS2)
- Describe phenomena using an understanding of the relationship between the speed and energy of an object, explain that energy can be transferred from place to place, describe the changes in energy that occur when objects collide, describe elements of a device that converts energy from one form to another, and describe phenomena using an understanding that food energy was once energy from the Sun. (PS3)
- Use models to describe phenomena using an understanding that waves can cause objects to move or that light allows objects to be seen; and describe a solution that uses patterns to transfer information. (PS4)
- Use models to describe phenomena using an understanding of the diversity and commonalities of the life cycles of organisms, make a claim about phenomena using an understanding that plants and animals have internal and external structures that support life functions, describe phenomena using an understanding that animals receive or process or respond to information from their senses, and make a claim about phenomena using an understanding that plants get the materials they need for growth chiefly from air and water. (LS1)

---

[1] "Most situations" refers to the following. Students at level X can be expected to be able to demonstrate knowledge and skills for most but not all.

- Make a claim about phenomena using an understanding that some animals form groups that help members survive; and use a model to describe phenomena using an understanding of the movement of matter among plants, animals, decomposers, and the environment. (LS2)
- Use data to describe phenomena using an understanding that plants and animals have inherited traits OR that variation of these traits exists in groups of similar organisms; and make a claim.
  about phenomena using an understanding that traits can be influenced by the environment. (LS3)
- Use fossil data to describe organisms and the environments in which they lived; describe phenomena using an understanding that variation among individuals of the same species is a survival advantage or that in a particular habitat some organisms survive well, some survive less well, and some cannot survive at all; and describe a solution to a problem caused by changes to the environment and the types of plants and animals that live there. (LS4)
- Describe phenomena using an understanding that patterns in rock formations or fossils in rock layers provide evidence to support changes in a landscape over time, describe phenomena using an understanding that differences in the apparent brightness of the Sun compared to other stars is due to their relative distances from Earth, and use data in graphical displays to reveal patterns of daily changes in shadows or day and night or the seasonal appearance of stars. (ESS1)
- Use data to describe typical seasonal weather conditions, describe climates in different regions of the world, describe phenomena using an understanding of the effects of weathering or the rate of erosion, use maps to describe patterns of Earth's features, use a model to describe phenomena using an understanding of how Earth's systems interact, and provide a description of the distribution of water on Earth. (ESS2)
- Describe a design solution or component of a design solution that reduces the impacts of a weather-related hazard, use information to determine that energy and fuels are derived from natural resources or how their uses affect the environment or ways individual communities protect Earth's resources and the environment, and describe a solution or component of a solution to reduce the impacts of natural Earth processes on humans. (ESS3)
- Define a simple design problem including at least one criteria for success or one constraint; generate a solution to a problem or compare two solutions to a problem; and use results of a fair test to identify one way to improve a model or prototype. (ETS1)

Novice

Students at the **Novice** level demonstrate evidence of emerging understanding and use of all three dimensions (science and engineering practices, crosscutting concepts, and disciplinary core ideas) to make sense of phenomena and/or to design solutions to problems in the physical, life, and Earth and space sciences.

- Combinations of disciplinary core ideas, practices, and crosscutting concepts

Science phenomena, in which students are required to apply their knowledge of science content and mastery of crosscutting concepts.

BIE Science Proficiency Level Descriptors

Grade 8

Policy PLDs

Policy PLDs define the knowledge and skill level expectations for all grades and content areas for the BIE Science Assessment.

Level 4. Advanced

Students demonstrate evidence of **thorough** understanding and use of college and career readiness knowledge, skills, and abilities.

Level 3. Proficient

Students demonstrate evidence of **satisfactory** understanding and use of college and career readiness knowledge, skills, and abilities.

Level 2. Nearing Proficiency

Students demonstrate evidence of **partial** understanding and use of college and career readiness knowledge, skills, and abilities.

Level 1. Novice

Students demonstrate evidence of **emerging** understanding and use of college and career readiness knowledge, skills, and abilities.

Borderline PLDs

Borderline PLDs describe the knowledge and skills that students, just barely within each proficiency level, are expected to be able to demonstrate. In line with the nature of the science standards, the statements combine science and engineering practices, disciplinary core ideas, and crosscutting concepts that students are expected to integrate and demonstrate.

Advanced

Students at the borderline of the **Advanced** level in most situations[1] can demonstrate evidence of thorough understanding and use of all three dimensions (science and engineering practices, crosscutting concepts, and disciplinary core ideas) to make sense of phenomena and/or to design solutions to problems in the physical, life, and Earth and space sciences. They also sometimes demonstrate the skills and understandings at the Proficient level rather than the skills and understandings of the Advanced level. Students performing at the borderline of the Advanced level can be expected in most situations[1] to be able to demonstrate knowledge and skills such as in the following examples, as evidence of just barely thorough understanding and use of the NGSS Standards:

- Develop, use, and analyze models to describe or explain phenomena using an understanding of the structure of matter; to predict, describe, and explain phenomena using an understanding of changes in particle motion and state; and to provide evidence for and describe phenomena using an understanding of conservation of mass during multiple physical and chemical changes. (PS1)
- Plan, carry out, and refine investigations to provide evidence to explain phenomena using an understanding of the effects of forces, interactions, and mass on the motion of objects, as well as analyze various data to evaluate claims about phenomena using an understanding of gravitational interactions in systems, and to design and/or compare multiple solutions to a problem using an understanding of systems of colliding objects. (PS2)
- Plan, carry out, and refine investigations, use and analyze data, and develop, use, and analyze models to explain phenomena using an understanding of various relationships involving kinetic and potential energy in systems, as well as apply such understanding to design, test, and evaluate devices to solve problems related to energy transfer and to support and/or evaluate claims about phenomena related to energy transfer. (PS3)
- Develop, use, and apply mathematical representations, patterns, and models to explain phenomena using an understanding of wave properties and relationships and wave interactions with various materials, and synthesize multiple sources of information using an understanding of signal types to evaluate claims about phenomena related to reliability of digital and analog signals. (PS4)
- Use multiple pieces of evidence from investigations of phenomena to explain that living things are made of cells; develop and use models of phenomena to describe the function of a cell and its parts and to describe how food is rearranged in organisms through chemical reactions to support growth and/or release energy; support arguments about phenomena using multiple pieces of evidence and an understanding that the body is a system of interacting subsystems composed of cells; support an explanation for how several animal behaviors and several specialized plant structures affect the probability of reproductive success; use multiple pieces of evidence and an understanding of environmental and genetic factors to explain phenomena about how those factors influence the growth of organisms; construct an explanation using multiple pieces of evidence to explain the role of photosynthesis in the cycling of matter and flow of energy into and out of organisms; synthesize multiple sources of information about phenomena and an understanding of behaviors of organisms to determine that sensory receptors

respond to stimuli by sending messages to the brain for immediate behavior or for storage as memories. (LS1)

- Analyze and interpret data about phenomena to provide evidence to explain multiple ways that resource availability affects populations, construct an argument supported by multiple pieces of evidence that populations are affected by changing physical and biological components of an ecosystem, develop and revise models to explain phenomena using an understanding of the cycling of matter and energy in an ecosystem, describe phenomena using an understanding of interactions among organisms and predict multiple patterns of interactions among organisms across multiple ecosystems, and evaluate multiple competing design solutions for phenomena that involve maintaining biodiversity in ecosystems. (LS2)

- Develop and use multiple models to explain phenomena using an understanding of how genetic mutations affect proteins resulting in harmful, beneficial, or neutral effects on an organism, and to explain phenomena using an understanding of how asexual reproduction results in offspring with identical genetic information and to explain how sexual reproduction results in offspring with genetic variation. (LS3)

- Analyze and interpret multiple pieces of data about phenomena using an understanding of the fossil record and modern organisms to show patterns in the change of life forms over time, apply multiple scientific ideas about phenomena to construct an explanation for the anatomical similarities and differences among modern and fossil organisms to infer evolutionary relationships, analyze pictorial data to compare similarities in embryological development across multiple familiar and unfamiliar species to identify evolutionary relationships, use multiple pieces of evidence and mathematical representations to explain phenomena using an understanding of how variation in genetic traits provides advantages to some individuals within a population and to support explanations of increases and decreases in specific traits over time, and explain phenomena by synthesizing multiple pieces of information about ways technologies have changed the way humans influence the inheritance of desired traits in organisms. (LS4)

- Develop, use, and revise a model of the Earth-Sun-Moon system to describe phenomena using an understanding of the cyclic pattern of the seasons and to describe phenomena using an understanding of the role of gravity in the motions within the solar system and galaxies, analyze and interpret data on multiple phenomena related to the scale properties of objects in the solar system, and use multiple pieces of evidence and an understanding of rock strata to explain phenomena about how the geologic time scale is used to organize Earth's history. (ESS1)

- Develop models to describe phenomena using an understanding of the flow of energy that drives the cycling of Earth's materials, use multiple pieces of evidence to explain phenomena using an understanding of how geoscience processes have changed Earth's surface at varying time and spatial scales, analyze and interpret multiple pieces of data to explain phenomena using an understanding of the evidence that supports past plate motions on Earth, develop models to describe phenomena using an understanding of the water cycle including energy and gravity, explain weather phenomena synthesizing and using evidence and an understanding of the interactions of air masses, and develop and use models to describe phenomena using an understanding of how unequal heating and Earth's rotation result in climate, atmospheric, and ocean circulation patterns. (ESS2)

- Use evidence to explain multiple phenomena using an understanding of how geoscience processes have resulted in uneven distribution of Earth's natural resources, analyze and interpret multiple pieces of data on natural hazard phenomena to forecast future catastrophic events and to inform the development of technologies to mitigate their effects, apply scientific principles to design a successful solution for monitoring and minimizing the human impacts on the environment, use multiple pieces of evidence to support an argument about phenomena

using an understanding of how increases in human population impact Earth's systems, and ask multiple questions about phenomena to clarify evidence of multiple factors that have caused the rise in global temperatures. (ESS3)

- Define the criteria and constraints of a design problem with sufficient precision to ensure an optimal solution and using an understanding of scientific principles and potential impacts on people and the environment and an understanding of how those impacts may limit possible solutions, use a systematic process to evaluate how well multiple competing design solutions meet required criteria and constraints, analyze data from tests of multiple different design solutions to identify the best characteristics of each solution that can be combined into a new solution that will better meet criteria for success, and develop a realistic model of a proposed object, tool, or process that generates data while it is repeatedly tested and modified until an optimal design is achieved. (ETS1)

Proficient

Students at the borderline of the **Proficient** level in most situations[1] can demonstrate evidence of satisfactory understanding and use of all three dimensions (science and engineering practices, crosscutting concepts, and disciplinary core ideas) to make sense of phenomena and/or to design solutions to problems in the physical, life, and Earth and space sciences. They also sometimes demonstrate the skills and understandings at the Nearing Proficiency level rather than the skills and understandings of the Proficient level. Students performing at the borderline of the Proficient level can be expected in most situations[1] to be able to demonstrate knowledge and skills such as in the following examples, as evidence of just barely satisfactory understanding and use of the NGSS Standards:

- Develop and use models to describe phenomena using an understanding of the structure of matter; to predict and describe phenomena using an understanding of changes in particle motion and state; and to describe phenomena using an understanding of conservation of mass during one or two physical and chemical changes. (PS1)
- Plan and carry out investigations to produce data and/or provide evidence about phenomena using an understanding of the effects of forces, interactions, and mass on the motion of objects, as well as use direct data to support claims about phenomena using an understanding of gravitational interactions in systems and to design a solution to a problem using an understanding of systems of colliding objects. (PS2)
- Plan investigations, use data, and develop or use models to describe phenomena using an understanding of relationships involving kinetic and potential energy in systems, as well as apply such understanding to design and test devices to solve problems related to energy transfer and to support claims about phenomena related to energy transfer. (PS3)
- Use mathematical representations and patterns, and develop and use models, to describe phenomena using an understanding of wave properties and relationships and wave interactions with various materials and use multiple sources of information and an understanding of signal types to support claims about phenomena related to reliability of digital and analog signals. (PS4)
- Use one or two pieces of evidence from investigations of phenomena to explain that living things are made of cells, develop and use models of phenomena to describe the function of a cell and its parts and to describe how food is rearranged in organisms through chemical reactions to support growth and/or release energy, support arguments about phenomena using one or two pieces of evidence and an understanding that the body is a system of interacting subsystems composed of cells, support an explanation for how some animal behaviors and some specialized

plant structures affect the probability of reproductive success, use one or two pieces of evidence and an understanding of environmental and genetic factors to explain phenomena about how those factors influence the growth of organisms, construct an explanation based on one or two pieces of evidence to explain the role of photosynthesis in the cycling of matter and flow of energy into and out of organisms, use information about phenomena and an understanding of behaviors of organisms to determine that sensory receptors respond to stimuli by sending messages to the brain for immediate behavior or for storage as memories. (LS1)

- Analyze and interpret data about phenomena to provide evidence to explain one or two ways that resource availability affects populations, construct an argument supported by one or two pieces of evidence that populations are affected by changing physical or biological components of an ecosystem, develop models to describe phenomena using an understanding of the cycling of matter and energy in an ecosystem, describe phenomena using an understanding of interactions among organisms and predict one or two patterns of interactions among organisms across multiple ecosystems, and evaluate two competing design solutions for phenomena that involve maintaining biodiversity in ecosystems. (LS2)

- Develop and use one or two models to explain phenomena using an understanding of how genetic mutations affect proteins resulting in harmful, beneficial, or neutral effects on an organism and to explain phenomena using an understanding of how asexual reproduction results in offspring with identical genetic information and how sexual reproduction results in offspring with genetic variation. (LS3)

- Analyze and interpret one or two pieces of data about phenomena using an understanding of the fossil record and modern organisms to show patterns in the change of life forms over time, apply one or two scientific ideas about phenomena to construct an explanation for the anatomical similarities and differences among modern and fossil organisms to infer evolutionary relationships, analyze pictorial data to compare similarities in embryological development across multiple familiar species to identify evolutionary relationships, use one or two pieces of evidence or mathematical representations to explain phenomena using an understanding of how variation in genetic traits provides advantages to some individuals within a population and to support explanations of increases and decreases in specific traits over time, and explain phenomena by synthesizing one or two pieces of information about technologies that have changed the way humans influence the inheritance of desired traits in organisms. (LS4)

- Develop and use a model of the Earth-Sun-Moon system to describe phenomena using an understanding of the cyclic pattern of the seasons and to describe phenomena using an understanding of the role of gravity in the motions within the solar system and galaxies, analyze and interpret data on one or two phenomena related to the scale properties of objects in the solar system, and use one or two pieces of evidence and an understanding of rock strata to explain phenomena about how the geologic time scale is used to organize Earth's history. (ESS1)

- Develop a model to describe phenomena using an understanding of the flow of energy that drives cycling of Earth's materials, use one or two pieces of evidence to explain phenomena using an understanding of how geoscience processes have changed Earth's surface at varying time and spatial scales, analyze and interpret one or two pieces of data to explain phenomena using an understanding of the evidence that supports past plate motions on Earth, develop a model to describe phenomena using an understanding of the water cycle including energy and gravity, explain weather phenomena using evidence and an understanding of the interactions of air masses, and develop and use a model to describe phenomena using an understanding of how unequal heating and Earth's rotation result in climate, atmospheric, or ocean circulation patterns. (ESS2)

- Use evidence to explain one or two phenomena using an understanding of how geoscience processes have resulted in uneven distribution of Earth's natural resources, analyze and interpret one or two pieces of data on natural hazard phenomena to forecast future catastrophic events and to inform the development of technologies to mitigate their effects, apply scientific principles to design a solution for monitoring and minimizing human impacts on the environment, use one or two pieces of evidence to support an argument about phenomena using an understanding of how increases in human population impact Earth's systems, and ask one or two questions about phenomena to clarify evidence of one or two factors that have caused the rise in global temperatures. (ESS3)
- Define the criteria and constraints of a design problem with sufficient precision to ensure a successful solution and using an understanding of scientific principles and potential impacts on people and the environment and an understanding of how those impacts may limit possible solutions, use a systematic process to evaluate how well two competing design solutions meet required criteria and constraints, analyze data from tests of two different design solutions to identify the best characteristics of each solution that can be combined into a new solution that will better meet criteria for success, and develop a model of a proposed object, tool, or process that generates data while it is repeatedly tested and modified until an optimal design is achieved. (ETS1)

Nearing Proficiency

Students at the borderline of the **Nearing Proficiency** level in most situations[2] can demonstrate evidence of partial understanding and use of all three dimensions (science and engineering practices, crosscutting concepts, and disciplinary core ideas) to make sense of phenomena and/or to design solutions to problems in the physical, life, and Earth and space sciences. They also sometimes demonstrate the skills and understandings at the Novice level rather than the skills and understandings of the Nearing Proficient level. Students performing at the borderline of the Nearing Proficiency level can be expected in most situations[1] to be able to demonstrate knowledge and skills such as in the following examples, as evidence of just barely partial understanding and use of the NGSS Standards:

- Use models to identify the structure of matter as it relates to phenomena; to describe basic phenomena using an understanding of changes in particle motion and state; and to identify that mass is conserved during physical or chemical changes that take place in various phenomena. (PS1)
- Identify or describe parts of investigations about phenomena using an understanding of the effects of forces, interactions, and mass on the motion of objects, describe aspects of phenomena using one or two pieces of data and an understanding of one gravitational interaction in a system, and identify a design or elements of a solution to a problem related to colliding objects. (PS2)
- Describe parts of investigations, use data, and use models to describe aspects of phenomena using an understanding of some relationships involving kinetic energy in systems; design and test a device for problems related to energy transfer; and identify principles that support claims about energy transfer in phenomena. (PS3)
- Use mathematical representations, patterns, and models to identify wave properties and wave interactions with various materials as they relate to phenomena and use one or two sources of

---

[2] "Most situations" refers to the following. Students at level X can be expected to be able to demonstrate knowledge and skills for most but not all:

information to identify that digital signals are more reliable than analog signals as demonstrated in various phenomena. (PS4)

- Use evidence from an investigation of a phenomenon to explain that living things are made of cells, use models of phenomena to describe the function of a cell and some of its parts and to describe that food is rearranged in organisms into new substances to support growth or to release energy, make claims about phenomena using evidence and a partial understanding that the body is a system of interacting subsystems composed of cells, describe some animal behaviors or specialized plant structures that may affect reproductive success, use evidence and a partial understanding of environmental or genetic factors to describe phenomena about how those factors influence the growth of organisms, construct an explanation based on one piece of evidence to describe the role of photosynthesis in the cycling of matter or flow of energy into and out of organisms, and use information about phenomena to describe that organisms use their senses to respond to stimuli immediately or to store information as memories. (LS1)

- Use data about phenomena to describe one way resource availability affects populations, make a claim supported by evidence that populations are affected by changing components of an ecosystem, use models to describe phenomena about the cycling of matter or energy in an ecosystem, use a partial understanding of interactions among organisms to describe one interaction between organisms and describe a design solution or components of a solution for phenomena that involve maintaining biodiversity in ecosystems. (LS2)

- Use models to partially explain phenomena using an understanding of one way that genetic mutations affect organisms and to describe phenomena using an understanding of asexual reproduction resulting in offspring with identical genetic information, or about sexual reproduction resulting in offspring with genetic variation. (LS3)

- Use data about phenomena using a partial understanding of the fossil record and modern organisms to show patterns in the change of life forms over time, apply a scientific idea about phenomena to describe an anatomical similarity or difference between modern and fossil organisms, use pictorial data to describe similarities in embryological development across multiple species, use evidence or one mathematical representation to describe phenomena using a partial understanding of variation in genetic traits among individuals within a population or to describe increases and decreases in specific traits over time, and describe phenomena using a partial understanding about technologies that have changed the way humans influence the inheritance of desired traits in organisms. (LS4)

- Use a model of the Earth-Sun-Moon system to describe phenomena using a partial understanding of the cyclic pattern of the seasons and to describe phenomena using a partial understanding of the role of gravity in the motions within the solar system or galaxies, use data on phenomena related to the scale properties of objects in the solar system, and use evidence and a partial understanding of rock strata to describe some aspects about how the geologic time scale is related to Earth's history. (ESS1)

- Use a model to describe phenomena using a partial understanding of the flow of energy that drives the cycling of Earth's materials, use evidence to explain phenomena using a partial understanding of how geoscience processes have changed Earth's surface, use data to explain phenomena using a partial understanding of the evidence that supports past plate motions on Earth, use a model to describe phenomena using an understanding of the water cycle, describe weather phenomena using evidence and a partial understanding of the interactions of air masses, and use a model to describe phenomena using an understanding of how unequal heating and Earth's rotation result in some climate, atmospheric, or ocean circulation patterns. (ESS2)

- Use evidence to explain a phenomenon using a partial understanding of how geoscience processes have resulted in uneven distribution of some of Earth's natural resources, use data on natural hazard phenomena to forecast future catastrophic events or to inform the development of one technology that could be used to mitigate their effects, identify human impacts on the environment or design parts of a solution for monitoring or minimizing the human impacts, use evidence to make a claim about phenomena using a partial understanding of how increases in human population impact Earth's systems, and ask one question about a phenomenon to clarify evidence of one factor that has caused the rise in global temperatures. (ESS3)
- Define one criterion or constraint of a design problem using an understanding of scientific principles and/or potential impacts on people and the environment and identify one way those impacts may limit possible solutions, use a systematic process to evaluate how well a design solution meets required criteria or constraints, analyze data from tests of a design solution to identify a characteristic of the solution that is necessary to meet the criteria for success, and develop a partial model of a proposed object, tool, or process that can be tested and modified. (ETS1)

Novice

Students at the **Novice** level demonstrate evidence of emerging understanding and use of all three dimensions (science and engineering practices, crosscutting concepts, and disciplinary core ideas) to make sense of phenomena and/or to design solutions to problems in the physical, life, and Earth and space sciences.

- Combinations of disciplinary core ideas, practices, and crosscutting concepts
- Science phenomena, in which students are required to apply their knowledge of science content and mastery of crosscutting concepts and practices.

BIE Science Borderline Proficiency Level Descriptors

Grade 11

Policy PLDs

Policy PLDs define the knowledge and skill level expectations for all grades and content areas for the BIE Science Assessment.

Level 4. Advanced

Students demonstrate evidence of **thorough** understanding and use of college and career readiness knowledge, skills, and abilities.

Level 3. Proficient

Students demonstrate evidence of **satisfactory** understanding and use of college and career readiness knowledge, skills, and abilities.

Level 2. Nearing Proficiency

Students demonstrate evidence of **partial** understanding and use of college and career readiness knowledge, skills, and abilities.

Level 1. Novice

Students demonstrate evidence of **emerging** understanding and use of college and career readiness knowledge, skills, and abilities.

Borderline PLDs

Borderline PLDs describe the knowledge and skills that students, just barely within each proficiency level, are expected to be able to demonstrate. In line with the nature of the science standards, the statements combine science and engineering practices, disciplinary core ideas, and crosscutting concepts that students are expected to integrate and demonstrate.

Advanced

Students at the borderline of the **Advanced** level in most situations[1] can demonstrate evidence of thorough understanding and use of all three dimensions (science and engineering practices, crosscutting concepts, and disciplinary core ideas) to make sense of phenomena and/or to design solutions to problems in the physical, life, and Earth and space sciences. They also sometimes demonstrate the skills and understandings at the Proficient level rather than the skills and understandings of the Advanced level. Students performing at the borderline of the Advanced level can be expected in most situations[1] to be able to demonstrate knowledge and skills such as in the following examples, as evidence of just barely thorough understanding and use of the NGSS Standards:

- Use an understanding of the periodic table to predict multiple relative properties of elements, plan and conduct an investigation of phenomena related to multiple bulk scale properties of substances and explain how these relate to the strength of electrical forces between particles, explain phenomena about the release or absorption of energy by a chemical system by developing models to show changes in total bond energy, use multiple pieces of evidence to explain phenomena using an understanding of how changes in temperature and concentration affect reaction rate, explain phenomena about chemical systems at equilibrium using an understanding of how multiple conditions on the system could be changed to produce more or fewer products or more or fewer reactants, use multiple mathematical representations to support claims that mass is conserved during a chemical reaction, and develop multiple models to describe phenomena using an understanding of the changes in the nucleus of an atom and the energy released during fission, fusion, and radioactive decay. (PS1)
- Analyze and use multiple pieces of data from phenomena to show that f = ma; use multiple mathematical representations of phenomena and an understanding of momentum to support the claim that the total momentum of a system is conserved when there is no net force on the system; apply multiple scientific and engineering ideas to design, evaluate, and refine multiple devices that minimize force on an object during a collision; use mathematical representations of Newton's law of gravitation and Coulomb's law to describe and make predictions about familiar and unfamiliar phenomena using an understanding of gravitational and electrostatic forces between objects; plan and conduct an investigation of phenomena to produce multiple pieces of evidence that prove an electric current produces a magnetic field and that a changing magnetic field produces electric current; and communicate multiple pieces of information about phenomena using an understanding of how the molecular structure of a material relates to its macroscopic properties and makes the material well suited for particular uses. (PS2)
- Create multiple computational models of phenomena to calculate changes in energy of a system when energy flows into and out of the system is known; develop and use multiple

models to

explain phenomena using an understanding of how energy at the macroscopic scale can be accounted for at the microscopic scale in energy associated with particle motion and relative position; design, build, and refine devices that convert one form of energy into another; plan and conduct an investigation of phenomena to provide multiple pieces of evidence using an understanding that when two components of different temperatures are combined within a closed system, both components eventually have the same temperature; and develop and use a model to explain phenomena related to the forces and the changes in energy between two objects interacting through electric fields and magnetic fields. (PS3)

- Explain phenomena using an understanding of multiple mathematical representations regarding relationships among frequency, wavelength, and speed of waves in various media; evaluate multiple questions about phenomena using an understanding of the advantages of using digital transmission and storage of information; use multiple phenomena to evaluate claims that electromagnetic radiation can be described using a wave or particle model; in the context of phenomena, evaluate multiple claims about the effects that different frequencies of electromagnetic radiation have on matter; and communicate technical information about phenomena using an understanding of how multiple specific technological devices use the principles of wave behavior and wave interactions to transmit and capture information and energy. (PS4)

- Use multiple pieces of evidence to explain phenomena using an understanding of how the structure of DNA determines the structure of proteins and how proteins carry out the functions of life through specialized cells; develop and use a complex model to describe phenomena using an understanding of the organization of interacting systems within multicellular organisms; plan and conduct an investigation to provide multiple pieces of evidence about phenomena that show that feedback mechanisms maintain homeostasis; use a complex model to describe phenomena using an understanding of how cell division and differentiation help produce and maintain complex organisms; use a complex model to describe phenomena using an understanding of how photosynthesis transforms light energy into stored chemical energy; use multiple pieces of evidence to construct and revise an explanation about phenomena using an understanding of how carbon, hydrogen, and oxygen from sugar molecules combine with other elements to form amino acids and other large carbon-based molecules; and use a complex model to describe phenomena using an understanding that cellular respiration is a chemical process that breaks the bonds in food and oxygen molecules and forms bonds in new compounds, which results in a net transfer of energy. (LS1)

- Use mathematical and computational representations to support explanations of phenomena using an understanding of factors that affect carrying capacity of ecosystems at different scales; use multiple pieces of evidence and mathematical representations to support and revise explanations of phenomena using an understanding of factors affecting biodiversity and populations in ecosystems of different scales and to support claims for the cycling of matter and flow of energy among organisms in an ecosystem; use evidence to construct and revise an explanation of phenomena using an understanding of the cycling of matter and flow of energy in aerobic and anaerobic conditions; develop models to describe phenomena using an understanding of the role of photosynthesis and cellular respiration in the cycling of carbon among Earth's spheres; evaluate multiple claims, pieces of evidence, and reasoning about phenomena involving complex interactions in ecosystems using an understanding that these interactions maintain relatively consistent numbers and types of organisms under stable conditions, but changing conditions may result in a new ecosystem; design, evaluate, and refine solutions for reducing impacts of human activities on the environment or biodiversity; and evaluate multiple pieces of evidence about phenomena using an understanding of the role of group behavior on individual and species' chances to survive and reproduce. (LS2)

- Ask multiple questions about phenomena to clarify relationships surrounding the role of DNA in chromosomes in coding the instructions for traits passed from parents to offspring; use

multiple pieces of evidence to make and defend a claim about phenomena using an understanding that inheritable genetic variations may result from new genetic combinations through meiosis, viable errors during replication, and/or mutations caused by environmental factors; and apply multiple concepts of statistics and probability to explain phenomena using an understanding of the variation and distribution of expressed traits in a population. (LS3)

- Communicate multiple pieces of scientific information about phenomena using an understanding that common ancestry and biological evolution are supported by multiple lines of empirical evidence; use multiple pieces of evidence to construct an explanation about phenomena using an understanding that the process of evolution primarily results from four factors: the potential for a species to increase in number, the heritable genetic variation of individuals in a species due to mutation and sexual reproduction, competition for limited resources, and the proliferation of those organisms that are better able to survive and reproduce in the environment; apply multiple concepts of statistics and probability to support explanations of phenomena using an understanding that organisms with an advantageous heritable trait tend to increase in proportion to organisms lacking the trait; use multiple pieces of evidence to construct an explanation of phenomena using an understanding of how natural selection leads to adaptation of populations; evaluate multiple pieces of evidence supporting claims about phenomena using an understanding that changes in environmental conditions may result in: increases in numbers of individuals of some species, the emergence of new species over time, and the extinction of other species; and create and revise a simulation to test a solution to mitigate adverse impacts of human activity on biodiversity. (LS4)

- Use multiple pieces of evidence to develop a model to describe phenomena using an understanding of the life span of the Sun and the role of nuclear fusion in the Sun's core to release energy that reaches Earth in the form of radiation; use multiple pieces of evidence and an understanding of the phenomena of light spectra, motion of distant galaxies, and composition of matter in the universe to construct an explanation of the big bang theory; communicate multiple scientific ideas about phenomena using an understanding of the way stars produce different elements over their life cycles; use mathematical and computational representations of phenomena to predict the motion of orbiting objects in the solar system; evaluate multiple pieces of evidence of phenomena using an understanding of past and current movements of continental and oceanic crust and the theory of plate tectonics to explain the ages of crustal rocks; and apply scientific reasoning and an understanding of multiple pieces of evidence from ancient Earth materials, meteorites, and other planetary surfaces to describe phenomena about Earth's formation and early history. (ESS1)

- Develop models to describe phenomena using an understanding of how Earth's internal and surface processes operate at different spatial and temporal scales to form continental and ocean floor features; analyze multiple types of geoscience data about phenomena to make a claim that one change to Earth's surface can create feedback that causes changes to other Earth systems; use multiple pieces of evidence to develop a model of Earth's interior to describe phenomena using an understanding of the cycling of matter by thermal convection; use models to describe phenomena using an understanding of how variations in the flow of energy into and out of Earth's systems result in changes in climate; plan and conduct investigations of phenomena related to the properties of water using an understanding of water's effects on Earth materials and surface processes; develop quantitative models to describe phenomena using an understanding of the cycling of carbon among the hydrosphere, atmosphere, geosphere, and biosphere; and use multiple pieces of evidence to construct an argument about phenomena using an understanding of the simultaneous coevolution of Earth's systems and life on Earth.

(ESS2)

- Use multiple pieces of evidence to construct an explanation about phenomena using an understanding of how the availability of natural resources, occurrence of natural hazards, and changes in climate have influenced human activity; using an understanding of cost-benefit ratios, evaluate multiple competing design solutions for developing, managing, and utilizing energy and mineral resources; create computational simulations of phenomena to show the relationships among management of natural resources, the sustainability of human populations, and biodiversity; using an understanding of human impacts on natural systems, evaluate and refine a technological solution that reduces these impacts; analyze multiple pieces of geoscience data and multiple global climate models of phenomena to make a forecast of the current rate of climate change and associated future impacts to Earth systems; and use computational representations to describe phenomena using an understanding of the relationships among Earth systems and how those relationships are modified due to human activity. (ESS3)

- Analyze a major global challenge to specify multiple qualitative and quantitative criteria and constraints for solutions that account for multiple societal needs and wants; design an engineering solution to multiple complex real-world problems by breaking them down into smaller, more manageable problems; evaluate and refine a solution to a complex real-world problem based on prioritized criteria and trade-offs that account for a range of constraints, including cost, safety, reliability, and aesthetics, as well as social, cultural, and environmental impacts; and use a computer simulation to model the impact of multiple proposed solutions to a complex real-world problem with numerous criteria and constraints on interactions within and between systems relevant to the problem. (ETS1)

Proficient

Students at the borderline of the **Proficient** level in most situations[1] can demonstrate evidence of satisfactory understanding and use of all three dimensions (science and engineering practices, crosscutting concepts, and disciplinary core ideas) to make sense of phenomena and/or to design solutions to problems in the physical, life, and Earth and space sciences. They also sometimes demonstrate the skills and understandings at the Nearing Proficiency level rather than the skills and understandings of the Proficient level. Students performing at the borderline of the Proficient level can be expected in most situations[1] to be able to demonstrate knowledge and skills such as in the following examples, as evidence of just barely satisfactory understanding and use of the NGSS Standards:

- Use an understanding of the periodic table to predict one or two relative properties of elements, plan and conduct an investigation of phenomena related to one or two bulk scale properties of substances and explain how these relate to the strength of electrical forces between particles, explain phenomena about the release or absorption of energy by a chemical system by developing a model to show changes in total bond energy, use one or two pieces of evidence to explain phenomena using an understanding of how changes in temperature or concentration affect reaction rate, explain phenomena about chemical systems at equilibrium using an understanding of how one or two of the conditions on the system could be changed to produce more products, use one or two mathematical representations to support claims that mass is conserved during a chemical reaction, and develop one or two models to describe phenomena using an understanding of the changes in the nucleus of an atom and the energy released during fission, fusion, and radioactive decay. (PS1)

- Analyze and use one or two pieces of data from phenomena to show that f = ma; use one or two mathematical representations of phenomena and an understanding of momentum to support the claim that the total momentum of a system is conserved when there is no net force on the system; apply one or two scientific and engineering ideas to design, evaluate, and refine a device that minimizes force on an object during a collision; use mathematical representations of Newton's law of gravitation and Coulomb's law to describe and make predictions about familiar phenomena using an understanding of gravitational and electrostatic forces between objects; plan and conduct an investigation of phenomena to produce one or two pieces of evidence that prove an electric current produces a magnetic field and that a changing magnetic field produces electric current; and communicate one or two pieces of information about phenomena using an understanding of how the molecular structure of a material relates to its macroscopic properties and makes the material well suited for particular uses. (PS2)

- Create a computational model of phenomena to calculate changes in energy of a system when energy flows into and out of the system is known; develop and use one or two models to explain phenomena using an understanding of how energy at the macroscopic scale can be accounted for at the microscopic scale in energy associated with particle motion and relative position; design, build, and refine a device that converts one form of energy into another; plan and conduct an investigation of phenomena to provide one or two pieces of evidence using an understanding that when two components of different temperatures are combined within a closed system, both components eventually have the same temperature; and develop and use a model to explain phenomena related to the forces and the changes in energy between two objects interacting through electric or magnetic fields. (PS3)

- Explain phenomena using an understanding of one or two mathematical representations regarding relationships among frequency, wavelength, and speed of waves in various media; evaluate one or two questions about phenomena using an understanding of the advantages of using digital transmission and storage of information; use one or two phenomena to evaluate claims that electromagnetic radiation can be described using a wave or particle model; in the context of phenomena, evaluate one or two claims about the effects that different frequencies of electromagnetic radiation have on matter; and communicate technical information about phenomena using an understanding of how one or two specific technological devices use the principles of wave behavior and wave interactions to transmit and capture information and energy. (PS4)

- Use evidence to explain phenomena using an understanding of how the structure of DNA determines the structure of proteins and how proteins carry out the functions of life through specialized cells; develop and use a model to describe phenomena using an understanding of the organization of interacting systems within multicellular organisms; plan and conduct an investigation to provide evidence about phenomena that show that feedback mechanisms maintain homeostasis; use a model to describe phenomena using an understanding of how cell division and differentiation help produce and maintain complex organisms; use a model to describe phenomena using an understanding of how photosynthesis transforms light energy into stored chemical energy; use evidence to construct and revise an explanation about phenomena using an understanding of how carbon, hydrogen, and oxygen from sugar molecules combine with other elements to form amino acids and/or other large carbon-based molecules; and use a model to describe phenomena using an understanding that cellular respiration is a chemical process that breaks the bonds in food and oxygen molecules and forms bonds in new compounds, which results in a net transfer of energy. (LS1)

- Use mathematical or computational representations to support explanations of phenomena

using an understanding of factors that affect carrying capacity of ecosystems at different scales; use one or two pieces of evidence and one or two mathematical representations to support and revise explanations of phenomena using an understanding of factors affecting biodiversity and populations in ecosystems of different scales and to support claims for the cycling of matter and flow of energy among organisms in an ecosystem; use evidence to construct or revise an explanation of phenomena using an understanding of the cycling of matter and flow of energy in aerobic and anaerobic conditions; develop a model to describe phenomena using an understanding of the role of photosynthesis and cellular respiration in the cycling of carbon among Earth's spheres; evaluate one or two claims, pieces of evidence, and reasoning about phenomena involving complex interactions in ecosystems using an understanding that these interactions maintain relatively consistent numbers and types of organisms under stable conditions, but changing conditions may result in a new ecosystem; design, evaluate, and refine a solution for reducing impacts of human activities on the environment or biodiversity; and evaluate one or two pieces of evidence about phenomena using an understanding of the role of group behavior on individual and species' chances to survive and reproduce. (LS2)

- Ask one or two questions about phenomena to clarify relationships about the role of DNA in chromosomes in coding the instructions for traits passed from parents to offspring; use one or two pieces of evidence to make and defend a claim about phenomena using an understanding that inheritable genetic variations may result from new genetic combinations through meiosis, viable errors during replication, and/or mutations caused by environmental factors; and apply one or two concepts of statistics and probability to explain phenomena using an understanding of the variation and distribution of expressed traits in a population. (LS3)

- Communicate one or two pieces of scientific information about phenomena using an understanding that common ancestry and biological evolution are supported by multiple lines of empirical evidence; use one or two pieces of evidence to construct an explanation about phenomena using an understanding that the process of evolution primarily results from four factors: the potential for a species to increase in number, the heritable genetic variation of individuals in a species due to mutation and sexual reproduction, competition for limited resources, and the proliferation of those organisms that are better able to survive and reproduce in the environment; apply one or two concepts of statistics and probability to support explanations of phenomena using an understanding that organisms with an advantageous heritable trait tend to increase in proportion to organisms lacking the trait; use one or two pieces of evidence to construct an explanation of phenomena using an understanding of how natural selection leads to adaptation of populations; evaluate one or two pieces of evidence supporting claims about phenomena using an understanding that changes in environmental conditions may result in: increases in numbers of individuals of some species, the emergence of new species over time, and the extinction of other species; and create or revise a simulation to test a solution to mitigate adverse impacts of human activity on biodiversity. (LS4)

- Use one or two pieces of evidence to develop a model to describe phenomena using an understanding of the life span of the Sun and the role of nuclear fusion in the Sun's core to release energy that reaches Earth in the form of radiation; use one or two pieces of evidence and an understanding of the phenomena of light spectra, motion of distant galaxies, and composition of matter in the universe to construct an explanation of the big bang theory; communicate one or two scientific ideas about phenomena using an understanding of the way stars produce different elements over their life cycles; use mathematical or computational representations of phenomena to predict the motion of orbiting objects in the solar system;

evaluate one or two pieces of evidence of phenomena using an understanding of past and current movements of continental and oceanic crust and the theory of plate tectonics to explain the ages of crustal rocks; and apply scientific reasoning and an understanding of one or two pieces of evidence from ancient Earth materials, meteorites, and other planetary surfaces to describe phenomena about Earth's formation and early history. (ESS1)

- Develop a model to describe phenomena using an understanding of how Earth's internal and surface processes operate at different spatial and temporal scales to form continental and ocean floor features; analyze one type of geoscience data about phenomena to make a claim that one change to Earth's surface can create feedback that causes changes to other Earth systems; use one or two pieces of evidence to develop a model of Earth's interior to describe phenomena using an understanding of the cycling of matter by thermal convection; use a model to describe phenomena using an understanding of how variations in the flow of energy into and out of Earth's systems result in changes in climate; plan and conduct an investigation of phenomena related to the properties of water using an understanding of water's effects on Earth materials and surface processes; develop a quantitative model to describe phenomena using an understanding of the cycling of carbon among the hydrosphere, atmosphere, geosphere, and biosphere; and use one or two pieces of evidence to construct an argument about phenomena using an understanding of the simultaneous coevolution of Earth's systems and life on Earth. (ESS2)

- Use one or two pieces of evidence to construct an explanation about phenomena using an understanding of how the availability of natural resources, occurrence of natural hazards, and changes in climate have influenced human activity; using an understanding of cost-benefit ratios, evaluate two competing design solutions for developing, managing, and utilizing energy and mineral resources; create a computational simulation of phenomena to show the relationships among management of natural resources, the sustainability of human populations, and biodiversity; using an understanding of human impacts on natural systems, evaluate or refine a technological solution that reduces these impacts; analyze one or two pieces of geoscience data and one or two global climate models of phenomena to make a forecast of the current rate of climate change and associated future impacts to Earth systems; and use a computational representation to describe phenomena using an understanding of the relationships among Earth systems and how those relationships are modified due to human activity. (ESS3)

- Analyze a major global challenge to specify one or two qualitative and quantitative criteria and constraints for solutions that account for one or two societal needs and wants; design an engineering solution to a complex real-world problem by breaking it down into smaller, more manageable problems; evaluate a solution to a complex real-world problem based on prioritized criteria and trade-offs that account for a range of constraints, including cost, safety, reliability, and aesthetics, as well as social, cultural, and environmental impacts; and use a computer simulation to model the impact of two proposed solutions to a complex real-world problem with numerous criteria and constraints on interactions within and between systems relevant to the problem. (ETS1)

Nearing Proficiency

Students at the borderline of the **Nearing Proficiency** level in most situations[3] can demonstrate

---

[3] "Most situations" refers to the following. Students at level X can be expected to be able to   demonstrate knowledge and skills for most but not all:

evidence of partial understanding and use of all three dimensions (science and engineering practices, crosscutting concepts, and disciplinary core ideas) to make sense of phenomena and/or to design solutions to problems in the physical, life, and Earth and space sciences. They also sometimes demonstrate the skills and understandings at the Novice level rather than the skills and understandings of the Nearing Proficient level. Students performing at the borderline of the Nearing Proficiency level can be expected in most situations[1] to be able to demonstrate knowledge and skills such as in the following examples, as evidence of just barely partial understanding and use of NGSS Standards:

- Use a partial understanding of the periodic table to predict one relative property of elements, conduct an investigation of phenomena related to bulk scale properties of substances or construct a partial explanation of how these relate to the strength of electrical forces between particles, describe phenomena about the release or absorption of energy by a chemical system, explain phenomena using a partial understanding of how changes in temperature or concentration affect reaction rate, explain phenomena about chemical systems at equilibrium using a partial understanding of how one condition on the system could be changed to produce more products, use a mathematical representation to support a claim that mass is conserved during a chemical reaction, and use models to describe phenomena using a partial understanding of the changes in the nucleus of an atom or the energy released during fission, fusion, or radioactive decay. (PS1)

- Use data from phenomena to show that f = ma, use a mathematical representation of a phenomenon and a partial understanding of momentum to support the claim that the total momentum of a system is conserved, apply a scientific or an engineering idea to design a device that minimizes force on an object during a collision, use mathematical representations of Newton's law of gravitation or Coulomb's law to describe phenomena using a partial understanding of gravitational and electrostatic forces between objects, conduct an investigation of phenomena to produce evidence that an electric current produces a magnetic field or that a changing magnetic field produces electric current, and communicate information about phenomena using a partial understanding of how the molecular structure of a material relates to its macroscopic properties or makes the material well suited for particular uses. (PS2)

- Use a computational model of phenomena to calculate changes in energy of a system when energy flows into and out of the system is known; use models to explain phenomena using a partial understanding of how energy at the macroscopic scale can be accounted for at the microscopic scale in energy associated with particle motion or relative position; design a device that converts one form of energy into another; conduct an investigation of phenomena that provides evidence using a partial understanding that when two components of different temperatures are combined within a closed system, both components eventually have the same temperature; and use a model to explain phenomena related to the forces or the changes in energy between two objects interacting through electric or magnetic fields. (PS3)

- Explain phenomena using an understanding of a mathematical representation regarding one relationship among frequency, wavelength, and speed of waves in various media; ask questions about phenomena using a partial understanding of the advantages of using digital transmission and storage of information; use one phenomenon to support a claim that electromagnetic radiation may be described using a wave or particle model; make a claim about phenomena related to the effects that different frequencies of electromagnetic radiation have on matter; and communicate information about phenomena using a partial

understanding of how specific technological devices use the principles of wave behavior or wave interactions to transmit or capture information and energy. (PS4)

- Use evidence to explain phenomena using a partial understanding of how the structure of DNA determines the structure of proteins; use a model to describe phenomena using a partial understanding of the organization of interacting systems within multicellular organisms; conduct an investigation to provide evidence about phenomena that show that some feedback mechanisms maintain homeostasis; use a model to describe phenomena using a partial understanding of how cell division or differentiation helps produce or maintain a complex organism; use a model to describe phenomena using a partial understanding of how photosynthesis transforms light energy into stored chemical energy; describe that carbon, hydrogen, and oxygen from sugar molecules combine with other elements to form amino acids or other large carbon-based molecules; and use a model to describe phenomena using a partial understanding that cellular respiration is a chemical process that breaks the bonds in food or oxygen molecules, forms bonds in new compounds, or results in a net transfer of energy. (LS1)

- Use mathematical or computational representations to support explanations of phenomena using a partial understanding of factors that affect carrying capacity of ecosystems; use evidence or mathematical representations to describe factors affecting biodiversity or populations in ecosystems and to support a claim for the cycling of matter or flow of energy among organisms in an ecosystem; use evidence to construct an explanation of phenomena using a partial understanding of the cycling of matter and flow of energy in aerobic or anaerobic conditions; use a model to describe phenomena using an understanding of the role of photosynthesis or cellular respiration in the cycling of carbon; evaluate a claim about a phenomenon involving interactions in ecosystems using a partial understanding that these interactions maintain relatively consistent numbers and types of organisms under stable conditions, but changing conditions may result in a new ecosystem; identify a solution for reducing impacts of human activities on the environment or biodiversity; and use evidence to describe the role of group behavior on individual and species' chances to survive and reproduce. (LS2)

- Ask a question about a phenomenon to clarify relationships about the role of DNA in chromosomes in coding the instructions for traits passed from parents to offspring; make a claim about phenomena using a partial understanding that inheritable genetic variations may result from new genetic combinations through meiosis, viable errors during replication, or mutations caused by environmental factors; and describe phenomena using an understanding of the variation and distribution of expressed traits in a population. (LS3)

- Communicate scientific information about phenomena using a partial understanding that common ancestry and biological evolution are supported by empirical evidence; use evidence to construct an explanation about phenomena using a partial understanding that the process of evolution primarily results from one or two of the following factors: the potential for a species to increase in number, the heritable genetic variation of individuals in a species due to mutation and sexual reproduction, competition for limited resources, and the proliferation of those organisms that are better able to survive and reproduce in the environment; apply concepts of statistics or probability to explain phenomena using a partial understanding that organisms with an advantageous heritable trait tend to increase in proportion to organisms lacking the trait; explain phenomena using a partial understanding of how natural selection leads to adaptation of populations; use evidence to support claims about phenomena using a partial understanding that changes in environmental conditions may result in one or two of the following: increases in numbers of individuals of some species, the emergence of new species

over time, or the extinction of other species; and use a simulation to test a solution to mitigate adverse impacts of human activity on biodiversity. (LS4)

- Use evidence to develop a model to describe phenomena using a partial understanding of the life span of the Sun or the role of nuclear fusion in the Sun's core to release energy that reaches Earth in the form of radiation; use evidence and a partial understanding of the phenomena of light spectra, motion of distant galaxies, or composition of matter in the universe to construct an explanation of the big bang theory; communicate a scientific idea about a phenomenon using a partial understanding of the way stars produce different elements over their life cycles; use a mathematical or computational representation of phenomena to predict the motion of orbiting objects in the solar system; use evidence of phenomena using a partial understanding of past and current movements of continental and oceanic crust or the theory of plate tectonics to explain the ages of crustal rocks; and apply scientific reasoning or a partial understanding of evidence from ancient Earth materials, meteorites, and other planetary surfaces to describe phenomena about Earth's formation and early history. (ESS1)

- Use a model to describe phenomena using a partial understanding of how Earth's internal and surface processes operate to form continental and ocean floor features; use geoscience data about phenomena to make a claim that one change to Earth's surface can create feedback that causes changes to other Earth systems; develop a model of Earth's interior to describe phenomena using an understanding of the cycling of matter by thermal convection; use a model to describe phenomena using a partial understanding of how variations in the flow of energy into and out of Earth's systems result in changes in climate; conduct an investigation of phenomena related to the properties of water using a partial understanding of water's effects on Earth materials or surface processes; use a quantitative model to describe phenomena using a partial understanding of the cycling of carbon among the hydrosphere, atmosphere, geosphere, and biosphere; and use evidence to make a claim about phenomena using a partial understanding of the simultaneous coevolution of Earth's systems and life on Earth. (ESS2)

- Use evidence to construct an explanation about phenomena using a partial understanding of how the availability of natural resources, occurrence of natural hazards, or changes in climate have influenced human activity; using a partial understanding of cost-benefit ratios, evaluate a design solution for developing, managing, or utilizing energy or mineral resources; use a computational simulation of phenomena to show some of the relationships among management of natural resources, the sustainability of human populations, and biodiversity; using a partial understanding of human impacts on natural systems, describe a technological solution that reduces these impacts; use geoscience data or global climate models of phenomena to make a forecast of the current rate of climate change or associated future impacts to Earth systems; and use a computational representation to describe phenomena using a partial understanding of the relationships among Earth systems or how those relationships are modified due to human activity. (ESS3)

- Analyze a major global challenge to specify a qualitative or quantitative criterion or constraint for a solution that accounts for a societal need or want; describe one or two ways a complex real-world problem could be broken down into smaller, more manageable problems that could be solved through engineering; explain how a solution to a complex real-world problem meets required criteria or explain one or two trade-offs of the solution; and use a computer simulation to model the impact of a proposed solution to a complex real-world problem with two or three criteria and constraints on interactions within or between systems relevant to the problem. (ETS1)

Novice

Students at the **Novice** level demonstrate evidence of emerging understanding and use of all three dimensions (science and engineering practices, crosscutting concepts, and disciplinary core ideas) to make sense of phenomena and/or to design solutions to problems in the physical, life, and Earth and space sciences.

- Combinations of disciplinary core ideas, practices, and crosscutting concepts
- Science phenomena, in which students are required to apply their knowledge of science content and mastery of crosscutting concepts and practices.

Test Specifications

Bureau of Indian Education (BIE) Science Assessment

Bureau of Indian Education

Purpose of the BIE Science Assessment

The BIE Science Assessment is the Bureau of Indian Education's summative assessment for Science, administered at the end of grades 5, 8, and 11. The BIE Science Assessment measures students' proficiency with the NGSS Science Standards. As the BIE Science Assessment is a single measure at the end of a grade band, interpretations and uses of BIE Science Assessment scores should be supplemented with additional measures, including information from classroom summative and formative assessments in science.

Claims/Score Interpretation and Use Statements

The BIE Science Assessment is designed to measure whether students are on track to be ready for college or career by showing they have mastered the NGSS Science Standards. The standards require integration of Science and Engineering Practices, Disciplinary Core Ideas, and Crosscutting Concepts to explain phenomena and solve problems.

In addition to overall scale score, student performance on three science subdomains is reported:
- Practices and Crosscutting Concepts in Physical Sciences
- Practices and Crosscutting Concepts in Life Sciences
- Practices and Crosscutting Concepts in Earth and Space Sciences

Test Specifications – Test Design

Assessable Standards

The BIE Science Assessment assesses the NGSS Science Standards as follows:
- Grade 5 test: All standards in grades 3, 4, and 5.
- Grade 8 test: All standards in the middle school grade band (6-8).
- Grade 11 test: All standards in the high school grade band (9-12).

Test Design

The BIE Science Assessment is administered in three sessions. The test is administered online as a computer-based test (CBT).

Online accommodations are available for the CBT. Paper, large-print, and Braille test forms are also provided.

No calculator is provided for the BIE Science Assessment, as no items require calculator use. A periodic table will be provided as a reference for high school (Grade 11).

The types of items on the BIE Science Assessment are item clusters (CL), 2-point machine-scored standalone items (MS-2), and 4-point open-ended standalone items (OE). Additional item type descriptions and sample items can be found in the item specifications section of this document.

Both core operational items (which count for a student's score) and a small number of matrix field test items (which are try-out items that do not count for a student's score) are included on the BIE Science Assessment.

The total numbers of test items, points, and estimated testing time for the BIE Science Assessment are shown in the following tables.

Student Testing Experience							
Grade 5	**Cluster/Passage Items**			**Standalone Items**		**Total Items**	**Total Points**
	Stim/Psg	**MS-1**	**MS-2**	**MS-2**	**OE**		
Core Operational Items	6	12	12	8	3	35	64
Matrix Field Test Items	2	4	4	4	1	13	24
Total Student Experience	8	16	16	12	4	48	88
				Estimated Testing Time (min)			**150**

Student Testing Experience							
Grade 8	**Cluster/Passage Items**			**Standalone Items**		**Total Items**	**Total Points**
	Stim/Psg	**MS-1**	**MS-2**	**MS-2**	**OE**		
Core Operational Items	6	12	12	8	3	35	64
Matrix Field Test Items	2	4	4	4	1	13	24
Total Student Experience	8	16	16	12	4	48	88
				Estimated Testing Time (min)			**150**

Student Testing Experience							
Grade 11	**Cluster/Passage Items**			**Standalone Items**		**Total Items**	**Total Points**
	Stim/Psg	**MS-1**	**MS-2**	**MS-2**	**OE**		
Core Operational Items	6	12	12	10	3	37	68
Matrix Field Test Items	2	4	4	5	1	14	26
Total Student Experience	8	16	16	15	4	51	94
				Estimated Testing Time (min)			**165**

Practice Test

A full-length practice test mirroring the operational test design is available for each grade. The practice tests and supporting materials can be accessed on the website.

Test Specifications – Reporting Categories

The reporting categories for BIE Science Assessment are based on the three content domains. Percentages for the distribution of operational (core) test points for each of the reporting categories reflect the distribution in the standards, so as not to over- or underrepresent content.

Based on this representativeness, the fourth content domain of Engineering, Technology, and Applications of Science is not reported as a subscore (as there are very few standards out of the total in each grade band). Items coded to these standards <u>do</u> count toward total test score.

Reporting Categories, Grade 5 BIE Science Assessment					
Reporting Category	**Typical Number of Clusters**	**Typical Number of Standalone MS-2**	**Typical Number of Standalone OE**	**Number of Core Points**	**Percent of Core Points (+/- 4%)**
Practices and Crosscutting Concepts in Physical Sciences	*2*	*4-6*	*1*	*24-28*	40%
Practices and Crosscutting Concepts in Life Sciences	*2*	*1-3*	*1*	*18-22*	30%
Practices and Crosscutting Concepts in Earth and Space Sciences	*2*	*1-3*	*1*	*18-22*	30%

Reporting Categories, Grade 8 BIE Science Assessment					
Reporting Category	**Typical Number of Clusters**	**Typical Number of Standalone MS-2**	**Typical Number of Standalone OE**	**Number of Core Points**	**Percent of Core Points (+/- 4%)**
Practices and Crosscutting Concepts in Physical Sciences	*2*	*2-4*	*1*	*20-24*	35%
Practices and Crosscutting Concepts in Life Sciences	*2*	*2-4*	*1*	*20-24*	35%
Practices and Crosscutting Concepts in Earth and Space Sciences	*2*	*1-3*	*1*	*18-22*	30%

Reporting Categories, Grade 11 BIE Science Assessment					
Reporting Category	**Typical Number of Clusters**	**Typical Number of Standalone MS-2**	**Typical Number of Standalone OE**	**Number of Core Points**	**Percent of Core Points (+/- 4%)**
Practices and Crosscutting Concepts in Physical Sciences	*2*	*3-5*	*1*	*22-26*	35%
Practices and Crosscutting Concepts in Life Sciences	*2*	*3-5*	*1*	*22-26*	35%
Practices and Crosscutting Concepts in Earth and Space Sciences	*2*	*1-3*	*1*	*18-22*	30%

Test Specifications – Cognitive Complexity

Because the items on the BIE Science Assessment are NGSS-aligned, the cognitive complexity of items is evaluated with a different framework than Depth of Knowledge.

For the items on the BIE Science Assessment, four indicators are used to classify the cognitive complexity of each item: stimulus, science and engineering practice, disciplinary core idea, and crosscutting concept. For each indicator, the classification in terms of high, medium, or low complexity is based on how the students are using the indicator to respond to the item – specifically, to what degree does students' engagement with the indicator contribute to the level of sense-making required by the item.

On the BIE Science Assessment, after summing the operational (core) test points at each cognitive complexity level across all four indicators, **at least 10% of the points should be high cognitive complexity** and **no more than 35% of the points should be low cognitive complexity**.

The descriptors for each indicator at the three complexity levels (high, medium, low) are presented in the following tables.

	STIMULUS
High	- Phenomenon is novel, complex, and/or unfamiliar to students - Students must synthesize multiple pieces of information and do a significant amount of "figuring out" to make sense of the phenomenon
Medium	- Phenomenon is somewhat novel, but may be analogous to what many students are familiar with - Students must use multiple pieces of information and do an intermediate amount of "figuring out" to make sense of the phenomenon
Low	- Phenomenon is familiar and/or more straightforward for students - Students only need to use simple/straightforward information, and/or a single piece of information, and do a minimal amount of "figuring out" to answer the question or contribute to making sense of the phenomenon

	SEP (SCIENCE AND ENGINEERING PRACTICE)
High	- Students must apply the SEP, or multiple SEPs, in a sophisticated way to make sense of the phenomenon (e.g., synthesis to perform more connections, steps, combination of SEP elements, such as having to combine data, produce a new graph or model as evidence, etc.) - Often little to no scaffolding that helps students apply the SEP
Medium	- Students must apply the SEP to make sense of the phenomenon - Typically some scaffolding that helps students apply the SEP
Low	- Students only need to use the SEP in a simple, mechanical way to answer the question or contribute to making sense of the phenomenon. - Often a large amount of scaffolding that helps students apply the SEP

	DCI (DISCIPLINARY CORE IDEA)

High	- Students must apply and connect DCIs in a sophisticated way to make sense of the phenomenon, i.e.,
	  o application of science ideas (often multiple, grade-band appropriate ideas) in unique ways or new combinations
	  o knowledge transfer to construct new understanding, make sense of novel phenomena.
	- Often little to no scaffolding that helps students apply the DCI

Medium	- Students must apply or reason with the DCI(s) to make sense of the phenomenon - Typically, some scaffolding that helps students apply the DCI
Low	- Students use the DCI in a simple, straightforward way (i.e., little to no application or reasoning) to answer the question or contribute to making sense of the phenomenon - Often a large amount of scaffolding that helps students apply the DCI

	CCC (CROSSCUTTING CONCEPT)
High	- Students must apply the CCC in an in-depth way to expand thinking and make non-typical connections to make sense of the phenomenon
Medium	- Students must use the CCC as specified by the CCC sub-bullet detail to make sense of the phenomenon
Low	- Students only use the CCC in a general way to answer the question or contribute to making sense of the phenomenon

Test Specifications – Fairness

Fairness is defined as the extent to which the test scores are valid for different groups of test takers. Consideration of universal design, bias, and sensitivity guidelines support the construction of fair, valid assessments.

Universal Design for Assessments

The concept of Universal Design for Assessments focuses on developing content and assessments that reach the widest population of students possible. Stimuli and items on the BIE Science Assessment are designed to simply and clearly present tasks and to provide maximum readability, comprehensibility, and legibility. The seven elements of Universal Design for Assessments are based on the original UDL guiding principles:

Universal Design for Assessments

Principle	Explanation
Inclusive Assessment Population	Tests designed for state, district, or school accountability. must include every student except those in the alternate assessment, and this is reflected in assessment design and field-testing procedures.
Precisely Defined Constructs	The specific constructs tested must be clearly defined so that all construct-irrelevant cognitive, sensory, emotional, and physical barriers are removed.
Accessible, Non-Biased Items	Accessibility is built into items from the beginning, and bias review procedures ensure that quality is retained in all items.
Amenable to Accommodations	Test design facilitates the use of needed accommodations (e.g., all items can be brailled).
Simple, Clear, and Intuitive Instructions and Procedures	All instructions and procedures are simple, clear, and presented in understandable language.
Maximum Readability and Comprehensibility	A variety of readability and plain language guidelines are followed (e.g., sentence length and number of difficult words kept to a minimum) for readable and comprehensible text.
Maximum Legibility	Characteristics that ensure easy decipherability are applied to text, tables, figures, and illustrations, and to response formats.

Bias

The concept of Bias is defined as the presence of some characteristic of an item that results in differential performance for two individuals of the same ability but from different ethnic, sex, cultural, or religious groups.

Bias can occur whenever content offends or disadvantages a student or group of students due to gender, race, regional background, socioeconomic status, or any other such classification.

Test developers take care to craft content in a way that does not misrepresent specific groups or rest on assumptions made about specific groups, that in turn could negatively impact how students interpret content.

- Stimulus and item content on the BIE Science Assessment should not present stereotypes or unfair representations of gender, race, ethnicity, disability, culture, or religion.
- Stimulus and item content on the BIE Science Assessment should not depend on overly-experiential information such as knowledge of technology, consumer goods, pop culture, geographic locations, or sports and extracurricular activities. While these topics are not completely excluded from use, care must be taken to ensure that the items are presented in a way that does not require a level of knowledge that would not be held by all students.

Sensitivity

Sensitivity refers to the presence of content that is contrary to the acceptable norms of the students, educators, parents, or other members of the community that may interact with the assessment. Sensitive subject matter can impact student performance or attitudes toward testing, and hence, their test scores.

Consideration of bias and sensitivity issues is very important when developing content for an assessment. Test developers must ensure that stimuli and items are free of content that will negatively affect a student's performance not because of what the student knows and can do but because the content evokes an emotional response from that student (or is in some other way distracting to the student).

Subjects/contexts that are likely to prompt emotional distress on the part of students cannot be used on the BIE Science Assessment (e.g., war, violence, human death or debilitating disease, animal-based medical research). Careful judgment should be applied to PEs that cover topics that may be considered controversial by some groups (e.g., evolution examples, population dynamics including death/extinction, environmental impact). Those PEs represent content knowledge to be assessed, but the assessment must be done in a sensitive, unbiased way.

Stimulus Specifications

All items for the BIE Science Assessment have a stimulus. For clusters, all items in the set are associated with a common stimulus that presents a science phenomenon or engineering design problem. For standalone items (MS-2, OE), the item includes a lead stimulus that provides a specific science phenomenon or engineering design problem, or context thereof. By phenomenon, we mean something observable that happens in the real world, whether natural or man-made. By engineering design problem, we mean a personal or societal need or want.

Specifications for Cluster Stimuli

1. The stimulus must present a single, rich science phenomenon or engineering design problem aligned to the PEs.

2. The stimulus may present any variety of elements to provide the necessary information to support sense-making (via the items) around the phenomenon or problem: text paragraphs, passages, graphs, data tables, models, drawings, etc.
3. The stimulus must be rich enough to support the development of enough items for the cluster, in the context of a storyline (sequence of sense-making) around the phenomenon or problem using the DCIs, SEPs, and CCCs of the targeted PEs.
4. All information in the stimulus should be necessary, but not conceptually sufficient, for students to respond (i.e., students must also use their own knowledge of the constructs in the PE(s) to answer the items, rather than simply identify given information).
5. The stimulus phenomenon or problem must be grade-appropriate, engaging, and relevant for students at that grade level.
6. The stimulus should adhere to the specifications in the following table regarding length, wording, and complexity.

Stimulus characteristic	Elementary School (Grades 3-5)	Middle School (Grades 6-8)	High School (Grades 9-12)
Text word count**	100-300 words	100-400 words	100-400 words
Vocabulary level *(excluding science content vocabulary)*	Grade 3	Grade 5 maximum	Grade 8 maximum
Readability/Lexile maximum	820L (Gr 3)	1010L (Gr 5)	1185L (Gr 8)
Qualitative text characteristics	Simple sentence structures, clear/uncomplicated graphics, lower vocabulary demands, use of only essential science vocabulary.	Slight mix of simple and more complex phrasing and sentence structure, average to moderately complex graphics, average vocabulary demands.	Mix of simple and more complex phrasing and sentence structure, average to moderately complex graphics, average vocabulary demands.

**Count should balance text and graphic load – in a stimulus with more and/or complex graphics, the word count should be lower; in a stimulus with few and/or very simple graphics, the word count could, if needed, be at the higher end of range.

Specifications for Standalone Item Stimuli

1. MS-2 items: The stimulus must present a hook or driving reason for the question being asked, and it must set a phenomenon- or problem-based context, aligned to the PE, for the item. The stimulus will typically not be as extensive as a stimulus for an item cluster.
2. OE items: The stimulus must present a hook or driving reason for the question being asked, and it must include a phenomena or problem, aligned to the PE, to drive the item. The stimulus for open-ended items will typically be more concise than for item clusters but more detailed than for MS-2 standalone items.

Item Specifications

Alignment

The items on the BIE Science Assessment are aligned to the NGSS Science Standards.

Each item is aligned to a performance expectation (PE) as well as dimensions of the performance expectation. All items must have either 2-dimensional or 3-dimensional alignment.

Item Types

The types of items on the BIE Science Assessment are item clusters, 2-point machine-scored standalone items (MS-2), and 4-point open-ended standalone items (OE):
- An item cluster is a set of items all associated with a common stimulus. Clusters contain four items. These items may be multiple choice, multiple select, or technology-enhanced, with two of the items being worth 1 point and two of the items being worth 2 points. The clusters typically align to two PEs, and all clusters measure all three dimensions of the PEs being assessed.
- Standalone MS-2 items are worth 2 points. These items have two parts (Part a and Part b) for students to answer, and 0, 1, or 2 points total can be earned across Part a and Part b. These items may be multiple choice, multiple select, or technology-enhanced (e.g., drag-and-drop, hot spot, drop-down selections).
- Open-ended items are worth 4 points. These items require students to write an extended response to a prompt. The prompt may be a single prompt, or more typically, the items are written with multiple, scaffolded parts for students to respond to. These items are hand-scored, with scorers using a rubric and scoring notes to evaluate responses on a scale from 0–4.

Samples of each of these item types are included on the following pages.

Clusters: Clusters are a set of 4 items all associated with an introductory passage, or "stimulus."
- The stimulus typically contains both text and graphics such as diagrams, tables, or graphs. An example stimulus from the *grade 5 practice test* is on the next page. The items associated with the cluster assess two Physical Sciences PEs:
  - 5-PS1-3: Make observations and measurements to identify materials based on their properties.
    SEP: Planning and Carrying Out Investigations
    DCI: PS1.A: Structure and Properties of Matter
    CCC: Scale, Proportion, and Quantity
  - 5-PS1-4: Conduct an investigation to determine whether the mixing of two or more substances results in new substances.
    SEP: Planning and Carrying Out Investigations
    DCI: PS1.B: Chemical Reactions
    CCC: Cause and Effect

Read the information. Then answer the questions that follow.

Investigating Gas Production

In class, a teacher demonstrates a chemical reaction by mixing vinegar and baking soda to produce bubbles of gas. Eliana wonders whether mixing other substances could also produce a gas. She decides to test different combinations of sugar, water, vinegar, and baking soda.

Some properties of these substances are shown in the table.

Properties of Substances

Substance	Color	Solid or Liquid	Attracted to a Magnet	Conducts Electricity
Sugar	White	Solid	No	No
Water	Clear	Liquid	No	Yes
Vinegar	Clear	Liquid	No	Yes
Baking soda	White	Solid	No	No

Investigation 1

Eliana mixes a small amount of each liquid and solid in a bowl and observes whether bubbles of gas are produced. Her observations are shown in the table.

Investigation 1 Observations

Liquid Used	Solid Used	Gas Produced
Water	Sugar	No
Water	Baking soda	No
Vinegar	Sugar	No
Vinegar	Baking soda	Yes

Investigation 2

Next, Eliana wonders whether changing the amount of baking soda would change the amount of gas produced. To investigate, she follows these steps:

1. Record the mass of a balloon.
2. Pour 50 milliliters of vinegar into a bottle.
3. Put 5 milliliters of baking soda inside the balloon. Hold the balloon so that the baking soda stays inside the balloon and attach the open end of the balloon to the top of the bottle.
4. Lift the balloon so that the baking soda falls into the bottle with vinegar.
5. Wait one minute.
6. Carefully remove the balloon from the bottle without allowing any gas to escape.
7. Measure the mass of the balloon filled with gas.
8. Calculate the mass of gas produced by subtracting the mass of the balloon from the mass of the balloon filled with gas.
9. Repeat steps 1–8 until three trials have been completed.
10. Repeat steps 1–9 with 10 milliliters and 15 milliliters of baking soda.

The results of one trial are shown in the diagram.

Gas in a Balloon

Balloon filled with gas

Bottle with baking soda and vinegar mixture

Table top
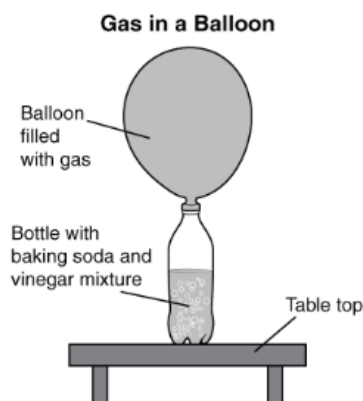
Eliana's data are shown in the table.

Investigation 2 Data

Amount of Baking Soda (milliliters)	Mass of Gas Produced (grams)			Average Mass of Gas Produced (grams)
	Trial 1	Trial 2	Trial 3	
5	1.0	0.8	1.2	1.0
10	1.5	1.9	1.4	1.6
15	2.4	1.9	2.6	2.3

- Two of the items in the cluster are machine-scored items worth 1 point each. These items may be multiple-choice, multi-select, or technology-enhanced items (e.g., drag-and-drop, hot spot, drop-down selections).

Which evidence from the investigations supports the claim that mixing vinegar and baking soda produces a new substance?

Hide All

👁 (A) A gas is produced when a liquid and a solid are mixed.

👁 (B) When a liquid and solid are mixed, the mass does not change.

👁 (C) The properties of substances stay the same when the substances are mixed.

👁 (D) Different amounts of baking soda can be mixed with the same amount of vinegar.

MS-1 cluster item, grade 5 practice test, aligned to PE 5-PS1-4: Conduct an investigation to determine whether the mixing of two or more substances results in new substances. The dimensions for the PE are SEP: Planning and Carrying Out Investigations; DCI: PS1.B: Chemical Reactions; CCC: Cause and Effect. This particular MS-1 item in the cluster assesses the DCI and CCC dimensions.

- The other two items in the cluster are machine-scored items worth 2 points each. These items have two parts, with Part a worth 1 point and Part b also worth 1 point. Each part of the item may be presented as multiple-choice, multi-select, or technology-enhanced (e.g., drag-and-drop, hot spot, drop-down selections).

This question has two parts. Be sure to answer both parts of the question.

Part a

Eliana claims that when baking soda and vinegar are mixed, a new substance forms.

Select the phrase that describes an observation from investigation 2 that supports her claim.

After baking soda and vinegar are mixed,

-Select an Answer-   ▾ .

Part b

Which observation is evidence that the new substance inside the balloon is a gas?

Hide All

👁 Ⓐ The new substance filled the balloon.

👁 Ⓑ The new substance has more mass than the baking soda.

👁 Ⓒ The new substance takes up less space than the vinegar.

👁 Ⓓ The new substance increases as the baking soda increases.

MS-2 cluster item, grade 5 practice test, aligned to PE 5-PS1-4: Conduct an investigation to determine whether the mixing of two or more substances results in new substances. The dimensions for the PE are SEP: Planning and Carrying Out Investigations; DCI: PS1.B: Chemical Reactions; CCC: Cause and Effect. This particular MS-2 item in the cluster assesses the DCI and CCC dimensions.

- The entire cluster is worth a total of 6 points. The diagram below summarizes the structure of a cluster.
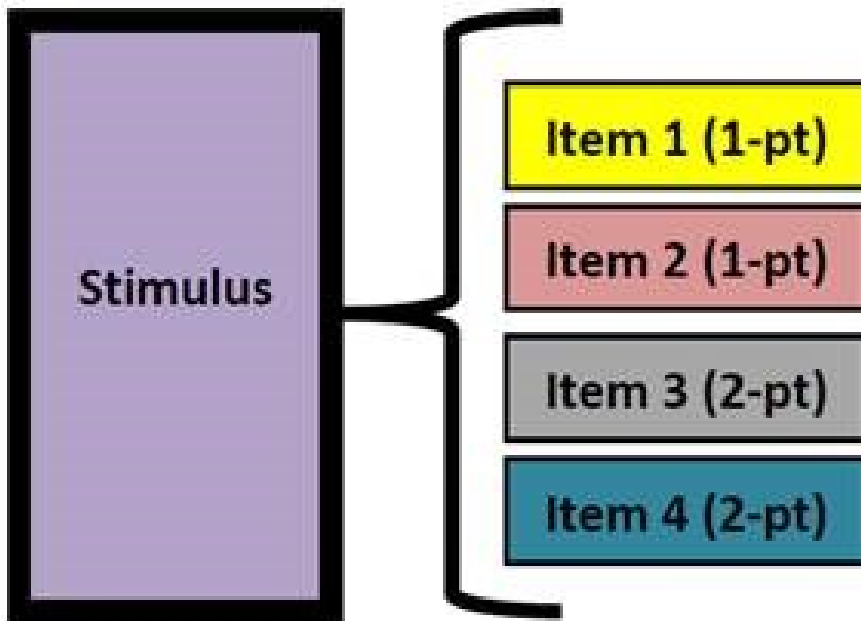
MS-2 Items: MS-2 items are standalone, or individual, machine-scored items.

- As in the cluster, the standalone MS-2 items are worth 2 points and have two parts, with Part a worth 1 point and Part b also worth 1 point. Each part of the item may be presented as multiple-choice, multi-select, or technology-enhanced (e.g., drag-and-drop, hot spot, drop-down selections).

This question has two parts. Be sure to answer both parts of the question.

Elaine is researching the characteristics of sound waves as she listens to music. She finds a diagram that shows two sound waves.

Sound Waves

Key
- - - - Wave 1
——— Wave 2
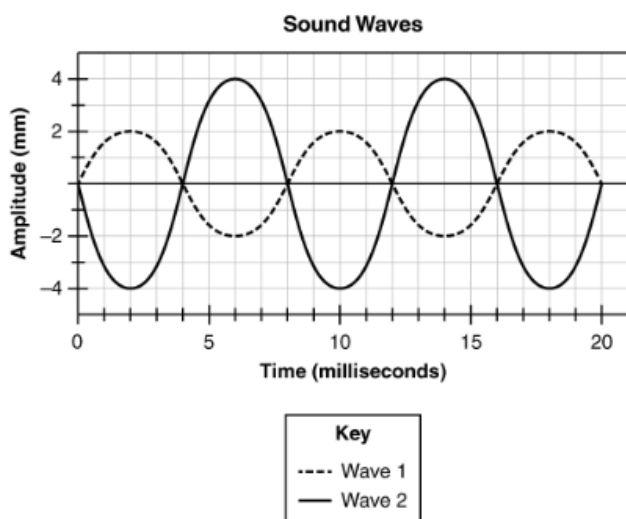
Part a
Select the number that describes the sound waves in the diagram.

Wave 1 has [-Select an Answer- ▾] the energy of Wave 2.

Part b
Elaine predicts that the energy in each sound wave would double if the frequency of the wave doubled.

Which statement describes her prediction?

Hide All

- Ⓐ Her prediction is correct because the energy of a wave is proportional to the wave's frequency.

- Ⓑ Her prediction is incorrect because the energy of a wave is proportional to the square of the wave's frequency.

- Ⓒ Her prediction is correct because the energy of a wave is proportional to speed and a wave with twice the frequency has twice the speed.

- Ⓓ Her prediction is incorrect because the energy of a wave is proportional to the wavelength and a wave with twice the frequency has half the wavelength.

MS-2 item, grade 8 practice test, aligned to PE MS-PS4-1: Use mathematical representations to describe a simple model for waves that includes how the amplitude of a wave is related to the energy in a wave. The dimensions for the PE are SEP: Using Mathematics and Computational Thinking; DCI: PS4.A: Wave Properties; CCC: Patterns. This particular MS-2 standalone item assesses the SEP, DCI, and CCC dimensions.

OE Items: OE, or open-ended, items are standalone items that require students to provide a written response to a prompt or question.

- The prompt or question may be a single prompt, or more typically, the item will be written with multiple, scaffolded parts for students to answer.
- The items are worth 4 points each and are hand-scored for 4, 3, 2, 1, or 0 points by trained scorers using a rubric and scoring notes.

This question has two parts. Be sure to answer both parts of the question.

Some students work at a local aquarium. One of their tasks is to care for mollusks and corals in ocean water in a tank at the aquarium. The students need to make sure that the ocean water has the right balance of calcium ions (Ca^{2+}) and carbonate ions (CO_3^{2-}) that the mollusks and corals need to build their shells and skeletons.

To do this, the students need to ensure that calcium and carbonate ions are continuously added to the ocean water in the tank. The students know that ocean water contains calcium carbonate, which naturally breaks down into calcium and carbonate ions. The equilibrium relationship between the components in the water is shown in the equation.

Equilibrium Equation

$$CO_2 + H_2O + CaCO_3 \rightleftarrows Ca^{2+} + 2\ H^+ + 2\ CO_3^{-2}$$
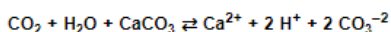
The students decide to test the equilibrium relationships in the equation. With ocean water as an input, the students remove calcium ions (Ca^{2+}) as the ions form in the water in the tank. The students observe that as they remove calcium ions, more calcium ions form in the tank. They realize that this is an example of Le Chatelier's principle that describes the equilibrium relationships in the water.

The people who work at the aquarium tell the students that ocean water contains carbon dioxide (CO_2) and that increasing amounts of CO_2 in ocean water can cause some of the calcium carbonate ($CaCO_3$) in the shells and skeletons of ocean organisms to dissolve.

The students want to solve this problem by decreasing the amount of carbon dioxide in ocean water.

a. Describe one way students could decrease the amount of CO_2 in ocean water by applying Le Chatelier's principle.

b. Describe one constraint on implementing the change described in Part (a).

OE item, grade 11 practice test, aligned to PE HS-PS1-6: Refine the design of a chemical system by specifying a change in conditions that would produce increased amounts of products at equilibrium. The dimensions for the PE are SEP: Constructing Explanations and Designing Solutions; DCI: PS1.B: Chemical Reactions and ETS1.C: Optimizing the Design Solution; CCC: Stability and Change. This particular OE item assesses the SEP, DCI, and CCC dimensions.

APPENDIX C
ACCOMMODATION FREQUENCIES

Only students who met the attemptedness rule (i.e., attempted 5 or more items) contributed to the frequencies in these tables.

Table C-1. Number of Students Taking BIE Science Assessment and Utilizing Accommodation(s) /Accessibility Feature(s), as a Function of Grade*

Accommodation/Accessibility Feature	Grades		
	5	8	11
AccAssistTech	--	--	6
AccReadAloudSelf	34	16	10
AccTactileGraphics	--	1	--
AccDirections	1	--	--
AccSpeechToText	31	16	4
AccPocketTranslator	1	--	--
AccComWordDictionary	1	15	--
AccHumanReaderENG	60	19	15
AccAssistDevice	2	--	5
AccHumanSigner	--	--	1
AccHumanScribe	16	6	--
AccWordPredictionEmbed	11	7	4
AccColorContrast	7	6	2
AccModeTesting	14	4	7
AccAltSetting	124	87	27
AccSCI_TTSENG	64	60	14
AccWordPrediction	6	3	--
AccPictureDictionary	1	1	--
AccHeadphones	34	28	4

Only students who met the attemptedness rule (i.e., attempted 5 or more items) contributed to the frequencies in these tables.

APPENDIX D
SCORER QUALIFICATION RATES

Table D-1 summarizes the qualification rates for the 2023 operational assessment for BIE Science. Rates of success during qualification varied. Multiple factors determine the success of a scorer during qualification. These include familiarity with the assessment, grade level, and variation of item types. Please note that not all scorers who failed Qual 1 attempted Qual 2.

Table D-1. Qualification Summary for BIE Science

Grade 5	661177 Qual 1	661177 Qual 2	Scorers Qualified 661177	666120 Qual 1	666120 Qual 2	Scorers Qualified 666120	697164 Qual 1	697164 Qual 2	Scorers Qualified 697164
Total Passed	19	4	23	29	3	32	15	5	20
Total Failed	4	0	0	3	0	0	5	0	0
Grade 8	663576 Qual 1	663576 Qual 2	Scorers Qualified 663576	697245 Qual 1	697245 Qual 2	Scorers Qualified 697245	717529 Qual 1	717529 Qual 2	Scorers Qualified 717529
Total Passed	33	5	38	21	2	23	27	5	32
Total Failed	9	4	4	6	4	4	7	2	2
Grade 11	666236 Qual 1	666236 Qual 2	Scorers Qualified 666236	710876 Qual 1	710876 Qual 2	Scorers Qualified 710876	735374 Qual 1	735374 Qual 2	Scorers Qualified 735374
Total Passed	15	2	17	18	1	19	18	1	19
Total Failed	3	1	1	1	0	0	3	2	2

APPENDIX E
CLASSICAL ITEM STATISTICS

Calculations are based on those students attempting 5 or more items on the form of the given BIE assessment. For 1-point items, the item-total correlation is the point-biserial. For 2 or more-point items, the item-total correlation is the point-polyserial.

All Classical Item Statistics listed in Appendix E were based on a national sample of students.

Table E-1. Classical Item Statistics for the Operational Items on Science Grade 5*

PsyItemNumber	Position	Item Type	Max Points	Item Mean	Item-Total Correlation
632724	1	MC	1	0.40	0.15
637807	2	MC	1	0.72	0.42
637951	3	MC	1	0.47	0.41
638656	4	MC	1	0.25	0.18
638658	5	MC	1	0.40	0.4
706138	6	MC	1	0.37	0.28
706149	7	MC	1	0.47	0.39
706792	8	MC	1	0.42	0.31
744445	9	MC	1	0.43	0.35
632837	10	OR	1	0.72	0.45
706801	11	OR	1	0.42	0.59
744451	12	OR	1	0.31	0.34
626442	13	OR	2	1.34	0.52
626478	14	OR	2	0.79	0.44
631577	15	OR	2	1.07	0.60
632426	16	OR	2	0.67	0.30
632454	17	OR	2	0.70	0.44
633906	18	OR	2	0.48	0.42
634127	19	OR	2	0.69	0.47
635879	20	OR	2	1.07	0.41
635886	21	OR	2	0.41	0.22
636189	22	OR	2	0.40	0.45
636211	23	OR	2	0.99	0.47
638354	24	OR	2	0.96	0.6
638526	25	OR	2	1.07	0.54
638558	26	OR	2	1.23	0.64
638639	27	OR	2	1.03	0.34
639474	28	OR	2	1.00	0.62
639510	29	OR	2	0.74	0.56
697027	30	OR	2	1.12	0.52
697044	31	OR	2	1.13	0.37
706119	32	OR	2	1.07	0.53
706135	33	OR	2	0.82	0.52
706765	34	OR	2	1.06	0.58
706847	35	OR	2	0.75	0.53
737916	36	OR	2	1.21	0.49
743165	37	OR	2	0.76	0.57
744455	38	OR	2	1.44	0.52
756457	39	OR	2	0.39	0.51
762758	40	OR	2	0.73	0.46
661177	41	OR	4	1.17	0.67
666120	42	OR	4	0.62	0.50
697164	43	OR	4	1.08	0.66

* Calculations are based on those students attempting 5 or more items on the form of the given BIE assessment. For 1-point items, the item-total correlation is the point-biserial. For 2 or more-point items, the item-total correlation is the point-polyserial.

Table E-2. Classical Item Statistics for the Operational Items on Science Grade 8*

PsyItemNumber	Position	Item Type	Max Points	Item Mean	Item-Total Correlation
636837	1	MC	1	0.39	0.44
636843	2	MC	1	0.41	0.33
641873	3	MC	1	0.35	0.33
641894	4	MC	1	0.27	0.27
709292	5	MC	1	0.32	0.25
709609	6	MC	1	0.29	0.36
758938	7	MC	1	0.48	0.31
709306	8	OR	1	0.16	0.26
709617	9	OR	1	0.38	0.41
713686	10	OR	1	0.47	0.42
713695	11	OR	1	0.24	0.45
758880	12	OR	1	0.49	0.44
631360	13	OR	2	1.05	0.39
636830	15	OR	2	1.10	0.47
636852	16	OR	2	0.79	0.43
637562	17	OR	2	0.99	0.54
637622	18	OR	2	1.04	0.52
637635	19	OR	2	1.09	0.33
640163	20	OR	2	0.78	0.32
640740	21	OR	2	1.02	0.47
641845	22	OR	2	0.48	0.36
641866	23	OR	2	1.48	0.47
642091	24	OR	2	0.59	0.32
642855	25	OR	2	1.10	0.62
643622	26	OR	2	0.58	0.49
707172	27	OR	2	0.78	0.37
709294	28	OR	2	0.36	0.30
709309	29	OR	2	0.85	0.51
709604	30	OR	2	0.40	0.24
709622	31	OR	2	1.09	0.57
712986	32	OR	2	0.90	0.42
713386	33	OR	2	0.86	0.51
713388	34	OR	2	0.84	0.32
716040	35	OR	2	1.10	0.48
731140	36	OR	2	0.98	0.51
741297	37	OR	2	0.86	0.58
752254	38	OR	2	0.91	0.50
758919	39	OR	2	0.62	0.54
763243	40	OR	2	0.90	0.41
663576	41	OR	4	0.57	0.68
697245	42	OR	4	0.63	0.62
717529	43	OR	4	0.87	0.63

* Calculations are based on those students attempting 5 or more items on the form of the given BIE assessment. For 1-point items, the item-total correlation is the point-biserial. For 2 or more-point items, the item-total correlation is the point-polyserial.

Table E-3. Classical Item Statistics for the Operational Items on Science Grade 11*

PsyItemNumber	Position	Item Type	Max Points	Item Mean	Item-Total Correlation
633246	1	MC	1	0.61	0.45
633315	2	MC	1	0.37	0.3
637994	3	MC	1	0.38	0.24
639319	4	MC	1	0.52	0.48
642634	5	MC	1	0.37	0.27
643598	6	MC	1	0.58	0.40
705787	7	MC	1	0.25	0.26
705815	8	MC	1	0.42	0.22
753774	9	MC	1	0.53	0.42
753780	10	MC	1	0.23	0.22
706534	11	OR	1	0.20	0.29
706670	12	OR	1	0.39	0.48
626027	13	OR	2	0.85	0.51
627081	14	OR	2	0.63	0.47
628033	15	OR	2	1.08	0.58
632730	16	OR	2	0.71	0.47
633116	17	OR	2	0.78	0.49
633266	18	OR	2	0.72	0.41
636258	19	OR	2	0.98	0.61
637608	20	OR	2	0.68	0.24
637610	21	OR	2	0.53	0.35
639344	22	OR	2	1.08	0.42
639346	23	OR	2	0.61	0.48
640447	24	OR	2	0.57	0.36
640641	25	OR	2	0.64	0.44
642245	26	OR	2	1.09	0.43
642377	27	OR	2	0.66	0.35
642454	28	OR	2	0.66	0.62
642533	29	OR	2	0.91	0.51
705738	31	OR	2	0.45	0.22
705807	32	OR	2	0.66	0.37
706468	34	OR	2	0.96	0.53
706583	35	OR	2	0.62	0.37
709638	36	OR	2	0.81	0.47
710663	37	OR	2	0.90	0.59
733448	38	OR	2	0.84	0.33
738031	39	OR	2	0.49	0.41
748109	40	OR	2	0.90	0.46
755860	41	OR	2	0.99	0.39
762012	42	OR	2	0.94	0.50
762916	43	OR	2	0.84	0.39
762922	44	OR	2	0.62	0.28
666236	45	OR	4	1.10	0.66
710876	46	OR	4	1.15	0.51
735374	47	OR	4	0.85	0.67

* Calculations are based on those students attempting 5 or more items on the form of the given BIE assessment. For 1-point items, the item-total correlation is the point-biserial. For 2 or more-point items, the item-total correlation is the point-polyserial.

APPENDIX F
ITEM RESPONSE THEORY PARAMETERS

Table F-1. IRT Parameters for Operational Items on the BIE Grade 5 Science Assessment Dichotomously-Scored Items

IREF	a	b	c		IREF	a	b	c
632724	0.44448	3.12171	0.33981		706149	0.41471	-0.19523	0
637807	0.62288	-1.35588	0		706792	0.83975	1.37537	0.26479
637951	0.94836	0.77058	0.22658		744445	0.57361	0.93911	0.15286
638656	1.01219	2.34572	0.1977		632837	0.80012	-0.94917	0
638658	0.63882	0.7963	0.09667		706801	0.99637	0.32987	0
706138	0.51892	1.59056	0.17098		744451	0.43881	1.26637	0

Table F-2. IRT Parameters for Operational Items on the BIE Grade 8 Science Assessment Dichotomously-Scored Items

IREF	a	b	c		IREF	a	b	c
636837	1.24502	0.88903	0.20065		758938	1.02121	1.12092	0.34469
636843	0.82189	1.19908	0.23931		709306	0.378	2.77656	0
641873	0.85801	1.37787	0.21168		709617	0.58397	0.54659	0
641894	0.28539	1.46677	0		713686	0.59822	0.11082	0
709292	0.73517	1.87278	0.21466		713695	0.73355	1.1572	0
709609	0.7008	1.40146	0.10364		758880	0.67003	0.04384	0

Table F-3. IRT Parameters for Operational Items on the BIE Grade 11 Science Assessment Dichotomously-Scored Items

IREF	a	b	c		IREF	a	b	c
633246	0.96478	-0.16949	0.2305		753774	0.71758	0.03096	0.13042
633315	1.07057	0.7227	0.26507		753780	1.02408	2.19229	0.17602
637994	0.70552	1.87717	0.27908		706534	0.40906	2.10731	0
639319	0.88077	0.0358	0.11812		706670	0.68128	0.33302	0
642634	0.58688	1.64355	0.2296					
643598	0.80955	0.10797	0.26409					
705787	0.51262	2.2002	0.11792					
705815	0.5655	1.8556	0.3039					

Table F-4. IRT Parameters for Operational Items on the BIE Grade 5 Science Assessment—Polytomously-Scored Items

IREF	a	b	c	d0	d1	d2
626442	0.73417	-0.85181	0	0.68121	-0.68121	0
626478	0.49019	0.78409	0	1.42087	-1.42087	0
631577	0.81699	-0.01798	0	0.80992	-0.80992	0
632426	0.41017	1.47637	0	1.5969	-1.5969	0
632454	0.47781	1.06192	0	1.36879	-1.36879	0
633906	0.52965	2.17526	0	1.83215	-1.83215	0
634127	0.60273	1.28859	0	1.79429	-1.79429	0
635879	0.55434	-0.23204	0	0.39888	-0.39888	0
635886	0.24564	4.63068	0	3.04072	-3.04072	0
636189	0.58582	2.44211	0	1.73361	-1.73361	0
636211	0.55585	0.0632	0	0.94133	-0.94133	0
638354	0.81846	0.1029	0	0.53019	-0.53019	0
638526	0.69458	-0.1343	0	0.73171	-0.73171	0
638558	1.03441	-0.45257	0	0.6017	-0.6017	0
638639	0.32562	-0.07699	0	1.3282	-1.3282	0
639474	0.89836	0.15411	0	0.79037	-0.79037	0
639510	0.74656	0.6709	0	0.97387	-0.97387	0
697027	0.64913	-0.42395	0	0.96675	-0.96675	0
697044	0.46279	-0.29765	0	1.78577	-1.78577	0
706119	0.61514	-0.17125	0	1.03488	-1.03488	0
706135	0.63263	0.51028	0	1.01185	-1.01185	0
706765	0.79008	-0.08108	0	0.71768	-0.71768	0
706847	0.66793	0.7281	0	0.98715	-0.98715	0
737916	0.59711	-0.43392	0	1.07544	-1.07544	0
743165	0.77342	0.79238	0	1.07838	-1.07838	0
744455	0.77711	-1.01047	0	0.80538	-0.80538	0
756457	0.69889	1.85712	0	1.19726	-1.19726	0
762758	0.51585	0.88948	0	1.26419	-1.26419	0
661177	1.01161	1.53694	0	2.34981	0.98451	-0.83217
666120	0.6236	3.29063	0	3.12968	1.13869	-0.96869
697164	0.93395	1.37909	0	1.89904	0.64813	-0.51044

Table F-5. IRT Parameters for Operational Items on the BIE Grade 8 Science Assessment—Polytomously-Scored Items

IREF	a	b	c	d0	d1	d2
631360	0.45462	-0.08329	0	1.44254	-1.44254	0
631831	0.75139	-0.25616	0	1.43286	-1.43286	0
636830	0.58204	-0.2841	0	1.11864	-1.11864	0
636852	0.48058	0.62908	0	1.14652	-1.14652	0
637562	0.74441	-0.05665	0	0.85441	-0.85441	0
637622	0.70843	-0.17387	0	0.83869	-0.83869	0
637635	0.34298	-0.32363	0	2.12381	-2.12381	0
640163	0.32268	1.00587	0	1.79723	-1.79723	0
640740	0.58809	-0.12792	0	0.84869	-0.84869	0
641845	0.40293	2.25513	0	1.71096	-1.71096	0
641866	0.7223	-1.48305	0	1.47667	-1.47667	0
642091	0.32482	2.04713	0	1.92193	-1.92193	0
642855	1.08455	-0.07076	0	0.33712	-0.33712	0
643622	0.60187	1.39915	0	1.22298	-1.22298	0
707172	0.51382	0.9717	0	2.13148	-2.13148	0
709294	0.40649	3.21875	0	2.05849	-2.05849	0
709309	0.64171	0.41131	0	1.05488	-1.05488	0
709604	0.33183	3.36256	0	2.11492	-2.11492	0
709622	0.81753	-0.21239	0	0.7944	-0.7944	0
712986	0.49914	0.28444	0	1.40211	-1.40211	0
713386	0.67185	0.37076	0	0.99105	-0.99105	0
713388	0.3141	0.88684	0	2.23302	-2.23302	0
716040	0.64656	-0.31981	0	1.23043	-1.23043	0
731140	0.66375	0.09718	0	0.7921	-0.7921	0
741297	0.81002	0.40976	0	0.92535	-0.92535	0
752254	0.64346	0.34647	0	1.08352	-1.08352	0
758919	0.79424	1.07421	0	1.24373	-1.24373	0
763243	0.47989	0.36736	0	1.4584	-1.4584	0
663576	1.17534	2.16833	0	1.95642	0.64121	-0.58147
697245	0.95052	2.31301	0	2.19963	0.67501	-0.74059
717529	0.94598	1.73954	0	2.08847	0.61388	-0.64822

Table F-6. IRT Parameters for Operational Items on the BIE Grade 11 Science Assessment—Polytomously-Scored Items

IREF	a	b	c	d0	d1	d2
626027	0.66291	0.03377	0	0.98072	-0.98072	0
627081	0.63793	0.93026	0	1.13172	-1.13172	0
628033	0.86686	-0.30662	0	0.76366	-0.76366	0
632730	0.63675	0.74758	0	1.19477	-1.19477	0
633116	0.59157	0.53132	0	1.17633	-1.17633	0
633266	0.44124	0.91661	0	1.37493	-1.37493	0
636258	0.89574	-0.05648	0	0.73651	-0.73651	0
637608	0.28352	1.7923	0	2.31758	-2.31758	0
637610	0.39515	2.07339	0	1.98026	-1.98026	0
639344	0.48912	-0.33432	0	1.39477	-1.39477	0
639346	0.57702	1.0448	0	1.11586	-1.11586	0
640447	0.41864	1.87538	0	2.15293	-2.15293	0
640641	0.53125	1.06443	0	1.30728	-1.30728	0
642245	0.51275	-0.44047	0	1.13713	-1.13713	0
642377	0.36397	1.57542	0	1.94544	-1.94544	0
642454	0.89617	0.64585	0	0.836	-0.836	0
642533	0.65514	0.10687	0	1.06042	-1.06042	0
697591	0.38747	1.52866	0	1.91576	-1.91576	0
705738	0.22567	4.72835	0	3.298	-3.298	0
705807	0.44018	1.57377	0	2.40979	-2.40979	0
706123	0.50857	0.84938	0	1.48219	-1.48219	0
706468	0.69722	0.001	0	1.04956	-1.04956	0
706583	0.40086	1.51632	0	1.74171	-1.74171	0
709638	0.56553	0.55154	0	1.31949	-1.31949	0
710663	0.82237	0.09484	0	0.67613	-0.67613	0
733448	0.33083	0.76398	0	2.09123	-2.09123	0
738031	0.48744	1.90326	0	1.53826	-1.53826	0
748109	0.55603	0.19015	0	1.35926	-1.35926	0
755860	0.44223	-0.02481	0	1.62641	-1.62641	0
762012	0.61362	0.01422	0	1.06998	-1.06998	0
762916	0.4261	0.53328	0	1.5248	-1.5248	0
762922	0.29663	2.33457	0	2.79378	-2.79378	0
666236	0.99709	1.45232	0	2.18447	1.06249	-0.50034
710876	0.52588	1.28889	0	1.79352	0.53457	-0.60506
735374	1.03598	1.60185	0	1.98687	0.45377	-0.22145

APPENDIX G
TEST CHARACTERISTIC CURVES
AND
TEST INFORMATION FUNCTIONS

SCI05 TIF
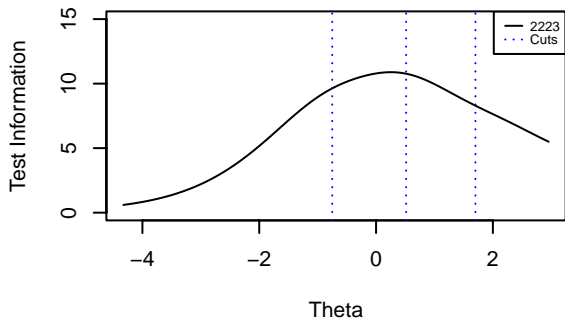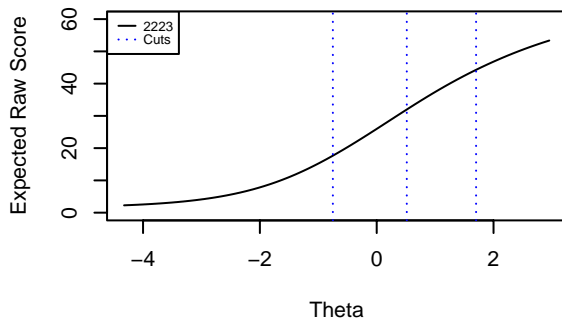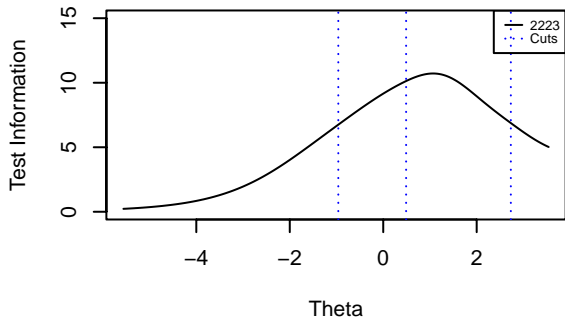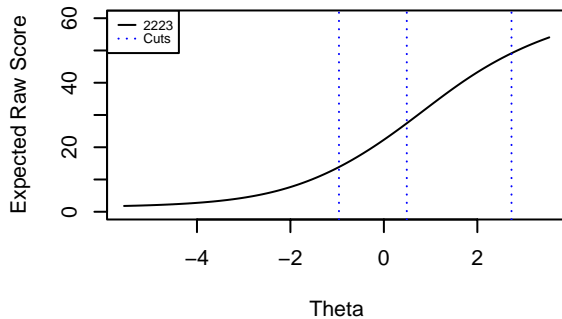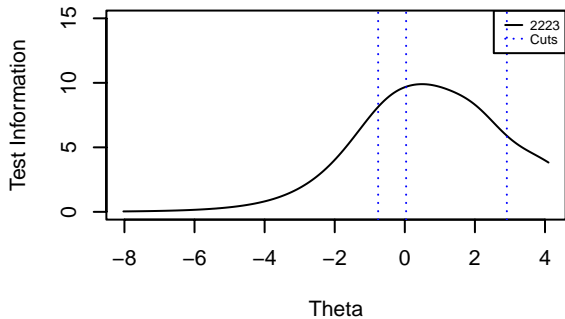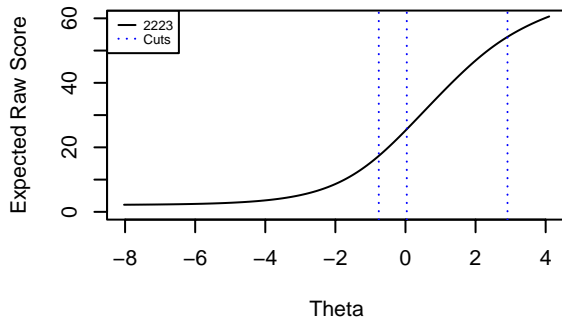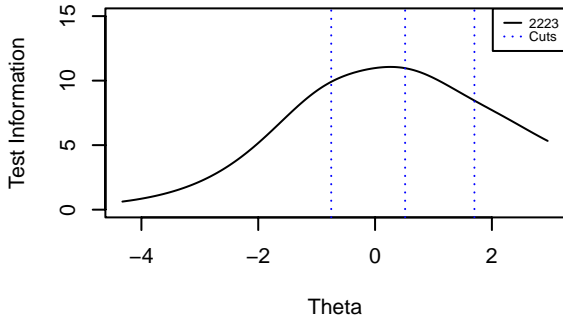
Test Information vs Theta, legend: 2223, Cuts

SCI05 TCC

Expected Raw Score vs Theta, legend: 2223, Cuts

SCI08 TIF

Test Information vs Theta, legend: 2223, Cuts

SCI08 TCC

Expected Raw Score vs Theta, legend: 2223, Cuts

SCI11 TIF

Test Information vs Theta, legend: 2223, Cuts

SCI11 TCC

Expected Raw Score vs Theta, legend: 2223, Cuts
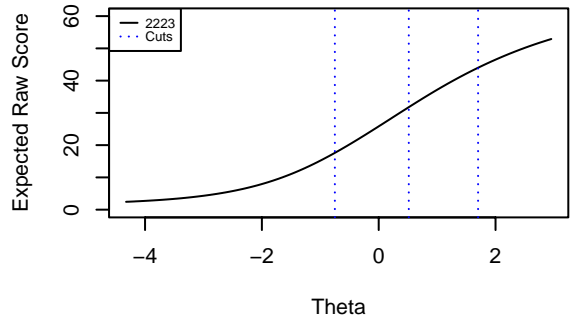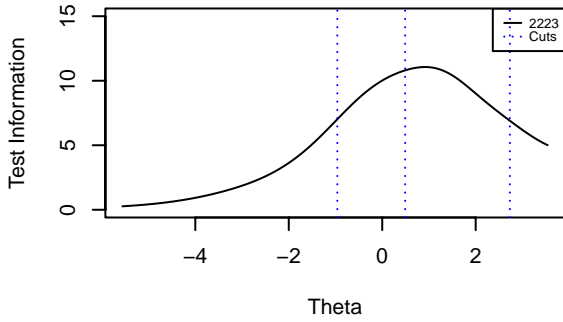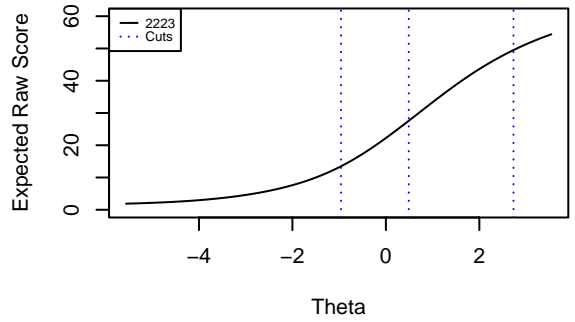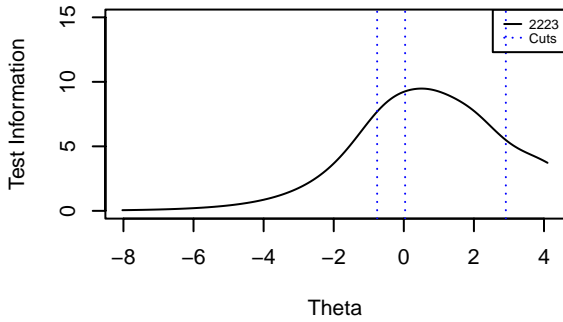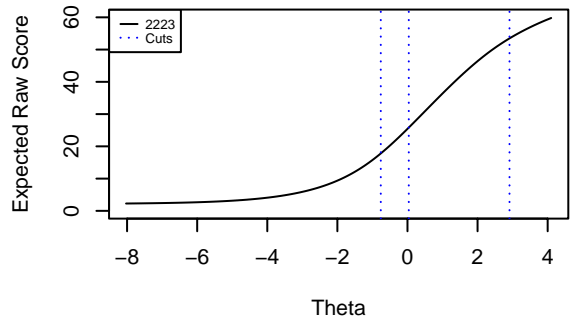
APPENDIX H
RAW TO SCALED SCORE LOOKUP TABLES

Table H-1. Raw to Scaled Score Look-up Table—2022 Science Grade 5 (Scale Form 1)

Raw Score	Theta	Scale Score	Performance Level	CSEM	Scaled CSEM
0	-4.32534	500	1	1.29035	16.12941
1	-4.08470	503	1	1.14454	14.30679
2	-3.84407	506	1	1.01520	12.68999
3	-3.60343	509	1	0.90099	11.26240
4	-3.04662	515	1	0.68794	8.59930
5	-2.67624	520	1	0.58032	7.25397
6	-2.39479	524	1	0.51390	6.42374
7	-2.16525	527	1	0.46836	5.85450
8	-1.96961	529	1	0.43508	5.43856
9	-1.79774	531	1	0.40975	5.12192
10	-1.64337	533	1	0.38994	4.87425
11	-1.50235	535	1	0.37415	4.67693
12	-1.37176	536	1	0.36142	4.51776
13	-1.24949	538	1	0.35106	4.38823
14	-1.13394	539	1	0.34257	4.28214
15	-1.02390	541	1	0.33558	4.19479
16	-0.91841	542	1	0.32980	4.12249
17	-0.81670	543	1	0.32499	4.06233
18	-0.71817	545	2	0.32095	4.01193
19	-0.62231	546	2	0.31755	3.96943
20	-0.52869	547	2	0.31467	3.93332
21	-0.43696	548	2	0.31220	3.90244
22	-0.34683	549	2	0.31008	3.87594
23	-0.25802	550	2	0.30826	3.85324
24	-0.17030	551	2	0.30672	3.83398
25	-0.08345	553	2	0.30544	3.81800
26	0.00273	554	2	0.30442	3.80530
27	0.08844	555	2	0.30368	3.79600
28	0.17387	556	2	0.30323	3.79033
29	0.25920	557	2	0.30308	3.78853
30	0.34463	558	2	0.30327	3.79091
31	0.43035	559	2	0.30382	3.79776
32	0.51656	560	3	0.30475	3.80934
33	0.60347	561	3	0.30607	3.82590
34	0.69128	562	3	0.30781	3.84759
35	0.78022	563	3	0.30996	3.87452
36	0.87050	564	3	0.31253	3.90669
37	0.96236	566	3	0.31552	3.94400
38	1.05602	567	3	0.31890	3.98626
39	1.15175	568	3	0.32265	4.03315
40	1.24977	569	3	0.32674	4.08426
41	1.35034	570	3	0.33113	4.13913
42	1.45373	572	3	0.33579	4.19732
43	1.56020	573	3	0.34068	4.25849
44	1.67006	573	3	0.34581	4.32258
45	1.78362	576	4	0.35119	4.38987
46	1.90126	577	4	0.35690	4.46124
47	2.02341	579	4	0.36306	4.53820
48	2.15062	580	4	0.36984	4.62298
49	2.28354	582	4	0.37747	4.71840
50	2.42301	584	4	0.38621	4.82767
51	2.57005	586	4	0.39631	4.95391
52	2.72588	588	4	0.40796	5.09956

Raw Score	Theta	Scale Score	Performance Level	CSEM	Scaled CSEM
53	2.89201	589	4	0.42126	5.26581
54	2.95466	590	4	0.42646	5.33071
55	2.95466	590	4	0.42646	5.33071
56	2.95466	590	4	0.42646	5.33071
57	2.95466	590	4	0.42646	5.33071
58	2.95466	590	4	0.42646	5.33071
59	2.95466	590	4	0.42646	5.33071
60	2.95466	590	4	0.42646	5.33071
61	2.95466	590	4	0.42646	5.33071
62	2.95466	590	4	0.42646	5.33071
63	2.95466	590	4	0.42646	5.33071
64	2.95466	590	4	0.42646	5.33071

Table H-2. Raw to Scaled Score Look-up Table—2022 Science Grade 5 (Scale Form 2)

Raw Score	Theta	Scale Score	Performance Level	CSEM	Scaled CSEM
0	-4.32534	500	1	1.26420	15.80254
1	-4.13329	502	1	1.15544	14.44306
2	-3.94124	504	1	1.05549	13.19363
3	-3.74919	507	1	0.96386	12.04826
4	-3.13684	514	1	0.72211	9.02638
5	-2.73597	519	1	0.60058	7.50720
6	-2.43538	523	1	0.52595	6.57440
7	-2.19298	526	1	0.47507	5.93832
8	-1.98834	529	1	0.43812	5.47645
9	-1.81002	531	1	0.41019	5.12739
10	-1.65095	533	1	0.38853	4.85659
11	-1.50645	535	1	0.37143	4.64286
12	-1.37326	536	1	0.35778	4.47224
13	-1.24899	538	1	0.34680	4.33504
14	-1.13190	539	1	0.33794	4.22419
15	-1.02062	541	1	0.33075	4.13436
16	-0.91411	542	1	0.32491	4.06140
17	-0.81151	543	1	0.32016	4.00199
18	-0.71218	545	2	0.31628	3.95345
19	-0.61554	546	2	0.31309	3.91363
20	-0.52117	547	2	0.31046	3.88075
21	-0.42868	548	2	0.30827	3.85342
22	-0.33777	549	2	0.30644	3.83056
23	-0.24815	550	2	0.30491	3.81138
24	-0.15959	552	2	0.30363	3.79536
25	-0.07187	553	2	0.30258	3.78219
26	0.01520	554	2	0.30174	3.77179
27	0.10181	555	2	0.30114	3.76421
28	0.18815	556	2	0.30077	3.75965
29	0.27440	557	2	0.30067	3.75838
30	0.36074	558	2	0.30086	3.76073
31	0.44737	559	2	0.30137	3.76707
32	0.53448	560	3	0.30222	3.77774
33	0.62227	561	3	0.30345	3.79308
34	0.71096	562	3	0.30507	3.81337
35	0.80078	564	3	0.30711	3.83884
36	0.89194	565	3	0.30957	3.86963
37	0.98470	566	3	0.31247	3.90582
38	1.07931	567	3	0.31579	3.94736

Raw Score	Theta	Scale Score	Performance Level	CSEM	Scaled CSEM
39	1.17603	568	3	0.31953	3.99413
40	1.27514	570	3	0.32367	4.04592
41	1.37693	571	3	0.32820	4.10249
42	1.48170	572	3	0.33309	4.16359
43	1.58978	573	3	0.33833	4.22914
44	1.70151	575	4	0.34394	4.29926
45	1.81730	576	4	0.34996	4.37456
46	1.93761	578	4	0.35650	4.45619
47	2.06298	579	4	0.36369	4.54609
48	2.19409	581	4	0.37176	4.64696
49	2.33179	583	4	0.38098	4.76223
50	2.47711	585	4	0.39166	4.89575
51	2.63137	586	4	0.40410	5.05119
52	2.79613	589	4	0.41850	5.23119
53	2.95466	590	4	0.43317	5.41463
54	2.95466	590	4	0.43317	5.41463
55	2.95466	590	4	0.43317	5.41463
56	2.95466	590	4	0.43317	5.41463
57	2.95466	590	4	0.43317	5.41463
58	2.95466	590	4	0.43317	5.41463
59	2.95466	590	4	0.43317	5.41463
60	2.95466	590	4	0.43317	5.41463
61	2.95466	590	4	0.43317	5.41463
62	2.95466	590	4	0.43317	5.41463
63	2.95466	590	4	0.43317	5.41463
64	2.95466	590	4	0.43317	5.41463

Table H-3. Raw to Scaled Score Look-up Table—Science Grade 8 (Scale Form 1)

Raw Score	Theta	Scale Score	Performance Level	CSEM	Scaled CSEM
0	-5.56012	800	1	2.09921	20.99208
1	-5.30607	802	1	1.89287	18.92873
2	-5.05201	805	1	1.70413	17.04127
3	-3.79247	817	1	0.99677	9.96774
4	-3.16684	823	1	0.76569	7.65686
5	-2.74405	828	1	0.64688	6.46877
6	-2.42020	831	1	0.57372	5.73718
7	-2.15445	834	1	0.52393	5.23930
8	-1.92666	836	1	0.48778	4.87779
9	-1.72554	838	1	0.46028	4.60283
10	-1.54412	840	1	0.43862	4.38617
11	-1.37783	841	1	0.42105	4.21050
12	-1.22355	843	1	0.40647	4.06465
13	-1.07902	844	1	0.39411	3.94112
14	-0.94256	846	2	0.38347	3.83470
15	-0.81291	847	2	0.37418	3.74176
16	-0.68907	848	2	0.36597	3.65969
17	-0.57024	849	2	0.35866	3.58662
18	-0.45578	851	2	0.35212	3.52116
19	-0.34512	852	2	0.34623	3.46230
20	-0.23780	853	2	0.34092	3.40923
21	-0.13340	854	2	0.33613	3.36127
22	-0.03155	855	2	0.33179	3.31786
23	0.06808	856	2	0.32785	3.27845
24	0.16577	857	2	0.32426	3.24258

Raw Score	Theta	Scale Score	Performance Level	CSEM	Scaled CSEM
25	0.26181	858	2	0.32099	3.20986
26	0.35645	859	2	0.31800	3.17999
27	0.44991	859	2	0.31528	3.15281
28	0.54243	861	3	0.31283	3.12835
29	0.63420	861	3	0.31067	3.10675
30	0.72545	862	3	0.30883	3.08831
31	0.81638	863	3	0.30734	3.07343
32	0.90720	864	3	0.30626	3.06258
33	0.99814	865	3	0.30562	3.05624
34	1.08943	866	3	0.30549	3.05490
35	1.18132	867	3	0.30590	3.05903
36	1.27407	868	3	0.30691	3.06907
37	1.36795	869	3	0.30854	3.08540
38	1.46327	870	3	0.31084	3.10838
39	1.56033	871	3	0.31383	3.13826
40	1.65948	872	3	0.31752	3.17520
41	1.76107	873	3	0.32192	3.21918
42	1.86548	874	3	0.32700	3.27002
43	1.97311	875	3	0.33274	3.32738
44	2.08438	876	3	0.33908	3.39078
45	2.19971	877	3	0.34598	3.45983
46	2.31958	878	3	0.35344	3.53442
47	2.44450	880	3	0.36150	3.61502
48	2.57508	881	3	0.37029	3.70293
49	2.71201	881	3	0.38002	3.80022
50	2.85617	884	4	0.39092	3.90923
51	3.00859	885	4	0.40315	4.03155
52	3.17047	887	4	0.41664	4.16643
53	3.34307	889	4	0.43096	4.30964
54	3.52775	889	4	0.44544	4.45437
55	3.53988	890	4	0.44634	4.46345
56	3.53988	890	4	0.44634	4.46345
57	3.53988	890	4	0.44634	4.46345
58	3.53988	890	4	0.44634	4.46345
59	3.53988	890	4	0.44634	4.46345
60	3.53988	890	4	0.44634	4.46345
61	3.53988	890	4	0.44634	4.46345
62	3.53988	890	4	0.44634	4.46345
63	3.53988	890	4	0.44634	4.46345
64	3.53988	890	4	0.44634	4.46345

Table H-4. Raw to Scaled Score Look-up Table—Science Grade 8 (Scale Form 2)

Raw Score	Theta	Scale Score	Performance Level	CSEM	Scaled CSEM
0	-5.56012	800	1	1.92916	19.29158
1	-5.42928	801	1	1.83121	18.31212
2	-5.29845	802	1	1.73778	17.37780
3	-3.97572	815	1	1.02957	10.29570
4	-3.29608	822	1	0.80823	8.08235
5	-2.82408	827	1	0.69122	6.91220
6	-2.45895	831	1	0.61289	6.12890
7	-2.16102	833	1	0.55457	5.54570
8	-1.90956	836	1	0.50916	5.09158
9	-1.69198	838	1	0.47302	4.73018
10	-1.49991	840	1	0.44385	4.43854

Raw Score	Theta	Scale Score	Performance Level	CSEM	Scaled CSEM
11	-1.32755	842	1	0.42007	4.20067
12	-1.17071	843	1	0.40050	4.00500
13	-1.02629	844	1	0.38429	3.84292
14	-0.89195	846	2	0.37079	3.70792
15	-0.76586	847	2	0.35950	3.59495
16	-0.64659	849	2	0.35000	3.50004
17	-0.53299	850	2	0.34200	3.41999
18	-0.42413	851	2	0.33522	3.35222
19	-0.31924	852	2	0.32946	3.29464
20	-0.21768	853	2	0.32456	3.24558
21	-0.11891	854	2	0.32036	3.20363
22	-0.02248	855	2	0.31677	3.16769
23	0.07202	856	2	0.31368	3.13682
24	0.16493	857	2	0.31103	3.11028
25	0.25655	858	2	0.30874	3.08743
26	0.34714	859	2	0.30678	3.06779
27	0.43694	859	2	0.30510	3.05100
28	0.52616	860	3	0.30368	3.03684
29	0.61502	861	3	0.30252	3.02523
30	0.70369	862	3	0.30162	3.01623
31	0.79235	863	3	0.30100	3.01004
32	0.88119	864	3	0.30069	3.00695
33	0.97039	865	3	0.30073	3.00733
34	1.06015	866	3	0.30116	3.01161
35	1.15067	867	3	0.30202	3.02025
36	1.24217	868	3	0.30337	3.03370
37	1.33488	868	3	0.30524	3.05243
38	1.42907	869	3	0.30769	3.07686
39	1.52502	870	3	0.31074	3.10737
40	1.62303	871	3	0.31442	3.14423
41	1.72342	872	3	0.31876	3.18758
42	1.82657	873	3	0.32374	3.23739
43	1.93283	874	3	0.32935	3.29345
44	2.04261	876	3	0.33554	3.35543
45	2.15634	877	3	0.34230	3.42298
46	2.27446	878	3	0.34959	3.49594
47	2.39751	879	3	0.35747	3.57467
48	2.52605	880	3	0.36603	3.66027
49	2.66079	881	3	0.37547	3.75471
50	2.80260	883	4	0.38605	3.86054
51	2.95251	885	4	0.39800	3.98003
52	3.11176	886	4	0.41138	4.11376
53	3.28177	888	4	0.42591	4.25907
54	3.46403	889	4	0.44098	4.40979
55	3.53988	890	4	0.44694	4.46942
56	3.53988	890	4	0.44694	4.46942
57	3.53988	890	4	0.44694	4.46942
58	3.53988	890	4	0.44694	4.46942
59	3.53988	890	4	0.44694	4.46942
60	3.53988	890	4	0.44694	4.46942
61	3.53988	890	4	0.44694	4.46942
62	3.53988	890	4	0.44694	4.46942
63	3.53988	890	4	0.44694	4.46942
64	3.53988	890	4	0.44694	4.46942

Table H-5. Raw to Scaled Score Look-up Table—Science Grade 11 (Scale Form 1)

Raw Score	Theta	Scale Score	Performance Level	CSEM	Scaled CSEM
0	-8.02951	1100	1	5.35410	40.15576
1	-6.91018	1108	1	3.54954	26.62157
2	-5.79086	1116	1	2.30127	17.25952
3	-4.67153	1125	1	1.46400	10.97997
4	-3.64720	1132	1	0.95846	7.18848
5	-3.08343	1137	1	0.75985	5.69887
6	-2.69071	1140	1	0.64845	4.86335
7	-2.38720	1142	1	0.57570	4.31772
8	-2.13822	1144	1	0.52398	3.92983
9	-1.92584	1145	1	0.48518	3.63887
10	-1.73958	1147	1	0.45500	3.41252
11	-1.57278	1148	1	0.43091	3.23186
12	-1.42094	1149	1	0.41133	3.08496
13	-1.28090	1150	1	0.39519	2.96390
14	-1.15032	1151	1	0.38176	2.86319
15	-1.02743	1152	1	0.37052	2.77887
16	-0.91087	1153	1	0.36107	2.70799
17	-0.79956	1153	1	0.35311	2.64830
18	-0.69261	1155	2	0.34640	2.59802
19	-0.58932	1155	2	0.34076	2.55571
20	-0.48909	1156	2	0.33603	2.52022
21	-0.39142	1157	2	0.33207	2.49055
22	-0.29589	1158	2	0.32878	2.46589
23	-0.20215	1158	2	0.32607	2.44552
24	-0.10989	1159	2	0.32385	2.42886
25	-0.01882	1159	2	0.32205	2.41541
26	0.07126	1160	3	0.32063	2.40475
27	0.16059	1161	3	0.31954	2.39655
28	0.24933	1162	3	0.31874	2.39054
29	0.33765	1162	3	0.31820	2.38653
30	0.42570	1163	3	0.31792	2.38437
31	0.51363	1164	3	0.31786	2.38396
32	0.60155	1164	3	0.31803	2.38520
33	0.68959	1165	3	0.31841	2.38805
34	0.77787	1166	3	0.31899	2.39244
35	0.86650	1166	3	0.31978	2.39832
36	0.95560	1167	3	0.32075	2.40562
37	1.04528	1168	3	0.32191	2.41430
38	1.13567	1168	3	0.32324	2.42432
39	1.22689	1169	3	0.32476	2.43566
40	1.31909	1170	3	0.32645	2.44840
41	1.41243	1170	3	0.32836	2.46267
42	1.50708	1171	3	0.33050	2.47874
43	1.60327	1172	3	0.33293	2.49699
44	1.70122	1172	3	0.33573	2.51800
45	1.80124	1173	3	0.33900	2.54246
46	1.90367	1174	3	0.34283	2.57125
47	2.00890	1175	3	0.34738	2.60533
48	2.11740	1176	3	0.35277	2.64575
49	2.22973	1176	3	0.35914	2.69356
50	2.34653	1177	3	0.36663	2.74971
51	2.46852	1178	3	0.37532	2.81492
52	2.59653	1179	3	0.38526	2.88946

Raw Score	Theta	Scale Score	Performance Level	CSEM	Scaled CSEM
53	2.73147	1180	3	0.39640	2.97298
54	2.87432	1180	3	0.40856	3.06423
55	3.02614	1182	4	0.42148	3.16109
56	3.18800	1184	4	0.43480	3.26103
57	3.36107	1185	4	0.44832	3.36243
58	3.54671	1186	4	0.46227	3.46705
59	3.74684	1188	4	0.47774	3.58301
60	3.96456	1189	4	0.49705	3.72785
61	4.10383	1190	4	0.51184	3.83878
62	4.10383	1190	4	0.51184	3.83878
63	4.10383	1190	4	0.51184	3.83878
64	4.10383	1190	4	0.51184	3.83878
65	4.10383	1190	4	0.51184	3.83878
66	4.10383	1190	4	0.51184	3.83878
67	4.10383	1190	4	0.51184	3.83878
68	4.10383	1190	4	0.51184	3.83878

Table H-6. Raw to Scaled Score Look-up Table—Science Grade 11 (Scale Form 2)

Raw Score	Theta	Scale Score	Performance Level	CSEM	Scaled CSEM
0	-8.02951	1100	1	4.31564	32.36731
1	-7.10334	1106	1	3.18264	23.86982
2	-6.17718	1113	1	2.32441	17.43304
3	-5.25102	1120	1	1.68303	12.62271
4	-4.06759	1129	1	1.10399	8.27995
5	-3.41312	1134	1	0.87194	6.53956
6	-2.95777	1138	1	0.73915	5.54365
7	-2.60754	1140	1	0.65057	4.87926
8	-2.32235	1142	1	0.58629	4.39719
9	-2.08127	1144	1	0.53723	4.02923
10	-1.87193	1146	1	0.49856	3.73918
11	-1.68640	1147	1	0.46743	3.50572
12	-1.51922	1148	1	0.44203	3.31522
13	-1.36652	1149	1	0.42111	3.15835
14	-1.22539	1151	1	0.40379	3.02842
15	-1.09364	1152	1	0.38939	2.92042
16	-0.96956	1152	1	0.37740	2.83049
17	-0.85180	1153	1	0.36741	2.75559
18	-0.73925	1154	2	0.35910	2.69324
19	-0.63103	1155	2	0.35219	2.64144
20	-0.52642	1156	2	0.34647	2.59850
21	-0.42479	1157	2	0.34174	2.56303
22	-0.32565	1157	2	0.33784	2.53383
23	-0.22855	1158	2	0.33466	2.50992
24	-0.13315	1159	2	0.33206	2.49045
25	-0.03912	1159	2	0.32997	2.47474
26	0.05381	1160	3	0.32830	2.46223
27	0.14587	1161	3	0.32700	2.45248
28	0.23727	1162	3	0.32603	2.44519
29	0.32821	1162	3	0.32535	2.44014
30	0.41885	1163	3	0.32496	2.43720
31	0.50935	1164	3	0.32484	2.43631
32	0.59987	1164	3	0.32499	2.43743
33	0.69054	1165	3	0.32541	2.44058

Raw Score	Theta	Scale Score	Performance Level	CSEM	Scaled CSEM
34	0.78152	1166	3	0.32610	2.44574
35	0.87293	1166	3	0.32706	2.45292
36	0.96493	1167	3	0.32828	2.46209
37	1.05766	1168	3	0.32976	2.47321
38	1.15126	1168	3	0.33150	2.48623
39	1.24589	1169	3	0.33349	2.50114
40	1.34173	1170	3	0.33573	2.51797
41	1.43896	1171	3	0.33825	2.53684
42	1.53780	1171	3	0.34107	2.55804
43	1.63849	1172	3	0.34427	2.58199
44	1.74132	1173	3	0.34791	2.60933
45	1.84662	1174	3	0.35212	2.64089
46	1.95477	1174	3	0.35702	2.67767
47	2.06625	1175	3	0.36277	2.72080
48	2.18157	1176	3	0.36952	2.77141
49	2.30136	1177	3	0.37740	2.83053
50	2.42631	1178	3	0.38652	2.89888
51	2.55719	1179	3	0.39688	2.97663
52	2.69483	1180	3	0.40841	3.06305
53	2.84009	1180	3	0.42084	3.15629
54	2.99383	1182	4	0.43377	3.25330
55	3.15687	1183	4	0.44673	3.35050
56	3.33002	1185	4	0.45937	3.44530
57	3.51418	1186	4	0.47183	3.53871
58	3.71067	1188	4	0.48509	3.63816
59	3.92171	1189	4	0.50127	3.75950
60	4.10383	1190	4	0.51857	3.88930
61	4.10383	1190	4	0.51857	3.88930
62	4.10383	1190	4	0.51857	3.88930
63	4.10383	1190	4	0.51857	3.88930
64	4.10383	1190	4	0.51857	3.88930
65	4.10383	1190	4	0.51857	3.88930
66	4.10383	1190	4	0.51857	3.88930
67	4.10383	1190	4	0.51857	3.88930
68	4.10383	1190	4	0.51857	3.88930

APPENDIX I
DECISION ACCURACY AND CONSISTENCY RESULTS

Calculations based on those students attempting 5 or more items on the given BIE assessment.
Statistic values are suppressed for those subjects/grades with fewer
than 50 students. The consistency statistics are in parentheses.

Table I-1. Decision Accuracy for BIE Science Forms, as a Function of Grade and Performance Level*

Grade	Form	Overall	Kappa	Performance Level 1	2	3	4**
5	1	0.84 (0.76)	0.60	0.87 (0.83)	0.81 (0.70)	0.76 (0.69)	0.73 (0.52)
	2	0.83 (0.76)	0.58	0.87 (0.82)	0.79 (0.71)	0.79 (0.66)	0.80 (0.72)
8	1	0.81 (0.73)	0.55	0.81 (0.75)	0.80 (0.74)	0.84 (0.00)	0.70 (0.00)
	2	0.83 (0.75)	0.58	0.79 (0.70)	0.84 (0.78)	0.84 (0.75)	0.85 (0.00)
11	1	0.79 (0.70)	0.47	0.83 (0.54)	0.67 (0.69)	0.85 (0.00)	0.80 (0.00)
	2	0.78 (0.71)	0.52	0.85 (0.79)	0.66 (0.55)	0.85 (0.74)	0.57 (0.00)

* Calculations based on those students attempting 5 or more items on the given BIE assessment. Statistical values are suppressed for those subjects/grades with fewer than 50 students. The consistency statistics are in parentheses.
**Too few BIE students scored at performance level 4 in 2023.

Table I-2. Decision Consistency for BIE Science Forms, as a Function of Grade and Cut Score*

Cut	Scores	1				2				3**			
Grade	Form	DA	DC	FP	FN	DA	DC	FP	FN	DA	DC	FP	FN
5	1	0.88	0.82	0.07	0.06	0.96	0.95	0.01	0.02	0.99	0.99	0.00	0.00
	2	0.87	0.81	0.07	0.07	0.97	0.96	0.02	0.01	0.99	0.99	0.00	0.00
8	1	0.86	0.81	0.05	0.08	0.95	0.92	0.03	0.02	--	--	--	--
	2	0.89	0.84	0.06	0.05	0.93	0.91	0.04	0.03	--	--	--	--
11	1	0.83	0.77	0.10	0.07	0.95	0.93	0.03	0.01	--	--	--	--
	2	0.84	0.79	0.08	0.07	0.94	0.91	0.04	0.02	--	--	--	--

* Calculations based on those students attempting 5 or more items on the given BIE assessment. Statistical values are suppressed for those subjects/grades with fewer than 50 students.
**Too Few BIE students scored at performance level 4 (or above cut 3) in 2023.

Appendix J
BIE Science Reporting Business Requirements

BIE Science Reporting Business Requirements

1/20/2023

Woreen-Ann Bogle

cognia™

Date	Updated Content Description	Updated By
1/20/2023	*Initial Creation*	*W.Bogle*
1/23/22	*PgM edits incorporated*	*W.Bogle*
5/12/23	*Addendum for Wingate High Students*	*W.Bogle*

Glossary	
CBT	Computer Based Test
PBT	Paper Based Test
HS	High School
MC	Multiple Choice
SRB	Student Response Booklet
EL	English Learner
OE	Open Ended also called Open Response items
FT	Field Test
BIE	Bureau of Indian Education

Table of Contents

Overview

Each year in the Spring, Cognia administers a Science assessment for BIE. The Science assessment utilizes items from Cognia's Secure Science Item Bank (SSIB). The online tests are administered in eMetric's iTester platform. The test is also administered on paper as needed. The test is administered across the country for students in grades 5, 8 and 11.

Points of Contact

Title	Name	Contact Email
Client Services Program Manager	Sara Hanlon	Sara.Hanlon@cognia.org
IT Project Manager	Sarah McCain	Sarah.McCain@cognia.org
Client Services Program Manager	Larry Ehret	Larry.Ehret@cognia.org
Client Services Program Manager	Mara Allaire	Mara.Allaire@cognia.org

Document References

- Student Results data file Layout
  - Used for Participation file in Preliminary Reporting
  - Student Results data file for both BIE and eMetric.
- Data Processing Specifications
  - Documents details around student and test data processing and staging of data for reporting purposes.

Change Log

Year to Year Changes

1. Logo Change on all reports
2. Test Report Code added for students who marked as participating in MSAA.
3. Alternate Setting is a new online accommodation.
4. The following accommodations were removed from the SRB:
   - Selected Response Human Scribe
   - Constructed Response Human Scribe
   - Human Reader English
   - Read Aloud to Self
   - Allow Accessibility Mode Testing
5. The following accommodations were added to the SRB:
   - Human Scribe
   - Speech to Text
   - Tactile Graphics
   - Word Prediction

General Information

Assessments

1. Assessments were administered to students beginning April 3, 2023, and ending on May 12, 2023.

2. Students are tested online (CBT) and on paper (PBT).

3. Tests are administered in grades 5, 8 and 11.

4. Contract-code is 800250.

Reporting Cycles

In 2023, there are 3 reporting cycles. Each cycle is described below.

1. The first reporting cycle is prior to the test administration. In this cycle the outbound rosters are produced. These rosters are produced using the Pre-ID file that contains paper testers. The roster is printed and shipped along with test materials for paper testers.

2. Post testing preliminary reporting there will be a participation file delivered to BIE. This file will follow the Student Results data file layout. However, the data file will have blank fields where the data is dependent on the results of the standards validation. For example, scaled score and performance level will not be available in this cycle. See Student Results data file layout for more details.

3. The final reporting cycle will include all reporting deliverables and will include the results from the standards validation.

Receivables

Data Files received will pass all validation rules and formats based on the layout and specification documents.

1. Cognia was provided a data file for the purpose of Pre-ID prior to the test administration window. This data was used to identify students who will test on paper. Paper testers receive test administration labels that are applied to the answer booklet in the schools.

2. Cognia will not receive a demographic file for reporting purposes.

Deliverables

Reports Produced

1. **Pre-test Administration**
   - Outbound Roster
     - Lists each student from the Pre-ID file provided by BIE. The file contains the student marked for paper testing by the school.

- o Demographic fields from the Pre-ID file are reported on the roster.
- o A Roster is produced for each school included in the Pre-ID file.

2. **Preliminary Reporting – Posttest administration, prior to the standards validation**
   No reports produced in this reporting phase.

3. **Final Reporting-Post-Standards Validation**

- Student Results Labels
  - o Contains the student's earned scaled score and overall performance level.
  - o Printed and shipped to schools.
  - o Only Tested students will receive a label.
  - o See Report Specific Rules section for more information.
- Student Report
  - o Contains the student's earned scaled score and overall performance level.
  - o Contains the performance level earned for each reporting category.
  - o Contains a comparison of the student's performance to the student's school, district and to all participating students.
  - o Printed and shipped to schools.
  - o PDFs provided to eMetric to be made available in the Download Hub.
  - o The PDFs will be accessible in the Download Hub at the organization level.
  - o Only Tested students will receive a student report.
  - o See Report Specific Rules section for more information.
- School Roster
  - o Contains a list of all students in the school.
  - o Printed and shipped to schools.
  - o The student's participation status is included on the report.
  - o If the student receives a scale score and achievement level the results are presented on the roster.
  - o See Report Specific Rules section for more information.

Data File Deliverables

1. **Pre-test Administration**
   No data file deliverables produced in the pretest administration period.

2. **Preliminary Reporting – Posttest administration, prior to the standards validation**

- Participation data file to BIE only
  - o Contains all students who participated in the assessment.
  - o Participation status reported based on test attemptedness rules below.
  - o Does not include scaled score and performance levels.

o   Follows the Student Results data file layout. The "Included in Participation File?" column indicates the fields that are populated in preliminary reporting.

3.   **Final Reporting-Post-Standards Validation**

- Student Results Data File

    o   Contains the results for each participating student.

    o   Participation Status for each student based on test attemptedness rules below.

    o   Follows the Student Results data file layout.

    o   Delivered to BIE and eMetric.

- eMetric Summary Data File

    o   Data file used to QC the summary data in DI.

    o   Summaries done for specified demographics.

- eMetric Metadata file for Student PDFs

    o   Data File used to describe the online set up and PDF files being handed off to eMetric to be loaded into the Download Hub.

    o   See Specific Reporting Rules section for specifications.

Pre-Test Administration

This section describes the data preparation for student records pre-test administration:

- The Pre-ID data file is used to provide answer booklet labels for students in the Pre-ID data file.

    o A total record count will be provided with the final label data to iCore Distribution.

    o Each student label has a unique Barcode associated with a Student ID

    o One student label will be printed for each booklet being administered.

- The Pre-ID data is used to produce the Outbound Rosters that accompany the answer booklet labels.

Forms

The test uses the items from Cognia's Secure Science Item Bank (SSIB). The test is composed of Operational and Field Test items. The Operational items count toward the students' overall score and are included in subscore calculations. Field Test items will not be scored.

1. For spring 2023, there are 2 main Core Forms: Core Form A and Core Form B. The form AX is the accommodated form used for paper replacement, Large Print, and Braille. Form A1 is used for English TTS

2. English Paper accommodated forms (represented by Form AX which is the paper replacement of the online form) are provided for students with the following accommodations:

    a. Large Print
    b. Braille

3. Paper accommodated forms contain replacement items for the TEIs on the online Form A1 test form that becomes Form AX.

4. The online accommodated form has the TTS-accessibility feature and is CBT Form A1.

5. Tests are provided only in English.

Test Mode	Grade	Forms Administered	Additional accommodated form
Online	05	A1+B4	Yes (Text-to-Speech (TTS) & Form X)
	08	A1+B4	Yes (Text-to-Speech (TTS) & Form X)
	11	A1+B5	Yes (Text-to-Speech (TTS) & Form X)
Paper	05, 08, 11	Form AX	Used for Large Print and Braille

Note: English accommodated computer-based test forms are Form 1 Text-To-Speech for each grade

Item Types

The following item types are administered in the test:
1.

Item Type	Definition	Valid Point Values	Scoring Method and Scoring Rules
MS-1	machine-scored item that may be multiple choice, multi-select, or TEI interaction	0,1	Machine scored; all or nothing scoring for the interaction, no partial credit
MS-2	machine-scored item with part a, part b; interactions may be any combination of multiple choice, multi-select, or TEI	0,1,2	Machine scored; part a and part b each worth one point; each part scored all or nothing (0,1); sum scoring for parts for total score of (0, 1, 2); each part scored independently
OE	hand-scored extended text interaction (traditional open response/constructed response item)	0,1,2,3,4	Hand scored; holistically scored; one rubric/one dimensional scoring

2. There is a Cluster/Passage which consists of two MS-1 items and two MS-2 items for a total of maximum 6 points.

Reporting Categories

1. For Disciplinary Core Ideas (DCI) reporting categories, the subscore reporting will be based on three disciplinary domains.

   - Physical Sciences (PS)
   - Life Sciences (LS)
   - Earth and Space Sciences (ESS)

2. The PrimaryContentStandard column in the NTS extract provides the coding for the DCI reporting categories.

3. The Standards Column in the NTS extract indicates the DCI.

4. DCI Reporting categories are defined as follows:

Code	DCI Reporting Category
LS	Life Science
PS	Physical Science
ES	Earth and Space Science

5. Psychometrics provides the performance indicators for student's performance on each reporting category. Values:

   1=Below Standard
   2=At/Near Standard
   3=Above Standard

Post Test Administration

Demographic Clean-up

Demographic cleanup is done in the iTester Portal for online testers. Paper testers are linked back to the pre-ID file provided when a valid ID is provided, or a valid label is used on the answer booklet. Data Processing Specifications contains more details on identification of students and linking to the pre-ID file.

Student Data Processing

1. Student Names will have all periods, commas and apostrophes removed.

    a. Middle Name is the first initial of the middle name or blank if not available.
    b. Special characters will be set to blank.
2. All records will be suppressed from processing if Name fields, Student ID and Test Items are all blank.

Test Data Processing

Duplicates may exist where there are more than one data record with the same Student ID, be the record online or paper.

1. Duplicate Test records with the same Student ID/Grade will be combined or otherwise suppressed. See *Data Processing Specifications* for resolution of duplicate tests.

    a. If there is a duplicate where the student takes one session in one test instance and another session in another test instance the 2 sessions will be combined/merged to created one complete test.
        i. If the schools differ between session 1 and session 2, the school from the last session taken will be used for reporting (if it can be determined by the session updated dates for online tests)
        ii. The record will be flagged in the data file as being a merged record.
        iii. If an online session is merged with a paper session, test mode flag is set to "both".
2. Duplicate Cross Grade tests are identified as more than one test taken with two different grades from the same student.

    a. Should the Student have no work in the off-grade book, or the book is void and work in the matching grade test, suppress the off-grade test.
    b. If both books have responses, send a report to Program Management for research and resolution.

Scan Paper Delivery and Data Denotation

Each Paper test is scanned and delivered immediately to the Reporting Data Processing team. At the time of receipt, Data Processing will perform procedures to accurately identify inaccuracies in the data. The data will be formatted as specified in the Scan Delivery Layout Format

1. All discrepancies with the Scan File be resolved accordingly.

2. Any Student Response Booklet where VOID is bubbled and there is at least one item is attempted will be researched via Webdesk system. See Data Processing Specifications for resolution of Void bubbles.

Scoring Data

Scoring division will provide Data Processing with the open response scores for all tests.

1. Every score record will contain valid scores for all items.

   a. A validation of score values will be performed.
   b. If a score value is found to be invalid, resolution will be attempted with the Scoring Division.

2. Each score record will be associated with a Booklet ID or a Test ID.

   a. If a score record is received without an associated Test or Booklet ID, resolution will be attempted with the Scoring Division.

3. All unresolved scoring records will be included in a report to the Scoring Division, for research and resolution.

4. The following values will be received from Scoring:

   B=Blank
   U=Unreadable with code number 51
   F=non-English with code number 53
   W=Wrong Location with code number 52
   O=Off Topic with code number 54

5. Score values of U and W will be blanked out and reported with a null/blank value.

6. Score values of B, F and O will be given a score of 0 for analysis purposes.

7. Only operational open response items are scored.

Student Participation and Exclusions

Test Attempt Rules

Test Attempted indicates that a student has answered at least five (5) operational items on the test.

1. If a session is voided, any items attempted will be blanked out and will not count toward test attemptedness.

2. A valid attempt to an open response item is determined if the score is a non-blank (not B=blank or NULL) score.

3. Only field test items can have a null score meaning that the item was not scored.

4. A student is classified into 2 possible attempt groups of Attempt Status:

   a. Attempt Status 0 is assigned to the test if the student did not provide a valid response to at least 5 operational items.
   b. Attempt Status 1 is assigned to the test if the student provided a valid attempt to at least 5 operational items on the test.

Not Tested Reasons

The below not tested reasons or Test Report Codes are the valid options on the scannables and in the online testing platform. School Administrators set the codes as appropriate. Test invalidations will be provided to Cognia by BIE, if appropriate.

- Valid Not Tested Reasons are:
  - Withdrew Before Test Completion (01)
  - Non-allowed Modification (02)
  - Medical Emergency (04)
  - Other non-completion (06)
  - Test Irregularities (07)
  - Absent (08)
  - COVID exemption (09)
  - Participating in MSAA (10)
  - Void or Do Not Report (bubbled on the scannable or DNR indicated in the online testing platform)

Student Participation Status

All student tests will reflect the Student Participation based on the test attempt status and any "Not Tested Reason" indicated.

1. Participation status is determined using both the test "Attempt Status" value and the "Not Tested Reason".

   a. If Attempt Status is 0 (less than 5 operational items with valid attempts),
      i. The test has a "Not Tested Reason", the "Not Tested Reason" is reported, otherwise, the test is reported as "Did Not Reach Minimum Attempt".
   b. If Attempt Status is 1 (the test has at least 5 operational items with valid attempts),
      i. The student is classified as Tested.
      ii. If the student has a Not Tested Reason, the Not Tested Reason is ignored.

2. Regardless of the test attempt status, if a student is on the test invalidation list from the client, their test will be marked as Invalidated. The test is suppressed from the reporting dataset and is not part of any reporting deliverable or analyses.

3. Only "Tested" students, that is, students who meet attemptedness will be included in analyses.

4. The client will indicate in the exception list any tests that should be invalidated.

5. The following hierarchy is followed if more than one not tested reason is marked and the student's attempt status is 0. The hierarchy is from highest to lowest priority:

   Invalidated test
   Void/DNR test
   Participating in MSAA
   COVID exemption
   Medical Emergency
   Absent
   Withdrew Before Test Completion
   Test Irregularities
   Non-Allowed Modification
   Other Non-Completion

Student Participation Summary

The participation summary table below defines the participation codes

Participation Status	Participation Code	Included in State File	Included in Aggregations	Receive a Student Report and Label	Included on Roster and DI
Tested	Z	Yes	Yes	Yes	Yes
Did Not Reach Minimum Attempt	B	Yes	No	No	Yes
Withdrew Before Test Completion	C	Yes	No	No	Yes
Non-Allowed Modification	D	Yes	No	No	Yes
Medical Emergency	F	Yes	No	No	Yes
Other Non-Completion	H	Yes	No	No	Yes
Test Irregularities	I	Yes	No	No	Yes
Absent	J	Yes	No	No	Yes
COVID Exemption	A	Yes	No	No	Yes
Participating in MSAA	M	Yes	No	No	Yes
Void Test	K	Yes	No	No	Yes
Invalidated Test	L	Yes	No	No	Yes

Calculations

Aggregations

Aggregation inclusion rules are summarized in the Student Participation Summary table above.

Scoring Method

1. Overall raw score is calculated by summing the student's score on each operational item.
2. NTS, Cognia's item bank, indicates which items are operational and count toward the student's overall raw score.

Performance Levels

1. In 2023, a standards validation will be conducted to ensure the standards set in the SSIB standard setting are a fit for the BIE population.
2. After the standards validation, Psychometrics will provide the scale to be used for scaling the BIE tests.
3. Scale scores and achievement levels will be assigned based on the students' overall raw score and applying the scale score lookup from Psychometrics.

Data Suppression Rules

See Report Specific Rules section for data suppression rules.

Specific Reporting Rules

LENS Online Reporting

Cognia's online reporting system LENS is not used for BIE. Online reporting is done in Data Interaction.

General rule: If the student's name is not available the name is reported as Name Not Provided. This applies to all deliverables.

Student Results Data File and Participation File

- Naming Convention of the data files: BIE2223sciStudentResults.csv for the State Results data file. BIE2223sciParticpationFile.csv for the participation file.

- Item level data will be provided according to the BIE2223sciItemLevelDataFilelayout.xlsx. The naming convention for the item level file is BIE2223sciStudentItemLevelResults.csv.

- If a student's test was merged to created one test, then the mergedtest flag is set to 1, otherwise it is set to 0.

- If the mergedtest flag is set to 1 and the student tested at 2 or more different locations, the student is reported back to where the last session was taken.

- A student appears in the file once with only one test.

- NumAttempted is the number of operational items in the test that met the item attemptedness rules described above.

- The Participation file produced prior to standard setting follows the same layout as the student results data file layout. Fields populated in the participation file are indicated with a Y in the "Included in Participation File" column in the layout. Scale scores and achievement level columns will be blank in the participation file.

- All grades are reported in the same file.

- Participation file is not handed off to eMetric.

- eMetric will receive the BIE2223sciStudentResults.csv after standards validation for loading DI. eMetric will use the 'Included in DI reporting" column in the file layout to determine what data to load to DI.

eMetric Metadata file for Student PDFs

- The column headings for the file are: ProgramName,ReportName,Year, Org_Num, PDF_name

- The file is a csv file

- The naming convention for the file is BIE2223sci_PDFmetadata.csv

- The file is posted to the ftp site for eMetric to access

- Org_Num=Districtcode-Schoolcode

- Year=2023
- ProgramName=Science (General Education)
- ReportName=Individual Student Report
- Web file naming convention:
  BIE2223sciStudentReport_<districtcode||schoolcode>.pdf

eMetric Summary Data File

- Data file used to compare summary results to aggregations in DI. Used for quality assurance checks by eMetric.
- The data file follows the BIE2223scieMetricSummaryDataTransfer layout.

Student Results Labels

- Student results labels are only produced for students who Tested (participation status='Z').
- The label contains the scale score and achievement level the student earned.
- The Test Date on the label is Spring 2023.
- There are 10 labels per page.
- Labels are produced by school and tested grade. Within each file the students in a school are sorted by student grade, lastname, firstname, mi, NASIS ID.
- The labels are shipped to the school.

Student Reports

- Student reports are only produced for students who Tested (participation status='Z').
- The student report contains the scale score and achievement level the student earned.
- The report is printed in color.
- 2 copies of the report are printed and shipped to the school.
- The reports are printed duplex on 8 ½ x 11 sheets of paper.
- Averages and percentages on the report are rounded to the nearest whole number.
- Percentages are formatted with an % sign.
- Orange color used on the report is used to indicate the student's earned scale score.
- The minimum N-size that must be reached for aggregations to be reported is 10. Any result including less than 10 students will be suppressed from the reports.
- Print the achievement level descriptor with examples on the second page corresponding to the student's earned achievement level.
- If the earned achievement level is not Advanced, print the achievement level descriptor without examples, for the achievement level directly above the earned achievement level. This appears adjacent to the earned achievement level descriptor

and is surrounded by a box. For example, the student earns Nearing Proficiency, the achievement level descriptor for Nearing Proficiency with examples is printed and the achievement level descriptor for Proficient is printed in a text box.

- Web versions do not include the slip sheets.

- Web reports are run by school.

- Within the school web files the students are sorted by student grade and lastname,firstname.

- Web file naming convention:
  BIE2223SciStudentReport_<districtcode||schoolcode>.pdf

Student Rosters

- The roster lists all students in the school in a specific tested grade.

- The report is printed duplex.

- Student names are formatted as Lastname, Firstname (title case).

- If a student is not tested, the scale score and achievement level will be '-' on the roster.

- Up to 14 students are listed on the front page.

- A roster may have multiple pages if there are more than 14 students in the school and tested grade.

- Students are listed alphabetically by Lastname, Firstname (title case).

- One copy of the report is printed.

- The report is printed landscape and simplex on 8 ½ x 11 paper in color.

- There is alternate row shading on the roster.

- If a student is Tested, the achievement level is highlighted in a shade of blue that matches the shade on the student report for that achievement level.

- The scale score and achievement levels are printed in bold font.

Shipping Product Code Summary

Reporting Products

Reports are being shipped to schools. Reports are packed by grade. Below is the definition of the 4 reporting products.

Contract Code: [800250] Description BIE Science 2023 Admin ID 1	Report Type	Report For	Grade(s)	Report Subtype	Content Code	Qty
Student Report Parent Copy	07	1	05,08,11	02	00	1
Student Report-School Copy	07	1	05,08,11	01	00	1
Student Label	07	1	05,08,11	03	00	1
Student Roster	07	1	05,08,11	12	00	1

Appendix

Addendum

A. 12th grade Wingate High School students registered and tested as 11th graders will be reported as follows:

1. They will receive a student report and label if they meet the attemptedness rule.
2. Tested Grade and Student Grade will be 11. The source of Student Grade is the registration file.
3. They will NOT be included in any aggregations regardless of attemptedness. However, aggregations will be reported on their student report for their school, district, and BIE.
4. They will be reported in DI (Dynamic Reporting Platform).
5. They will be included in the results data file to BIE.
6. They will be included on the printed Roster.
7. Data Processing will assign a participation status using their student IDs.
8. If they meet attemptedness their participation status is set to N. If they do not meet attemptedness their participation status will be B (Did not Meet Minimum Attempt).

APPENDIX K
SCORE REPORT INTERPRETATION QUICK GUIDE

BIE Science Assessment – Report Interpretation Quick Guide

The Bureau of Indian Education (BIE) Science Assessment is based on Next Generation Science Standards (NGSS), and is administered to students in grades 5, 8, and 11. The standards focus on important disciplinary core ideas, scientific and engineering practices, and crosscutting concepts that apply across scientific disciplines. The assessment provides information regarding each student's progress toward achievement of essential knowledge and skills that will help them explain and make sense of phenomena in the world around them, solve problems, and apply their scientific literacy to understand that scientific dilemmas they may face as adults.

This document outlines how to interpret the information on the individual student report. The report provides general information about the BIE Science Assessment, resources for parents/guardians, and the student's results, including the student's scale score and achievement level. The report also provides information on how the student performed compared to other BIE students.

Sample Individual Student Report (page 1)

BIE Science Assessment
Individual Student Report
Spring 2023

Dear Parent or Guardian:

We are excited that your student participated in the high-quality Science Assessment aligned to the Next Generation Science Standards (NGSS) as required for all Bureau Funded Schools as part of *Every Student Succeeds Act (ESSA)*. This is your student's performance report based on the Spring Science Assessment results.

As your student grows and learns during the school year, the Bureau of Indian Education (BIE) will administer the science assessment annually for grades 5, 8, and 11 to measure how well your student is performing on the Next Generation Science Standards. It will provide a better understanding of what your student knows and is able to do.

These assessment results will allow you to gain critical insights to your student's progress, support your understanding of his/her strengths and areas identified as needing improvement. Additionally, the information will support instruction, identify resources, and provide enrichment activities to meet the needs of all learners.

We thank you for your continued support and cooperation.

The Bureau of Indian Education

(A)

Student Name:	LASTNAME177, FIRST177
NASIS ID:	D00177
Tested Grade:	05
Student Grade:	05
District:	Demonstration District A
School:	Demonstration School 2

(B)

RESOURCES

U.S. Department of the Interior/Bureau of Indian Education: To learn more about the Bureau of Indian Education, please navigate to https://www.bie.edu or scan the QR code.

Assessment & Accountability: Integrated content and assessments ensure broad access to accurate and timely student and school-level information which is used for educational decision making. For more information on BIE Assessments and Accountability, please navigate to https://www.bie.edu/landing-page/assessments-and-accountability or scan the QR code.

Standards: The BIE Science assessment is aligned to the Next Generation Science Standards. To read the standards, navigate to https://www.nextgenscience.org/search-standards or scan the QR code.

(C)

Achievement Level Descriptors

500 Novice 544	Nearing Proficiency 560	Proficient 574	Advanced 590
Students demonstrate evidence of **emerging** understanding and use of college and career readiness knowledge, skills, and abilities.	Students demonstrate evidence of **partial** understanding and use of college and career readiness knowledge, skills, and abilities.	Students demonstrate evidence of **satisfactory** understanding and use of college and career readiness knowledge, skills, and abilities.	Students demonstrate evidence of **thorough** understanding and use of college and career readiness knowledge, skills, and abilities.

(A) This section provides assessment information and a note to parents/guardians around the use of the report information in schools.

(B) This section provides links to additional resources for the BIE Science Assessment, including standards, general BIE information, and specific assessment information. Links can be accessed by typing the link into a browser or scanning the QR code with a smart phone/tablet.
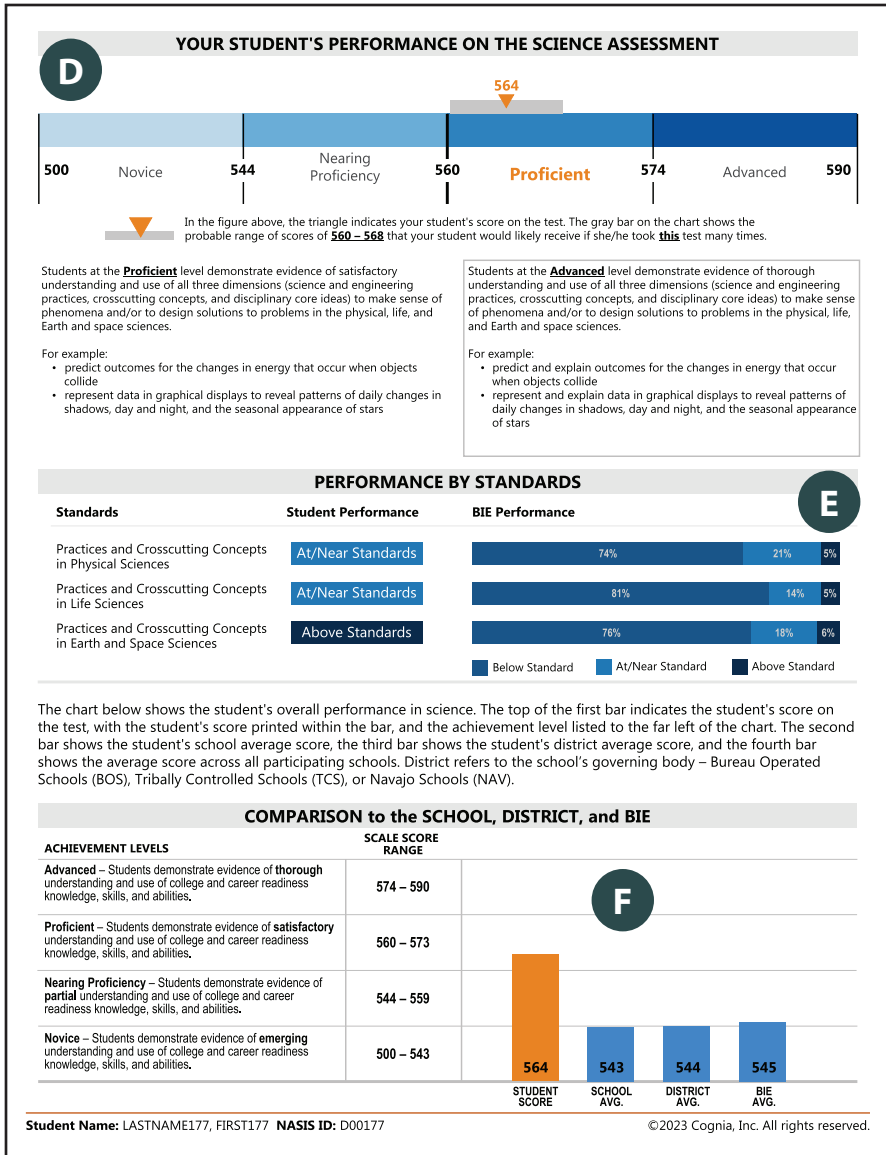
(C) This section describes the four possible achievement levels a student can reach – Novice, Nearing Proficiency, Proficient, and Advanced.

While page one outlines general assessment and report information, page two provides specific information on the student's performance, both individually and in relation to their peers.

Sample Individual Student Report (page 2)

YOUR STUDENT'S PERFORMANCE ON THE SCIENCE ASSESSMENT

D

564

| 500 | Novice | 544 | Nearing Proficiency | 560 | Proficient | 574 | Advanced | 590 |

In the figure above, the triangle indicates your student's score on the test. The gray bar on the chart shows the probable range of scores of **560 – 568** that your student would likely receive if she/he took **this** test many times.

Students at the **Proficient** level demonstrate evidence of satisfactory understanding and use of all three dimensions (science and engineering practices, crosscutting concepts, and disciplinary core ideas) to make sense of phenomena and/or to design solutions to problems in the physical, life, and Earth and space sciences.

For example:
- predict outcomes for the changes in energy that occur when objects collide
- represent data in graphical displays to reveal patterns of daily changes in shadows, day and night, and the seasonal appearance of stars

Students at the **Advanced** level demonstrate evidence of thorough understanding and use of all three dimensions (science and engineering practices, crosscutting concepts, and disciplinary core ideas) to make sense of phenomena and/or to design solutions to problems in the physical, life, and Earth and space sciences.

For example:
- predict and explain outcomes for the changes in energy that occur when objects collide
- represent and explain data in graphical displays to reveal patterns of daily changes in shadows, day and night, and the seasonal appearance of stars

PERFORMANCE BY STANDARDS

E

Standards	Student Performance	BIE Performance
Practices and Crosscutting Concepts in Physical Sciences	At/Near Standards	74% 21% 5%
Practices and Crosscutting Concepts in Life Sciences	At/Near Standards	81% 14% 5%
Practices and Crosscutting Concepts in Earth and Space Sciences	Above Standards	76% 18% 6%

Below Standard / At/Near Standard / Above Standard

The chart below shows the student's overall performance in science. The top of the first bar indicates the student's score on the test, with the student's score printed within the bar, and the achievement level listed to the far left of the chart. The second bar shows the student's school average score, the third bar shows the student's district average score, and the fourth bar shows the average score across all participating schools. District refers to the school's governing body – Bureau Operated Schools (BOS), Tribally Controlled Schools (TCS), or Navajo Schools (NAV).

COMPARISON to the SCHOOL, DISTRICT, and BIE

ACHIEVEMENT LEVELS	SCALE SCORE RANGE
Advanced – Students demonstrate evidence of **thorough** understanding and use of college and career readiness knowledge, skills, and abilities.	574 – 590
Proficient – Students demonstrate evidence of **satisfactory** understanding and use of college and career readiness knowledge, skills, and abilities.	560 – 573
Nearing Proficiency – Students demonstrate evidence of **partial** understanding and use of college and career readiness knowledge, skills, and abilities.	544 – 559
Novice – Students demonstrate evidence of **emerging** understanding and use of college and career readiness knowledge, skills, and abilities.	500 – 543

F

STUDENT SCORE	SCHOOL AVG.	DISTRICT AVG.	BIE AVG.
564	543	544	545

Student Name: LASTNAME177, FIRST177 **NASIS ID:** D00177

©2023 Cognia, Inc. All rights reserved.

D This section describes your student's performance on the overall assessment. Their scale score is provided on the chart, indicating their achievement level. Additional information regarding the achievement level and the skills demonstrated at that level are outlined below. To the right, information is included on what skills the student would need to develop to reach the next achievement level.

E This section breaks down the student performance by each science discipline. The student performance is displayed along with a bar graph of overall BIE performance.

F This section displays the student's performance in relation to peers in his or her school, district, and across the BIE. District, as used on this report, refers to the school's governing body – Bureau Operated Schools, Tribally Controlled Schools, and Navajo Schools.

For additional information regarding the student report, contact your school test coordinator (STC). School inquiries should be directed to the BIE.

APPENDIX L
CUMULATIVE SCALED-SCORE DISTRIBUTIONS

BIE Science Cumulative Scale Score Distributions

Legend: 05 (solid red), 08 (dashed green), 11 (dashed blue)

Y-axis: Cumulative Percentage

X-axis: Scale Score
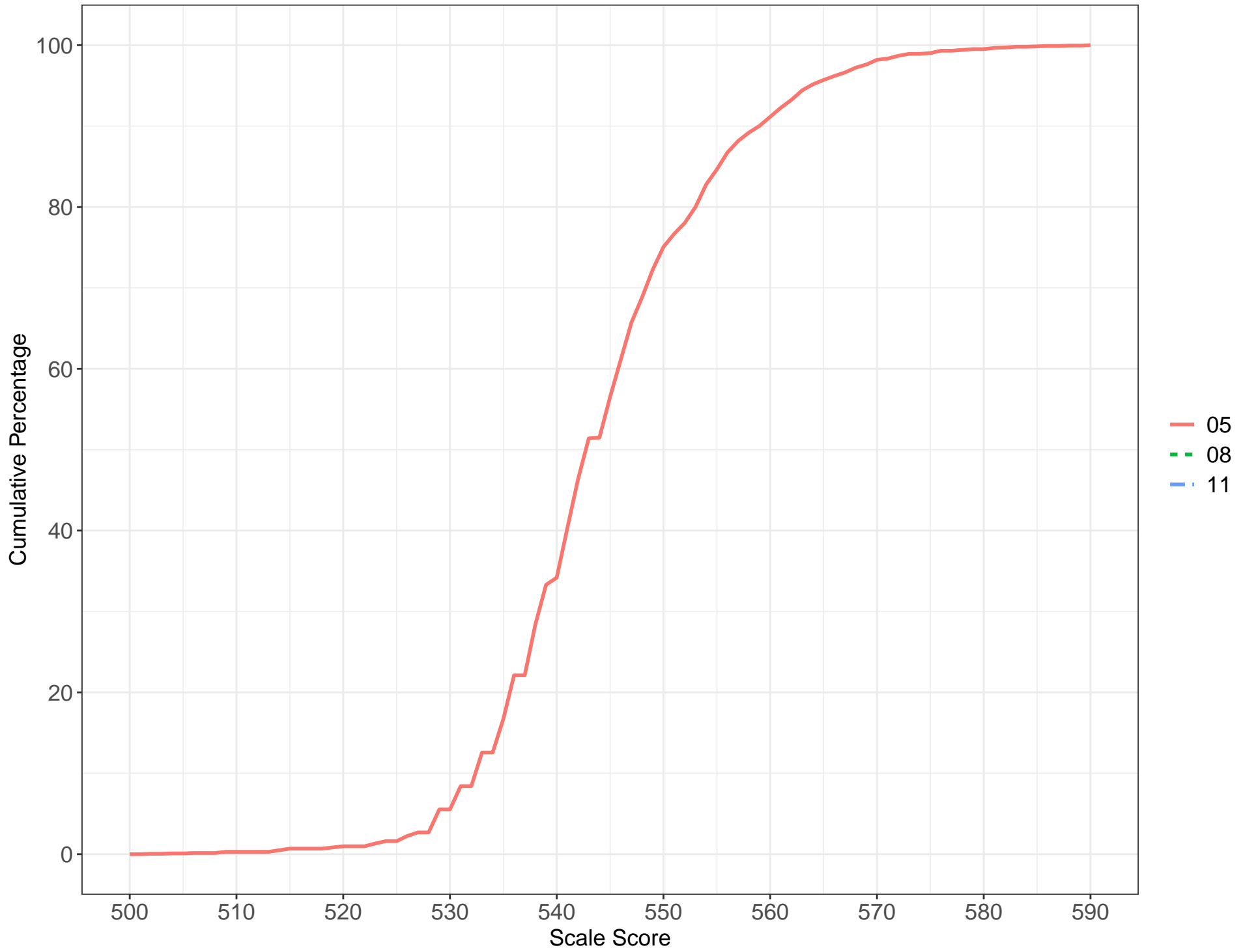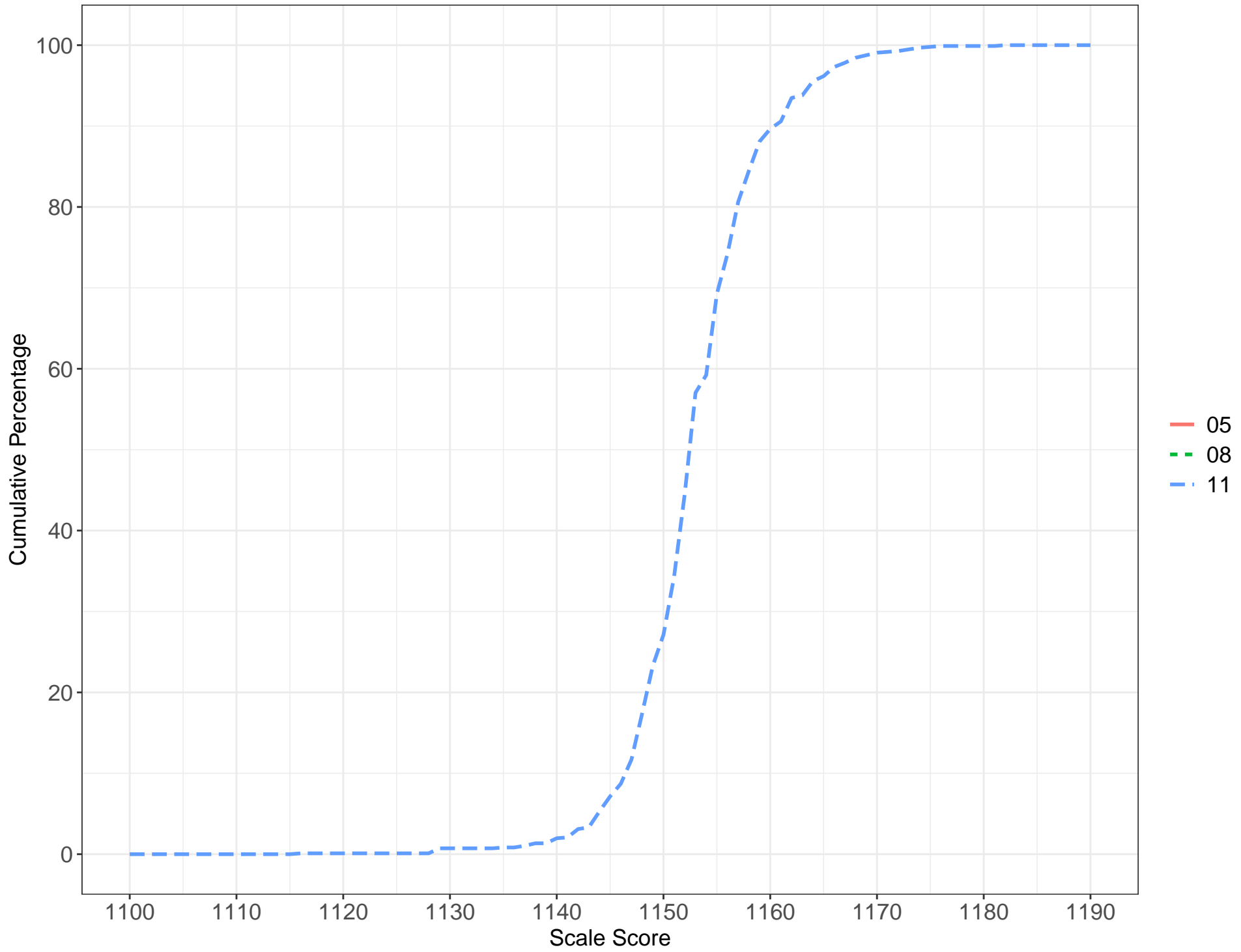
BIE Science Cumulative Scale Score Distributions
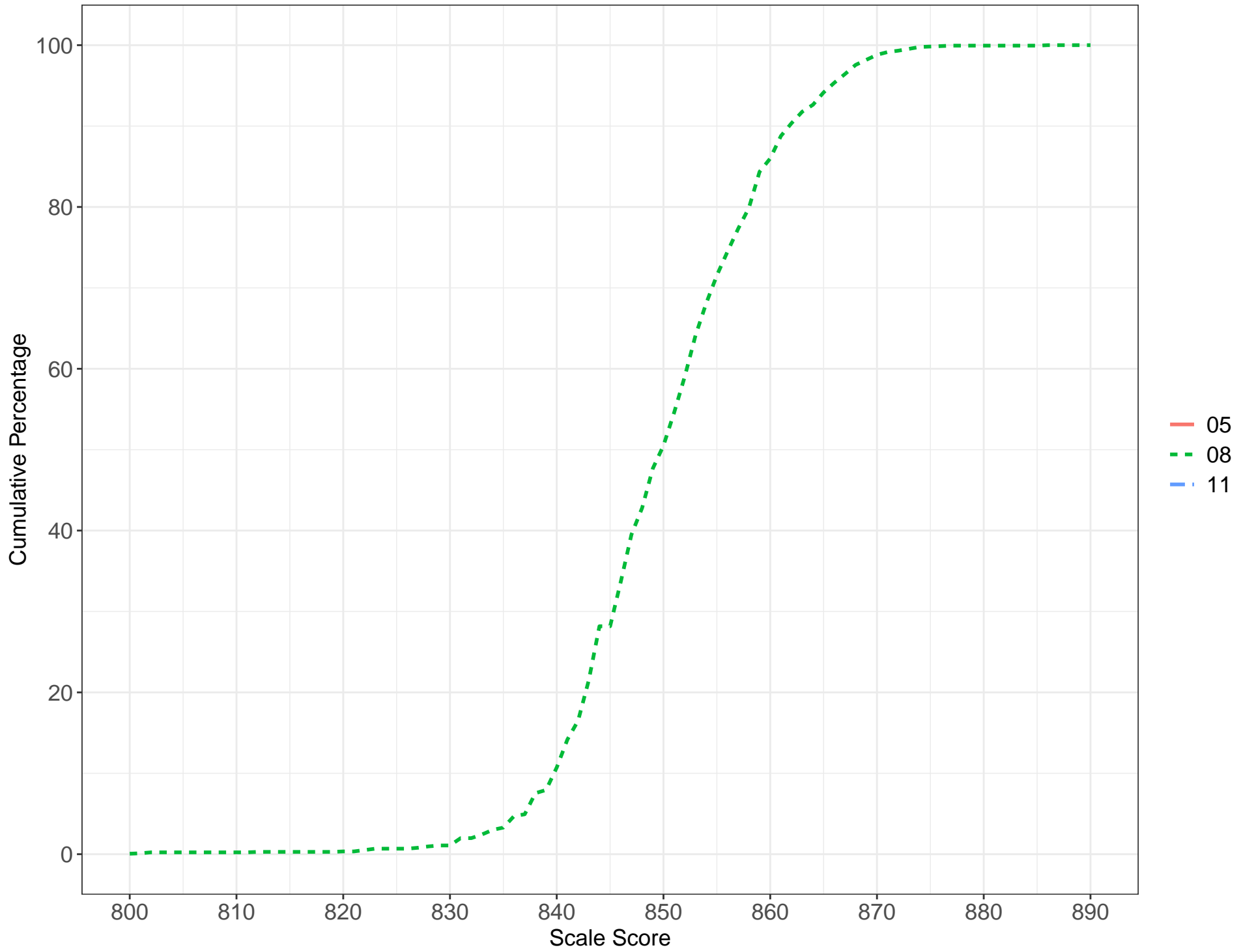
BIE Science Cumulative Scale Score Distributions

APPENDIX M
SCALED SCORE DESCRIPTIVE STATISTICS

Calculations based on those students attempting 5 or more items on the given BIE assessment. Statistic values are suppressed for those subjects/grades with fewer than 50 students.

Table M-1. Scaled Score Descriptive Statistics for BIE Science Grade 5, as a Function of Subgroup*

Subgroup	Number of Students	Mean	Median	SD	Skewness	Kurtosis
Overall	2,045	544.86112	543	10.77502	0.47440	4.07870
Female	973	544.94347	543	10.34487	0.45977	4.43253
Male	1,068	544.77996	543	11.16869	0.48736	3.80316
Currently receiving LEP services	329	543.77204	543	9.28871	0.37810	3.95324
Not receiving LEP services	341	546.03226	545	11.08877	0.50340	3.19035
Special Ed	337	540.07715	539	8.97019	0.49042	6.06616
Non-Special Ed	1,702	545.81022	545	10.86483	0.43341	3.92003
Gifted Students	110	550.39091	547.5	12.66295	0.42628	3.46705
Non-gifted Students	101	544.71287	543	10.16498	0.52393	4.02055
Title1 Students	1,972	544.84736	543	10.71851	0.50976	4.04472
Non-title1 Students	46	546.54348	544	11.17280	0.09274	3.18649

Calculations based on those students attempting 5 or more items on the given BIE assessment. Statistical values are suppressed for those subjects/grades with fewer than 50 students.

Table M-2. Scaled Score Descriptive Statistics for BIE Science Grade 8, as a Function of Subgroup*

Subgroup	Number of Students	Mean	Median	SD	Skewness	Kurtosis
Overall	1,764	850.55782	850	9.19244	-0.25073	4.70411
Female	830	850.51325	851	8.73835	-0.26320	4.36100
Male	934	850.59743	850	9.58242	-0.24312	4.85288
Currently receiving LEP services	215	850.13488	850	7.63021	0.43641	2.97333
Not receiving LEP services	299	851.50836	851	8.72140	0.08258	2.85570
Special Ed	271	845.99262	846	7.43465	-0.47931	8.24037
Non-Special Ed	1,487	851.39812	851	9.25380	-0.32378	4.60958
Gifted Students	139	855.72662	856	10.89063	-0.71693	6.38906
Non-gifted Students	97	850.60825	851	7.79096	0.07560	2.47643
Title1 Students	1,706	850.53751	850	9.25114	-0.25002	4.69682
Non-title1 Students	38	850.34211	852	7.92286	-0.15014	3.01905

Calculations based on those students attempting 5 or more items on the given BIE assessment. Statistical values are suppressed for those subjects/grades with fewer than 50 students.

Table M-3. Scaled Score Descriptive Statistics for BIE Science Grade 11, as a Function of Subgroup*

Subgroup	Number of Students	Mean	Median	SD	Skewness	Kurtosis
Overall	966	1153.38509	1153	6.19985	-0.05892	6.38686
Female	482	1153.42324	1153	5.99197	-0.44964	7.52801
Male	484	1153.34711	1153	6.40614	0.26188	5.48254
Currently receiving LEP services	167	1154.92216	1155	5.48052	0.24224	3.36050
Not receiving LEP services	223	1154.20179	1153	5.91072	0.92619	5.25569
Special Ed	140	1149.93571	1150.5	5.53109	-0.40727	5.76873
Non-Special Ed	804	1153.94776	1153	6.17106	-0.04038	6.74198
Gifted Students	62	1154.61290	1155	7.76574	-0.46041	4.42715
Non-gifted Students	93	1153.43011	1153	6.13153	0.88424	4.74156
Title1 Students	906	1153.43157	1153	6.26772	-0.08563	6.42916
Non-title1 Students	34	1152.14706	1153	5.46134	0.85477	3.94429

Calculations based on those students attempting 5 or more items on the given BIE assessment. Statistical values are suppressed for those subjects/grades with fewer than 50 students.